ENCYCLOPÆDIA BRITANNICA

Fourteenth Edition, Revised

1966

v21: SOR-TEX

THE UNIVERSITY OF CHICAGO

The Encyclopædia Britannica
is published with the editorial advice of the faculties
of The University of Chicago and of a
committee of members of the faculties of Oxford, Cambridge
and London universities and of a committee
at The University of Toronto

*

"LET KNOWLEDGE GROW FROM MORE TO MORE
AND THUS BE HUMAN LIFE ENRICHED."

A New Survey of Universal Knowledge

ENCYCLOPÆDIA BRITANNICA

Volume 21

SORDELLO *TO* TEXTBOOKS

ENCYCLOPÆDIA BRITANNICA, INC.

WILLIAM BENTON, PUBLISHER

CHICAGO · LONDON · TORONTO · GENEVA · SYDNEY · TOKYO

©

1966

BY ENCYCLOPÆDIA BRITANNICA, INC.

COPYRIGHT UNDER INTERNATIONAL COPYRIGHT UNION

ALL RIGHTS RESERVED UNDER PAN AMERICAN AND UNIVERSAL COPYRIGHT CONVENTIONS BY ENCYCLOPÆDIA BRITANNICA, INC.

PRINTED IN THE U. S. A.

Library of Congress Catalog Card Number: 66-10173

FOUNDED A.D. 1768

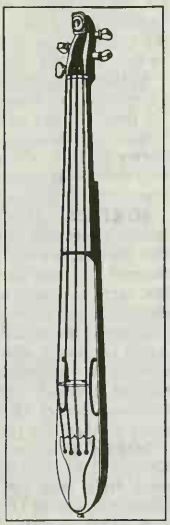
ENCYCLOPÆDIA BRITANNICA

Volume 21

SORDELLO *TO* TEXTBOOKS

SORDELLO, a 13th-century Italian troubadour, born at Goito (*c.* 1200), who is praised by Dante in the *De vulgari eloquentia,* and in the *Purgatorio* is made the type of patriotic pride. He is the hero of a well-known poem by Robert Browning. The real Sordello was the most famous of the Italian troubadours. About 1220 he appeared at Florence in a tavern brawl; and in 1226, while at the court of Richard of Bonifazio at Verona, he abducted his master's wife, Cunizza, at the instigation of her brother, Ezzelino da Romano. The scandal resulted in his flight (1229) to Provence, where he seems to have been for some time. He entered the service of Charles of Anjou, and probably accompanied him (1265) on his Naples expedition; in 1266 he was a prisoner in Naples. The last documentary mention of him is in 1269, and he is supposed to have died in Provence. His didactic poem, *L'Ensenhamen d'onor,* and his love songs and satirical pieces have little in common with Dante's presentation, but the invective against negligent princes which Dante puts into his mouth in the 7th canto of the *Purgatorio* is more adequately paralleled in his *Serventese* (1237) on the death of his patron Blacatz, where he invites all Christian princes to feed on the heart of the hero.

SORDINO, SORDONI, SORDUN, musical terms somewhat promiscuously applied (1) to contrivances for damping or muting wind, string and percussion instruments (*Sordini*); (2) to a family of obsolete wind instruments resembling the bassoon, blown by means of a double reed (*Sordoni* or *Sordun*); (3) to a stringed instrument. To these is added the *Sardellina* or *Sordellina,* a kind of musette invented (*see* BAGPIPE) in Naples in the 17th century.



BY COURTESY OF MESSRS. HOOPER AND JACKSON

SORDINO, AN OBSOLETE STRING INSTRUMENT

SOREL, AGNÈS (*c.* 1422–1450), mistress of Charles VII of France, was born of the lesser nobility at Fromenteau in Touraine. She was attached to the service of Isabella of Lorraine, wife of René of Anjou, then to that of the queen of France, Mary of Anjou. From 1444 she was the acknowledged mistress of the king, the first to hold that semi-official position. Her ascendancy dated from the festivals at Nancy in 1444, the first brilliant court of Charles VII. Charles's gift to her of the castle of Beauté-sur-Marne led to her being nicknamed "Reine de Beauté." Her death from dysentery, shortly after the birth of her fourth child in 1450, was attributed, apparently without foundation, to poison.

See G. du Fresne de Beaucourt, *Histoire de Charles VII,* vol. iii (Paris, 1881–91); P. Champion, *Agnès Sorel, la dame de Beauté* (Paris, 1931).

SOREL, ALBERT (1842–1906), French historian, was born at Honfleur on Aug. 13, 1842. He was of a characteristically Norman type and remained all his life a lover of his native province. He studied law in Paris and entered the foreign office (1866). In 1870 he was chosen as secretary by M. de Chaudordy, who had been sent to Tours as a delegate in charge of the diplomatic side of the problem of national defense; in these affairs he proved himself a most valuable collaborator. After the war of 1870–71, when Boutmy founded the *École libre des sciences politiques,* Sorel was appointed to teach diplomatic history (1872), a post in which he achieved great success.

Some of Sorel's courses have formed books: *Le Traité de Paris du 20 novembre 1815* (1873); *Histoire diplomatique de la guerre franco-allemande* (1875); also the *Précis du droit des gens* which he published (1877) in collaboration with his colleague Théodore Funck-Brentano. In 1875 Sorel left the foreign office and became general secretary of the *Présidence du sénat.*

His duties left him sufficient leisure for the great work of his life, *L'Europe et la révolution française* (8 vols., 1885–1904). His object was to do over again the work already done by Sybel, but from a less restricted point of view and with a clearer and more calm understanding of the chessboard of Europe. He spent almost 30 years in the preparation of this history; the analysis of the documents, mostly unpublished, on French diplomacy during the first years of the Revolution, which he published in the *Revue historique* (vol. v–vii, xi–xiii), shows with what scrupulous care he read the innumerable dispatches which passed under his notice. Sorel was elected a member of the Académie française (1894). He died in Paris on June 29, 1906.

Sorel's other works include: *La Question d'Orient au XVIIIe siècle, les origines de la triple alliance* (1878); *Montesquieu* (1887) and *Mme. de Staël* (1891) in the *Grands écrivains* series; *Bonaparte et Hoche en 1797* (1896) and *Recueil des instructions données aux ambassadeurs*, vol. i only (1884). Most of his essays and articles contributed to various reviews and to the *Temps* have been collected.

SOREL, CHARLES, Sieur de Souvigny (1602–1674), French writer, author of *Histoire comique de Francion* (1623), a picaresque novel of the road in Burgundy and Paris, remarkable chiefly for its sense of life and vigour of language, which went through 30 editions before the end of the century. He also wrote *Le Berger extravagant* (1627), *Polyandre* (1648) and many shorter satires and scholarly and religious works, including one of the first histories of French literature, *La Bibliothèque Françoise* (1664).

Sorel was born in Paris. Guy Patin in his letters described him as a fat little man with a sharp nose, leading a quiet bachelor's life with his sister, planning more works than his health would allow him to write. His chief glory is to have inspired more than one scene of Molière. He died on March 7, 1674.

Bibliography.—Biography by E. Roy (1891); *Francion*, ed. by E. Roy, 3 vol. (1928); G. Reynier, *Le Roman réaliste au 17e siècle* (1914). (W. G. Me.)

SORGHUM, a cereal, forage and sirup crop plant grown in many countries and known botanically as *Sorghum vulgare*. Sorghum probably originated in Africa. The types grown for grain (grain sorghum) are called by various names, including durra, Egyptian corn, great millet or Indian millet. On the subcontinent of India it is known as jowar, cholum or jonna, and in the West Indies as petit mil or Guinea corn. In China and Manchuria it is called kaoliang. It is a strong grass, growing to a height of from 2 to 8 ft. or even 16 ft. The stalks and leaves are coated with a white waxy bloom. The pith in the stalks of different varieties may be juicy or rather dry. The juice may be sweet or nonsweet. The leaves are sheathing, solitary and about 2 in. broad and 2½ ft. in length; the panicles or flower clusters are loose, contracted or dense. Self-pollination of the flowers is common but considerable cross-pollination occurs. The grains may be either free or retained in the hulls after threshing. Many varieties are awned. The seeds are ellipsoid, rounded or flattened and of varied sizes somewhat smaller than wheat grains. The seeds may be white, yellow, red or brown. The hulls are mostly straw-coloured, red, brown or black.

W. H. HODGE

SORGHUM (S. VULGARE)

Sorghum is the leading cereal grain in Africa and is important also in the United States, India, Pakistan, north China and Manchuria. It is grown to some extent in the U.S.S.R., Iran, Arabia, Argentina, Australia and southern Europe, as well as in other regions. It is best adapted to warm conditions and is very resistant to drought and heat. Hundreds of varieties are grown. The grain is similar in composition to that of maize except in being higher in protein and lower in fat. It replaces maize as a feed grain in hot, dry regions. For food it usually is ground into a meal and made into porridge, bread or cakes. Natives of south Africa refer to the product as "mealies." Whole grains sometimes are popped or puffed. The grain also is used in making starch, dextrose, paste and alcoholic beverages. The stalks provide fodder and building materials. The sweet sorghums (sorgos) are grown chiefly in the United States and south Africa for forage or for sirup manufacture. The sweet stalks are chewed by peoples of various countries. The broomcorn plant, belonging also to the species *S. vulgare*, is similar to other sorghums in adaptation and many plant characteristics. (J. H. Mn.)

SORIA, a province of Spain, formed in 1833 out of Old Castile. Pop. (1950 census) 164,575; area, 3,977 sq.mi. Soria is a bleak region, bounded on three sides by mountains. A range of sierras culminating in the peaks of Urbion (7,310 ft.) and Cebollera (7,026 ft.) on the north and the great Sierra del Moncayo (7,588 ft.) on the east separate the valley of the Duero (Douro) from the Ebro. Almost the whole of the province belongs to the region watered by the Duero and its affluents. There are forests of pine, oak and beech and large tracts of pasture land. The climate is cold and dry, and the scenery austere.

SORIA, the capital of the Spanish province of Soria; on the Duero (Douro) river. Pop. (1950) 16,328. The churches of Santa Domingo and San Nicolas, the collegiate church of San Pedro, the cloisters of the convent of San Juan and several other ecclesiastical buildings are fine specimens of Romanesque work of the 12th and 13th centuries. Near the Duero are the ruins of the old citadel, and the 13th century walls remain.

SORIANO, a department in southwest Uruguay bordering on the Uruguay river. Area 3,414 sq.mi. Pop. (1954 est.) 110,939. Wheat, flax, corn, barley are grown, and poultry, swine and bees are raised. The northern region of Soriano, a sheep and cattle area, is famous for its rich pastures. The departmental capital is Mercedes (*q.v.*), an agricultural, livestock and communication centre on the Negro river. Dolores, a port on the San Salvador river with a population of about 19,000 (1954 est.), becomes active during the harvest. (M. I. V.)

SORIN, EDWARD FREDERICK (1814–1893), French-U.S. Roman Catholic priest, educator, founder and first president of the University of Notre Dame, was born at Ahuillé, near Laval, France, on Feb. 6, 1814. Ordained priest in 1838, he joined in 1840 the Congregation of Holy Cross, a group of priests and brothers organized at Le Mans by Abbé Basil Antoine Moreau.

At the invitation of Bishop Celestine Hailandière of Vincennes, Ind., Sorin and six brothers went to Vincennes in 1841, at first settling at St. Peter's in Daviess county. In 1842 Hailandière offered Sorin land near South Bend, in St. Joseph county, where Father Stephen Theodore Badin had formerly conducted the mission Sainte Marie des Lacs. Sorin arrived there on Nov. 26, and in 1844 he obtained from the general assembly of Indiana a charter for the University of Notre Dame. He was president of the university until 1865, was provincial superior of his community in the United States until 1868, and was from then until his death, on Oct. 31, 1893, the superior general.

In 1843 Sorin established at Bertrand, Mich., near Notre Dame, a community of French Sisters of Holy Cross, and in 1854 secured for them the site adjacent to Notre Dame on which St. Mary's college was founded. He was instrumental in bringing to the Sisters of Holy Cross Mother Angela (Eliza Maria Gillespie), who guided the community for nearly 30 years. In 1865 Sorin began publishing *Ave Maria* magazine. (T. T. McA.)

SORITES. In traditional logic (*see* Logic), a chain of successive categorical syllogisms in the first figure may be so related that the conclusion of each except the last is the minor premiss of the next. If then the intermediate conclusions are suppressed, so that there are stated only the two premisses of the first syllogism, the major premisses of the remaining syllogisms, and the final conclusion, the resulting argument is a valid inference from the stated premisses, which may be considered independently of its analysis into syllogisms, and which is called a *sorites*. The distinction between the so-called *Aristotelian* sorites and the *Goclenian* sorites (Rudolph Goclenius, 1598) concerns only the order in which the premisses are stated. (Ao. C.)

SOROCA, a town of the Moldavian S.S.R., U.S.S.R., on the right bank of the Dniester, 81 mi. N.N.W. of Kishinev. Corn, wool, fruit, wine and cattle are exported. Soroca was the old Genoese colony of Olchiona and still has the ruins of a 13th-century Genoese castle. The Moldavians erected a fortress in the 15th century.

Soroca changed hands many times between Poland, Russia and

Turkey; in 1940 Rumania ceded it to the U.S.S.R. but regained it the next year. After 1944 it was in the Moldavian Soviet Socialist Republic.

SOROLLA Y BASTIDA, JOAQUÍN (1863–1923), Spanish painter, whose style was a conservative variant of Impressionism, was born at Valencia on Feb. 27, 1863, and studied at the academy there, in Italy and in Paris, where he especially interested himself in the works of J. Bastien-Lepage and A. von Menzel; he was also influenced by the north European realists, particularly A. Zorn. His early paintings were of history and social realism (one of the latter kind, "Otra Margarita," 1892, being his earliest success), but he later became well known for brightly lit scenes with Valencian peasants and fisherfolk and children playing in the surf, his style after about 1903–04 becoming impressionistic and summary, with heavily impasted pigments. Between 1910 and 1920 he painted portraits of Spanish writers and a "Panorama of the Forty-nine Provinces of Spain" for the Hispanic Society of America. Sorolla died in Madrid on Aug. 10, 1923.

SORORATE. This term was introduced by Sir James Frazer to designate all marriages with a wife's sister, whether in the lifetime of the first wife or after her death. In his view, generally accepted today, it is complementary to the custom of the levirate (*q.v.*). The concept already appears in E. B. Tylor's "On a Method of Investigating the Development of Institutions," where levirate and sororate correctly figure as correlates of the postulate that matrimony is a bond between families rather than individuals. A. R. Radcliffe-Brown subsumes both institutions under the principle of the social equivalence of brothers and sisters, respectively. These relationship terms are to be understood in a classificatory sense, *i.e.*, a more remote relative of the same sex may serve as secondary mate instead of a blood sibling. Though related in principle, the levirate and sororate are not invariably associated, but they usually are, and appear to be the commonest of preferential secondary marriages. Either may be permissive rather than obligatory. The Maricopa (Arizona) insist on a widow's marrying a husband's relative, whereas the replacement of a deceased wife by a kinswoman of hers is customary, but not compulsory.

Though successive and simultaneous marriage of two or more sisters falls under the same principle, some tribes, *e.g.*, the Kazak, favour one, while tabooing the other practice. Hence it has become necessary to distinguish between sororal polygyny and sororate. The typical rule for the former is that the husband of the eldest girl in a family marries her juniors as they come of age; L. H. Morgan found this usage in at least 40 North American tribes; and even recently Navaho men often were simultaneously married to two sisters, occasionally to three. Australian aborigines recognized the same pre-emptive claim, but in many tribes the husband contented himself with the two oldest girls, conveying his claims on their younger sisters to his junior brother. With remarkable unanimity aborigines explain sororal polygyny on the ground that sisters are unlikely to quarrel as cowives.

The effect of both the sororate and sororal polygyny is to have children extend the term "mother" to the maternal aunt, but this terminological trait is more probably directly correlated with unilinear descent.

See also MARRIAGE, PRIMITIVE.

BIBLIOGRAPHY.—A. R. Radcliffe Brown, *The Social Organisation of Australian Tribes* (1931); R. H. Lowie, *Social Organization* (1948); George P. Murdock, *Social Structure* (1949). (R. H. LO.)

SORORITY, a social, professional or honorary organization of women, usually a secret society whose name consists of a series of Greek letters and which is connected with a college or university. *See* FRATERNITY AND SORORITY.

SORREL, the common name applied to various species of *Rumex* (family Polygonaceae), especially to *R. acetosa*, the garden sorrel; *R. acetosella*, sheep sorrel; *R. paucifolius*, mountain sorrel; and *R. hastatulus*, heartwing sorrel. The leaves of garden sorrel are used in soups, salads and sauces and as a potherb. French sorrel, *R. scutatus*, is a hardy perennial, distributed throughout Europe, but not native in Great Britain.

The species of the genus *Oxalis* (*q.v.*), (family Oxalidaceae), are commonly called wood sorrel or lady's sorrel, and the sourwood (*Oxydendrum arboreum*) of the heath family is known as the sorrel tree. It is an attractive deciduous tree, native to the southeastern United States; its leaves turn a brilliant scarlet in the autumn. (J. M. BL.)

SORRENTO, a city of Campania, Italy (ancient *Surrentum*). Pop. (1961) 12,531 (commune). Sorrento stands on cliffs about 160 ft. high, between the Bay of Naples and the Bay of Salerno. It is a summer and winter resort, its northerly aspect rendering it comparatively cool.

At Sorrento, Bernardo Tasso wrote his romantic poem Amadigi. His son, the poet Torquato Tasso (1544–95) was born there.

The most important temples of the ancient city were those of Athena and of the Sirens, the latter the only one in the Greek world in historic times. The place was famous for its wine, its fish and its red Campanian vases. It was protected by deep gorges, except for a distance of 300 yd. on the southwest, defended by walls, the line of which is followed by those of the modern town. The arrangement of the modern streets also preserves that of the ancient town.

On the east the most important ancient ruin is the reservoir of the subterranean aqueducts, which had 27 chambers. There are also remains of villas, including that of Pollius Felix, the friend of Statius.

SORSOGON, the name of a municipality, provincial capital and the southernmost province of Luzon, Philippines. Area of province, 793 sq.mi.; pop. (1960) 347,771. The province is bordered on the south by San Bernardino strait and comprises largely volcanic cones with broad rich level areas between; the highest peak is Mt. Bulusan (5,118 ft.), a recently active cone. Rainfall varies from 100 in. to more than 160 in. with no dry season and maximum fall in the winter.

Rice, sweet potatoes and cassava are the main food crops produced, but the province is best known as an abacá (Manila hemp) centre. Before World War II slightly over 40% of the cultivated area was devoted to abacá, but by the early 1960s this area was halved. Coconuts continue to occupy a significant proportion of the farmed area.

Sorsogon (1960 pop. 35,548) is a trading town, processing station for abacá and copra; and port on the northeastern shore of Sorsogon bay. (R. E. HE.)

SOSIGENES, Greek astronomer and mathematician, probably of Alexandria, flourished in the 1st century B.C. According to Pliny's *Natural History,* he was employed by Julius Caesar in the reform of the Roman calendar (46 B.C.), and wrote three treatises. From another passage of Pliny it is inferred that Sosigenes maintained the doctrine of the motion of Mercury around the sun, which is referred to by Cicero and was also held by the Egyptians.

Sosigenes was the tutor of Alexander of Aphrodisias. He wrote *Revolving Spheres*, from which important extracts are preserved in Simplicius' commentary on Aristotle's *De caelo.*

SOSITHEUS (*c.* 280 B.C.), Greek tragic poet, of Alexandria Troas, a member of the Alexandrian "pleiad." He must have resided at some time in Athens, since Diogenes Laërtius tells us (vii, 5, 4) that he attacked the Stoic Cleanthes on the stage and was hissed off by the audience. Suïdas calls him a Syracusan. According to an epigram of Dioscorides in the Greek anthology (*Anth. Pal.* vii, 707) he restored the satyric drama in its original form. Part of his pastoral play, *Daphnis* or *Lityerses,* is extant.

See O. Crusius *s.v.* Lityerses in Roscher's *Lexikon der griechischen und römischen Mythologie.* The fragment in Nauck's *Tragicorum graecorum fragmenta* apparently contains the beginning of the drama.

SOSNOWIEC, a town of Poland, in Katowice province. Pop. (1960) 131,600. It owes its importance to its position in the centre of Dabrowa coal field, near Bedzin, Dabrowa and Katowice. The towns of this region are almost continuous, extending from Kielce into the provinces of Kraków and Slask. Sosnowiec is also a railway junction. Situated on the Warsaw-Vienna railway, it is a junction for the Kielce and Radom, Kraków and Lvov and Katowice and Breslau lines. Electric power stations were established. Iron foundries and textile factories, as well as coal mines,

employ large numbers of workmen.

Sosnowiec was seized by Germany in World War II and was returned to Poland in 1945.

SOSTENUTO, musical term signifying that the passage so marked is to be played in a "sustained" manner.

SOTER, SAINT, pope from about 166 to about 175. He wrote to the church of Corinth and sent it aid. His letter is mentioned in the reply given by Dionysius, bishop of Corinth, and A. Harnack thought that it could be identified with the second so-called epistle of Clement. St. Soter's feast day is April 22.

SOTHERN, EDWARD HUGH (1859–1933), U.S. actor, was born at New Orleans, La., on Dec. 6, 1859, the son of Edward Askew Sothern, noted English comedian. His first stage appearance was in a small part with his father's company at the Park theatre in New York city in 1879. He toured England in 1882–83, became leading comedian in John McCullough's company in 1883, and under Daniel Frohman was leading man at the Lyceum theatre in New York city. He married Virginia Harned in 1896, and in 1899 formed his own company with her as his leading lady.

In 1900 he appeared in the title role of *Hamlet*, in 1901 in that of *Richard Lovelace* and in 1902–03 as Villon in *If I Were King*, three of his greatest roles. In 1904 Sothern played opposite Julia Marlowe (*q.v.*) for the first time in *Romeo and Juliet* at Chicago, Ill. Thereafter, except for two years, 1907–09, they appeared together on the stage almost continuously until their retirement. They were married in 1911. Besides *Romeo and Juliet* they cooperated in *Much Ado About Nothing*, *Taming of the Shrew*, *Merchant of Venice*, *Twelfth Night*, *Macbeth*, *Jeanne D'Arc*, *John the Baptist*, *When Knighthood Was in Flower* and *The Sunken Bell*. Although noted chiefly as a Shakespearean actor, Sothern had a repertory of over 125 diverse parts. He won wide popularity as a dashing, romantic hero in melodramas such as *The Prisoner of Zenda*. Sothern wrote an autobiography, *The Melancholy Tale of Me* (1916).

SOTHIC PERIOD, in ancient Egyptian chronology, the period in which the year of 365 days circled in succession through all the seasons. The tropical year, determined as it was in Egypt by the heliacal rising of Sirius (Sothis), was almost exactly the Julian year of precisely 365¼ days (differing from the true solar year, which was 11 minutes less than this). The sothic period was thus 1,461 years. *See* CALENDAR; EGYPT: *History: Ancient Civilization and Culture.*

SOTHO, a powerful nation of Bantu-speaking peoples which inhabits the British colony of Basutoland (*q.v.*) in South Africa. It is made up of a large number of different tribes, which were welded together early in the 19th century by the great chief Moshesh.

SOTO, FERDINANDO DE: *see* DE SOTO, HERNANDO OR FERNANDO.

SOTTO VOCE (It.), lit. "under the voice," that is, an undertone. Term applied both to music and speech.

SOUBISE, BENJAMIN DE ROHAN, DUC DE (?1589–1642), Huguenot leader, younger brother of Henri de Rohan, inherited his title through his mother Catherine de Parthenay. He served his apprenticeship as a soldier under Prince Maurice of Orange-Nassau in the Low Countries. In the religious wars from 1621 onwards his elder brother chiefly commanded on land and in the south, Soubise in the west and along the sea-coast. Soubise's chief exploit was a singularly bold and well-conducted attack (in 1625) on the Royalist fleet in the river Blavet (which included the cutting of a boom in the face of superior numbers) and the occupation of Oléron. He commanded at Rochelle during the famous siege. When surrender became inevitable he fled to England, which he had previously visited in quest of succour. He died in 1642 in London.

SOUBISE, CHARLES DE ROHAN (1715–1787), peer and marshal of France, the grandson of the princesse de Soubise known to history as one of the mistresses of Louis XIV. He accompanied Louis XV. in the campaign of 1744–48. Soon after the beginning of the Seven Years' War, through the influence of Mme. de Pompadour, he was put in command of a corps of 24,000 men, and was defeated at Rossbach (1757). He continued in the service until the peace of 1763. He died in Paris on July 4, 1787.

SOUHAM, JOSEPH, COUNT (1760–1837), French soldier, was born at Lubersac on April 30, 1760, and became a general of division in 1793. He was disgraced with Moreau and Pichegru for alleged participation in the conspiracy of Cadoudal. He regained his rank in 1809, took a notable part in Gouvion St. Cyr's operations in Catalonia, and won the title of count. In 1812 Masséna, in declining the command of Marmont's army recommended Souham for the post. The latter was thus pitted against Wellington, and by his skilful manoeuvres regained the ground lost at Salamanca. At the fall of the First Empire he deserted the emperor, and was well received by Louis XVIII., who gave him high commands. He retired in 1832, and died on April 28, 1837.

SOULOUQUE, FAUSTIN ÉLIE (1789?–1867), Negro emperor of Haiti, was born a slave about 1789 while Haiti was under French rule. He participated in the successful Haitian revolt against the French in 1803 and thereafter continued as an officer in the Haitian army. He was made president of Haiti in 1847 because the mulatto leaders who had dominated the government under several figurehead Negro presidents thought that his illiteracy and ignorance would make him easy to control. He soon turned against his would-be advisers and ruled as a cruel and corrupt despot, proclaiming himself Emperor Faustin I in 1849 and creating a numerous nobility. Many of the mulatto leaders were killed or exiled.

Soulouque made several costly and unsuccessful attempts to conquer the Dominican Republic, until the United States and France and Great Britain in 1851 demanded that he desist. He later renewed his attacks and in 1855 was defeated by the Dominican army.

In 1859 he was ousted after the chief of his general staff, realizing that the emperor suspected his loyalty, led a revolt. Soulouque escaped and went into exile. He died in 1867.

See also HAITI: *History.* (D. G. Mo.)

SOULT, NICOLAS JEAN DE DIEU, Duke of Dalmatia (1769–1851), marshal of France, was born at Saint-Amans-la-Bastide (now in department of the Tarn) on March 29, 1769, the son of a notary. He was intended for the bar, but on his father's death in 1785 he enlisted as a private in the French infantry, and rose rapidly in the army. He laid the foundations of his military fame by his conduct in Masséna's great Swiss campaign (1799), and especially at the battle of Zürich. He acted as Masséna's principal lieutenant through the protracted siege of Genoa, and after many successful actions he was wounded and taken prisoner at Monte Cretto on April 13, 1800. The victory of Marengo restoring his freedom, he received the command of the southern part of the kingdom of Naples, and in 1802 he was appointed one of the four generals commanding the consular guard. Despising Napoleon, Soult affected devotion, being appointed in 1803 to the command at Boulogne and in 1804 to be one of the first marshals of France. He commanded a corps at Ulm, and at Austerlitz (*q.v.*) he led the decisive attack. After the peace of Tilsit he was created (1808) duke of Dalmatia. In the following year he was given a command in Spain after the battle of Gamonal and he pursued Sir John Moore to Corunna.

For the next four years Soult remained in Spain, and his military history is that of the Peninsular War (*q.v.*). In 1812 he was obliged, after Wellington's victory of Salamanca, to evacuate Andalusia, and was soon after recalled from Spain at the request of Joseph Bonaparte, with whom he had always disagreed. In March 1813 he assumed the command of the IV. corps of the *Grande Armée*, but he was soon sent to the south of France to repair the damage done by the defeat of Vittoria. His campaign there is the finest proof of his genius as a general, although he was repeatedly defeated by Wellington, for his soldiers were raw conscripts, facing Wellington's veterans.

Marshal Soult's political career was less creditable, and it has been said of him that he had character only in front of the enemy. After the first abdication of Napoleon he declared himself a Royalist, received the order of St. Louis, and acted as minister for war (Dec. 3, 1814–March 11, 1815). When Napoleon returned from Elba Soult declared himself a Bonapartist, was

made a peer of France and acted as major-general (chief of staff) to the emperor in the Waterloo campaign.

At the second Restoration he was exiled, but was recalled in 1819 and in 1820 again made a marshal of France and in 1827 a peer. After the revolution of 1830 he made out that he was a partisan of Louis Philippe, who revived for him the title of marshal-general. He was minister for war, 1830–34 and 1840–44, and ambassador extraordinary to London for the coronation of Queen Victoria in 1838. In 1848, when Louis Philippe was overthrown, Soult again declared himself a republican. He died at his castle of Soultberg, near his birthplace, on Nov. 26, 1851. Soult published a memoir justifying his adhesion to Napoleon during the Hundred Days, and his notes and journals were arranged by his son Napoleon Hector (1801–57), who published the first part (*Mémoires du maréchal-général Soult*) in 1854. Le Noble's *Mémoires sur les opérations des Français en Galicie* are supposed to have been written from Soult papers.

See A. Sallé, *Vie politique du maréchal Soult* (1834); A. de Grozelier, *Le Maréchal Soult* (1851); A. Combes, *Histoire anecdotique du maréchal Soult* (1869).

SOUND. This article is divided into the following sections:

I. What Is Sound?
  1. Motion and Sound
  2. Wave Motion; Velocity and Intensity
  3. Reflection and Refraction of Sound
  4. Diffraction and Scattering
  5. Harmonic Sound Waves; Pitch, Frequency and Wave Length
  6. Wave Fronts and Rays
II. Analysis of Sound Propagation
  1. Wave Functions
  2. Differential Equation of Wave Motion and Its Deduction
  3. Velocity of Sound Waves in Fluids
  4. Mathematical Representation of Wave Fronts
  5. Sound Rays; Refraction
  6. Refraction in a Moving Medium
  7. Huygens' Principle; Reflection and Refraction at an Interface
  8. Energy Propagation in Sound Waves; Intensity
  9. Intensity of Spherical Sound Waves
  10. Flow of Acoustic Energy Across a Boundary
  11. Acoustic Impedance
  12. Selective Acoustic Transmission; Filtration
  13. Absorption and Dispersion of Sound in Fluids
  14. Diffraction and Scattering of Sound
  15. Sound Waves in Finite Spaces; Standing Waves in a Tube
  16. Wave Pulses; Group Velocity
  17. Radiation Pressure
  18. Shock Waves
III. Sources of Sound
  1. Introduction
  2. Mechanical Vibrations, Damped and Forced
  3. Coupled Systems
  4. Vibrations of Strings; the Helmholtz Resonator
  5. Vibrations of Membranes
  6. Longitudinal Vibrations of Rods
  7. Flexural and Torsional Vibrations of Rods
  8. Vibrations of Plates and Diaphragms
  9. Radiation Efficiency of a Sound Source
  10. Beams From Sound Sources
  11. Practical Sound Sources
    Voice and Speech
    Sirens
    Jet-Edge Tones and Organ Pipes
    Electroacoustic Sources
  12. Doppler Effect in Acoustics
IV. Reception of Sound
  1. Introduction
  2. The Ear and Hearing
  3. Microphones
V. Applications of Sound
  1. Sound Measurements
    Velocity
    Frequency
    Acoustic Intensity
    Acoustic Impedance
  2. Recording and Reproduction of Sound
  3. Architectural Acoustics
    Requirements for Good Room Acoustics
    Sound Distribution in a Room
    Optimum Acoustical Conditions
    Acoustic Absorbing Materials
    Sound-Amplification Systems
    Noise Control in Rooms
  4. Musical Acoustics
  5. Military Acoustics
    Sound Ranging in Air
    Underwater Sound
  6. Ultrasonics
  7. Noise and Its Control

I. WHAT IS SOUND?

When a person opens his mouth and utters speech he is said to emit sound. Anyone in the vicinity with normal ears (or in default thereof a suitable hearing aid) is said to hear the sound. This common experience is a physical phenomenon of the greatest significance to human beings, who may be considered to be immersed in a world of sound having notable influence on most of their everyday activities.

The importance of this particular sensation in human life is reflected in the large number of words descriptive of the different kinds of sounds which come to our attention. Inanimate nature produces the thunder of the storm, the roar and pounding of the surf, the whistling of the wind, the whispering of the trees, the patter of rain, the rippling and gurgling of running water, the humming of wires, the creaking of snow. Even more rich is the vocabulary descriptive of the sounds of living things: the barking and snarling of dogs, mewing of cats, crowing of fowls, roaring of lions, hissing of snakes, the lowing and bellowing of cattle, blatting of goats, chirping of birds and insects, screaming of gulls, crying of infants, etc.

Finally, man, not content with all the racket around him over which he has little or no control, has contrived to produce sounds of almost infinite variety causing both pleasure and pain to countless millions. The boom of cannon, the crack of the pistol, the rattle of musketry and machine-gun fire, the whine of the shell and the blast of its explosion are unpleasantly familiar. The whir of machinery and the ticking of clocks have more agreeable connotations though they might be distracting to some people. But the melody and harmony of music are generally admitted to contribute aesthetic enjoyment.

What is this sound which forms so large a part of man's waking experience? The purpose of this article is to give the physicist's answer to this question; it will be shown that there is far more to sound than "meets the ear" and that the sounds which are not heard are in some respects the most important ones for modern physics and its technical applications.

Sound is a physical phenomenon and the branch of physics which describes it is called acoustics (from the Greek word meaning "hearing"). Physics is an abstract science and its method of description employs mathematical analysis freely. Without such analysis a physicist considers a deep understanding of a physical phenomena unattainable. On the other hand, the general reader understandably prefers to be told the results of scientific description in the language of everyday speech. This article introduces the subject, therefore, with a general nonmathematical survey of sound. This is followed by a more analytical account.

1. Motion and Sound.—It is obvious to even the most casual observer that a thorough analysis of all that is involved in the emission of any one of the sounds mentioned above must be complex. To make clear what really goes on, for example, in the head and larynx when one speaks would undoubtedly require the combined efforts of a physiologist, a psychologist and a physicist (with, of course, the aid of a mathematician). Nevertheless, it is generally agreed that the ultimate result is some motion of the air in front of the mouth. It is also generally accepted that when the normal person hears a sound there is some motion of the air at the entrance to his ear.

A careful examination of all sound-producing phenomena always shows the existence of motion of some medium. This is rather satisfying to physicists, who have long sought to explain most natural phenomena in terms of motion, because they believe they understand what the latter is. To those whose ideas of motion are confined to such solid things as automobiles and airplanes, the association of motion with sound may provide some difficulty. Realizing that in open, still air it is possible to hear a cricket chirp at a distance of half a mile, the sceptic asks how it is possible that such a small insect can move the mass of air in a hemisphere with a radius of half a mile (over 1,000,000 tons) so that the air in the

vicinity of one's ear may move sufficiently to lead to hearing. The fallacy of this reasoning lies in the assumption that the sound producer must move *all* the surrounding air *all at once* in order to produce the sensation of sound. It ignores the fact that air is a compressible elastic fluid, *i.e.*, can be squeezed, so that it is possible for motion to take place in one part without appearing simultaneously everywhere else. One of the interesting properties of such a medium, however, is that if a "squeeze" is produced in one small region, it does not stay, as it were, "frozen" there. Once squeezed the medium tends to expand again and in so doing compresses (*i.e.*, moves) an adjoining portion of the medium, which in turn repeats the process, with the result that portions of the medium far from the source ultimately get squeezed (of course at a time later than the original disturbance). This kind of motion, communicated in time from one part of a medium to another, is called wave motion. We say that sound travels in air as a compressional or squeeze wave.

2. Wave Motion; Velocity and Intensity.—The standard pictorial illustration of wave motion is provided by the ripples produced on the surface of water when a stone is dropped into it. One of the striking features of this phenomenon is the fact that the water itself does not move outward from the centre of the disturbance, whereas the distortion of the surface which constitutes the ripple does precisely this. The distance traveled by the ripple per unit time is called the wave velocity. All waves studied by physicists have a finite wave velocity, although its magnitude varies greatly from type to type. Thus ripples on the surface of water may move only a metre or so per second. Light in free space, on the other hand, travels at the enormous speed of 3×10^8 m. per second. What can we say of the velocity of sound?

The speed of sound in air increases (as we shall see later) with the temperature, but at room temperature it is about 344 m. (1,125 ft.) per second. Its relatively small value compared with that of light is responsible for the common experience in which the sound of a distant boat whistle is heard some time after the puff of steam is seen.

Not all sound waves travel with the speed of sound in air. Thus sound in fresh water has a velocity of about 1,500 m. per second at room temperature. More details on this are presented in the analytical discussion in section II.3 below.

In everyday language we distinguish between loud and soft sounds. This suggests attention to the magnitude of the motions of air involved in sound propagation. In normal conversational speech the pressure of the air in front of the mouth of the speaker is changed at the most by only about one-millionth of the standard atmospheric pressure. At the same time the accompanying motion that produces the compression leads to an air-flow velocity with a maximum value of only two one-hundredths of a centimetre per second. That such small changes in the air when reproduced at the ear cause the sensation of hearing should strengthen respect for the sensitivity of that sense organ. Since sound involves propagated motion and motion implies energy, we can look upon sound in air as equivalent to the transmission of mechanical energy through the air. It then turns out that an average rate of transfer of only 10^{-16} (1 divided by 10 raised to the 16th power) watt per square centimetre of acoustic power is sufficient to produce the sensation of hearing in a normal youthful ear. The average power transmission in a sound wave per unit area is called its intensity. The modern unit for this quantity is the decibel (abbreviated to db.). Strictly this measures the intensity of a sound relative to a standard intensity. (*See* section II.8 below.) For convenience the latter is often chosen as that corresponding to minimum audibility (maximum excess pressure 2×10^{-4} dynes/cm.^{2}). On this basis the sound of conversational speech has an intensity of about 50 db. at a distance of a few feet and that of traffic at a busy intersection 70 db.; that of a boiler factory can attain the value 110 db.

Everyone has noticed that the farther off he is from a source of sound in the open, as, for example, an airplane, the less distinctly he hears it. In technical terms, the intensity of sound in an unenclosed space decreases with the distance from the source. This is really merely a matter of geometry. The airplane may be thought of as emitting the same amount of energy every second in every direction, but as the distance increases this energy passes through surfaces of larger and larger area, and hence the average flow per unit area is diminished. In fact, if the sound disturbance travels with the same velocity in every direction, it will at any given instant reach the surface of a *sphere* with the source at the centre and with radius equal to the product of the velocity of sound in the medium and the elapsed time. Hence the intensity will decrease in the same ratio as the surface of the sphere increases with its radius; *i.e.*, with the square of the distance from the source. (*See* section II.9 below.) .The situation is unfortunately more complicated for a source of sound on the ground or under water (near the surface), and even more involved when the source is enclosed in a room. Nature is rarely simple! And the only way to begin to understand its ways is to pick rather idealized situations.

Even the inverse-square law just mentioned does not tell the whole truth about the change of sound intensity with distance from the source in an unbounded three-dimensional medium. It is observed actually that the drop is greater than is predicted by this law. Why is this? Apparently some of the energy represented by the sound disturbance gets "lost" in the process of propagation. We say it is absorbed and changed into heat. This absorption of sound is particularly noticeable in a viscous liquid like glycerin. In air it is accentuated by reflection and refraction and scattering (*see* sections I.3 and I.4 following).

3. Reflection and Refraction of Sound.—If sound is a wave motion it must manifest other properties of waves besides velocity and intensity. We know that light waves are reflected from surfaces separating media of different properties, and refracted (*i.e.*, bent from the original direction of propagation) in crossing such surfaces obliquely. There are sound phenomena with precisely these characteristics. A familiar example is the echo produced when a loud, sharp sound is emitted near a high wall or cliff. This is most readily interpreted as caused by the reflection of the original sound from the solid surface, verified by the experimental observation that the time interval elapsing between the initial sound and the echo is the time taken for the sound to reach the cliff and the reflected disturbance to return.

Important also is the concept of acoustic image. Like its optical counterpart this is an imaginary source of sound which, if it were located exactly as far behind the reflecting surface as the initial source is in front of it and if the surface were then removed, would give the observed echo at the proper time and with the proper intensity.

It is much easier for a speaker talking with a certain strength of voice to reach an auditor at a given distance if the two are in a closed room rather than out in the open air. Curiously enough this also is an important illustration of the reflection of sound. For the sound from the speaker reaches the auditor not merely in the direct line between the two but also by many paths involving reflection by the walls, floor and ceiling of the room. Consequently much of the energy which would be dissipated in all directions in space out of doors is, so to speak, trapped in the room and helps to build up the sound intensity at every point. In a not-too-large room the echoes from the various reflecting surfaces merge to form a continuous sound or reverberation, which lasts longer than the direct sound and can impair the distinctness of hearing if it is too long. We then say that the room needs acoustic treatment to cut down the effect of wall reflection. This is done by the introduction of suitable sound-absorbing materials (*see* section V.3 below).

The refraction of sound is less obvious than refraction of light but produces well-recognized effects. Since sound travels faster in warm air than cold and since the atmosphere almost always manifests a vertical temperature gradient, it is to be expected that sound will rarely travel through it in straight lines from the source. Actually it is observed that in the presence of a negative temperature gradient (the normal condition) the sound rays (the significance of a ray of sound will be explained fully in section I.6) are bent upward and hence the range of hearing in the horizontal direction is materially reduced. On the other hand when the temperature gradient is positive, with the temperature increasing upward (so-

called temperature inversion, a rather infrequent phenomenon in most localities, but one which occurs occasionally in the early daylight hours after a clear, wind-free night), the upward-moving rays tend to be bent downward, assuring longer range of transmission. Just as the refraction of light through air of varying temperature leads to the phenomenon known as the mirage, so it is also possible to have acoustic mirages under similar conditions, and sound may appear to come from a quite different direction than it would through a homogeneous medium. All these effects of refraction caused by temperature are further complicated by refraction caused by the motion of the medium; *i.e.*, winds. It is well known that sound in air travels better with the wind than against it, and the effective propagation velocity is materially changed by vector addition with the wind velocity.

The refraction of sound in air by wind and temperature plays a very significant and somewhat embarrassing role in gun ranging, the acoustical method of locating guns by the sounds produced by their firing (section V.5 below). The reflection and refraction of sound in water have taken on great practical importance in the echo-ranging method of detecting underwater objects in which a beam of sound is projected in water and a portion of it reflected by a solid object (*e.g.*, a submarine) which it strikes. The detection of the reflected sound leads to an estimate of range and bearing of the object being sought after. But refraction or bending of the sound paths caused by temperature gradients in the water introduces obvious complications (section V.5 below).

4. Diffraction and Scattering.—One of the most striking properties of sound is its ability to bend around obstacles. It is because of this that one can hear around a corner. This is an illustration of the characteristic of all waves known as diffraction. Even light displays it but obviously not to such a pronounced degree as sound, since we cannot see around a corner (without a periscope). The reason for this difference is discussed in section I.5 below.

Diffraction accounts for the inability of objects of ordinary size to form sharp shadows of audible sound. Here again further discussion is necessary since not all sounds bend easily around obstacles (section II.14 below).

An interesting example of sound diffraction is provided by the human head. Our hearing of sound is definitely conditioned by the fact that our ears are imbedded in a roughly spherical, more or less solid sphere. The ticking of a watch sounds different in front of the head than it does at the same distance directly behind. This is due to diffraction. Theory and experiment alike indicate that because of diffraction a source of sound produces a greater intensity at the head if it lies on an extension of the line joining the ears than if it is at the same distance directly in front of the head. This effect has significant bearing on the testing of microphones.

When the obstacle diffracting the sound wave is relatively small, *e.g.*, a fog droplet, the wave is said to be scattered, since the bending makes the sound turn in all directions from the original one. Examples of this are presented later, in section II.14.

5. Harmonic Sound Waves; Pitch, Frequency and Wave Length.—In the preceding discussion of simple acoustical phenomena, little has been said about one striking difference between sounds from different sources. This is the difference in pitch. Instinctively when we hear two successive sounds we classify one as higher or lower than the other. This characteristic is obviously quite different from loudness, previously described in terms of the intensity of the sound. What physical effect is associated with pitch? Here it is necessary to invoke actual experiments. For rough qualitative considerations it will suffice to call attention to the ordinary air siren, which produces sound by interrupting jets of air at different rates by means of a rotating disk with orifices in its periphery (section III.11 below). The more rapidly the interruption takes place, *i.e.*, the greater the number of separate puffs of air per unit time, the higher the pitch. But a source of sound in which the disturbance repeats itself regularly like the puffs of air in the siren is said to produce a harmonic sound wave characterized by a definite frequency. This is defined as the number of times per second the disturbance in the sound wave at any point repeats itself.

Actually the relation between pitch and frequency in a complicated sound wave like that associated with the human voice or ordinary musical instruments is not at all simple. Hence it is advantageous at this point to confine attention to a pure harmonic wave like that emitted by a carefully made tuning fork. This is said to produce a pure tone of definite single frequency.

In a harmonic sound wave in air, then, the "squeeze" responsible for the sound varies periodically at every point. The time for one complete cycle of pressure change at any point is called the period of the wave. The period in seconds is equal to the reciprocal of the frequency, where the latter is represented by the number of complete cycles per second.

One of the most important facts about harmonic sound waves is that we do not hear sounds of all frequencies equally well. If the hand is waved back and forth periodically with a frequency less than 15 cycles per second no sound is heard. An object vibrating above this frequency will produce an audible sensation if the intensity is sufficiently great. However, when the frequency is increased to above 20,000 cycles per second, audibility vanishes for most people. In fact, with advancing age the frequency threshold for hearing diminishes decidedly. Of course the sound intensity plays a role here and the matter is not simple. Nevertheless the significant thing is that it is possible to have sounds of such high frequency that they are inaudible no matter how great their intensity. These sounds are called ultrasonic and they play an important role in the modern study of acoustics. (*See* ULTRASONICS.) As far as the physics of sound is concerned, the sounds which are not heard attract greater interest than those which are. One reason is that ultrasonic waves tend to travel in beams like light, while low-frequency sound waves tend to spread in every direction from the source. Obviously such sounds need special devices for their production and detection. Much of the progress of acoustics after about 1925 was due to the development of such equipment, which made possible extensive investigations with sounds of frequencies up to 100,000,000 cycles per second (*see* section V.6 below).

To come back to harmonic waves in general, they have another useful characteristic—the wave length, which is the distance between successive points in a spreading wave at which the disturbance is exactly the same and doing the same thing (*i.e.*, getting either larger or smaller). At two such points the phase of the wave is said to be effectively the same (section II.4 below). Since the disturbance in a wave travels a distance of one wave length in one period, the product of wave length and frequency equals the wave velocity. This important relation is true for waves of all kinds. It means that high-frequency acoustic waves in a given medium have smaller wave length than low-frequency sounds. Thus if we are talking about water, a wave of frequency 1,000 cycles (1 kc.) has a wave length of approximately 1.5 m., whereas for a frequency of 1,000,000 cycles (1 mc.) the corresponding wave length is only 1.5 mm.

Many of the properties of sound discussed above are definitely dependent in their behaviour and magnitude on the frequency. Very few sounds arising in practice are pure tones characterized by a single frequency. Most actual sounds insofar as they are periodic can be considered as more or less complicated combinations of component harmonic sounds, each of definite frequency. One or more frequencies will often predominate and help give the sound its observed quality. In general a high-frequency sound wave is bent by diffraction less than a low-frequency wave is, and hence enables an obstacle to cast a sharper acoustic shadow. (There is a corresponding phenomenon in the case of visible light, where the frequency is very much higher than that of any sound wave yet produced.) On the other hand when sound is scattered by objects whose dimensions are small compared with the wave length, the high-frequency components are scattered much more effectively than the low. This is quite analogous to the scattering of sunlight by the small dust and molecular particles of the atmosphere; the short wave length or high-frequency light (*i.e.*, the blue) is scattered much more than the long wave length (red) and so the sky appears blue.

6. Wave Fronts and Rays.—We have already commented on

the fact that an audible sound wave tends to spread out in all directions from the source. On the other hand it is possible to produce a beam of sound if the frequency is high enough. To describe this situation more precisely the concept of wave front is introduced. This is a surface of such a kind that at a particular instant the disturbance characterizing the sound wave is the same at every point.

An example is the airplane propeller radiating sound into the surrounding air. At any instant of time the state of "squeeze" of the air in this sound wave will be the same on the surface of a sphere with centre at the propeller and with radius equal to the time taken by the squeeze to travel at the speed of sound from the propeller to the sphere. We shall call this sphere a wave front of the sound from the propeller and can think of the propagation of the sound as the motion of such a wave front carrying with it a definite state of squeeze, and moving with the speed of sound. The expanding circular ripples on water produced by a dropped stone form an excellent two-dimensional analogy; in the present example, however, the wave fronts are spherical. On the other hand it is possible to confine a sound wave to a tube of constant cross section, as for example a speaking tube. Though the propagation of the disturbance here is indeed a somewhat complicated affair because of the reflections from the walls of the tube, it is approximately correct to assume that at any instant at any place the squeeze is the same over the plane at that place perpendicular to the direction of propagation. In other words, the transmission proceeds by the motion of a *plane* wave front.

Plane and spherical wave fronts are the most important types encountered in sound propagation. Of the two, plane waves are the more important for at great distances from the source the effective part of a spherical wave is very nearly plane.

Lines perpendicular to a wave front are known as rays. These indicate the direction in which the wave front and hence the sound is moving at any place at any instant. For many purposes it is more convenient to describe sound transmission in terms of rays just as in the analogous case of light. However, it must be remembered that the wave length of even ultrasound is long compared with that of visible light and consequently ray acoustics is in general by no means as accurate a representation as ray optics (*see* section II.5 below). It is more difficult to produce a beam of sound than a beam of light, but it can be done if the wave length is sufficiently small compared with the size of the sound source. The mechanism by which this takes place is called interference and is explained in section III.10 below. Such sound beams are very useful in underwater detection.

II. ANALYSIS OF SOUND PROPAGATION

1. Wave Functions.—It is now necessary to investigate more carefully the qualitative considerations of the foregoing and introduce some quantitative description. The mathematical treatment of sound waves in general is very complicated, so we start by idealizing the problem. It is simplest to begin with the kind of wave one can produce in a long, perfectly flexible string by flicking one end. The kink so produced as a disturbance in the original straight position of the string travels along it in an easily visualized fashion. (*See* fig. 1.)

Here we have a wave traveling in one direction only. Since the disturbance is at right angles to the direction of propagation it is called a transverse wave. In this respect it is different from a sound wave. In the latter the disturbance, being a pressure squeeze, is in the same direction as the wave transmission and the wave is called longitudinal. Each particle of the fluid in this case is displaced from its equilibrium position in the direction of the wave propagation. However, the mathematical analysis in its more fundamental aspects is the same for both kinds of waves.

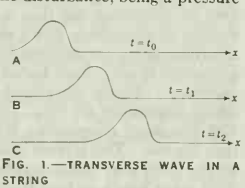


FIG. 1.—TRANSVERSE WAVE IN A STRING

If we were to take a snapshot picture of the kinked string at a given instant, say $t = t_0$, it might look like A in fig. 1. Here the disturbance, which is measured by the distance each part of the string has been displaced from its original undisturbed position, can be represented by a mathematical function $f(x)$, if we denote distance along the string by x. At the later time, $t = t_1$, the disturbance has moved a distance $V(t_1 - t_0)$ along the string, where V is the velocity of the wave, so the same function $f(x)$ reappears, but displaced (B in fig. 1). And so likewise for $t = t_2$, as indicated in C. The problem then is to represent mathematically a function of x which moves along the positive x axis with velocity V. It turns out that this is

$$\xi = f(x - Vt) \tag{II.1-1}$$

The Greek letter ξ is used to denote the displacement from equilibrium representing the disturbance. Moreover, if the wave travels in the negative x direction the corresponding wave function, as it is called, is

$$\xi = g(x + Vt) \tag{II.1-2}$$

It will be noted that there is nothing essentially periodic about the wave functions (1)[1] and (2). However, as was stressed in section I.5, periodic waves are the most interesting kind, and hence we proceed at once to specialize the above to waves characterized by frequency and wave length. The simplest of all periodic waves is the harmonic, in which (1) takes the form

$$\xi = A \sin(\omega t - kx) = -A \sin k(x - \omega/k \cdot t) \tag{II.1-3}$$

This is evidently in the proper form for a wave function if we interpret ω/k to be equal to the velocity V. But what are ω and k? If we fix our attention on a particular point $x = x_0$ of the string traversed by such a wave and take a motion picture of the way the displacement there varies with time we find it is sinusoidal with period $P = 2\pi/\omega$ or frequency $\nu = \omega/2\pi$. Hence this fixes the frequency of the wave. On the other hand, if we take a snapshot picture of the string at the instant $t = t_0$, we find that the displacement varies sinusoidally with x and the distance between successive points of maximum displacement in the same direction is $2\pi/k$. But we have already defined this as the wave length λ of the wave. Hence we have

$$\nu = \omega/2\pi, \lambda = 2\pi/k \tag{II.1-4}$$

Since $\omega/k = V$, we also obtain the important relation already commented on in section I.4 above

$$\nu\lambda = V \tag{II.1-5}$$

This relation holds for any harmonic wave. It is worth noting that if we used the cosine function in place of the sine function in (3) the physical meaning would not be altered. The quantity ω is often referred to as the angular frequency of the wave.

2. Differential Equation of Wave Motion and Its Deduction.—If we differentiate ξ in equation (II.1-3) twice with respect to x (keeping t constant), and then twice with respect to t (keeping x constant) we can show, keeping in mind equation (II.1-5), that

$$\frac{\partial^2 \xi}{\partial t^2} = V^2 \frac{\partial^2 \xi}{\partial x^2} \tag{II.2-1}$$

As a matter of fact the existence of this equation also follows from the general wave functions (II.1-1) and (II.1-2). It is called the general differential equation for wave motion in the x direction. Its general solution is the sum of wave functions (II.1-1) and (II.1-2) representing general waves progressing in both positive and negative x directions. The advantage of a partial differential equation like (1) from a physical point of view is that it contains the description of so much in so compact a form. The task of theoretical physics in every branch may be said to be the development of such equations, which by their very generality include within themselves a vast amount of information about physical phenomena. The whole theory of the propagation of sound may be considered to be implied in the above equation and its generalization to three dimensions (in which any point in

[1]For simplicity references to equations in the same section are given in terms of the last digit in the equation number. Reference to an equation in a different section employs the whole equation number.

space has the co-ordinates x, y, z), viz.,

$$\frac{\partial^2\xi}{\partial t^2} = V^2\left(\frac{\partial^2\xi}{\partial x^2} + \frac{\partial^2\xi}{\partial y^2} + \frac{\partial^2\xi}{\partial z^2}\right) \qquad \text{(II.2-2)}$$

The physicist therefore feels that he understands something about the propagation of sound waves if he can show that an equation like (1) or (2) is a logical consequence of the behaviour of a compressible fluid when it is squeezed.

The behaviour of an ideal fluid when disturbed from its state of rest is determined (*see* MECHANICS, FLUID) by three fundamental equations: (1) equation of continuity; (2) equation of motion; and (3) equation of state. The first says that no matter how the fluid moves, as long as it hangs together as a continuous fluid there must be conservation of mass. The second, or equation of motion of the ideal fluid, says that the total time rate of change of velocity of a particle of fluid must equal the force per unit mass acting on the particle, the latter for purely compressional disturbances being proportional to the negative of the rate of change of the excess pressure with distance. Finally the third, or equation of state, connects the change in pressure with the change in density. This says that the excess pressure p_e produced by any disturbance in the fluid (*i.e.*, the difference between the actual pressure and the equilibrium pressure prevailing before the disturbance) is directly proportional to the excess density ρ_e (*i.e.*, the difference between the actual density in the disturbed fluid and the equilibrium density). The coefficient of proportionality is usually written as c^2. It should be noted that this is an idealized static equation of state and takes no account of the fact that in an actual fluid the imposition of a given excess pressure needs time to produce the excess density indicated.

It is an interesting fact that the mathematical combination of the three equations just described leads to an equation of precisely the same mathematical form as (1). This should not be too surprising when it is recalled that when a fluid is disturbed mechanically the pressure is bound to change; and this leads to motion of the fluid as well as a squeezing or rarefying action (viz., change in density). This must take place in such a way that no mass is lost (continuity), and if the fluid is truly elastic the squeeze will not stay put but will inevitably move through the fluid as a wave. The result of the combination can be expressed most conveniently in terms of the excess pressure p_e and then takes the form

$$\partial^2 p_e/\partial t^2 = c^2\partial^2 p_e/\partial x^2 \qquad \text{(II.2-3)}$$

which is of precisely the same mathematical form as equation (1) with p_e in place of ξ and c in place of V. Physically it therefore means that the acoustical disturbance represented by the excess pressure is propagated as a wave along the x axis with velocity c. Actually it is not difficult to show that the displacement ξ and the excess density ρ_e are propagated in the same way.

It must indeed be emphasized that in the mathematical deduction leading to (3) it has been assumed that the disturbance is not too extreme. Specifically this means that the ratio of the excess density to the equilibrium density (usually called the condensation) is very small compared with unity.

This also means that the gradient of the particle displacement in the fluid (*i.e.*, $\partial\xi/\partial x$) is much smaller than unity in magnitude. Moreover the flow or particle velocity associated with the disturbance must be very small indeed compared with c (the wave velocity). These conditions prevail for all acoustic phenomena associated with speech, sound reproduction and even for most acoustic transmission, as in underwater sound signaling, etc. However, they no longer hold for sounds produced by explosions, or by the rush of gas in the jet of a jet plane. The same is true for sound which can cause cavitation (*i.e.*, the appearance of bubbles) in liquids (*see* section V.6 below). High-intensity sound, to which may be attached the designation macrosonics, is of importance in the use of sound for processing. Under these conditions the acoustic wave equation no longer has the forms (1) or (3). In place of (1) we now must write

$$\partial^2\xi/\partial t^2 = \frac{V^2\partial^2\xi/\partial x^2}{(1+\partial\xi/\partial x)^{\gamma+1}} \qquad \text{(II.2-4)}$$

in which the quantity γ in the exponent, if the medium is a gas, is the ratio C_p/C_v, where C_v = specific heat of gas at constant volume and C_p = specific heat of the gas at constant pressure. The equation can also hold for solid and liquid media, though here γ has a different meaning; in the simplest cases it has the value unity, but it can have higher values. The solution of (4) is by no means so simple as that of (1), to which it reduces for $\partial\xi/\partial x < < 1$. For $\gamma = 1$, for example, the solution can be put into the form

$$\partial\xi/\partial t = f[x + \xi - (V + \partial\xi/\partial t)\, t] \qquad \text{(II.2-5)}$$

where f is an arbitrary function. The physical meaning of this rather complicated expression is that the various parts of the wave hump or profile (*see* fig. 1) do not move forward with the same velocity. Rather, the top of the hump moves faster than the bottom and hence a disturbance which starts out with a symmetrical profile as in fig. 2(A) will develop an asymmetric form as in fig. 2(B) in which the front of the profile will be steeper than the rear. Eventually the front will become vertical and then lean over so that its top is farther to the right than its bottom. This state of affairs, however, cannot be of physical significance—in any actual case the wave "breaks," as in water waves at the seashore. The only way a macrosonic wave can maintain its profile or wave form is through some kind of damping mechanism, as, for example, viscosity (*see* section II.18).

FIG. 2

3. Velocity of Sound Waves in Fluids.—As was pointed out in section II.2, the velocity of sound in an ideal fluid is given by c in the expression

$$c = \sqrt{p_e/\rho_e} \qquad \text{(II.3-1)}$$

The evaluation of the radical for a liquid follows from the definition of the bulk modulus

$$B = \frac{p_e}{\rho_e/\rho} \qquad \text{(II.3-2)}$$

whence

$$c = \sqrt{B/\rho} = \sqrt{1/\rho K} \qquad \text{(II.3-3)}$$

where K is the compressibility. Since liquids have a relatively small compressibility and it changes but little with the density, it is in general adequate to insert ρ_0 (the mean density) for ρ in (3).

For gases the situation is rather different. Thus for an ideal gas the equation of state takes the form

$$p/\rho = RT \qquad \text{(II.3-4)}$$

in which R is the so-called gas constant per unit mass and T is the absolute temperature. We therefore obtain

$$p_e/\rho_e = RT = p/\rho \qquad \text{(II.3-5)}$$

if the temperature remains constant. Hence if the compressional disturbance in the gas takes place isothermally, the velocity has the form

$$c = \sqrt{p/\rho} \qquad \text{(II.3-6)}$$

This formula was first derived by Newton. It gives, however, a value definitely smaller than the measured one for the nearly ideal gases like oxygen and nitrogen. On the other hand, if we assume that the disturbance constituting the sound takes place adiabatically, we must supplement (4) by

$$p/\rho^\gamma = p_0/\rho_0^\gamma \qquad \text{(II.3-7)}$$

in which $\gamma = C_p/C_v$ is the ratio of the specific heat of the gas at constant pressure to that at constant volume. It follows then that

$$c = \sqrt{\gamma\, p/\rho} \qquad \text{(II.3-8)}$$

This result agrees so well with experiment that the measurement of sound velocity became a standard scheme for evaluating γ, and

it is generally assumed that sound disturbances in an ideal fluid take place adiabatically. It must not be assumed offhand that the same is true of a viscous fluid, for example, or one displaying conductivity of heat. For liquids (equation [3]) we must take the adiabatic compressibility in computing c, but the difference between adiabatic and isothermal compressibilities here is generally small.

For an ideal gas it follows from (4) and (7) that the velocity of sound is independent of the pressure. On the other hand c does depend on the temperature. In fact

$$c = \sqrt{\gamma RT} \qquad \text{(II.3-9)}$$

If the velocity at 0° C. is c_0, that at temperature t° C. is clearly

$$c_t = c_0\sqrt{1 + t/273} \qquad \text{(II.3-10)}$$

For dry air under standard conditions $c_0 = 331.3$ m. per second. This agrees rather well with actual measurements made in the open air, but it must be remembered that the latter are extremely difficult to perform precisely since atmospheric conditions are rarely uniform, and sound transmission in air from the standpoint of precision is extremely complicated because of nonhomogeneity and wind disturbances.

For real gases (9) must be modified, since the equation of state is no longer in the form (4). It is now more convenient to revert to equation (3). Realizing that the adiabatic bulk modulus must be used for B, we write $B_{\text{ad}} = \gamma B_{\text{iso}}$ and by definition

$$B_{\text{iso}} = \rho(\partial p/\partial \rho)_T \qquad \text{(II.3-11)}$$

The expression for the velocity c then becomes

$$c = \sqrt{\gamma(\partial p/\partial \rho)_T} \qquad \text{(II.3-12)}$$

For the ideal gas (4), this formula yields (8). One useful form of the equation of state of a real gas is

$$p/\rho = RT(1 + \beta\rho + C\rho^2 + \ldots) \qquad \text{(II.3-13)}$$

The quantities β and C are known as the second and third virial coefficients per gram of gas respectively and are functions of the temperature. Substituting into (12) yields

$$c = \sqrt{\gamma RT(1 + 2\beta\rho + 3C\rho^2 + \ldots)} \qquad \text{(II.3-14)}$$

In this form it is clear that the velocity of sound depends on the pressure as well as the temperature. The effect in air is small except at very low temperatures. Thus for air at 90.1° K. the velocity at 1 atm. pressure is about 188 m. per second (A. van Itterbeek and W. van Doninck, *Annales de Physique*, 19:88 [1944]), whereas at 0.2 atm. it is 191 m. per second. (Note that at this temperature β is negative.)

The experimental determination of the velocity of sound can be used as a method of measuring γ as well as the virial coefficient β. It is also possible to calculate theoretically the velocity of sound in a mixture of two gases and then use the formula to determine the relative concentrations of the components by a sound velocity measurement. (H. B. Dixon and G. Greenwood, *Proc. Roy. Soc. Lon.* (A), 109:561 [1925].)

As might be expected, the velocity of sound is much more sharply dependent on pressure in the neighbourhood of the critical point. Thus in carbon dioxide at 31° C., for example, the value of c drops from about 260 m. per second at 20 atm. pressure to a minimum of 150 m. per second at 71 atm. and rises again steeply thereafter to 330 m. per second at 90 atm.

The formulas so far presented involve no dependence of the velocity of sound on the frequency. This corresponds to experience as far as the so-called permanent gases are concerned at ordinary pressure and low frequencies; they show no dispersion. However as the frequency-pressure ratio increases all gases ultimately exhibit dispersion, and indeed the velocity is found to increase with increasing ν/p. Thus for hydrogen, c increases about 8% above its low-frequency standard-pressure value of 1,284 m. per second as ν/p is raised from 1 mc. per atmosphere to 30 mc. per atmosphere. The reason for this will be discussed in section II.13.

The velocity of sound in gases also depends on the purity of the gas. Small traces of foreign gases can produce significant changes in c. Thus for pure carbon dioxide (with only a trace of neon and argon) the velocity for $\nu/p = 3.2 \times 10^4$ cycles per atmosphere corresponds to a value of γ in (8) of approximately 1.34. With the addition of 2.8% water vapour, however, this same value of γ is associated with $\nu/p = 1.6 \times 10^6$ cycles per atmosphere. The whole dispersion curve is thus displaced. Alternatively put, the sound velocity at a given ν/p value is much decreased by the addition of the water vapour. The velocity of sound is of course affected by the motion of the medium. Thus in the open air with the wind blowing the velocity of sound with respect to the ground will be greater with the wind than against it.

Coming now to the velocity of sound in liquids we can still use equation (3) but must necessarily employ the adiabatic bulk modulus or compressibility. The change in compressibility of a liquid with temperature is much more complicated than that of a gas and it is practically impossible to develop helpful general formulas. The case of water has been carefully studied and the following equation (G. W. Willard, *J. Acous. Soc. Amer.*, 19:235 [1947]) represents rather accurately the dependence of sound velocity in metres per second on temperature in degrees centigrade

$$c_t = 1{,}557 - 0.0245(74 - t)^2 \qquad \text{(II.3-15)}$$

This indicates that c_t increases with temperature up to a maximum value at $t = 74$° C., and thereafter decreases. For practically all other liquids which have been studied c decreases with temperature over the whole range in which the material can stay liquid. Salt solutions obey an equation like (15) with different values of the temperature for maximum c_t. The case of sea water is of particular interest because of the importance of underwater sound. Here matters are complicated by the fact that changes in salinity and pressure (with depth in the ocean) as well as temperature are effective in changing velocity. The following empirical formula (L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, p. 435 [1950]) holds fairly well over the range of variables usually encountered

$$c_t = 1.41 \times 10^6 + 4.21 \times 10^2 t - 3.7t^2 + 1.1 \times 10^2 S + 1.8 \times 10^{-2} d \qquad \text{(II.3-16)}$$

where t is in degrees centigrade, S is the salinity in parts per 1,000 and d is the depth in centimetres. The velocity is given in centimetres per second.

As is suggested by (16), sound velocity in solutions is markedly dependent on the concentration of the solute. For NaCl, for example, the velocity in a 10% solution at room temperature is 1,600 m. per second while for a 20% solution it rises to 1,720 m. per second. Not all solutions show a rise over the velocity in the pure solvent, however, and the problem is complicated.

As (16) also indicates, increase in pressure on a liquid raises the sound velocity more or less linearly. Thus for benzene P. Biquard found that the velocity is increased by about 17% in going from atmospheric pressure to 500 atm.

Unlike the situation in gases, no definite indication of sound dispersion in liquids has been found experimentally, though theoretical considerations suggest that it should exist for sufficiently high frequencies, probably of the order of several hundred megacycles.

4. Mathematical Representation of Wave Fronts.—In section I.6 it was seen that to describe adequately the propagation of acoustic waves it is necessary to introduce the concept of wave front as a surface at every point of which at a given instant the propagated disturbance is the same and doing the same thing (*i.e.*, either increasing or decreasing). We now give this analytical precision, confining our attention to harmonic waves. Previously (equation II.1–3) we dealt with harmonic wave functions of the form

$$\xi = A \sin(\omega t - kx)$$

corresponding to propagation along the positive x axis only. But we have already pointed out that it is the general tendency for sound waves in free space to spread in all directions and hence we must generalize our analysis. To simplify matters, when we consider wave propagation in three-dimensional space, we shall use

the excess pressure as the measure of the acoustical disturbance. Consider the function

$$p_e = A(x, y, z) \sin \omega[t - \psi(x, y, z)/c_0] \qquad \text{(II.4-1)}$$

in which A is the amplitude of the pressure wave and may be a function of position in space. The function $\psi\ (x, y, z)$ is assumed to have the dimensions of distance only, and c_0 is a constant having the dimensions of velocity.

It turns out by careful examination that (1) is a genuine harmonic wave function (*i.e.*, a solution of the general three-dimensional wave equation of the form of [II.2-2]) under certain conditions. The phase of the wave, viz., $\omega[t - \psi(x, y, z)/c_0]$, has the constant value ωt_0 at the time t over the surface with the equation

$$\psi(x, y, z) = c_0(t - t_0) \qquad \text{(II.4-2)}$$

Hence this is the general equation for the wave front at time t. In the passage of time the wave front equation will alter just as if the wave front itself were moving through space, carrying with it the same value of the disturbance.

If

$$\psi(x, y, z) = \alpha x + \beta y + \gamma z \qquad \text{(II.4-3)}$$

equation (2) represents a plane wave front with the direction cosines of the normal to the front equal to α, β, γ. In this case $A(x, y, z)$ reduces to a constant.

If

$$\psi(x, y, z) = \sqrt{x^2 + y^2 + z^2} \qquad \text{(II.4-4)}$$

equation (2) represents a spherical wave front. It develops that (1) is then a harmonic pressure wave spreading out from the origin if $A(x, y, z) = A_0/r$, where

$$r = \sqrt{x^2 + y^2 + z^2} \qquad \text{(II.4-5)}$$

and A_0 is constant. Plane and spherical waves are the two most important types for sound propagation. Both are obviously somewhat idealized; a spherical wave front demands a point source and there is no such thing as a precise point source of sound. A plane wave front demands an infinitely extended plane source, which likewise does not exist. Nevertheless actual wave fronts often approximate these ideal types sufficiently well to justify their use. In general the smaller the dimensions of a source compared with the wave length of the sound the more nearly the wave front approximates the spherical or hemispherical shape. The larger the dimensions of a plane source compared with the wave length, the more nearly is a plane wave front realized.

5. Sound Rays; Refraction.—For many purposes the propagation of light can be effectively studied in terms of rays—curves which are always perpendicular to the family of wave fronts associated with a particular propagation. The ray concept is also useful in acoustics but only if the wave length is sufficiently small or, more accurately, if the change in wave length (due to change in velocity from one point to another in the medium) over a distance of the order of one wave length is very small compared with the wave length itself.

Moreover, the ray must not bend too rapidly as the sound progresses. Finally, the change in amplitude over a like distance must be small compared with the initial amplitude. If these conditions are satisfied it is meaningful to write the differential equations of the ray paths in any medium in which the wave velocity c varies from point to point. Actually, in analogy with optics it is customary to define the ratio c_0/c as the index of refraction n of the medium. The physical meaning of c_0 is the nonvarying velocity in some homogeneous medium bounding the given nonhomogeneous medium. Let the direction cosines of the ray path be α, β, γ. These are in general functions of position, *i.e.*, of the co-ordinates x, y, z. If ds is the element of distance along the ray, the differential equations of the ray paths take the form

$$d(\alpha n)/ds = \partial n/\partial x, \quad d(\beta n)/ds = \partial n/\partial y, \quad d(\gamma n)/ds = \partial n/\partial z \qquad \text{(II.5-1)}$$

As soon as n is known as a function of x, y, z, equations (1) can be integrated to give the rays. As an illustration, suppose $n =$ constant, so that there is no variation of sound velocity in the medium. Equations (1) then lead to

$$\alpha = K_1, \beta = K_2, \gamma = K_3 \qquad \text{(II.5-2)}$$

where K_1, K_2, K_3 are constants, giving then the direction cosines of a straight line. The rays are straight lines, irrespective of the character of the wave fronts (*e.g.*, the latter may be plane or spherical).

One typical case of ray propagation occurs when the medium is stratified; *i.e.*, the change in velocity takes place in one direction only. Let us take this as the z axis. It will be sufficient to consider the ray path in the xz plane. In fig. 3 let AB be the element ds of the ray and let the tangents at A and B, respectively, make the angles Θ and $\Theta + d\Theta$ with the x axis. We now have

$$\partial n/\partial x = \partial n/\partial y = 0,$$
$$\alpha = \cos \Theta, \beta = 0, \gamma = \sin \Theta$$

Equations (1) now yield

$$d(n \cos \Theta)/ds = 0, \quad d(n \sin \Theta)/ds = dn/dz \qquad \text{(II.5-3)}$$

The first equation says that along the ray

$$n \cos \Theta = \text{constant} = n_1 \cos \Theta_1 \qquad \text{(II.5-4)}$$

where Θ_1 is the direction of the ray at the place where the index is n_1. If we write $\Phi = \pi/2 - \Theta$, where Φ is the angle the ray makes with z, (4) becomes

$$\sin \Phi/\sin \Phi_1 = n_1/n = c/c_1 \qquad \text{(II.5-5)}$$

This has the familiar form of Snell's law for the refraction of light rays. Here, however, it is of much more general form since the variation of velocity is continuous and does not take place abruptly at a boundary plane as in the usual elementary optical case. The abrupt change involves a discontinuity in n and therefore demands special treatment, discussed in section II.7. It is interesting to observe that the second equation in (3) makes possible the evaluation of the curvature $(d\Theta/ds)$ of the ray at any point and hence ultimately the equation of the ray path. By straightforward analysis (3) leads to

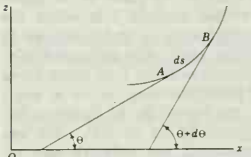


FIG. 3

$$d\Theta/ds = -\,dc/dz \;.\; \cos \Theta_1/c_1 \qquad \text{(II.5-6)}$$

As an illustration, suppose

$$c = c_1 + az \qquad \text{(II.5-7)}$$

i.e., the velocity is a linear function of the depth z (a is a constant), with c_1 = velocity at $z = 0$, which might, for example, be the surface of the sea. Then (6) says that the curvature is constant. This means that the path must be a circle with radius $= |c_1/a \cos \Theta_1|$. If a is positive, the velocity increases with depth and the curvature is negative and the path bends upward toward the surface $z = 0$. If a is negative the velocity decreases with depth (a more common situation in the ocean) and the circular path bends downward. These results can readily be followed graphically.

6. Refraction in a Moving Medium.—The progagation of sound through a moving medium (*e.g.*, the atmosphere with the wind blowing) is extremely complicated and its analysis will not be given here. However, the law of refraction for a stratified medium turns out to be a rather simple modification of Snell's law (II.5–5). If the velocity of the medium is confined to the x direction only and is equal to u, which is a function of z (*e.g.*, height above the ground, taken as the xy plane, in the case of transmission through the atmosphere), the law of refraction becomes

$$c/\cos \Theta - c_1/\cos \Theta_1 = u_1 - u \qquad \text{(II.6-1)}$$

Here u_1 is the medium velocity at the point where the sound velocity is c_1 and the ray direction is given by Θ_1. Equation (1) reduces to (II.5–5) for $u = u_1 = 0$. It is clear from (1) and

also from relatively simple graphical constructions that the effect of the motion of the medium is to bend the wave front and ray in the direction in which the motion takes place. Thus if the wind velocity increases upward from the ground, sound wave fronts are lifted to windward and depressed to leeward. Other things being equal, this decreases the range to windward and increases it to leeward. These considerations were once of considerable importance in the acoustical detection of aircraft, but were superseded by radar. They remain of great importance, however, in sonic gun ranging (*see* section V.5 below).

7. Huygens' Principle; Reflection and Refraction at an Interface.—The propagation of an acoustical wave front through a medium can also be studied by means of a principle developed by Christiaan Huygens. According to this principle, in its elementary form, each point on a wave front (like F_1 in fig. 4, for example) may be assumed to be the source of a hemispherical wavelet which moves outward from F_1 in the direction of propagation; the new wave front after a short time is the mathematical envelope of all these wavelets; *e.g.*, F_2 in the figure. This principle can be demonstrated by considering the solution of the general wave equation. A simple illustration of the principle is the establishment of the laws of reflection and refraction at a plane interface separating fluid media with different properties; *e.g.*, different mean densities and sound velocities.

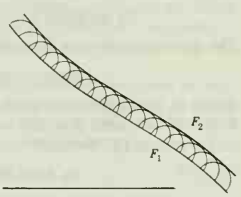


FIG. 4

Let AB in fig. 5 represent the trace in the plane of the paper of the plane interface between a fluid medium I in which the velocity of sound is c_1 and the mean density ρ_1, and a medium II in which the corresponding quantities are c_2 and ρ_2 respectively. Let CD denote the trace of a plane wave front (the latter being perpendicular to the plane of the paper) incident on AB at angle Θ_i. This means that the incident ray travels in the direction HC perpendicular to CD. We can construct the reflected and refracted wave fronts by means of Huygens' principle. At the instant the disturbance at one end of the wave front has reached C on the interface, the disturbance at the other end is still at D in medium I. In the time $t_1 = DE/c_1$ which it takes the disturbance at D to reach E on the interface, the disturbance will travel out in all directions from C. In particular in I it will traverse a distance equal to DE.

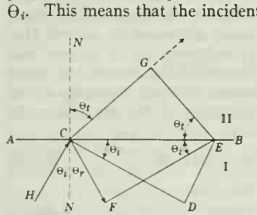


FIG. 5

If we draw with C as centre a semicircle with radius DE and construct the straight line from E tangent to this we obtain FE which from geometry makes the angle $\Theta_r = \Theta_i$ with AB. That this is the reflected wave front can be confirmed by making a Huygens construction for other points on the incident wave front. We see that the reflected wave front makes the same angle with the interface as the incident wave front. This, combined with the readily established additional fact that both wave fronts are perpendicular to the same plane (namely the plane of the paper) constitutes the law of reflection. It can of course be equally well expressed in terms of the incident and reflected rays, and is the same law that holds for light wave fronts and rays.

Precisely the same kind of analysis shows that after transmission across the interface the incident wave front CD is changed to the refracted wave front GE making the angle Θ_t with AB, such that

$$\sin \Theta_t / \sin \Theta_i = c_2/c_1 = n_{21} \qquad \text{(II.7-1)}$$

This is Snell's law again, with n_{21} the index of refraction of medium II with respect to medium I. If the media are in motion with respect to each other, this result must be modified (*cf.* equation II.6–1). The above formula accounts well for the passage of sound from air to water or vice versa.

8. Energy Propagation in Sound Waves; Intensity.—In section I.2 the intensity of sound was discussed, and its relation with energy was implied. We now look upon sound propagation as essentially a form of transmission of energy through a fluid. When a fluid is disturbed by squeezing it at any place, work is done, representing the expenditure of energy. The reappearance of the disturbance at a distant point by wave propagation corresponds to the transfer of the original energy from the point where it was put into the fluid to the distant point. The greater the amount of energy transported on the average per unit area of wave front per unit of time, the greater will be the intensity of the corresponding sound wave.

To be more precise, the average rate of flow of energy in the sound wave per unit time per unit area of wave front is simply calculated as the average of the product of the excess pressure p_e and the flow velocity $\partial\xi/\partial t$ in the medium. Analysis shows that the intensity I of a plane wave computed in this way can be expressed in the following way

$$I = p_0^2/2\rho_0 c \qquad \text{(II.8-1)}$$

This important formula (G. W. Stewart and R. B. Lindsay, *Acoustics*, pp. 62 ff. [1930]) says that the intensity of a plane wave is equal to the square of the excess pressure amplitude p_0 divided by twice the product of mean density and sound velocity. Significantly it is independent of frequency.

The standard absolute unit of intensity is the erg/sec. cm.^{2} More practical is the watt/cm.^{2} or, in the metre-kilogram-second system favoured by some engineers and physicists, the watt/metre^2. However, as noted in section I, the decibel is now used as the standard measure of sound intensity. This is a relative measure. If two sound waves have absolute intensities I_1 and I_2 respectively they are said to differ in intensity level by D decibels (or db.), where

$$D = 10 \log_{10} I_2/I_1 \qquad \text{(II.8-2)}$$

Hence if $I_2 = 2I_1$, $D = 3.01$ db., or doubling the intensity means an increase of intensity of a little more than 3 db. To use this method of measuring intensity requires the choice of a standard level. This is often taken as that corresponding to an average excess pressure of 1 dyne/cm.^{2} in air (close to the normal level for conversational speech). The minimum audible sound lies approximately 70 db. below this level, while the threshold of feeling, when hearing becomes painful, lies about 70 db. above this level. These figures are approximate and actually depend on frequency.

From equation (1) it is clear that the intensity of sound at given pressure depends on the medium and in particular on the product $\rho_0 c$. This quantity is known as the specific acoustic resistance of the medium (for a plane wave). For water, for example, it is about 3,800 times its value for air at standard pressure. Consequently the same excess pressure in water produces a much smaller intensity than in air and it might seem that it is harder to produce an intense sound in water than in air. This, however, neglects considerations connected with the source of sound, which are discussed later (section III.9). It turns out that at given frequency and pressure a solid radiating source is actually more efficient in water (and in liquids generally) than in air.

9. Intensity of Spherical Sound Waves.—Though equation (II.8–1) was cited as applying to a plane wave, a similar formula applies to a spherical wave in the corresponding medium. Here, to be sure, the pressure and particle velocity expressions are different. Whereas for a plane wave the excess pressure and flow velocity (sometimes called particle velocity) amplitudes are constant in a nonabsorbing medium, in the case of a spherical wave they both fall off with the distance from the source of the radiation (section II.4 above). The analytical expressions are complicated and will not be reproduced here. When the multiplication and averaging processes are carried out as in section II.8 above, we obtain for the intensity in the spherical wave at distance r from the source

$$I = p_0^2/2\rho_0 c r^2 \qquad \text{(II.9-1)}$$

Both (II.8–1) and (II.9–1) can clearly be put in the same form if

we write

$$I = \overline{p_e^2}/\rho c \qquad \text{(II.9-2)}$$

where $\overline{p_e^2}$ is the average of the square of the excess pressure at the point where the intensity is computed.

10. Flow of Acoustic Energy Across a Boundary.—We have already noted in section I that when a sound wave strikes the boundary separating one medium from a second with different physical properties, *e.g.*, going from air to water or vice versa, some of the sound is reflected while the rest is transmitted (and in general refracted). This means that the transmission of acoustic energy is affected by the presence of the boundary, a common phenomenon which deserves consideration.

In fig. 6, OA denotes the plane boundary separating the two media I and II, the x axis being perpendicular to the boundary. In medium I the mean equilibrium density is ρ_1 and the sound velocity c_1. The corresponding quantities in medium II are ρ_2 and c_2. The arrow i denotes a plane harmonic sound wave traveling in I from left to right in the x direction and striking the boundary at $x = 0$. Similarly r denotes the wave in I reflected from the boundary and t denotes the wave transmitted into medium II across the boundary. For simplicity we assume perpendicular incidence. Let the amplitude of the excess pressure in the incident wave be p_i, that in the reflected wave p_r and that in the transmitted wave p_t. It is clear that if we can set up two algebraic relations connecting these three quantities, we can express the reflected and transmitted pressure amplitudes in terms of the incident, and by doing the same thing for the particle velocity amplitudes, ultimately get an estimate of the relative amount of energy transmitted and reflected.

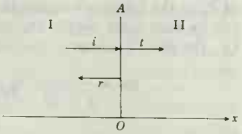


FIG. 6

The two relations just mentioned are called the boundary conditions prevailing at the interface separating the two media. They express the fact that there must be no discontinuity in the excess pressure and particle velocity at the boundary. Otherwise put, the sum of the pressure amplitudes of the incident and reflected waves at the boundary must be equal to the transmitted pressure amplitude, and similarly for the particle velocity amplitudes. The result of applying these two conditions is

$$\frac{p_r}{p_i} = \frac{1 - \rho_2 c_2/\rho_1 c_1}{1 + \rho_2 c_2/\rho_1 c_1} \qquad \text{(II.10-1)}$$

and

$$\frac{p_t}{p_i} = \frac{2}{1 + \rho_1 c_1/\rho_2 c_2} \qquad \text{(II.10-2)}$$

Use of equation (II.8–1) yields for the ratio of the intensity in the transmitted wave to that in the incident wave, which we may call the power transmission ratio P_r,

$$P_r = I_t/I_i = \frac{4\rho_2 c_2/\rho_1 c_1}{(1 + \rho_2 c_2/\rho_1 c_1)^2} \qquad \text{(II.10-3)}$$

It will be noted that P_r depends wholly on the ratio $\rho_2 c_2/\rho_1 c_1$, or the ratio of what we have above called the specific acoustic resistances of the two media. If the latter are equal, in spite of the boundary, $P_r = 1$, but for any other ratio $P_r < 1$. The specific acoustic resistance ρc therefore controls the transmission of energy at normal incidence from one medium to another. If the one medium is a gas and the second a solid, it is clear from the large difference between the ρc values for the two that P_r will be small, suggesting that in this case the reflection coefficient or ratio of reflected to incident intensity will be high. As a matter of fact the reflection coefficient R is given by

$$R = 1 - P_r = \left(\frac{1 - \rho_2 c_2/\rho_1 c_1}{1 + \rho_2 c_2/\rho_1 c_1}\right)^2 \qquad \text{(II.10-4)}$$

Equations (3) and (4) account adequately for the commonly observed fact that the transmission of sound from air to water or from air to a solid wall is relatively poor and the reflection correspondingly high. In fact for air to water transmission the value of P_r is only 0.12%.

The more general case of oblique incidence should not be overlooked. We have already established Snell's law as descriptive of the geometrical refraction associated with oblique transmission (sections II.5 and II.7 above). As might be expected the power transmission ratio now depends on the angles of incidence and refraction Θ_i and Θ_t respectively. Thus

$$P_r = \frac{4\rho_2/\rho_1 \cdot \cot \Theta_t/\cot \Theta_i}{(\rho_2/\rho_1 + \cot \Theta_t/\cot \Theta_i)^2} \qquad \text{(II.10-5)}$$

It is to be noted that P_r here is the actual ratio of transmitted power (*e.g.*, watts) to incident power. It is not the ratio of transmitted intensity (*e.g.*, watts/cm.^{2}) to incident intensity. To obtain the latter we must divide P_r by the ratio S_t/S_i, where S_i and S_t are respectively the areas of the incident and transmitted wave fronts. From fig. 5,

$$S_t/S_i = \cos \Theta_t/\cos \Theta_i \qquad \text{(II.10-6)}$$

Actually P_r is in general the more important quantity. The reflection coefficient R in this case (the wave fronts are here equal in area) can be calculated as in (4) and becomes

$$R = 1 - P_r = \left(\frac{\rho_2/\rho_1 - \cot \Theta_t/\cot \Theta_i}{\rho_2/\rho_1 + \cot \Theta_t/\cot \Theta_i}\right)^2 \qquad \text{(II.10-7)}$$

When $\Theta_i = 0$ (normal incidence), (7) reduces to (4). On the other hand when $\Theta_i = 90°$ (grazing incidence), $P_r = 0$ and there is complete reflection. Actually if $c_2 > c_1$, total reflection ensues for angles of incidence greater than the critical angle Θ_c, where

$$\Theta_c = \arcsin c_1/c_2 \qquad \text{(II.10-8)}$$

in complete analogy with geometrical optics. On the other hand, there is complete transmission and no reflection if

$$\cot \Theta_t/\cot \Theta_i = \rho_2/\rho_1 \qquad \text{(II.10-9)}$$

This takes place for angle of incidence Θ_i such that

$$\cot^2 \Theta_i = \frac{c_1^2/c_2^2 - 1}{\rho_2^2/\rho_1^2 - c_1^2/c_2^2} \qquad \text{(II.10-10)}$$

demanding

$$\rho_2/\rho_1 > c_1/c_2 > 1 \text{ or } \rho_2/\rho_1 < c_1/c_2 < 1 \qquad \text{(II.10-11)}$$

There are important applications of these formulas to sound transmission in the atmosphere between layers of air at different temperature and humidity, and to transmission in sea water containing temperature gradients (*see* section V.5 below).

11. Acoustic Impedance.—The importance of the product ρc (usually written $\rho_0 c$) in the transmission of a plane wave at normal incidence across the boundary separating two media, as shown in the preceding section, suggests more attention to this quantity and in particular to the reason for its name of specific acoustic resistance.

Inevitably this suggests an electrical analogy. In fact, it is readily verified that the ratio of the excess pressure to the flow or particle velocity in a plane progressive acoustic wave in a fluid is precisely $\rho_0 c$. Now in the case of direct currents in electricity, the ratio of the applied potential difference between two points in a conductor to the current through the conductor is called the resistance of the conductor between the two points. If we think of acoustic excess pressure as analogous to potential difference and acoustic particle velocity as analogous to electric current, the significance of acoustic resistance becomes clear. As a matter of fact, however, it has been thought wiser to be more general. Since the acoustical quantities vary with time and space (sinusoidally in the most commonly used cases) it appears more relevant to make the analogy with alternating electric current. In the latter the ratio of alternating potential difference (or electromotive force, E.M.F.) to alternating current is called the electrical impedance. Moreover $\frac{\partial \xi}{\partial t}$ appears in the guise of an acoustic current density. Hence we generalize by defining

$$p_e/\frac{\partial X}{\partial t} = Z = \text{acoustic impedance} \qquad \text{(II.11-1)}$$

where $\frac{\partial X}{\partial t} = S\frac{\partial \xi}{\partial t}$, the so-called acoustic volume current, S being the area of the acoustic wave front.

In alternating current theory the impedance Z is in general made up of a resistive component and a reactive component, the former yielding that part of the current which is in phase with the alternating electromotive force and the latter that part of the current which is out of phase with the electromotive force by 90°. For an alternating current circuit containing resistance R in series with inductance L and capacitance C, the ratio of the amplitude of the applied electromotive force to the amplitude of the resulting current is

$$Z = \sqrt{R^2 + (\omega L - 1/\omega C)^2} = \sqrt{R^2 + \chi^2} \qquad \text{(II.11-2)}$$

where $\omega (= 2\pi\nu)$ is the angular frequency of the current and χ is equivalent reactance. The current lags behind the E.M.F. by the phase angle

$$\alpha = \arctan\left(\frac{\omega L - 1/\omega C}{R}\right) = \arctan\frac{\chi}{R} \qquad \text{(II.11-3)}$$

The electrical current theory just mentioned applies to oscillations, while in the acoustical analogy being developed we are talking about waves. Nevertheless the analogy is helpful. There is no phase lag between excess pressure and volume current in a plane acoustic wave. Hence the corresponding phase angle in this case is zero. This means that the equivalent acoustic reactance vanishes and the acoustic impedance is a pure resistance. Hence the name acoustic resistance for $\rho_0 c/S$. Calling $\rho_0 c$ itself the specific acoustic resistance (denoted by Z_s) also follows electrical analogy, since the specific electrical resistance of a conductor is the resistance *per* unit length and *for* unit area of cross section.

Though the specific acoustic impedance of a plane wave is a pure specific resistance, this is not true of a spherical wave, which has both reactive and resistive components in its impedance. The physical meaning of this is, of course, connected with the spreading of the wave in all directions. The reactance of a diverging spherical wave is positive and behaves like an electric inductance, whereas the reactance of a spherical wave converging to a point is negative and thus behaves like an electric capacitance.

The impedance notation proves of great importance in acoustics, since by its use acoustic transmission calculations can be replaced to advantage by analogous electric transmission computations. Moreover, many acoustical effects can be readily estimated in terms of known values of acoustic impedance at strategic points. For example, the effectiveness of a horn (*i.e.*, a tube of variable cross section) both as a receiver and a transmitter of sound is measured (for given volume current) by the value of the acoustic resistance at the throat. The effective acoustic transmission through a tube in which a branch line is inserted at some point can be shown to depend in a relatively simple manner on the acoustic impedance at the branch point. Problems in architectural acoustics involving the reflection of sound from the walls of a room can often be simplified by the use of the impedance notation.

12. Selective Acoustic Transmission; Filtration.—The transmission of acoustic energy in a harmonic wave across the interface separating two media is independent of frequency as equations (II.10–3) and (II.10–5) indicate. Here both media have been considered semi-infinite in extent (medium I infinite to the left of the interface and medium II infinite to the right). However, a different situation arises if one of the media is finite in extent in the direction of wave travel.

In fig. 7 let medium II, of length l, be inserted between media I and III, which are semi-infinite as in the earlier illustration. The mean densities in the three media are ρ_1, ρ_2, ρ_3 respectively and the corresponding sound velocities c_1, c_2, c_3. The analysis of section II.10 above may then be applied, with special care paid to the fact that because of the reflections at interfaces 1 and 2 there will be waves traveling in both positive and negative directions in medium II, even though the radiation comes from the left in I and goes away to the right in III. Employing the appropriate boundary conditions as before, we get for the power transmission ratio P_r (ratio of transmitted power in III to incident power in I) for normal incidence

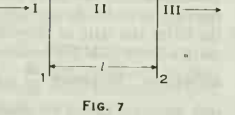


FIG. 7

$$P_r = \frac{4R_3/R_1}{(R_3/R_1+1)^2} \cdot \frac{1}{1 - \dfrac{(R_3^2/R_2^2-1)\,(R_2^2/R_1^2-1)}{(R_3/R_1+1)^2} \cdot \sin^2 k_2 l} \qquad \text{(II.12-1)}$$

In this expression R_1, R_2, R_3 are the specific acoustic resistances of media I, II, III respectively, and $k_2 = \omega/c_2$. The dependence of P_r on the frequency occurs through the term $\sin^2 k_2 l$. The effect can be seen most simply for the special case in which $R_1 = R_3$; *i.e.*, the third medium is the same as the first. Then

$$P_r' = \frac{1}{1 - \dfrac{(2 - R_2^2/R_1^2 - R_1^2/R_2^2)}{4} \sin^2 k_2 l} \qquad \text{(II.12-2)}$$

If $\sin^2 k_2 l = 0$, $P_r' = 1$. If $\sin^2 k_2 l = 1$, P_r' passes through the minimum value $4/(2 + R_2^2/R_1^2 + R_1^2/R_2^2)$, which is the smaller, the greater the difference between R_1 and R_2. The plot of P_r' as a function of frequency then appears as in fig. 8. The transmission is said to be selective with respect to frequency, with the frequencies $\nu_n = nc/2l$, where n is integral, yielding 100% transmission, and the frequencies $\nu_n = (2n+1)\,c/2l$ yielding minimum transmission. In no case does the value of P_r' become precisely zero. The situation may be described by saying that the medium II acts as if it preferred to have harmonic waves of frequencies $\nu_n = nc/2l$ traveling back and forth in it and would not tolerate others so readily. The full significance of this will become apparent in section II.15 below, in which standing waves in a finite medium are discussed.

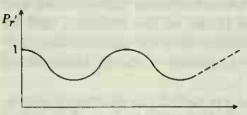


FIG. 8

The above discussion suggests the possibility of arranging a collection of media which will act as an acoustic filter; *i.e.*, transmit waves of certain frequencies well but refuse entirely to pass waves of other frequencies. If such could be realized the curve of P_r v. frequency would appear, for example, as in fig. 9. The transmission here has unit value between O and A, B and C, D and E, etc., and we should then call these frequency intervals pass bands, whereas since in AB, CD, etc., P_r drops to zero, they correspond to attenuation bands.

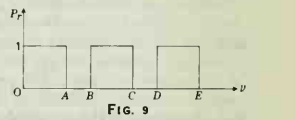


FIG. 9

Theoretical study indicates that a structure with these properties can be approximated by an iteration of the two media in the first illustration (*see* fig. 10). The larger the number of pairs in the iteration (or sections of the filter) the more closely will the transmission curve correspond to the sharp alternation of pass and attenuation bands as in fig. 9. In any case the structure is called a low-pass filter, since low frequencies from zero to A are passed.

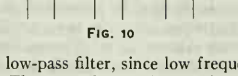


FIG. 10

The type of acoustic filter just discussed depends on the alternation of properties of two different media. This is not necessary to achieve the same general result.

Fig. 11 exhibits schematically a succession of cylindrical tubes of cross-sectional areas S_1, S_2 respectively and of lengths $AB = CD = EF$, etc., for the wide tubes and $BC = DE$, etc., for the narrow tubes. Such a structure acts as a low-pass filter for harmonic sound radiation—

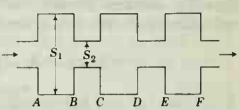


FIG. 11

indeed the transmission pattern of pass and attenuation bands is very similar to that in fig. 9. Here the ratio S_2/S_1 of the cross-sectional areas of the constricted and expanded tubes serves the same role as the ratio of the specific acoustic resistances in the iterated media type of filter discussed just before. The lengths of the sections, of course, have a decided bearing on the frequency extent of the transmission and attenuation bands. Many other types of acoustic filters are possible and have been constructed to meet requirements involving the suppression of certain frequency bands in blowers, mufflers and other equipment producing noise by the flow of air and other gases. The same type of theory applies to the filtration of elastic waves in solids; *e.g.*, rods. Many musical instruments involve acoustical filtration in their behaviour.

It should be emphasized that the selective transmission of sound discussed in this section is not due to a genuine frequency-dependent absorption in the structure with resulting transformation of acoustic energy into heat. The filtration results effectively because sound of certain frequency ranges is literally refused transmission, or, in other words, is not accepted from the source.

13. Absorption and Dispersion of Sound in Fluids.—The decrease in sound intensity from a localized source (*i.e.*, approximately at a point) because of spherical spreading has already been commented on (section II.9). But it is a well-known fact that there is always an additional loss in intensity which cannot be accounted for by this means. This loss is manifested even when sound passes through a more or less homogeneous medium where selective transmission or filtering action such as was described in the previous section does not exist. Such loss is ascribed to genuine absorption of the sound energy and its ultimate transformation into heat. An obvious source of such absorption is to be found in the internal friction of the fluid medium through which the radiation passes. This friction is exemplified by the so-called shear viscosity of the fluid, which is a measure of the tendency of relative motion of adjacent layers of the fluid to be decreased. For the simplest type of viscous fluid (*e.g.*, air and water) the shearing stress or drag per unit area of parallel layers of fluid on each other is directly proportional to the rate of change of the flow velocity (assumed parallel to the layers) with respect to distance normal to the flowing layers; the coefficient of proportionality is called the shear viscosity, which we shall denominate by η.

Analysis shows that for a plane harmonic acoustic wave of angular frequency ω progressing in the x direction in a viscous medium the displacement ξ can be written in the form

$$\xi = \xi_0 e^{-\alpha x} \cos(\omega t - kx) \qquad \text{(II.13-1)}$$

in which α is known as the amplitude absorption coefficient. It represents an exponential decrease in the amplitude of the wave with distance traveled, corresponding to the absorption of the acoustic energy by the medium. Expressed in terms of the viscosity of the medium and the frequency, the absorption coefficient can be written

$$\alpha = \omega^2 \tau V / 2c^2 (1 + \omega^2 \tau^2) \qquad \text{(II.13-2)}$$

where

$$V = \sqrt{2}c/\omega\tau \cdot [(1 + \omega^2\tau^2)(\sqrt{1 + \omega^2\tau^2} - 1)]^{\frac{1}{2}} \qquad \text{(II.13-3)}$$

and

$$\tau = 4\eta/3\rho_0 c^2 \qquad \text{(II.13-4)}$$

Here $V = \omega/k =$ the phase velocity of the wave. The fact that V in (3) depends on the frequency indicates that there is dispersion of the wave. This recalls the corresponding situation in light in which the velocity in a material medium normally decreases as the frequency increases. From equation (3) we see that at zero frequency $V = c = \sqrt{B/\rho_0}$, where B is the adiabatic bulk modulus of the fluid, and as ω increases, V increases. For real fluids, even with considerable viscosity (*e.g.*, glycerin with $\eta = 5$ poises at room temperature), the value of τ (often referred to as a relaxation time) is extremely small and hence $\omega\tau$ is negligibly small compared with unity except for extremely high frequencies. As a matter of fact, formulas (2) and (3) for α and V respectively are not in themselves competent to explain the observed absorption and dispersion of sound in fluids even at low frequencies. Other mechanisms must therefore enter into the attenuating process.

Another obvious cause of sound absorption in fluids is the conduction and radiation of heat. In the passage of a compressional wave the temperature is increased at points of maximum excess density (condensation). In the simple theory of wave energy propagation set forth earlier in this article, the heat energy corresponding to this temperature (strictly extra kinetic energy of the molecules of the fluid) is passed on to the adjacent portion of the fluid to produce a corresponding compression there. This, however, neglects the fact that whenever a temperature difference is set up in any substance, heat conduction and radiation ensue to the surroundings. This means that some of the energy which is necessary to propagate the disturbance from place to place gets lost in transit, effectively corresponding to absorption of the sound. The analysis is complicated but leads finally to the result that the absorption coefficient due to the combined effects of viscosity and heat conduction at not too high values of the ratio of frequency to pressure (*i.e.*, under 1,000 mc. per atmosphere) becomes

$$\alpha = 2\pi^2\nu^2/\gamma p_0 c \cdot [4\eta/3 + (\gamma - 1)\kappa/c_p] \qquad \text{(II.13-5)}$$

where ν is the actual frequency, p_0 the mean equilibrium pressure, c_p the specific heat at constant pressure, κ the thermal conductivity, γ the ratio of the specific heat at constant pressure to that at constant volume and the other symbols have their previously assigned meanings. The two terms in the brackets in (5) correspond to the contributions of viscosity and heat conduction respectively. Analysis discloses that for monatomic gases the contribution of heat conduction to α is about 70% that due to viscosity. In fact, the ratio of the heat conduction to the viscosity contribution is approximately $\gamma - 1$, so that for polyatomic gases the heat conduction contributes relatively much less.

For liquids other than mercury and liquefied gases the contribution of heat conduction is such a small fraction of that due to viscosity that it may be safely neglected. The part played by radiation in both gases and liquids turns out to be negligible except possibly at very low pressures or very high temperatures.

Formula (5), the so-called "classical" absorption formula, agrees very well with experimental values for monatomic gases (*e.g.*, the rare gases, helium, argon, etc.). In fact, the more precise formula of which this is an approximation has been found to agree in the case of helium down to pressures of the order of 1 mm. of Hg and up to frequencies of 1 mc. However, for diatomic and other polyatomic gases and liquids it breaks down badly in that there are some frequency regions in which the absorption is much greater than that predicted. For the vast majority of liquids the absorption over wide frequency ranges is found experimentally to vary more or less as the square of the frequency as equation (5) indicates. However, the experimental values are almost uniformly in excess of those predicted. Thus α/ν^2 for water at room temperature over the range from 1 to 50 mc. is about 25×10^{-17} sec.^{2} cm.$^{-1}$, whereas the formula yields only 8.5×10^{-7} sec.^{2} cm.$^{-1}$. The discrepancy is even greater (*i.e.*, approaching a factor of 100 or more) for many organic liquids like benzene.

The deviations just mentioned have led to considerable theorizing to provide alternative mechanisms for the explanation of sound absorption. One of these is based on the idea that when a fluid is squeezed in the passage of a compressional wave the work that is done does not go merely to increase the average translational velocity of the molecules; some of it, through molecular collisions, enters into other states of energy; *e.g.*, rotational and vibrational in the case of diatomic and other polyatomic molecules. Calculation shows that there is a lag in this process of energy transfer; *i.e.*, a certain time on the average is required for the energy to come back to translational form after it has got into the rotational or vibrational form. This time is called the relaxation time. Because of this time or phase lag the changes in translational energy of the molecules fail to keep in step with the propagation of the wave and thus some of the original energy of the disturbance is made unavailable during each cycle for transmission. It is therefore lost in dissipation and the result is effectively absorption. Actually shear viscosity itself can be considered a kind of relaxation process and the corresponding relaxation time is the τ given by equation (4). It is evidently not an adequate mechanism. In the case of

polyatomic gases vibrational (and to a certain extent rotational) relaxation has proved rather successful as an explanation of the observed excess absorption. This type of relaxation is usually termed thermal, since it corresponds effectively to an apparent increase in specific heat.

If a thermal relaxation process is at work in a gas, when the product of absorption coefficient α and the wave length λ is plotted against the frequency, the result is a curve which rises to a maximum and then falls off again. The frequency at which the maximum takes place is called the relaxation frequency and it is equal to $1/2\,\pi\tau$, where τ is the relaxation time. Sound absorption measurements in gases thus serve to indicate the existence of a relaxation effect if it is present and to evaluate τ. In many gases which have been studied, the relaxation frequency is in the ultrasonic range. For example, for hydrogen it is 10 megacycles at 1 atmosphere. For lower pressures it is proportionally reduced. Measurements of relaxation frequency in gases have shed considerable light on the internal energy states of the molecules of which these gases are composed and in particular on the probability of transitions between these states, a quantity proportional to the reciprocal of the relaxation time.

These acoustic absorption measurements can also be of practical engineering value. Thus in the high-speed flow of gases and vapours which impinge on obstacles (*e.g.*, steam flowing against turbine blades), sudden compressions and rarefactions are produced accompanied by temperature changes. The time during which such changes take place is controlled by the dimensions of the obstacles and the flow velocity. If the time interval is of the same order as the relaxation time for transitions to and from internal energy states of the molecules of the gas, loss of translational energy will take place (similar to the absorption of sound in the gas) and will result in a measurable decrease of impact pressure on the obstacle. Measurement of the relaxation time in the relevant pressure and temperature range by acoustical means is the best way to estimate the likelihood of this presumptive process.

The absorption situation with regard to liquids is also interesting, though not so satisfactory theoretically as for gases. For many organic liquids (*e.g.*, benzene and carbon tetrachloride) the excess absorption over that provided by classical viscosity and heat conduction can probably be adequately explained by thermal relaxation. This is, however, not the case for water. Here the best theory so far advanced is due to L. H. Hall (*Phys. Rev.*, 73:775 [1948]) and assumes that the molecules of water can take on two kinds of arrangement. In one they are arranged more or less tetragonally as in ice, while in the other they are packed more closely together (cubic close-packed structure). These two arrangements can be thought of as analogous to states of translational and internal energy of the molecules of a gas and a relaxation theory can be worked out for them. This is appropriately termed structural relaxation. Implied in it is a kind of relaxational compressibility which is the difference between the static compressibility as ordinarily measured and what may be called the instantaneous compressibility corresponding to very-high-frequency sound. This leads to a compressibility which varies with frequency and hence yields dispersion as well as absorption. By reasonable choice of parameters the theory accounts for the observed excess of absorption over that classically computed and also gives the proper variation of the absorption coefficient with temperature over the range from 0° C. to 100° C. The predicted dispersion is extremely small and has not been measured. Whether the theory of structural relaxation can be applied to water solutions like electrolytes is uncertain. Relaxation due to chemical dissociation probably plays a role here also. As in the case of gases, it is likely that much useful information about the liquid state (the most difficult state of matter to understand) could be gained from the study of the absorption and dispersion of sound, mainly of course in the ultrasonic range.

The relaxation time theory of sound absorption in liquids has suggested the introduction of the new concept of bulk or volume viscosity. This is associated with a resistance related to the rate of volume compression as the more common shear viscosity is related to the rate of shearing strain. According to the theory of Hall, mentioned above, this bulk viscosity for water is several times as great as the shear viscosity. It decreases as the temperature increases. A bulk viscosity of this magnitude corresponds to a very high relaxation frequency. In water, for example, this is of the order of 10^{12} cycles/sec. and hence outside the range of direct experiment.

The dispersion of sound in fluids has been mentioned only briefly in the earlier part of this section and demands a little more attention. It is essentially a high-frequency or ultrasonic phenomenon. More strictly, in gases it shows up at large ratios of frequency to pressure, *i.e.*, it has been observed at relatively low frequencies (though not in the audible range) if the pressure is low enough. It must be observed, however, that the dispersion of sound is a second order phenomenon compared with absorption, that is, the velocity of sound differs from its value $c = \sqrt{\gamma p/\rho}$ in the simple classical theory (section II.3 above) only by terms of the order $c\omega^2\tau^2$, whereas the absorption coefficient depends on the first power of τ. This is clear from equations (2) and (3) above for the case of shear viscosity, and the result is not materially different in the case of relaxation absorption and dispersion. Hence dispersion due to relaxation may be more difficult to detect than the associated absorption.

Acoustic dispersion in liquids has been particularly difficult to detect experimentally, since the velocity is very sensitive to temperature, and it is hard to maintain constant temperature in sound transmission. Up to the mid-1950s no unambiguous case of dispersion in a pure liquid had been exhibited, though dispersion has been measured in liquid suspensions.

It is commonly assumed in elementary treatments of wave propagation that liquids like gases have no shear elasticity and hence can transmit compressional waves only. This has been indeed the theoretical basis for the analysis presented so far in this article. However, it is in reality not borne out in the case of certain viscous liquids like polybutylene, gels and rubber solutions. These are called visco-elastic substances, possess shear elasticity and can propagate shear waves. Since the effective rigidity modulus for such liquids increases with the frequency, the velocity of shear waves in this case is also frequency dependent. For certain polybutylene solutions the rigidity may rise from 3×10^7 dynes/cm.^{2} at 300 kc. per second to 5×10^9 dynes/cm.^{2} at about a megacycle. It is interesting to note that the viscosity of such liquids can be determined by purely acoustical measurements.

Of all liquids, probably liquid helium exhibits the most interesting and anomalous physical properties. Thus below 2.18° K. (its so-called lambda point) it loses practically all shear viscosity and gains large thermal conductivity. This modification is known as helium II and is called a superfluid. Its acoustical characteristics also show anomalies. Above the lambda point the velocity of sound increases as the temperature is lowered, reaching a maximum of 220 meters/sec. at 2.5° K. It then falls sharply to about 216 meters/sec. at the lambda point and thereafter rises again as the temperature is lowered, appearing to level off to about 239 meters/sec. as absolute zero is approached. Sound attenuation in liquid helium above the lambda point is explainable in terms of viscosity and heat conduction and no thermal relaxation appears to be involved. However, the attenuation reaches a maximum (for given frequency) at the lambda point and then declines very rapidly to a minimum at 2° K. It then rises again as the temperature is lowered to absolute zero. This peculiar behaviour is evidently due to some kind of unusual relaxation mechanism but it had not been elucidated by the mid-1950s.

Liquid helium II can transmit a peculiar kind of thermal wave, *i.e.*, a wave in which pressure and density remain constant throughout the liquids, but a temperature fluctuation is propagated. The velocity of this wave is zero at the lambda point but increases as the temperature falls, reaching a maximum of 20 m. per second at about 1.75° K., and then decreasing again as the temperature is lowered further. This wave phenomenon is called "second sound," though this is a rather questionable terminology.

The source of the thermal radiation is usually a thermophone, which is a platinum ribbon to which an alternating current is ap-

plied, producing a harmonically varying source of heat. The waves are detected by a resistance thermometer. For more information on the properties of helium at low temperatures, *see* LOW-TEMPERATURE PHYSICS.

Liquid helium is a valuable adjunct in low-temperature acoustics, being used as a coolant in transmission studies at very low temperatures, especially in solids (*cf.* section III.6).

The absorption and dispersion of sound in solid media are discussed in section III.6 below (*see* also section V.6 below.)

14. Diffraction and Scattering of Sound.—It has already been mentioned that when a sound wave strikes an obstacle of finite size not only is some of the sound reflected from the obstacle and some transmitted through it, but in general some gets around it or is said to be diffracted. As a result of the presence of the obstacle the original sound beam is deflected and the pattern of sound intensity distribution is altered. It is customary to say that the sound is scattered. The resultant intensity pattern is a superposition of the original sound beam without the presence of the obstacle and the scattered sound due to reflection, transmission and diffraction.

The general method of theoretically estimating harmonic sound scattering by a completely rigid obstacle is straightforward. One assumes that from the obstacle there spreads out in every direction a scattered wave of arbitrary amplitude and phase. When the particle displacement due to this wave is added to that corresponding to the primary undisturbed wave the result must be zero everywhere on the surface of the rigid obstacle. (This is, of course, idealized since there is really no such thing as a completely rigid obstacle into which sound energy cannot penetrate. Nevertheless there are examples for which this approximation is a fair one, such as the human head.) The application of this boundary condition suffices to evaluate the unknown amplitude of the scattered wave and hence to give the scattering pattern. This general method can, to be sure, be applied with any degree of ease only to obstacles of simple geometrical shape, such as cylinders and spheres. But many practical obstacles can be approximated reasonably well by such surfaces.

The scattering pattern depends on the relation between the wave length of the primary wave and the dimensions of the obstacle. The two limiting cases are readily visualized. Consider for simplicity a sphere of radius a as the obstacle. If the wave length is very small compared with $2\pi a$, the scattered wave on the side of the sphere opposite to the direction of the incident primary wave combines with the latter so as to produce zero intensity (destructive interference) and hence the obstacle forms a shadow the sharpness of which is the greater the smaller the ratio $\lambda/2\pi a$. Since this is analogous to the shadows formed in optics by objects with dimensions much larger than the wave length of light (the case of geometrical optics), we may refer to this situation in sound as geometrical acoustics. For audible sounds (wave length in air greater than about 1.7 cm.) this type of scattering is important only for relatively large obstacles. However, for ultrasonic (high-frequency) sounds it can be quite significant. In other words, geometrical acoustics in the audible region is not as applicable as is geometrical optics for visible light, the reason lying in the much smaller wave length of the latter.

To consider the opposite extreme, if the wave length is very large compared with $2\pi a$, the primary wave is scattered in all directions and there is no distinct shadow. The intensity of the scattered wave is directly proportional to the square of the volume of the scattering particle (in this case a sphere) and inversely proportional to the fourth power of the wave length. This is the Rayleigh scattering law in acoustics, the optical analogue of which leads to the explanation of the blue colour of the sky. For in the scattered radiation the higher frequencies will produce the high intensities. Sound scattered by a grove of trees appears to be raised in pitch. A similar phenomenon results when underwater sound is scattered by air bubbles.

The middle range in which the wave length and the dimensions of the scattering obstacle are comparable is much more difficult to handle computationally. In general the intensity of the scattered radiation at any particular distance from the scattering obstacle goes through a series of maxima and minima as the azimuth is varied. Relatively near the obstacle the diffraction pattern tends to be very complex, resembling the well-known Fresnel pattern in optics. Farther away it reduces to the smoother Fraunhofer diffraction pattern.

The special case of diffraction by the human head is extremely interesting since it can easily become evident to any careful observer. For given audible wave length the intensity of the sound from a speaker's mouth is greatest in the direction directly in front. It falls off to a smaller value in the directions at right angles to this but rises again in the direction back of the head. This secondary maximum is only comparable with that directly in front at distances large compared with the wave length. Very close to the head, however, the variation in intensity with azimuth is marked.

The effect of the head on sound received at the ears is also a well-marked phenomenon. It develops that the resultant apparent intensity at the ears is a maximum when the line connecting the ears lies in the direction of the sound and falls in general to a minimum when the direction of the sound lies at right angles to this line.

This result is obtained theoretically by the use of the so-called reciprocal theorem of Helmholtz which may be stated in the form (quoting from Rayleigh, *Theory of Sound*, vol. ii, para. 294 [1945]): "If in a space filled with air which is partly bounded by finitely extended fixed bodies and is partly unbounded, sound waves be excited at any point A, the resulting velocity potential φ at a second point B is the same both in magnitude and phase as it would have been at A had B been the source of sound." Though stated for the velocity potential (equal in acoustical waves to the negative of the time integral of the excess pressure divided by the mean density), the theorem also applies to the intensity of sound and plays an important role in the calibration of acoustical instruments used in the measurement of radiation.

15. Sound Waves in Finite Spaces; Standing Waves in a Tube.—So far we have dealt primarily with spreading sound waves or waves progressing in one direction only. Strictly speaking, this behaviour is possible only in an infinitely extended medium. As soon as the medium is surrounded by a barrier so as to produce a closed finite space, reflection must in general take place, leading to the existence of waves in several directions. This has interesting consequences.

For the sake of simplicity consider the example of a cylindrical tube of length l and cross-sectional area S and closed rigidly at both ends. A disturbance produced anywhere in the air in the tube will travel in both directions, be reflected from both ends and thus produce in general a series of waves traveling in both directions along the tube. These waves, from the geometry of the situation and the finite constant value of the velocity of sound, must be periodic waves the frequencies of which cannot take all possible values but are fixed by the boundary conditions prevailing at the ends of the tube; *i.e.*, zero resultant value. The allowed frequencies of the waves in the tube are those satisfying the equation

$$\sin kl = 0 \qquad \text{(II.15-1)}$$

i.e., the frequencies

$$\nu_n = nc/2l \qquad \text{(II.15-2)}$$

where n is any integer and c is the sound velocity in the tube. These are the frequencies of the harmonic waves which can exist in the tube and still satisfy the boundary conditions at the two ends. They are called the characteristic frequencies or normal modes of vibration of the air column. The fundamental frequency is

$$\nu_1 = \nu_f = c/2l \qquad \text{(II.15-3)}$$

The higher frequencies, called harmonics or overtones, are multiples of the fundamental. It is now customary to refer to the fundamental as the first harmonic. For $n = 2$ we have the second harmonic or first overtone, etc. Approximately the same set of characteristic frequencies hold for a cylindrical tube open at both ends, though here the boundary conditions are different. Ideally at the open end of such a tube it is plausible to assume that the acoustic excess pressure remains zero since the tube is open to the

surrounding air. This is by no means obvious, however, since we can readily have nonvanishing excess pressure in a wave progressing through the open air. One may indeed reason that at such an open end there is nothing to check the air displacement along the length of the tube and hence the particle displacement or velocity will be a maximum there. This means $\partial\xi/\partial x = 0$, which is equivalent to the vanishing of the excess pressure. The use of a probe microphone (*see* section IV.3 below) confirms this approximately by direct experimentation. It yields, as stated, the normal modes given by (2) and these are in approximate agreement with experiment.

It is not surprising that exact agreement is not obtained when it is recalled that the air at the end of the tube possesses inertia and must be disturbed by the wave traveling along the tube. Moreover, the disturbance at the end of the tube acts effectively as a source of sound radiating into the open air, leading to loss of energy. Actually, this latter phenomenon has only a small effect on the characteristic frequencies of the tube, but the former can be quite noticeable. Suppose we have a tube closed at one end and open at the other. Application of the ideal boundary conditions just mentioned then leads to the set of characteristic frequencies

$$\nu_n = (n+\tfrac{1}{2})\ c/2l \qquad \text{(II.15-4)}$$

but because of the open end effect these are actually modified to

$$\nu_n' = (n+\tfrac{1}{2})\ c/2(l+S/C) \qquad \text{(II.15-5)}$$

where S is the area of cross section of the tube and C is a quantity of the dimensions of a length and known as the acoustic conductivity of the opening. Unfortunately it is difficult to compute C theoretically but for a circular opening it is of the order of magnitude of the diameter. Hence the term S/C in this case is approximately $\pi a/2$, where a is the radius of the circular tube. The characteristic frequencies are thus those of a tube of equivalent length $l + \pi a/2$, which for a short wide tube can be rather different from l. In general the quantity S/C is known as the end correction. It can of course be evaluated by actual measurement of the normal modes. This is indeed a useful method of obtaining C (*see* section III.4).

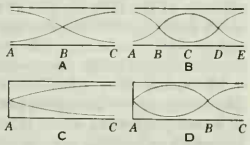


FIG. 12.—STANDING WAVES. (A) Fundamental or first harmonic in a tube open at both ends; (B) second harmonic in tube open at both ends; (C) fundamental or first harmonic in tube open at only one end; (D) second harmonic in tube open at only one end

To come back to the actual wave disturbance corresponding to a normal mode in a tube open at both ends, it develops that there are certain positions in the tube, namely for

$$\cos n\pi x/l = 0 \qquad \text{(II.15-6)}$$

at which the displacement is zero for all time. This situation obviously cannot take place in a progressive wave. For this reason the wave disturbance corresponding to a normal mode is known as a standing wave. The positions of null disturbance or displacement for all time are known as nodes, while the positions for which $\cos n\pi x/l = 1$ are called antinodes or loops, corresponding to maximum displacement. The distance between successive nodes is equal to a half wave length of the particular mode. From the relation between displacement and excess pressure it follows that a displacement node is an excess pressure antinode and vice versa. Standing waves are usually pictured by means of simple diagrams of which fig. 12 is an illustration. The tube open at both ends is typified by (A) and (B). Here in (A) A and C are antinodes and B (at centre) is a node. In (B) A, C and E are antinodes and B and D ($\frac{1}{4}$ and $\frac{3}{4}$ of the distance along the tube) are nodes. The tube open at one end and closed at the other is exemplified in (C) and (D). In (C) A is a node and C an antinode. (All this refers to displacement, of course.)

The resultant particle displacement in the standing wave produced by arbitrary initial conditions in a tube of length l open at both ends is taken to be the sum over all the normal modes. That is,

$$\xi = \sum_{n=0}^{\infty}(A_n \cos n\pi c/l \cdot t + B_n \sin n\pi c/l \cdot t) \cos n\pi x/l \qquad \text{(II.15-7)}$$

The assumption is known as the principle of superposition and was enunciated by D. Bernoulli in 1755. It is at the basis of what is known as the Fourier analysis of an arbitrary disturbance in a finite medium into harmonic modes.

From (7) it follows that the initial displacement at any point x in the tube is

$$\xi_0(x) = \sum A_n \cos n\pi x/l \qquad \text{(II.15-8)}$$

whereas the initial velocity is (using the dot to denote partial differentiation with respect to time)

$$\dot{\xi}_0(x) = \sum n\pi c/l \cdot B_n \cos n\pi x/l \qquad \text{(II.15-9)}$$

It will be noted that $\xi_0(x)$ and $\dot{\xi}_0(x)$, though presumably arbitrary, are here represented as sums of harmonic components in x. The evaluation of the coefficients A_n and B_n proceeds according to the Fourier rule and we find

$$A_n = 2/l \cdot \int_0^l \xi_0(x) \cos n\pi x/l \cdot dx \qquad \text{(II.15-10)}$$

$$B_n = 2/n\pi c \cdot \int_0^l \dot{\xi}_0(x) \cos n\pi x/l \cdot dx \qquad \text{(II.15-11)}$$

This of course assumes the mathematical validity of the expansion (*see* FOURIER SERIES). Actual physically possible forms for $\xi_0(x)$ and $\dot{\xi}_0(x)$ in general satisfy the mathematical conditions.

Although this section of the article is limited to sound disturbances propagated in a cylindrical tube, the same type of analysis can be applied to a sound disturbance in a confined space of any kind; *e.g.*, the air in a room with or without open windows. Here the reflection of the disturbance from the boundaries of the room produces a three-dimensional standing wave pattern which can be analyzed into a set of normal modes or characteristic frequencies depending on the geometrical shape and dimensions of the room. Some of these modes may be much more prominent than others and have a good deal to do with the hearing of speech or music in the room. Of course, absorption at the walls and in the air itself plays an important role. These considerations enter vitally into architectural acoustics (*see* section V.3 below).

16. Wave Pulses; Group Velocity.—In the study of acoustic wave propagation so far we have restricted our attention to continuous progressive waves like

$$\xi = A \cos(\omega t - kx) \qquad \text{(II.16-1)}$$

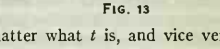


FIG. 13

in which for given x there exist nonvanishing displacements no matter what t is, and vice versa, or standing waves like

$$\xi = A \sin \omega t \cdot \sin kx \qquad \text{(II.16-2)}$$

where again the displacement takes place for all x and all t. These are of course ideal situations. Instead of encountering the wave in (1) in practice, we actually meet with waves in which the disturbance at any instant is localized in a bounded region of space as in fig. 13, which pictures the disturbance at the instant t as a function of x. From infinity on the left up to A, $\xi = 0$; from A to B, ξ is given by (1), and from B to infinity on the right ξ is again zero. This is known as a finite wave train or wave pulse. As time passes, the train moves from left to right carrying the disturbance with it. It is clear that in all practically producible wave disturbances this is the sort of thing realized.

It turns out to be possible by sufficiently elaborate mathematics to represent a wave pulse by a superposition of continuous and infinite harmonic wave trains (like [1]), provided that the frequencies do not form a discrete set like the normal modes in a tube but rather are themselves infinitely many and continuously distributed. The amplitude of each harmonic component depends on its frequency or wave length and may become negligibly small for sufficiently great and small frequencies; so effectively the band of frequencies necessary for the harmonic decomposition of the pulse will be finite if the pulse is finite in length. It is easiest to understand this by an example. Take the pulse

$$\xi\,(x-ct)=\cos k_0(x-ct),\quad |x-ct|<L/2$$
$$\xi\,(x-ct)=0,\quad |x-ct|>L/2 \qquad \text{(II.16-3)}$$

which is indeed the situation pictured in fig. 13. The amplitude $A(k)$ of the harmonic component corresponding to wave parameter k then is

$$A(k)\doteq\frac{\sin\,(k_0-k)L/2}{\pi(k_0-k)} \qquad \text{(II.16-4)}$$

By plotting $A(k)$ as a function of (k_0-k) it is clear that the range of wave parameter k necessary to represent the traveling pulse adequately is inversely proportional to L, the length of the pulse. A very long wave train therefore needs only a very small range of parameter to represent it and in the limit of $L\rightarrow\infty$, the train becomes $\cos k_0(x-ct)$ precisely.

On the other hand, if the pulse is very short k_0-k becomes correspondingly large.

The immediately preceding considerations apply, indeed, only if the wave velocity c is independent of the frequency or wave parameter. If there is dispersion (*see* section II.13 above) the situation is more complicated. It then turns out on the assumption that only a finite range Δk of wave parameter about a mean parameter k_0 is necessary for the adequate representation of the pulse, that the wave train can be expressed approximately as a single sinusoidal wave of angular frequency ω_0 with an amplitude which is a function of time and space through the quantity

$$x-(d\omega/dk)_0 t \qquad \text{(II.16-5)}$$

The meaning of $(d\omega/dk)_0$ is this: we express the dispersion of the medium by writing ω as a function of k, viz.,

$$\omega=\omega(k) \qquad \text{(II.16-6)}$$

We then differentiate ω with respect to k and take the value of the derivative corresponding to $k=k_0$. If we proceed in the x direction with velocity

$$U=(d\omega/dk)_0 \qquad \text{(II.16-7)}$$

we shall always be accompanied by the same amplitude and if we travel at this velocity with the maximum amplitude we shall be riding a point where the individual harmonic components in the original wave train heap up, so to speak, to form what may be called a wave group. The velocity of progress of the group is called the group velocity and is given by (7).

The group velocity can be rather conveniently expressed alternatively by

$$U=\frac{c_0}{1-(\nu/c_0)\,(dc/d\nu)_0} \qquad \text{(II.16-8)}$$

where c, the phase velocity, is expressed as a function of the frequency ν, and c_0 is the mean phase velocity in the group. If $c=c_0$, a constant independent of frequency, $U=c_0$, and the group velocity coincides with the constant phase velocity. Since the dispersion of sound in fluids is generally very small over obtainable frequency ranges (save for hydrogen, as pointed out in section II.13 above), the group velocity will not in general differ much from the phase velocity. But it is well to remember at any rate that in the use of short pulse techniques for measuring sound velocity in fluids it is strictly the group velocity which is measured. In the dispersion of sound in fluids associated with viscosity, heat conduction as well as various relaxational effects, the velocity in general increases with the frequency. Hence $dc/d\nu>0$ and hence $U>c_0$; *i.e.*, the group velocity is greater than the average phase velocity. It is interesting to note that this is opposite to the case of normal dispersion of light.

The advantage of the use of short-wave pulses in the measurement of sound properties is obvious, since undesirable reflections can be separated out from the main beam. The echo method of acoustic depth recording as well as gun ranging and the detection of submerged objects in water (sonar) depend on the use of pulse techniques (*see* section V.5 below).

From the definition of group velocity it is plausible to assume that this is the velocity with which the energy in the wave train is transmitted. This is approximately correct, though the matter demands more careful consideration than can be given here.

It is often desirable to amplify an acoustic pulse by transforming it into an electrical pulse in an electrical oscillating circuit and then feeding it into an amplifier. If the original shape of the pulse is to survive amplification unaltered it is essential that the uniform frequency response of the amplifier cover a sufficiently large band of frequencies. Otherwise, distortion will result. Thus, for the pulse just considered in the example given above the frequency band must correspond to a wave parameter band satisfying the approximate condition

$$(k_0-k)\;L/2\doteq\pi \qquad \text{(II.16-9)}$$

17. Radiation Pressure.—So far when we have talked about the pressure associated with an acoustic wave we have had in mind the excess pressure produced by the acoustic disturbance and propagated through the medium. For a plane harmonic progressive wave the excess pressure is a sinusoidally varying function of time and space and hence, if averaged over the time at any particular place, vanishes. It does of course possess a nonvanishing mean square value. However, it can be shown that associated with a progressive acoustic wave in any physically realizable medium there is also a pressure which does not vanish on the average and which therefore makes a positive net addition to the static pressure prevailing in the medium. This is the so-called radiation pressure. It is a so-called second-order quantity because it is much smaller than the root-mean-square (square root of the average of the square) acoustic excess pressure. In the ordinary acoustic equations, as we have developed them, it is therefore usually neglected. However, it is always present physically because of the essential nonlinearity of the acoustical equations if higher order terms are included. This means that if the particle displacement in an acoustic wave has the form (II.1–3), the corresponding acoustic excess pressure will be made up not merely of a corresponding linear term proportional to $\sin(\omega t-kx)$ but also a term proportional to $\sin^2(\omega t-kx)$, with indeed a much smaller coefficient of proportionality. This latter term, when averaged over time, gives rise to the steady radiation pressure.

The simplest expression for the radiation pressure is

$$p_r=\frac{p_e\partial\xi/\partial t}{c} \qquad \text{(II.17-1)}$$

where it must be understood that p_e is the root-mean-square excess acoustic pressure and likewise $\partial\xi/\partial t$ is the root-mean-square particle velocity. Since the latter is always smaller than the wave velocity c, the statement that $p_r<<p_e$ follows. In spite of its small value relative to p_e, p_r produces some important consequences. Its existence implies that any solid object placed in the path of an acoustic beam will experience a steady pressure. This may be used to estimate the intensity of the sound, since p_r is directly proportional to the intensity (*see* section V.1 below). Moreover, under certain conditions p_r is associated with the production of steady streaming of the fluid medium through which the sound wave is passing. This streaming may produce interesting physical effects (section V.6 below).

18. Shock Waves.—In section II.2 the fact was noted that in a macrosonic or high-intensity sound wave, a disturbance which starts out with a symmetrical profile (fig. 2) will very soon develop a steepening of the wave front. This process is opposed in the case of a fluid, like air or water, by the attenuation mechanisms already mentioned in section II.13.

The result can be the formation of a saw-toothed curve as shown in fig. 14, in which at points A, B, C, etc. there are considerable changes in pressure over a very small space. Ideally there is a genuine discontinuity in pressure, density and flow velocity at A, B, C, etc. In actual fact there is no real discontinuity, but the drop in pressure or density takes place over an interval of about 10 mean free paths of the fluid if it is a gas, *i.e.*, of the order of 10^{-4} cm. in a gas un-

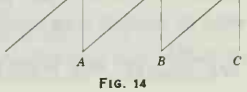


FIG. 14

der standard conditions. This transition state moves through the medium with a velocity in excess of the normal velocity of sound. It is called a shock wave and can result physically from an explosion or from the motion of a solid object (*e.g.*, a missile or airplane) moving through a fluid with a velocity greater than that of sound. Shock waves can be studied experimentally in a tube by arranging for the sudden rupture of a membrane separating a region of high pressure from one of low pressure. The thickness of a shock front, as it is called, can be estimated by the reflectivity of light from it due to the change in the index of refraction or alternatively by measuring the change in temperature through it. Very instructive pictures of the travel of a shock front can be obtained by the use of an optical interferometer; the rapid change in density across the front produces a shift in the interference fringes. The *strength* of a shock is defined as the ratio of the pressure just ahead of the front to that just behind it.

A shock wave is reflected when it strikes an obstacle, but the law of reflection for ordinary low amplitude sound is followed only for angles less than about 40° for strong shocks. For larger angles the reflection is a complicated phenomenon. Similar considerations apply to refraction and diffraction.

III. SOURCES OF SOUND

1. Introduction.—In the preceding sections sound has been considered as a propagated elastic disturbance in a material medium. It is clear that such disturbances must be produced in the first place before they can be propagated. Any disturbance involving a change in the density at a point in a material elastic medium may be said to be a source of sound. This accounts for the relatively great variety of sound sources, since there are very many ways of producing a density change in an elastic solid, liquid or gas, involving mechanical, thermal, electrical, magnetic, electromagnetic or chemical actions.

The most common sound waves are either periodic in character or for convenience are always thought of as compounded of harmonic components. Therefore the simplest and by all odds most important sources of sound are vibrating portions of solids, liquids and gases.

2. Mechanical Vibrations, Damped and Forced.—In any mechanical vibrator a portion of the mechanical system involved, when displaced from equilibrium, has a tendency to return to its undisplaced condition. Such a vibrator is characterized by three physical elements: a mass or inertia factor m; a stiffness factor f, representing the elastic tendency to return to the equilibrium state; and a damping factor R, representing the tendency for the motion resulting from the original disturbance to decay instead of continuing forever. The simplest illustration is a weight suspended from one end of an ideally weightless spring, the other end of which is fastened to a rigid ceiling. When the weight is displaced from its rest position and let go, it strives to return to the original position and motion results. The inertia factor is here the mass of the weight, the stiffness factor is simply the stiffness of the spring (*i.e.*, applied force per unit stretch) and the damping factor is provided in part by the internal friction of the spring, in part by the viscosity of the medium in which the motion takes place, and finally by the loss of energy to the surrounding medium in the form of sound radiation. This is not generally an efficient sound source. A membrane presenting a larger vibrating and radiating area to the surrounding medium is more practical. However, the assignment of the three vibrational elements is more difficult in this case. For example, though the equivalent mass of the vibrator is proportional to the actual mass of the membrane it is not equal to it (*see* section III.5 below). However, there is great analytical advantage in treating any vibrator in terms of its three elements.

The equation of motion of a vibrator whose displacement from equilibrium is denoted by ξ is

$$m\ddot{\xi}+R\dot{\xi}+f\xi=0 \qquad \text{(III.2-1)}$$

where we again are using the dot to denote time differentiation; *i.e.*, $\ddot{\xi}=\frac{d^2\xi}{dt^2}$. The form of the solution depends on the relation between m, R and F. If

$$f>R^2/4m \qquad \text{(III.2-2)}$$

actual vibrations result and the solution is

$$\xi=Ae^{-Rt/2m}\cdot\cos(\omega t+\alpha) \qquad \text{(III.2-3)}$$

where

$$\omega=\sqrt{f/m-R^2/4m^2}=2\pi\nu \qquad \text{(III.2-4)}$$

and ν is the frequency of the vibrations. A is called the initial amplitude and α is the initial phase angle or epoch of the motion. Because of the presence of the term $e^{-Rt/2m}$, the amplitude of the oscillations steadily diminishes with time and ξ plotted as a function of t appears as in fig. 15. This is called a damped oscillation; equation (3) describes its so-called "free" vibration; *i.e.*, uninfluenced by any outside force. The extent of the damping is described by $2m/R$, which is called the decay modulus. Alternatively, the logarithmic decrement δ is also used; it is the natural logarithm of the ratio of two successive maxima of the ξ, t curve, and is given by

$$\delta=RP/2m \qquad \text{(III.2-5)}$$

where P = period of the vibrations = time for one complete to-and-fro motion of the vibrating mass = time from one maximum of the curve in fig. 14 to the next = $1/\nu$. The larger R for given P the larger is δ and the more rapid the decay.

In many applications in electroacoustics in which electric circuits are used as the source of mechanical oscillations, the so-called quality factor Q is used to describe the damping. In terms of the logarithmic decrement this takes the form

$$Q=\pi/\delta \qquad \text{(III.2-6)}$$

Thus a highly damped oscillator is said to have a low Q and vice versa.

The quantities A and α depend on the initial conditions of the motion. If the value of ξ at $t=0$ is ξ_0 and that of the displacement velocity $\dot{\xi}$ is $\dot{\xi}_0$, we can write

$$A=\sqrt{\xi_0^2+(\dot{\xi}_0+R\xi_0/2m)^2/\omega^2} \qquad \text{(III.2-7)}$$

$$\alpha=\arctan\left(-\frac{\dot{\xi}_0+R\xi_0/2m}{\omega\xi_0}\right) \qquad \text{(III.2-8)}$$

If $f<R^2/4m$, *i.e.*, relatively large damping, there are no vibrations and ξ merely makes a single swing and thereafter decays to zero. Such overdamped systems are not considered in steady state acoustics, though they can serve as sources of sound pulses.

The average rate of dissipation of energy of the vibrating system due to the damping force is

$$\bar{\dot{E}}=-\overline{R\dot{\xi}^2} \qquad \text{(III.2-9)}$$

where the bar indicates the average taken over an arbitrary time interval. If $\overline{\xi^2}_{\max}$ denotes the mean square amplitude over the time interval in question, we have

$$\bar{\dot{E}}\doteq-R\omega^2\overline{\xi^2}_{\max}/2 \qquad \text{(III.2-10)}$$

If the damping were due entirely to the radiation of acoustical energy from the vibrator into the surrounding medium, (10) would give the acoustically radiated power output. This is never the case, however; in general the power output is some proper fraction of (10), measuring thereby the acoustical efficiency of the source. Damped vibrations provide only a transient sound source. To maintain steady oscillations and a steady sound output, the imposition of an external force is necessary. If the external force varies harmonically with the time with angular frequency ω_0 and amplitude F_0 the system characterized by the elements m, f and R will get ultimately into a steady state of vibration with the same angular frequency as the force, which of course in general will not be the same as the free vi-

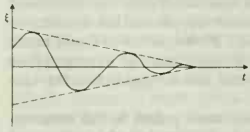


FIG. 15

bration frequency (4). The steady state displacement has the form

$$\xi=\frac{F_0\cos(\omega_0 t-\beta)}{\sqrt{\omega_0^2R^2+(f-m\omega_0^2)^2}} \qquad \text{(III.2-11)}$$

where

$$\beta=\arctan(\omega_0 R/[f-m\omega_0^2]) \qquad \text{(III.2-12)}$$

is the phase angle between the force and the displacement. Other things being equal, the smaller the damping factor R, the more nearly ξ is in phase with the force or 180° out of phase with it.

In some ways the velocity is more significant. Here

$$\dot{\xi}=\frac{F_0\cos(\omega_0 t-\gamma)}{\sqrt{R^2+(m\omega_0-f/\omega_0)^2}} \qquad \text{(III.2-13)}$$

where

$$\gamma=\arctan\left(\frac{m\omega_0-f/\omega_0}{R}\right) \qquad \text{(III.2-14)}$$

is the phase angle between force and velocity; clearly $\beta=\pi/2-\gamma$. Both ξ and $\dot{\xi}$ are harmonic in time with frequency $\nu_0=\omega_0/2\pi$. The amplitude of $\dot{\xi}$ is a maximum for

$$\omega_0=\sqrt{f/m} \qquad \text{(III.2-15)}$$

and this is called the resonance (angular) frequency of the system. In words, when the frequency of the force is adjusted to $1/2\pi\cdot\sqrt{f/m}$, the system is said to resonate with the force and maximum steady state velocity oscillations ensue. The displacement, to be sure, does not attain maximum steady state amplitude at resonance but at the frequency

$$\nu_1=1/2\pi\cdot\sqrt{f/m-R^2/2m^2} \qquad \text{(III.2-16)}$$

If $R^2/2m^2 << f/m$, this is of course close to the resonance frequency and indeed to the free oscillation frequency (4).

The analogy with electrical oscillations of an L, R, C circuit is rather striking. The denominator in the right-hand side of (13) is called the mechanical impedance of the vibrator. We write

$$Z=\sqrt{R^2+(m\omega_0-f/\omega_0)^2} \qquad \text{(III.2-17)}$$

This is a minimum for resonance and equal to the damping factor R.

Since the harmonic force maintains the system in a steady state of vibration, the average rate of energy flow from the force to the system just equals the average rate of dissipation (equation 9) of energy through the damping. This average flow of energy is a maximum at resonance.

If the amplitude of $\dot{\xi}$ is plotted as a function of $\omega_0-\omega_{res}$, a so-called resonance curve results. $\dot{\xi}$ has its maximum for $\omega_0-\omega_{res}=0$ and decreases on either side of the origin. The nature of the resonance peak is controlled by the damping factor R. If R is small, the peak is high and sharp. If R is large the peak is low and broad. In the first case the response of the system to the force is more selective with respect to frequency than in the second.

3. Coupled Systems.—As has already been emphasized, the vibrations of a single mass particle form only an idealized representation of actual sound source motions. Actual sound sources are extended bodies or collections of particles, like strings, membranes, rods and plates. Such sources may be thought of as collections of vibrating particles coupled together. A simple illustration is provided by two mass particles attached to a weightless, stretched, flexible, horizontal string of finite length. If the particles are allowed to move in a vertical plane only, it is clear without going through the analysis that the motion of each is affected by that of the other through the tension of the string joining them. It is not surprising that the vibrations of such a system correspond to the superposition of component vibrations of two different frequencies, even if the masses are the same and the attachment to the string is perfectly symmetrical. The generalization of this situation is that of n particles attached at regular intervals to the stretched string (the so-called "loaded" string). There are now n characteristic frequencies and the motion of each particle is a superposition of n harmonic components with these n different frequencies. Finally, as n increases indefinitely we expect that the system will act more and more like a continuous string. This is best handled mathematically by treating the string not as a collection of discrete particles but as a continuous medium, albeit a linear one, somewhat like a one-dimensional continuous elastic fluid whose motion, when disturbed from equilibrium, is essentially wave motion.

4. Vibrations of Strings; the Helmholtz Resonator.—If an indefinitely long horizontal flexible string of mass ρ_l per unit length is stretched with tension T along the x axis any disturbance represented by a transverse displacement ξ from the equilibrium (rest) position is propagated in accordance with the wave equation (R. B. Lindsay, *Concepts and Methods of Theoretical Physics*, p. 324 [1951]):

$$\frac{\partial^2\xi}{\partial t^2}=T/\rho_l\cdot\frac{\partial^2\xi}{\partial x^2} \qquad \text{(III.4-1)}$$

the solution of which corresponds to transverse waves propagated along the string with velocity

$$c=\sqrt{T/\rho_l} \qquad \text{(III.4-2)}$$

If the string is finite, of length l, and fastened at both ends, the solution of (1) satisfying the boundary conditions $\xi=0$ at $x=0$ and $x=l$ is of the same form as that discussed in section II.15 above. Thus the motion of the string is compounded of a series of harmonic standing waves or normal modes of vibration with frequencies given by

$$\nu_n=nc/2l \qquad \text{(III.4-3)}$$

where n is integral. The relative importance of the various modes in any given motion of the string depends on the initial conditions; *i.e.*, the methods used to excite the string into vibration. Commonly these are plucking (harp, mandolin), striking (piano) and bowing (violin). Each of these methods in turn permits considerable variation in the nature of the sound output depending on where the initial disturbance is produced. When a string is plucked at the centre, for example, all the even harmonics are absent, seriously affecting the musical quality of the emitted sound. Plucking near one end restores all the harmonics and improves the quality. The nature of the plucking instrument also is of importance. The struck string involves more elaborate considerations, since the impact of the hammer is a complicated function of the time. One general result of interest is more or less obvious: all harmonics having a node at the point where the string is struck will have zero amplitude. Bowing is the most complicated of all ways of exciting a string, since the communication of energy to the string is more or less continuous. The oscillation is really a special kind of forced vibration which may be properly called self-maintained. These considerations belong properly to the theory of musical instruments.

The transmission of energy to a string by bowing is not precisely like the forced vibrations considered in section III.2 above, since the force in this case is not harmonic. The string certainly abstracts energy from the bowing, but the frequency of vibration is conditioned by the string itself and its mounting, as well as the position of the bow. The resulting oscillations are often termed auto-oscillations or self-maintained oscillations. Their mathematical study can be facilitated by treating the force as a kind of negative damping. Another acoustic illustration is provided by the blown organ pipe (jet-edge tones; *see* section III.11 below). The electrical oscillations produced by an ordinary simple vacuum tube circuit are also of this character.

The actual output of sound energy from a vibrating string depends on its surroundings; *i.e.*, the presence of other resonating systems such as air cavities and solid baffles.

The Helmholtz resonator is a good example of a resonating air cavity. In its ideal form it is simply a hollow sphere with a circular opening (with a diameter of the order of one-tenth that of the sphere) and usually equipped with a very short neck. A source of sound brought near the resonator can set the air in the opening in vibration. The action then is effectively that of a simple vibrating system of mass equal to that of the moving plug of air in the orifice, of stiffness provided by the elasticity of the air cushion inside the sphere and of resistance largely due to the radiation of

sound energy from the opening. Analysis indicates that the appropriate elements are respectively

$$\begin{aligned} m &= \rho_0 S^2/C \\ f &= \rho_0 c^2 S^2/V \\ R &= \rho_0 \omega^2 S^2/2\pi c \end{aligned} \qquad \text{(III.4-4)}$$

In the above expressions ρ_0 is the density of the air, V the volume of the resonator sphere, S the cross-sectional area of the opening, C the acoustic conductivity of the opening and c the velocity of sound in air. The resonance angular frequency of the resonator is approximately

$$\omega_{res} = c\sqrt{C/V} \qquad \text{(III.4-5)}$$

The amplification A of the resonator is defined to be the ratio of the square of the excess pressure amplitude in the opening to that present at the same point in the absence of the resonator. At resonance this has the form

$$A = 4\pi^2 V/C^3 \qquad \text{(III.4-6)}$$

Resonators serve as useful acoustic amplifiers (for single frequencies, of course) for numerous types of sound sources and receivers (*see* section IV.3 below).

5. Vibrations of Membranes.—A vibrating circular membrane is a more practical source of sound of high intensity than a vibrating string, as naturally follows from its two-dimensional character. The equation of motion of a perfectly flexible membrane stretched with surface tension S (force per unit length across the membrane) and surface density ρ_s (mass per unit area) is of the form

$$\frac{S}{\rho_s}\left(\frac{\partial^2\xi}{\partial x^2}+\frac{\partial^2\xi}{\partial y^2}\right)=\ddot{\xi} \qquad \text{(III.5-1)}$$

if the equilibrium position of the membrane is in the xy plane and ξ denotes a displacement perpendicular to this plane. If the membrane is circular it is more simple to replace x and y by the polar co-ordinates r and θ and write (1) in the form

$$\frac{\partial^2\xi}{\partial r^2}+\frac{1}{r}\frac{\partial\xi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\xi}{\partial\theta^2}=\frac{1}{c^2}\ddot{\xi} \qquad \text{(III.5-2)}$$

with the velocity

$$c=\sqrt{S/\rho_s} \qquad \text{(III.5-3)}$$

The solutions of the differential equation (2) are complicated and will not be given here. It is of interest, however, to note the normal modes of oscillation of a membrane of radius a rigidly clamped around its periphery, since many membrane sound sources are used in this way. The three lowest harmonics have frequencies

$$\nu_1=\frac{0.77c}{2a},\ \nu_2=\frac{1.22c}{2a},\ \nu_3=\frac{1.76c}{2a} \qquad \text{(III.5-4)}$$

The fundamental is, of course, ν_1. It will be noticed that the higher modes do not form a harmonic series in the sense of the characteristic frequencies of the vibrating string or tube. The standing waves corresponding to the normal modes of the clamped vibrating membrane have nodal lines similar to the nodal points in the string or tube. These are both circles and radii whose existence can be experimentally demonstrated by covering the vibrating membrane with lycopodium powder which tends to collect along the nodal lines.

For practical purposes it is often sufficiently accurate to consider a circular membrane clamped at the periphery and vibrating in its fundamental mode as equivalent to a vibrating piston in which all points have the same displacement and velocity at the same instant. This piston may be thought of as possessing effective mass m and effective stiffness f, so that its resonance frequency is given by equation (III.2-15). If one uses the approximation that in the fundamental mode the shape of the displaced membrane is approximately paraboloidal, it turns out that the effective mass is one-third of the actual mass and that the effective stiffness is 2π times the surface tension S. The fundamental frequency given by (III.2-15) in this case is only about 1% higher than $0.77c/2a$ mentioned above.

6. Longitudinal Vibrations of Rods.—For a straight rod of uniform cross section, the length of which is very large compared with its diameter, the differential equation of motion for a small extensional disturbance ξ (*i.e.*, directed along the rod) is

$$\ddot{\xi} = Y/\rho_0 \cdot \partial^2\xi/\partial x^2 \qquad \text{(III.6-1)}$$

in which Y is Young's modulus and ρ_0 is the actual volume density. The disturbance thus travels along the rod as a longitudinal wave with velocity

$$c=\sqrt{Y/\rho_0} \qquad \text{(III.6-2)}$$

For most hard solids this velocity is of the order of 3,000 m. per second, though it ranges from about 1,000 m. per second to 6,000 m. per second, depending on density and elasticity.

The specific acoustic resistance for a longitudinal wave in a solid rod comes out to be $\rho_0 c$, as in the case of compressional waves in a fluid.

A finite vibrating rod, serving as a source of sound, necessarily has to satisfy certain boundary conditions. These will lead to the production of standing waves precisely as in the case of air vibrations in a tube (section II.5 above). Thus if the rod vibrates so that both ends are rigidly clamped, we must have $\xi = 0$ at $x = 0$, l (if l is the length of the rod). This leads to the normal modes or characteristic frequencies

$$\nu_n = nc/2l \qquad \text{(III.6-3)}$$

the fundamental corresponding to $n = 1$.

On the other hand if one end of the rod only is clamped and the other end is left free, the corresponding frequencies are

$$\nu_n' = (n+\tfrac{1}{2})\,c/2l \qquad \text{(III.6-4)}$$

A common case is that in which the rod is clamped at the middle. The fundamental frequency is still given by $c/2l$, since this provides for a displacement node at the centre. However, it is clear that in this case the even harmonics are absent since they do not have nodes at the centre, and rigid clamping at the centre must always correspond to a node there.

The modes of oscillation of a longitudinally vibrating rod are usually studied by sprinkling the rod with lycopodium powder which gathers at the nodes, thus making possible the measurement of wave length and through the knowledge of the driving frequency ultimately the longitudinal wave velocity. If the transverse dimensions of the rod are not negligible compared with the length, wave transmission along the rod is much more complicated and the simple formula (2) no longer holds. We realize that when an elastic rod is stretched there also must occur a change in the cross-sectional area. This leads to radial vibrations and the coupling between these and the longitudinal variety may be expected to lead to complicating effects, especially when the wave length of the longitudinal wave is of the order of magnitude of the transverse diameter. For very low frequencies with corresponding long wave lengths the coupling between the two types of modes is weak and the velocity (2) is observed. But as the frequency increases, say to 200 kc. for a rod in which the diameter is 1 cm., the effective longitudinal velocity drops noticeably from the value given by (2). This is a type of dispersion; *i.e.*, change of velocity with frequency. It may be called configurational, since it is associated with the finite transverse dimensions of the rod rather than with internal molecular effects. As the frequency is further increased, say to 500 kc. in the case of the rod just mentioned, the velocity falls asymptotically to the value for shear or torsional waves in the rod, viz., $c=\sqrt{\mu/\rho_0}$, where μ is the rigidity or shear modulus. Further increase of frequency produces little change. A complete explanation of these effects demands a complicated analysis of elastic wave propagation in the solid rod as influenced by the appropriate boundary conditions.

Just as in elastic wave propagation in a fluid, so also in a solid there are dissipative effects which lead to absorption of the wave energy with a loss of intensity. In analogy with the discussion in section II.13 above, we can write for the displacement from equilibrium in a harmonic longitudinal disturbance propagated along the rod

$$\xi = \xi_0 e^{-\alpha x}\cos(\omega t - kx) \qquad \text{(III.6-5)}$$

where $\omega(=2\pi\nu)$ is the angular frequency of the wave and α is the absorption coefficient. For most solid rods at all frequencies above a few cycles per second, α is not proportional to the square of the frequency as in the case of liquids and gases but more nearly approximates a linear dependence. For polycrystalline media (with metallic grains) this is corrected by an additional term depending on the fourth power of the frequency. Thus for polycrystalline aluminum, W. P. Mason and H. J. McSkimin found

$$\alpha = a_1\nu + a_2\nu^4 \qquad \text{(III.6-6)}$$

with $a_1 = 0.845 \times 10^{-9}$ sec./cm., $a_2 = 3.74 \times 10^{-30}$ sec.^{4}/cm. The linear dependence on the frequency is thought to have its origin in internal friction or solid viscosity (which evidently must operate in a different way from the viscosity of fluids), whereas the fourth power dependence is associated with a scattering of the elastic wave by the metallic grains of which the polycrystalline material is composed. This scattering is analogous to that which in the case of light is considered to account for the blue colour of the sky. The above formula was found to work well in the frequency range from 2 to 15 mc. per second.

Equation (5) expresses the absorption of sound in a longitudinal harmonic wave in a solid in terms of the amplitude decay with distance of travel. The absorption may also be expressed in terms of decay in time. This is indeed a more natural way to express it in a finite rod, set freely vibrating by some energy source and then set free from the energy source and allowed to die down. In this case the mathematical expression for the displacement at any place x at time t in a progressive wave in the positive x direction is

$$\xi = \xi_0 e^{-\beta t} \cos(\omega t - kx) \qquad \text{(III.6-7)}$$

in which β is a kind of temporal absorption coefficient. At any particular point on the rod (definite value of x) the displacement behaves like a damped harmonic oscillator (*see* section III.2 above). The logarithmic decrement is

$$\delta = \beta P = \pi/Q \qquad \text{(III.6-8)}$$

where P is the period of the wave $(=1/\nu)$. We have already introduced the quality factor Q in section III.2. Still another way of referring to the attenuation is to use the average fractional loss of energy W in the radiation per unit volume for one cycle of the harmonic disturbance. Recalling that the energy per unit volume is proportional to the square of the amplitude, we see that

$$W = 2\beta P = 2\delta \qquad \text{(III.6-9)}$$

It can be proved that independent of the precise attenuating mechanism we can also express the fractional energy loss in the form

$$W = 2\alpha\lambda \qquad \text{(III.6-10)}$$

where λ is the wave length. This makes it possible to express the decrement and Q value in terms of the absorption coefficient α.

Besides damping due to internal friction, there is also a contribution to the attenuation due to the friction drag of the surrounding fluid medium on the vibrating surface of the rod. This is generally smaller than the internal effects. The fact that the rod radiates sound energy to the surrounding medium also accounts for loss of energy in the vibrations. This effect, however, is also very small except at very high frequency. In other words, at low frequencies a vibrating rod is not a very efficient source of sound, unless it is coupled to a membrane or diaphragm with a larger radiating surface.

In the 1940s and 1950s there was great interest in the study of acoustical transmission in solid rods. Studies were conducted, mainly at ultrasonic frequencies, for the purpose of learning more about the physical properties of the solid material, *e.g.*, its elastic constants. Thus equation (2) permits the evaluation of Young's modulus Y from a knowledge of the density ρ and the velocity c of longitudinal waves in the rod, provided the frequency used is such that the wave length is considerably larger than the width or diameter of the rod (if circular). High-frequency sound is particularly suitable for this purpose, since the wave length can be long enough to satisfy the requirement just stated and still short enough to permit its measurement by a powder pattern on the rod produced by standing waves.

For optically transparent solids the determination of elastic constants is easiest by an optical technique based on the Debye-Sears effect. If a transparent solid cube is filled with an ultrasonic standing wave pattern produced by a piezoelectric quartz oscillator (*see* section III.11) cemented to one of its faces, and a beam of light is then passed through it, an optical diffraction pattern is formed. If the solid is isotropic the pattern consists of rings which elasticity theory readily associates with the bulk and shear elastic moduli as well as Poisson's ratio σ. For transparent crystalline solids the same method is applicable, though the patterns are more complicated.

For the study of sound transmission through solid rods and other extended solids, pulse techniques (section II.16) are often more practical than standing wave patterns, especially if the solids are optically opaque. With appropriate electronic equipment it is possible to produce ultrasonic pulses of only a few milliseconds duration and measure with great accuracy their transit time through a solid specimen; in this way velocity can be readily obtained. It must be remembered, indeed, that it is the group velocity which is obtained in this way. The method is very simple in principle but demands many precautions against errors if high precision is sought.

Sound attenuation in a solid can also be measured by the pulse method by noting the relative amplitudes of the initial pulse and that returning to the source after a double transit of the specimen under study. When rubberlike materials are investigated by this method, considerable velocity dispersion is found. Thus for natural rubber at 0° C. the velocity ranges from about 1,800 m. per second at 0.1 megacycle to about 2,700 m. per second at 10 megacycles. The dispersion decreases as the temperature is raised. Attenuation measurements in such materials indicate the presence of relaxation mechanisms similar to those already mentioned in liquids and gases.

In the 1950s considerable interest developed in the study of the properties of solids at very low temperatures, *i.e.*, in the region between 20° K. and absolute zero, available experimentally through the use of liquid helium. Particular attention has been paid to those solids which become superconducting at low temperatures, such as lead, tin, indium, etc. These metals appear to lose all their electrical resistance at temperatures around 4° K. Thus tin becomes superconducting below 3.73° K. its so-called "transition" temperature. If ultrasonic radiation (*e.g.*, at 10 megacycles) is passed through a rod of such a superconducting metal, the attenuation above the transition temperature increases as the temperature decreases until the transition temperature is reached, when a sharp discontinuity in slope appears, and as the temperature is lowered the attenuation decreases rapidly. On the other hand, if a magnetic field is imposed on the specimen, the superconductivity is destroyed and the attenuation continues to increase as the temperature falls below the normal transition point. It thus appears that the acoustic attenuation is due to a mechanism closely related to the mechanism responsible for the superconductivity. It should be remembered that a metal crystal consists of a more or less permanent and stable lattice of positive ions and a collection of readily mobile, more or less free conduction electrons which are responsible for the electrical conductivity of the metal. The lattice ions are indeed never wholly motionless, but vibrate about their mean positions and above the transition temperature can exchange energy with the conduction electrons. When a high-frequency sound wave passes through the metal this energy exchange is increased, and this leads to attenuation. Evidently this cannot happen in the superconducting state. It was believed in the late 1950s that in this state it becomes more difficult to transfer energy from the lattice to the electrons and hence the attenuation mechanism ceases to work. This relation between acoustic attenuation and superconductivity is a good example of what is known as physical acoustics of the solid state.

At temperatures well above the superconducting transition and for nonsuperconductors there are other relaxation mechanisms

leading to acoustic attenuation in solids. These are connected with imperfections in the metal lattice of which dislocations (*i.e.*, displacements of one layer of metal ions relative to the adjoining one) are good examples.

7. Flexural and Torsional Vibrations of Rods.—Of equal and indeed possibly more practical importance than longitudinal vibrations of rods are the bending variety in which the propagated disturbance is a flexure at right angles to the length of the rod.

The disturbance in this case is no longer propagated in accordance with the simple wave equation (II.2–1). Nevertheless, it can be shown that the flexure can move along the rod as a simple harmonic wave provided that the velocity is a function of frequency; *i.e.*, the bending rod acts as a dispersive medium. The phase velocity is found to have the form (*see* H. Lamb, *Dynamical Theory of Sound*, p. 125 [1925])

$$c = \frac{kK}{\sqrt{1+k^2K^2}} \cdot \sqrt{\frac{Y}{\rho_0}} \qquad \text{(III.7-1)}$$

where Y is Young's (or the stretch) modulus of the rod and K is the radius of gyration of the cross-sectional area about an axis through its centre of symmetry normal to the plane of flexure. As usual, k is the wave parameter. Wave energy is transmitted by flexure along the rod with the group velocity (*see* section II.16 above).

$$U = 2c - \frac{k^3K^3\sqrt{Y/\rho_0}}{(1+k^2K^2)^{3/2}} \qquad \text{(III.7-2)}$$

For very low frequencies for which $k^2K^2 << 1$, c is a small fraction of $\sqrt{Y/\rho_0}$, the velocity of longitudinal stress waves in the rod, and U is very nearly twice c. On the other hand for very high frequencies for which $k^2K^2 >> 1$, $c \to \sqrt{Y/\rho_0}$ and U approximates the same value. To be sure, this neglects radial effects; *i.e.*, the change of cross section accompanying bending as well as extension. When these are taken into account, it is found that as the frequency is increased c rises steadily from zero to an asymptotic value equal again to the velocity of torsional waves in the rod ($\sqrt{\mu/\rho_0}$). This was experimentally verified, among others by S. K. Shear and A. B. Focke. (*Phys. Rev.*, 52:532 [1940].)

The greatest interest in the flexural vibrations of rods rests on their use as vibrators for sound sources. Here we have to do with a finite rod clamped in some specified way. Application of the appropriate boundary conditions yields the fundamental frequencies and overtones. A few samples are indicated here (*see* P. M. Morse, *Vibration and Sound*, 2nd ed., pp. 158, 162 [1948]).

1. Rod clamped at one end and free at the other
   Fundamental frequency $\nu_f = 0.5596\ K/l^2 \cdot \sqrt{Y/\rho_0}$
   First overtone frequency $= 6.267\ \nu_f$
   Second overtone frequency $= 17.55\ \nu_f$
2. Rod clamped at both ends
   Fundamental frequency $\nu_f = 3.56\ K/l^2 \cdot \sqrt{Y/\rho_0}$
   First overtone frequency $= 2.76\ \nu_f$
   Second overtone frequency $= 5.40\ \nu_f$

It is clear that the overtones are not harmonics.

A torsional wave results when a twist in a solid circular rod or tube at right angles to the length is propagated lengthwise. If the twist at distance x from a chosen origin is denoted by θ, *i.e.*, the angle through which a circular section is turned from its equilibrium position, the wave equation for the torsional wave becomes

$$\ddot{\theta} = \mu/\rho_0 \cdot \partial^2\theta/\partial x^2 \qquad \text{(III.7-3)}$$

where μ is again the shear modulus. The velocity of torsional waves is therefore

$$c = \sqrt{\mu/\rho_0} \qquad \text{(III.7-4)}$$

which is less than that for longitudinal waves, though greater in general than for flexural waves. The damping effect of the friction drag of the surrounding medium on flexural and torsional vibrations is much greater than on longitudinal vibrations. As a matter of fact the decay in torsional oscillations of a cylinder in a gas is commonly used to measure the viscosity of the gas.

Wholly aside from the dissipative attenuation produced in elastic waves in solid rods, it is possible to produce wave transmission which is selective with respect to frequency; *i.e.*, filtration, by the placing of appropriate equally spaced and iterated loads along the rod. Such structures are usually band-pass solid acoustic filters with the lowest frequency band a pass band.

8. Vibration of Plates and Diaphragms.—A membrane is assumed to have perfect flexibility, but a solid sheet or plate with its greater elastic constants provides as a rule much greater stiffness. Consequently, the velocity of propagation of elastic disturbances in a plate is usually greater than in a membrane. The theory of the vibrations of a plate is very difficult. We merely note here the normal mode frequencies of a circular plate of radius a clamped around the periphery. The fundamental frequency is

$$\nu_f = 1.62\ c^2/a^2 \qquad \text{(III.8-1)}$$

where

$$c = [Yl^2/12\rho_0(1-\sigma^2)]^{\frac{1}{4}} \qquad \text{(III.8-2)}$$

and l is the thickness of the plate, ρ_0 the density of the material of which it is made, and σ is Poisson's ratio (the ratio of the lateral contraction to lengthwise elongation of a bar of the material composing the plate, with a value ranging from 0.2 to 0.4 for most solids). For example, for an iron plate for which $a = 2.2$ cm. and $l = 0.02$ cm., $\nu_f \approx 10^3$ cycles per second. Actually the higher characteristic frequencies are not harmonic. Thus the first few are given by P. M. Morse (*Vibration and Sound*, p. 210) as

$$\nu_{11} = 2.09\ \nu_f,\ \nu_{21} = 3.43\ \nu_f,\ \nu_{02} = 3.91\ \nu_f,$$
$$\nu_{12} = 5.98\ \nu_f$$

Comparison with the corresponding case of the flexible membrane shows that the overtones for the plate are more widely spread out than for the membrane.

The vibration theory of plates is so difficult that it is often customary modern practice to replace the circular diaphragm by an equivalent piston vibrator, all parts of which move in phase. This technique, which was mentioned in section III.5 in connection with membranes, is strictly justified only when the diaphragm vibrates below its fundamental frequency, so there are no nodal lines or circles. This is further discussed in the following section in connection with radiation from piston sources.

9. Radiation Efficiency of a Sound Source.—If a vibrating system is to serve as a source of sound it must set the surrounding medium (a liquid or a gas) into vibrating motion and hence radiate sound energy. It seems clear that the radiation efficiency of such a source will depend on the properties of the medium as well as those of the source itself. This is most simply understood for the ideal case of a spherical shell vibrating radially. If the outside radius of the shell is a (with surface area $S = 4\pi a^2$) and it vibrates harmonically with angular frequency ω, it may be shown by a study of spherical waves that the radiation from the sphere exerts a reaction force on it which has two principal effects: (1) to increase the effective mass of the vibrating sphere; and (2) to increase its effective damping coefficient. Thus, if first we consider the sphere as an oscillator in a vacuum, it will be characterized by a mechanical mass m, a damping factor R and a stiffness f. As a result of the fact that it is actually radiating into a material medium of density ρ_0, its mass is effectively increased by the amount

$$m_r = Z_2 S/\omega \qquad \text{(III.9-1)}$$

and its damping coefficient is effectively increased by the amount

$$R_r = Z_1 S \qquad \text{(III.9-2)}$$

In these expressions

$$Z_1 = \frac{\omega\rho_0 k a^2}{1+k^2a^2} \qquad \text{(III.9-3)}$$

and is called the specific radiation resistance at the surface of the sphere. Also

$$Z_2 = \frac{\omega\rho_0 a}{1+k^2a^2} \qquad \text{(III.9-4)}$$

and is called the specific radiation reactance at the surface of the sphere. They are the components of the specific acoustic radiation

impedance, in the electrical analogy already discussed in section II.11. The quantity m_r in (1) is usually referred to as the radiation inertia, while R_r is called the radiation resistance.

For low frequencies or long wave lengths for which $ka << 1$ the radiation inertia becomes the mass of a volume of the medium three times the volume of the vibrating sphere and hence can be appreciable (affecting, for example, the resonance frequency of the vibration), while the radiation resistance is relatively small. On the other hand for high frequencies or short wave lengths ($ka >> 1$) the radiation inertia becomes negligibly small and the radiation resistance approaches $4\pi a^2\rho_0 c$. The specific radiation resistance indeed becomes $\rho_0 c$ or precisely the value for a plane wave. The result of these considerations is that for given velocity amplitude the pulsating sphere is a more powerful sound radiator at high frequencies than at low, for the dissipation of energy from the radiator in the form of radiation increases with the frequency as the behaviour of the radiation resistance shows. This is one of the reasons for the interest in ultrasonics or high-frequency sound.

The pulsating sphere usually is not a practical sound source. However, similar considerations to those above apply to the common membrane or diaphragm source used in loud-speakers, and lead to the result that for sound sources of all kinds, other things being equal, radiation efficiency is greater at high frequencies than at low. Moreover, at all frequencies the average rate of radiation depends directly on the density of the medium into which the source is radiating. This means it is increasingly difficult to get radiation from any source into a medium of much smaller density. This is the real explanation of the famous bell-in-evacuated-jar experiment, rather than the common statement that sound is not transmitted through a "vacuum"; sound will still travel without abnormal dissipation in gases at low pressure, the principal difficulty being to get the sound into the gas in the first place.

10. Beams From Sound Sources.—A relatively sharp beam can be produced from a light source by the use of appropriate devices such as mirrors and lenses. Can a beam of sound be similarly produced? Common experience indicates that most ordinary audible sound sources radiate, though perhaps not equally well, in all directions. This is particularly noticeable in the human voice. However, ultrasonic or high-frequency inaudible sound sources can

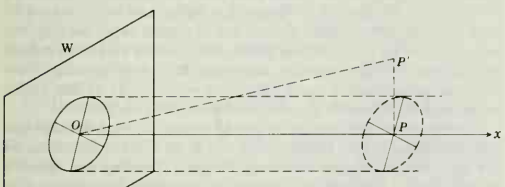


FIG. 16

produce beam radiation. We can see in a qualitative way how this comes about by considering a plane wave of sound traveling in the x direction and striking the infinite rigid baffle W placed at right angles to the direction of propagation. Let a hole of radius a be cut in the baffle with centre at O on the x axis, so that some of the sound can go through. By Huygen's principle the sound field to the right of the barrier will depend on the wave front at the hole. In other words, every point in the hole will act effectively as an individual source of sound, sending out waves whose superposition at any point, *e.g.*, P, will determine the intensity of the sound there. In calculating this superposition we have to consider that the effect of each part of the sound disturbance in the hole at the point P will depend on the distance of this part from P. The reason is not simply that intensity falls off with distance, but more particularly that wave disturbances reaching P from different parts of the hole can arrive out of phase with each other. Thus if the distances to P from two points of the hole differ by a half wave length or any odd multiple thereof, two disturbances starting out together from these points will be effectively a half wave length apart on arrival and hence will cancel each other out or, put in more technical language, will interfere destructively with each other. On the other hand, if the two distances in question differ by a whole number of wave lengths, the disturbances will arrive in phase and constructive interference will ensue; *i.e.*, there will be a reinforcement of the resulting disturbance at P. The mathematical analysis is somewhat involved, but we can see qualitatively at any rate that if the wave length is small compared with the diameter of the hole, the magnitude of the radiated disturbance will tend to be larger in the right circular cylinder whose axis is the x axis, and that at points like P' well outside this cylinder the intensity will be decidedly less and grow more so as P' recedes from the axis. This conclusion will not follow if the wave length is large compared with the size of the hole. In fact, if it is large enough radiation will tend to spread more or less uniformly in all directions to the right of the baffle; *i.e.*, act effectively as a hemispherical wave. Thus with short wave length or high-frequency radiation we can actually form beams of sound something like beams of light, though since even high-frequency sound has frequencies much smaller than those of visible light, ultrasonic beams are by no means so sharp as light beams.

The above considerations hold also if the hole is filled with a vibrating piston source of sound. If the amplitude of the velocity at each point of the piston has the constant value ξ_0, and the wave length λ and radius a of the piston are connected by the relation $a = m\lambda$ (where m is any number), the intensity at any point P on the x axis distant x from the piston is given by

$$I = 2\rho_0 c\xi_0^2 \sin^2(\pi m^2\lambda/2x) \qquad \text{(III.10-1)}$$

Thus the intensity on the axis runs through a succession of maxima and minima over the region for which $m^2\lambda > x > m\lambda$. However, when $x > m^2\lambda$, I effectively falls off inversely as the square of the distance. This shows that no matter what the wave length is, cylindrical beam radiation will not persist indefinitely far from the source. Rather there always ensues conical divergence after $x = m^2\lambda$. We can calculate the order of magnitude of the conical or solid angle in this case. It is π/m^2. More exactly, for x twice the effective diameter of the piston the circle at the periphery of which the intensity falls to 1/10 that at the axis in the same plane will subtend at the centre of the source a solid angle corresponding to a plane angular spread of 2 arc tan $(.45\lambda/a)$. Thus for a piston source with $a = 10$ cm. emitting radiation of frequency 50 kc. the plane angular spread of the beam is about 15°.

An interesting special case is that of a concave piston with radial motion of its parts. Here there exists the possibility of getting focused radiation with consequent increased intensity at some points.

11. Practical Sound Sources.—*Voice and Speech.*—The human voice operates by the forcing of air from the trachea between the vocal cords, which then vibrate. This in turn sets into vibration the air in the cavities of the throat and mouth and the resulting disturbance emerges from the lips. The part that the vocal cords play in the production of musical and ordinary speech sounds has been clearly established by motion pictures of their vibration in human subjects.

Work on the acoustical power output of conversational speech (for example, that by H. Fletcher) showed that in the case of an average speaker the over-all level directly in front of the mouth can vary from 81 db. at 15 cm. from the mouth to 65 db. at 100 cm. These decibel levels are relative to the standard reference level (adopted by the American Standards association) of 10^{-16} w./cm.^{2}. The mouth is not precisely a point source giving rise to spherical waves, for at 100 cm. directly back of the head the intensity in the case mentioned was 62 db. This effect is due to diffraction by the head and is indeed present to a greater or lesser extent in all sound sources which have appreciable size.

The power in speech sounds varies greatly with the nature of the sound, being much larger for the vowels than for the consonants. On the other hand, articulation in speech is due primarily to the consonants. The nature of the ordinary speech sound is very com-

plex since it does not consist of a simple sinusoidal wave, but of a superposition of many waves of different frequencies. Moreover, the recognition of speech by the ear is not at all as simple as the study of ordinary sound sources by suitable receivers.

Sirens.—One of the simplest sources of intense sound is the siren, which in principle consists of a disk containing a number of small holes placed at regular intervals around the same periphery. If this is rotated past an orifice lined up with the holes and compressed air is driven through the orifice, the result will be a succession of puffs of air. If the number of puffs per second (governed by the speed of rotation of the disk) is sufficiently large the result will be a periodic sound whose frequency is equal to the number of puffs per second. By varying the rotational speed of the disk, variations in frequency can be produced. Sirens have been built to produce both sinusoidal and nonsinusoidal outputs. In one type of siren acoustic power outputs as high as 176 w. with an air pressure of 0.2 atm. have been achieved in the frequency range from 3 kc. to 19 kc. (about 150 db. relative to the minimum audible level at 3 kc. at a distance from the siren of 25 cm. on its axis). By stepping up the air pressure to 2 atm., a power output of 2,000 w. becomes possible with an over-all efficiency of 20%.

Jet-Edge Tones and Organ Pipes.—Another standard sound source operated by blown air is the jet-edge system coupled to an air column as in the organ pipe. In the familiar type of pipe a narrow jet of air impinges on a narrow edge of the bottom of the pipe. The resulting vibrations are coupled to those of the air in the pipe. The actual frequency of the sound emitted depends on the velocity of efflux of the air. The natural frequency of the edge tone uncoupled to the pipe increases approximately linearly with the velocity, but the coupling forces the fundamental tone of the pipe on the system for small velocities. When the flow velocity increases to the value which would correspond to a natural isolated edge tone of frequency equal to the second harmonic of the pipe, the emitted tone rises to this harmonic and the pipe is said to be "overblown." The theory of edge tones is rather uncertain, though their production is clearly associated with the presence of turbulence (vortex motion) in the air flow in the vicinity of the edge.

Closely allied to the jet edge as a producer of sound is the Hartmann oscillator in which a high-velocity jet is directed at the orifice of a small bottle-type resonator. Usually the output is mainly ultrasonic radiation of high intensity.

Electroacoustic Sources.—The most important sound sources in practical use are those which involve the employment of electrical energy. These may be divided into two main classes: irreversible and reversible. In sources of the first class the energy for the sound radiation is supplied by a flow of fluid (air or water) under pressure, and the electrical energy is used for control purposes. For example, a variable magnetic field can produce vibrations in a reed or diaphragm which in turn will control the flow of compressed air through variable orifices. Such a device can produce very intense sound with high efficiency (*i.e.*, high acoustic energy output relative to the electrical and mechanical energy input). In this respect it resembles the siren, though it is not a single frequency source.

A reversible electroacoustic source transforms electrical energy directly into mechanical energy and thence into acoustical radiation. As its name implies, it can be employed in the reverse manner to transform acoustical energy into electrical via some mechanical structure. Such sources can therefore be used equally well for receivers. There is indeed a theorem known as the reciprocal or reciprocity theorem (*see* section II.14 above) which states that under rather carefully specified conditions a good reversible source of sound will be a good receiver. It was shown by W. Schottky that the ratio of the reception efficiency to the emission efficiency decreases with the square of the frequency. This means that low frequencies are relatively better received than emitted.

Any system that transforms electrical energy to acoustical energy is called an electroacoustic transducer, whether used for emission or reception. The important characteristics of a transducer used as a source are: (1) its acoustical output; (2) its efficiency; and (3) its departure from linearity, which means the extent to which the mechanical driving force producing the vibrations is not strictly proportional to the electrical field or voltage applied in the input. Such lack of proportionality or nonlinearity leads to distortion in the radiated output. Another characteristic which is often significant is the directivity; that is, the intensity of the radiation in different directions from the source.

All electroacoustic transducers involve the production of electrical voltage which ultimately produces an oscillating mechanical force operating on a mechanical vibrator; *i.e.*, a rod or diaphragm. The force can be produced in a variety of ways. In the electromagnetic transducer, an alternating current of the frequency desired for the sound output is passed through a coil which actuates an electromagnet and produces an alternating force on a diaphragm made of iron or some magnetic alloy. This is the type employed in the normal telephone. In the more common electrodynamic transducer, the oscillating current passes through a coil suspended in a permanent magnetic field and experiences the usual ponderomotive force. The coil (called the voice coil) is fastened mechanically to the radiating surface. In the well-known loud-speaker the latter is usually a cone attached to the voice coil at or near its vertex. The cone is the acoustical element, radiating directly to the surrounding medium. In earlier types the conical diaphragm was coupled to the air by means of a flaring horn, a method that continues to be used in large speakers for auditorium use. Mounting the cone on a baffle or in a cabinet can also materially improve its radiating efficiency. As already noted in section III.8, the cone as a diaphragm is a resonant system and this can lead to distorted output at certain frequencies unless mechanical or electrical filters are employed. Certain compromises can be reached with respect to the size of the conical diaphragm. If it is small, the number of resonance frequencies in the audible range is small; on the other hand it takes more power to drive it for a desired output because of the small radiating surface. A larger cone is a more efficient radiator, although it naturally has more resonance frequencies. If the latter are numerous enough they may be distributed over the usable frequency range so as to give good reproduction. All the general properties of acoustic radiators discussed in section III.9 above apply here. Other things being equal, the efficiency of a dynamic loud-speaker varies directly as its linear dimensions. For speakers in common use (radio sets, etc.) the efficiency runs from 2% to 5%. By the use of horns this figure can be increased to 25% or even 50%. These figures are of course dependent on the frequency range. In a typical small cone speaker 10 cm. in radius, the efficiency has the value 2% only over the frequency range from 75 cycles per second to 1,000 cycles per second. At 10,000 cycles per second its efficiency has dropped practically to zero.

The total acoustical power output of a dynamic loud-speaker can have maximum values ranging from 0.5 w. to 150 w., depending on size and the associated electrical circuit. Driving the speaker at high power involves the danger of nonlinear behaviour; *i.e.*, large amplitude motion of the diaphragm which no longer is proportional to the electrical voltage.

Another type of reversible electroacoustic transducer is the electrostatic or condenser type. In this an alternating electrostatic field of large intensity (*i.e.*, many volts per centimetre) is set up between a fixed metal plate and a thin metal diaphragm separated from it by a very small distance. The diaphragm is set into vibration by the alternate attraction to and repulsion from the fixed plate. By proper choice of mounting and dimensions this transducer can be made to give rather uniform output over a wide frequency range. However, its radiation efficiency is low except at very high (ultrasonic) frequencies, and it is more commonly used as a receiver (microphone) than a radiator.

When high-frequency radiation is desired as in ultrasonic research, the most important electroacoustic radiators are based on the piezoelectric (sometimes called electrostrictive) and magnetostrictive effects. Certain asymmetric crystals like quartz and Rochelle salt have electrical charges of opposite sign appear on opposite faces when they are subjected to dilatational stress. This is the piezoelectric effect. In the converse effect if such a crystal is made the dielectric in a condenser subjected to an alternating electrical field the crystal will vibrate with the frequency of the

field. When a plate of quartz is cut with its faces parallel to the optic axis of the crystal (so-called X cut) the highest fundamental frequency will correspond to vibrations in the direction of its thickness. This frequency is given approximately by the empirical formula

$$\nu = \frac{286}{d} \times 10^3 \qquad \text{(III.11-1)}$$

in cycles per second, where d is the thickness in centimetres. This shows clearly the availability of the quartz oscillator as an ultrasonic radiator. By making the thickness small enough, very high frequencies may be generated. Moreover, harmonics of the fundamental can provide even higher frequencies, and indeed this plan is often used, since there is a practical lower limit to d. The resonance in a quartz crystal is very sharp.

For ultrasonic radiation usually a number of X-cut crystals are arranged in the form of a mosaic so as to provide a larger radiating surface. The piezoelectric effect can also be used in a loud-speaker in the audio range. In this case it is customary to cement two or more crystal plates together and arrange the electrodes so that the whole structure will bend or twist when electrically excited. Large amplitudes can be obtained in this way, although the system is not so practicable as the electrodynamic speaker.

For the vacuum tube oscillators and associated circuits used in driving piezoelectric crystals *see* E. G. Richardson's *Ultrasonic Physics*, ch. i (1952). Another source of information about piezoelectric oscillators in acoustics is W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (1950).

The magnetostrictive effect is the change in length of a rod of magnetic material (iron, nickel or ferromagnetic alloy) when the magnetic field to which it is exposed varies in magnitude. If an alternating current passes through a coil surrounding such a rod, the latter will vibrate lengthwise with a frequency double that of the current if the rod was originally unmagnetized and with a frequency equal to that of the current if the rod has some magnetization to start with. If the frequency of the current coincides with one of the resonant modes of the rod, particularly large vibrations result and sound of this frequency is radiated. As an example, a nickel rod 10 cm. long emits a fundamental frequency of 24.3 kc. Magnetostrictive sources are less efficient than piezoelectric largely because of the loss of energy due to the hysteresis in the magnetized rod. This can be avoided to a certain extent by using as a sender laminated nickel sheets fastened together as in the core of a transformer. Such transducers are used in producing underwater sound beams, though usually somewhat above the audible range.

12. Doppler Effect in Acoustics.—In the discussion of sound sources in the preceding sections, it has been tacitly assumed that the source is at rest relative to the medium. There are, however, obvious cases of moving sources which are important in application; *e.g.*, airplanes, submarines and projectiles. A significant effect of such motion on the reception of sound is the apparent change in frequency, in accordance with the Doppler principle. If a harmonic source of actual frequency ν moves with velocity v_s with respect to a stationary medium along the straight line joining the source to a receiver which is itself moving with velocity v_r along the same line, the receiver will estimate the frequency as ν' where

$$\nu' = \nu(c - v_r)/(c - v_s) \qquad \text{(III.12-1)}$$

where c is the velocity of sound. Here v_r and v_s are reckoned as positive in the direction from the source to the receiver and negative otherwise. For the special case of a receiver at rest and the source alone in motion, this says that the apparent frequency as measured at the receiver is less than the actual frequency if the source is receding from the receiver, and is greater than the actual frequency if the source is approaching the receiver. Knowledge of ν and measurement of ν' can thus lead to the determination of the velocity of the source, a matter of some significance in military applications.

IV. RECEPTION OF SOUND

1. Introduction.—In previous sections the production of sound by various sources and the mechanism of its propagation through material media have been discussed. To be of ultimate human interest, it is essential that the propagated sound energy be detected; *i.e.*, transformed in some way so as to make it perceptible to one of the senses. It is the function of a sound receiver to do this, and any scheme for transforming the energy of elastic deformation of a medium into another form of energy can serve this purpose. It should be clear that every type of sound source can in principle be used as a sound receiver. Nevertheless receivers are worthy of special attention because of the variety of uses to which they are put in practice.

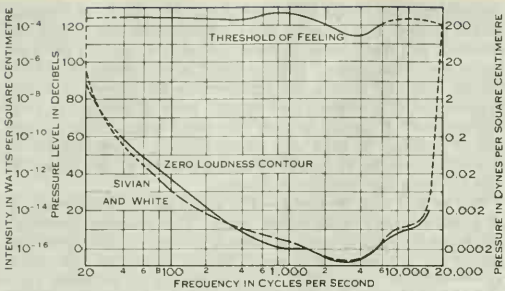
2. The Ear and Hearing.—The ear is the most important of all sound receivers to the human being. This somewhat tautological statement gains in meaning when it is reflected that the normal human ear is competent to detect sounds of intensity as low as 10^{-16} w./cm.^{2}, while it can stand intensities up to about 10^{-4} w./cm.^{2} before pain ensues. The actual figures depend on the frequency, and those cited represent minima and maxima respectively over the range of audibility, which runs approximately from 16 cycles per second to 20,000 cycles per second. The presence of two ears leads to the binaural effect, or the ability to detect the direction of approach of the sound from a distant source. Even when damaged or diseased the ear can often still be used successfully as a sound receiver by the employment of an auxiliary hearing aid, a product of acoustical research.

A complete account of the anatomy of the ear will be found in Ear, Anatomy of. Here we merely note the outer ear or auricle, the influence of which on "collecting" sound is negligible for most audible frequencies and becomes noticeable only for frequencies over 10 kc., the ear canal terminated by the eardrum or tympanum, leading into the middle ear with its ossicles, and finally the inner ear with the cochlea.

Sound passing through the ear canal sets the eardrum into vibration, and this motion is communicated via the ossicles (the hammer, anvil and stirrup; a kind of solid acoustic filter) to the oval window leading into the inner ear, filled with liquid, the most important part of which from the standpoint of hearing is the cochlea, a spiral cavity with two and three quarter turns. In the cochlea is the flexible basilar membrane which can vibrate under the impact of motions of the cochlear fluid. On top of the basilar membrane is the organ of Corti containing fine hairs which project into the cochlea. These hairs are nerve terminals connected with the auditory nerve.

There have been numerous theories of hearing; *i.e.*, of the action of the various organs in the ear when a compressional wave strikes the tympanum and the brain recognizes the sensation of sound. The more plausible of these are generalizations of the original idea of Helmholtz that the fibres terminating in the hair cells on the surface of the basilar membrane act like resonators tuned to various frequencies over the audible range. Experiments indicate that the stimulation of the nerve endings in the basilar membrane produces electrical potential impulses of the same frequency as the exciting sound radiation. These travel along the auditory nerve to the brain. In the so-called space pattern theory of hearing, the temporal pattern of the sound impulse is transformed into a space pattern in the basilar membrane and the nervous impulses transmitted from the variously stimulated parts of the membrane tell the brain, so to speak, what the sound is like. In the time pattern theory it is assumed that the temporal sequence of disturbance in the radiation is communicated directly to the brain as if the basilar membrane acted much like an electroacoustic transducer. Evidence seems to favour a combination of these two points of view.

The thresholds of hearing and feeling as a function of frequency are shown in fig. 17, based on the original work of H. Fletcher but confirmed by numerous other investigators. The lower curve is a plot against frequency of the minimum sound intensity necessary to stimulate hearing. It is seen that this rises to very large values at the two ends of the audible frequency range and has a minimum at about 3,000 cycles per second. The upper curve is the corresponding plot of the threshold of feeling; intensities greater than those along this curve produce pain. The total audible range is therefore that included between the two curves. It must be emphasized that individual ears show considerable varia-

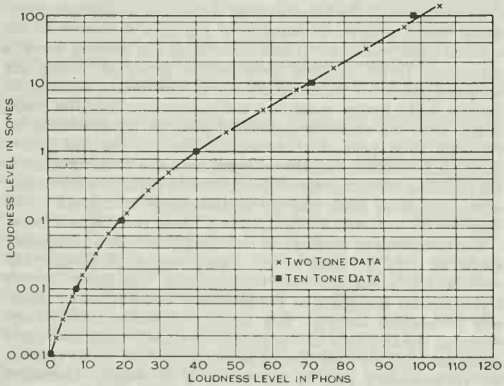


FROM "SPEECH AND HEARING IN COMMUNICATIONS" BY HARVEY FLETCHER. COPYRIGHT D. VAN NOSTRAND COMPANY, INC., 1953

FIG. 17.—AUDITORY AREA BETWEEN THRESHOLD OF FEELING AND THRESHOLD OF HEARING

tion, so that the curves presented represent averages of what is supposed to be a fairly representative youthful population. Advance in age results in general in hearing loss, which would be measured in the diagram on fig. 17 by a rise in the lower curve, though its shape in general remains roughly the same. The tendency is for the upper frequency limit to decrease. In fig. 17 two lower curves are actually presented but for purposes of this article the difference between them may be neglected. *See* Harvey Fletcher, *Speech and Hearing in Communication*, 2nd ed. (1953).

Loudness is a sensation which, though connected with the intensity of the sound incident on the ear, is not directly proportional to it. The relation has been the subject of long and somewhat controversial study which is not surprising, since while intensity is an objective physical quantity, loudness is subjective and any definition of it must necessarily apply to the "average" listener. Any method of measuring loudness, and there are several, involves somehow the subjective estimate of the relation between sounds of given intensity. In one common method a reference tone of variable intensity is provided and raised to such a level that it appears to the listener to be as loud as the sound under test. The number of decibels by which the reference tone has been raised above the minimum audible threshold of 2×10^{-4} dyne/cm.^{2} is said to be the loudness level (intensity) in *phons*. This applies, strictly speaking, only if the reference tone has the frequency 1,000 cycles/sec. and is at a distance greater than 1 m. from the head of the listener. This choice of frequency for the reference tone is arbitrary but dictated to a certain extent by the fact that it is in the midrange of audible frequencies and its range of auditory intensity is a maximum (fig. 17).



FROM "SPEECH AND HEARING IN COMMUNICATIONS" BY HARVEY FLETCHER. COPYRIGHT D. VAN NOSTRAND COMPANY, INC., 1953

FIG. 18.—RELATION BETWEEN LOUDNESS LEVEL (INTENSITY) IN PHONS AND THE TRUE LOUDNESS LEVEL IN SONES

The unit of loudness is the *sone*, which is defined to be the loudness produced by a tone at 40 decibels above the standard reference level (the minimum audible threshold as given above). By the statistical study of numerous experiments in hearing, a loudness scale in sones has been established. One version of the relation between the loudness level (intensity) in phons and the true loudness in sones is shown in the graph in fig. 18. A survey by S. S. Stevens in the mid-1950s indicated that the best interpretation of the statistical data yielded the following relation between the loudness, L (in sones), and intensity, P (in phons),

$$\log_{10} L = 0.03P - k \qquad \text{(IV.2-1)}$$

where k is a constant depending on the frequency. For a reference tone of 1,000 cycles/sec., $k = 1.2$. This means that L is proportional to the intensity raised to the power 0.3.

Hearing with two ears leads to the binaural effect mentioned above—the ability to detect the direction from which sound is approaching the head. Thus a source of sound on the right of the plane bisecting the line joining the ears and perpendicular to it will be recognized as on the right. It was originally believed that this was due to the difference in intensity at the two ears, but this theory proved to be an inadequate explanation of the localizing ability. Rather, the binaural effect is due to difference in phase at the two ears. This effect has proved very useful in sound detectors for gun ranging and underwater ranging.

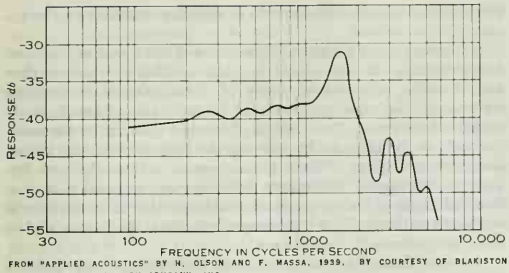
Since the eardrum is an asymmetrical vibrator (*i.e.*, its displacement for given applied pressure is not the same in the positive as in the negative direction), the ear can hear summation and difference tones. If two tones of frequencies ν_1 and ν_2 are presented to the ear and if $\nu_1 - \nu_2$ is small compared with either frequency, the usual phenomenon of beats ensues; *i.e.*, the intensity of the resulting sound fluctuates $\nu_1 - \nu_2$ times a second. On the other hand, if ν_1 and ν_2 are rather widely different a new tone of frequency $\nu_1 - \nu_2$ will be heard. So will a tone of frequency $\nu_1 + \nu_2$. The difference tone is sometimes called a beat frequency tone. And in fact tones of frequency $j\nu_1 \pm k\nu_2$, where j and k are arbitrary small integers, have been detected. These effects are all predicted theoretically. (*See* also HEARING.)

3. Microphones.—Any physical system which is competent to transform a pressure change at a point in a material medium into some other form of physical disturbance, *e.g.*, electrical or optical, can be used as a receiver of sound. An early example is the sensitive flame, formed by burning the gas issuing at high speed from a very small orifice and keeping it just below the point of flaring. This responds by jumping when sounds of rather high frequency are incident on it. In the phonodeik of D. C. Miller the incident sound vibrates a fine membrane which communicates its motion to a movable mirror from which a spot of light is reflected.

Practically all modern sound receivers are electroacoustic transducers, which when used for reception are known as microphones (sometimes termed hydrophones when used in water). We have already noted in section III.11 that reversible transducers may be used interchangeably as sources and receivers. When used as a microphone such a transducer transforms pressure changes into alternating electrical currents which can be amplified and rendered visible on a cathode-ray or other type oscillograph.

One of the important properties of any microphone is its sensitivity or response—defined as the ratio of the electrical output (usually measured in terms of open-circuit voltage) to the excess sound pressure incident on the microphone. The free-field response is that which refers to the sound pressure existing at a definite point in the sound field before the microphone is inserted. The term pressure response is sometimes used to refer to the case in which the pressure is that acting directly on the microphone when it is in place. These differ because of the sound diffraction by the microphone, which serves to distort the sound field prevailing before the insertion of the microphone. The response is commonly expressed in terms of decibels relative to some arbitrary reference level, which is often chosen to be 1 v. per dyne/cm.^{2} excess pressure.

Since the microphone diaphragm possesses mechanical elements its vibration depends on the frequency of the incident radiation



FROM "APPLIED ACOUSTICS" BY N. OLSON AND F. MASSA, 1939. BY COURTESY OF BLAKISTON CIV., MCGRAW-HILL BOOK COMPANY, INC.

FIG. 19.—RESPONSE FREQUENCY CHARACTERISTIC OF SINGLE-BUTTON CARBON MICROPHONE FOR CONSTANT SOUND PRESSURE OF ONE BAR IN FREE SPACE. O DB = 1 V. OPEN CIRCUIT. RESISTANCE OF BUTTON = 100 OHMS

(*see* section III.2 above) and the same is true of the response. A typical response *v.* frequency curve for the simple carbon microphone used widely in telephony is shown in fig. 19. In this figure zero decibel corresponds to 1 v. per dyne/cm.^{2} Hence -40 db. means 40 db. in intensity level below 1 v. per dyne/cm.^{2} Since power is proportional to the square of voltage, the number of decibels corresponding to the voltage ratio e_1/e_2 is $20 \log_{10} e_1/e_2$. Hence -40 db. relative to 1 v. per dyne/cm.^{2} is 10^{-2} v. per dyne/cm.^{2} From the figure the response stays around this value from 100 to 1,000 cycles per second. Between 1,000 and 2,000 cycles per second it rises to between -30 db. and -35 db. and thereafter falls off sharply. The uneven response is due to the various modes of vibration of the microphone diaphragm.

The response discussed here has reference to the normal incidence of a plane acoustic wave. Another important characteristic of a microphone is its directivity pattern. This is the plot of the free-field response at a given frequency as a function of the angle of incidence of the radiation. If a microphone is completely nondirectional, the directivity pattern for every plane will be a circle. On the other hand the ideal pattern for the so-called ribbon microphone (*cf.* below) has the appearance of fig. 20. In the x or $-x$ direction the response is a maximum, falling to zero or near zero in the y or perpendicular direction.

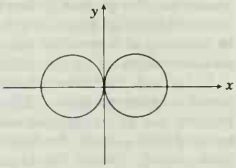


FIG. 20

Electroacoustic microphones may be divided into two main classes: pressure and pressure gradient or velocity microphones. The carbon microphone referred to above is a good illustration of the pressure type. Another is the condenser microphone, already mentioned in section III.11 as the electrostatic or condenser transducer. By appropriate mounting of the condenser diaphragm a very uniform response characteristic with frequency can be obtained, though the level is considerably below that of the carbon microphone and for practical use the instrument needs an associated amplifier to increase the output voltage. Other examples of the pressure type are the various forms of crystal microphones and the dynamic or moving conductor microphone, analogous to the electrodynamic loud-speaker. In the pressure gradient type of microphone, the electrical voltage produced corresponds to the difference in pressure between two points in the instrument. When this distance is a small fraction of the acoustical wave length, the pressure gradient corresponds to the particle velocity in the medium. Hence such a microphone, in which a loosely stretched metallic ribbon is suspended in the air gap between two magnetic pole pieces, may be considered a velocity microphone.

The ribbon vibrates under the influence of the pressure difference produced by the sound wave on its two sides. Its vibration in the magnetic field induces an electromotive force between its two ends. Its response is best at low frequency. Its directivity pattern has already been presented in fig. 20, where the x axis is normal to the plane of the ribbon.

Another interesting example of a velocity microphone is the hot-wire variety. In this a fine wire is heated electrically. When a sound wave impinges on it the cooling effect changes the electrical resistance in two ways, viz., (1) a steady change and (2) an alternating change of twice the frequency of the sound radiation. The sensitivity is materially increased by mounting the wire in the opening of a Helmholtz resonator (*see* section III.4 above).

Great increase in the directionality of a microphone may be obtained by mounting it at the focus of a suitably large parabolic reflector.

Sound received by a microphone can be recorded in a variety of ways. The voltage produced can be fed directly into a suitable voltmeter via an amplifier and measured directly in volts or in decibels. Special continuously recording meters have been devised for this purpose. Often it is desirable to reproduce the signal from the microphone in direct visual form and in this case a cathode-ray oscilloscope may be used. This is particularly helpful when the radiation being received is in the form of a succession of pulses. Considerable electronic circuitry is usually employed in modern acoustical reception techniques. (*See* also MICROPHONE.)

V. APPLICATIONS OF SOUND

1. Sound Measurements.—All acoustical phenomena can be studied through the measurement of certain fundamental quantities. These are velocity, frequency, sound pressure or intensity and impedance. To this list might be added the analysis of complex sounds although this really involves the measurement of the previously mentioned quantities performed in a special way. Common methods of measuring these quantities are discussed below.

Velocity.—The velocity of sound in air was the earliest acoustical quantity to be measured. Marin Mersenne is usually credited with making the first estimation (1640) by timing with a pendulum the interval between the uttering of a sound and the return of the echo from a distant reflector. He found the value 316 m. per second; there is no record of the mean temperature. This is an example of what may be called the direct method of velocity measurement, which essentially reduces to the careful measurement of the time it takes a given sound to travel a known distance. Another illustration is provided by the cannon flash and sound scheme used by many investigators in the late 17th and early 18th centuries. By this method G. A. Borelli and V. Viviani in Florence in 1660 obtained the value of the velocity of sound in air as 361 m. per second. The personal equation involved in this scheme rendered its accuracy questionable. Moreover, all results of measurement of the velocity in air are subject to variation due to atmospheric conditions of temperature, chemical composition, wind and humidity. Automatic time recording was introduced by H. V. Regnault and F. P. le Roux in the 1860s in the course of measuring the velocity in air confined in pipes with rigid walls. This velocity is always somewhat less than that in the free atmosphere.

The direct method has also been applied to the measurement of velocity in liquids and solids. In the 20th century the accepted method employs sound signals in the form of pulses produced by magnetostriction or piezoelectric oscillators. In the simplest scheme a radio frequency pulse of, say, 6 mc. per second and around 10 microseconds in duration excites a quartz crystal which radiates a corresponding acoustic pulse into the medium which it traverses until it strikes a reflector placed at a given distance from the crystal. In the meantime another pulse is emitted by the crystal. If the first echo coincides with this pulse, the time for the first pulse to travel to the reflector and back equals the time between successively emitted pulses. Since the latter time can be accurately measured (the so-called pulse repetition rate) the velocity is readily determined simply as the distance traveled by the pulse divided by the time.

In a more elaborate variant of this method the emitted pulse is received at a second crystal placed at a fixed distance from the emitting crystal, and is there reconverted into an electric signal and sent back through an electronic circuit to the oscillator driving the emitting crystal. There it triggers another pulse which repeats the process, driving a succession of pulses around the loop consisting of an electronic path and an acoustic path. Each trigger-

ing pulse is counted and the total time for a large number of round trips is measured by an electronic clock. The total time for the given number of pulses to traverse the acoustic path yields the velocity. Of course the time delay in traversing the electronic path must be measured or calculated. It can indeed be eliminated by having the same train of pulses pass through two acoustic paths of the same material but of different lengths. With a method of this kind velocity measurements in liquids and solids can be carried out to an accuracy of 1 part in 10^5. It must be emphasized that it is the group velocity which is measured by this scheme. (*See* section II.16 above.)

Among indirect methods of measuring the velocity of sound may be cited first the use of standing waves. From section II.15 above it follows that if a standing wave vibration mode is set up in a column of gas, liquid or solid of a fixed length the distance between two successive nodes is a half wave length. Measurement of this quantity together with measurement of the frequency of the vibration leads at once to the velocity through the product of the two. The familiar Kundt's tube is an illustration of the application of this scheme in gases. Here the nodes are indicated by piles of cork dust which arrange themselves in layers when the tube is excited to a particular resonance frequency. In solids the nodes are usually indicated by patterns made with lycopodium powder.

Another example of the indirect method is provided by the acoustic interferometer. Sound produced usually by a piezoelectric source is radiated continuously into a fluid medium and strikes a movable reflector placed accurately parallel to the source. It is then reflected back to the source. As the distance between source and reflector is varied the current in the plate circuit of the electronic oscillator driving the source varies periodically, the period corresponding to changes in the source-reflector distance equal to one-half wave length. From a knowledge of the oscillator frequency the velocity is then immediately determined. This method is usually confined to liquids and gases. It has been brought to a high degree of accuracy, although probably it is not susceptible of quite the same precision as the pulse technique. (*See* INTERFEROMETER.)

An interesting optical technique for measuring sound velocity was devised by P. Debye and F. W. Sears, who discovered that a train of plane sound waves progressing through a liquid can act as a diffraction grating for a beam of monochromatic light traveling through the liquid at right angles to the direction of sound propagation. The distance between the lines of the acoustic grating, l, is the wave length of the sound and this can be determined by the usual grating formula for the diffraction angle θ of the light:

$$\sin\theta = \pm\lambda_0/2l \qquad \text{(V.1-1)}$$

where λ_0 is the wave length of the light. The method has been applied to transparent liquids and solids. It can even be applied indirectly to optically opaque solids.

Frequency.—Since harmonic waves are the most used in practice, the measurement of frequency is of cardinal importance. The problem is one of creating a frequency standard with respect to which any periodic sound can be compared.

One common frequency standard is a tuning fork driven by a vacuum tube circuit. The fork can be calibrated by having it drive a synchronous motor. The latter in turn rotates a stroboscopic disk illuminated by a flashing lamp controlled by a modulated radio signal, as from the station maintained by the national bureau of standards in Washington, D.C. Here the final control lies in the radio transmitter whose accuracy ultimately goes back to the measurement of time. (*See* TIME MEASUREMENT.) The accuracy of the radio signals is estimated to be better than 1 part in 10^7. They are modulated to produce sounds of 1,440 and 4,000 cycles per second respectively, and can be heard throughout the day at radio carrier frequencies of 10, 15, 20, 25, 30 and 35 mc.

In practical cases it is sufficient to compare the sound of unknown frequency (assumed to be a pure tone; *i.e.*, of single frequency only) with a calibrated oscillator capable of producing sounds of any frequency in a given range. It is usually possible to guess the approximate frequency of the unknown. The oscillator is then set to produce a sound close to this frequency. From the beats associated with the two sounds the unknown can be accurately determined.

Still more precise is the transformation of the sound of unknown frequency into an equivalent electrical oscillation which is then placed across one pair of plates of a cathode-ray oscilloscope while across the other pair is placed an oscillation of continuously variable frequency. The resulting Lissajous pattern on the oscilloscope can be adjusted by varying the frequency of the known electrical oscillation. When it becomes an ellipse the two frequencies are equal. The carrier frequency of the usual radio station ordinarily is not so precisely controlled that it can safely be employed to monitor the frequency of electrical oscillators used to drive electroacoustic transducers.

Acoustic Intensity.—It has already been pointed out that the intensity of acoustic radiation may be given either in terms of the acoustic excess pressure p_e or the particle velocity ξ in the medium (section II.8 above). Hence the measurement of either of these quantities suffices to yield the intensity. Properly calibrated pressure or velocity microphones can therefore serve as intensity meters. However, they must be checked against absolute standards. An absolute standard for the determination of particle velocity is the Rayleigh disk, which is a light circular plate suspended by a fine fibre in the path of the sound radiation, usually in a tube. If the equilibrium position of the disk in the absence of sound is such that the normal to its plane makes an angle θ with the axis of the tube, the torque on the disk in the presence of sound for which the mean square particle velocity is $\overline{\xi^2}$ becomes

$$L = \frac{4}{3}\rho_0 a^3\,\overline{\xi^2}\sin 2\theta \qquad \text{(V.1-2)}$$

where a is the radius of the disk and ρ_0 is the mean density of the medium. Usually θ is chosen as 45° and the torque, measured by a torsion head on the suspension, suffices to give $\overline{\xi^2}$ and hence the intensity in absolute measure.

Among numerous direct attempts to measure acoustic excess pressure may be mentioned the pinhole probe of C. Barus. He found that if a pinhole orifice in the shape of a cone is inserted in an air column at a pressure loop and the other end is connected to one arm of a sensitive manometer, the excess pressure can be measured directly by an optical interferometer technique.

E. Gerlach devised a sensitive electrical method for measuring acoustic excess pressure. He compensated the pressure on a diaphragm due to an incident sound wave by measurable electrodynamic forces so that the diaphragm remains motionless under the joint action. Since the required electrodynamic effect can be measured very accurately this provides a precise measure of sound pressure.

In addition to the oscillating acoustic excess pressure in a sound wave, to which reference has so far been confined, sound also produces a radiation pressure on any object on which it impinges (section II.17 above). As has already been mentioned this is a steady pressure like that produced by light and other electromagnetic radiation, and has a magnitude directly proportional to the intensity of the radiation. Actually in practice it is necessary to distinguish between the so-called Rayleigh radiation pressure and the Langevin radiation pressure. Suppose the mean equilibrium hydrostatic pressure in a fluid at rest is p_0. If $\bar{p}$ is the time average pressure in this fluid due to the passage of a sound wave, then $\bar{p} - p_0$ is the Rayleigh radiation pressure. For an elastic liquid it is zero, whereas for an ideal gas it becomes $(\gamma + 1)/2.\ I/c$, where I is the intensity of the sound, c its velocity and γ the ratio of the specific heats of the gas. By the Langevin radiation pressure is meant the difference between the average acoustic pressure at a wall or obstacle (whether absorbing or totally reflecting) and that in the medium at rest behind the obstacle. Calculation shows this to be equal to I/c in every case (section II.17 above). This is the radiation pressure which is usually measured experimentally, as for example in the balance method of measuring acoustic absorption in liquids, in which a vertical beam of sound impinges on an obstacle attached to one arm of a balance and can be thereby directly "weighed." Differences in radiation pressure at different points in a sound beam yield corresponding differences in sound

intensity and hence to the decay of intensity with distance and ultimately to the absorption coefficient.

The absolute measurement of the amplitude of particle displacement under the action of a sound wave would also provide an accurate measure of intensity. However, this is a difficult task. E. N. C. Andrade attempted to do this by observing the motion of smoke particles in a sound field, and predicting from the theory of the motion of a solid particle suspended in a viscous fluid the relation between the motion of the fluid (*i.e.*, the sound-induced motion) and the corresponding motion of the smoke particle. Experiments made with this method show that sound amplitudes can be measured thereby to an accuracy of about 2% up to a frequency of 5 kc.

The reciprocity theorem already mentioned in sections II.14 and III.11 can be used in the calibration of a microphone as an acoustic pressure measuring device. Let A and B be any two points in a homogeneous isotropic medium of very large extent (a so-called free field) and let a transducer T be located at A while a receiving microphone is placed at B (*see* fig. 21). If an E.M.F. E applied to T produces an acoustic excess pressure p_e at R, and if at the same time an acoustic source of strength Q (volume current flow X as in section II.11 above) placed at B produces a current I in the transducer at A, the following equation holds, having due regard for consistent units,

FIG. 21

$$E/p_e = Q/I \qquad \text{(V.1-3)}$$

With the use of this relation and the assumption that the transducer and hydrophone are actually identical instruments, one can obtain the open circuit sensitivity of this device in terms of purely electrical measurements; *i.e.*, current and voltage. This constitutes the calibration of the instrument as an intensity-measuring device. Since it is rare to have identical transducers, three transducers A, B and C are employed in practical calibration. A is used only as a transmitter, B only as a receiver and C is used once as a transmitter and once as a receiver. Three measurements are performed using the pairs AB, AC and CB respectively (the first named is always the transmitter). By successive application of the reciprocity theorem the sensitivity of each transducer can be determined in terms of purely electrical quantities.

Acoustic Impedance.—This quantity, usually denoted by Z, was defined in section II.11 above as the ratio (generally complex) of the acoustic pressure to the acoustic volume current. Its analogy with electrical impedance renders it a useful quantity in acoustical calculations since a knowledge of it in every part of an acoustical system makes it possible to replace the latter with an equivalent electrical circuit or transmission line. This is common practice in acoustical engineering. Knowing the impedance of any element in an acoustical circuit, *e.g.*, that of a side tube or orifice in an acoustic transmission line like a horn or filter or that of the termination of a tube as acoustical absorbing material, is equivalent to knowing what is practically important about the acoustical behaviour of the element. The usual electroacoustic analogy referred to here relates acoustic excess pressure to electromotive force and acoustic volume current to electric current. This is the one commonly employed in acoustical engineering problems. However, it sometimes proves advantageous to use the inverse analogy in which excess pressure is taken to be analogous to electric current and volume current to electromotive force. This plan was introduced by F. A. Firestone and has been applied successfully, particularly to magnetostrictive elements.

Strictly speaking, the direct way of measuring Z would be to measure the pressure and volume current independently at the same place. This is rarely practical, nor is it necessary. In most methods advantage is taken of theoretical expressions giving the impedance at a point in the medium, that is, at some acoustical absorbing material, in terms of the pressure in phase and magnitude at various other points. This is the standard practice in the so-called transmission line method to measure the impedance of material placed at the end of a long tube or that of a branch line inserted at a point in the side of the tube. Often the pressure measurements are those of maximum and minimum values in the standing wave setup in the impedance tube.

2. Recording and Reproduction of Sound.—The recording and reproduction of sound as exemplified by the various types of phonographs and other sound recording devices became a major industry, dating from Thomas A. Edison's first invention of the cylindrical drum phonograph record in 1877. But the phonograph is only one application of sound recording, which covers all cases in which a transient or permanent record of a given sound sequence is desired in a form which is suitable for "playing back." Actually we should include in sound recording the transformation of sound into a visible record as in an oscilloscope, whence it can be preserved by photographic techniques. The analysis of complex sounds has been carried out in this way.

Three types of recording systems are in common use: (1) mechanical, as in the disk phonograph; (2) optical, as in the sound film; and (3) magnetic, as in the tape or wire recorder. In the mechanical system the record is cut into a rotating disk of wax or foil with a metal underlay by means of a cutting stylus with a sapphire point. The stylus is usually driven by electroacoustic coupling with the source of the sound being recorded; *e.g.*, piezoelectric or electrodynamic. The record then consists of a spiral groove extending from the circumference of the disk to the centre. In most cases the stylus is actuated to vibrate transversely to the groove. Some experimentation continues to be carried out, indeed, with the "hill-and-dale" or vertically vibrating cutter originally employed in Edison's phonograph.

In the standard 78 r.p.m. records there are roughly four grooves to the millimetre and the maximum excursion from the average line is about 0.06 mm. In the long-playing 33⅓ r.p.m. microgroove record there are about eight grooves to the millimetre with the maximum excursion of 0.035 mm. From the original cutting a metal reproduction is made, and after various processes, copies can be turned out in large numbers on shellac or plastic disks.

In the reproduction of the sound from the disk record, the disk must be rotated at the same speed at which the recording was made and a needle must follow the groove. This needle is part of a pickup, almost always electroacoustic in character; *i.e.*, transforming mechanical vibrations into electrical oscillations and then back again into mechanical vibrations in a loud-speaker. An important requirement in both cutting the original groove and in reproducing the sound from the disk is the uniformity of frequency response to avoid distortion in the process. The modern high-fidelity phonographic recording system provides fairly uniform frequency response over the range from 25 to 10,000 cycles per second.

All optical systems of sound recording employ the modulation of the intensity of a beam of light by means of the sound. A photographic record of the light then constitutes the sound track —the basic plan used in sound motion pictures. The sound may either modulate the brightness of the light source (rarely used) or control the beam electromechanically by a chopper. The sound track may be either of the variable density or variable area type. In sound movies it is usually placed directly on the film. Reproduction is carried out by illuminating the sound track by a beam of light which, after passing through the track, falls on a photocell whose output after amplification is fed to a loud-speaker.

In magnetic recording a steel wire or film strip coated with a paste of iron oxide powder is moved past an electromagnet (the so-called sound "head") through which oscillatory current is passed modulated by a sound wave picked up by a microphone. The current through the head magnetizes the wire or tape and the remanent magnetization provides a more or less permanent record of the sound. For reproduction the tape is passed by another sound head in which it induces a variable current which when amplified can be fed into a loud-speaker. The record on the wire or tape can be readily "erased" by using a larger demagnetizing field. The same tape can therefore either be stored as a record or used over again for the recording of different sounds. Though much used in commercial broadcasting and in many industrial operations, such as recording results of calculations on high-speed computers, the magnetic tape system had not by the latter 1950s competed successfully with the ordinary phonograph for home reproduction.

Since the hearing of sounds distributed in space, as for example the different instruments in a large orchestra, depends on the direction the sound takes (the binaural effect as discussed in section IV.2 above), consideration of direction is important in achieving the most natural reproduction of a composite sound from a distributed source. This is the aim of the so-called stereophonic reproduction.

The basic idea is illustrated by the experiment of picking up an extended sound source by several microphones spaced at suitable intervals and feeding the signal from each to a separate loud-speaker correspondingly spaced in the listening room. Even two microphones with two loud-speakers can achieve a much more natural effect than one of each. In the actual reproduction a sound track is produced by the sound from each microphone and then each played back through a separate loud-speaker.

Besides its importance in the entertainment field the recording of sound has many technical and scientific uses. Whenever it is desirable to preserve a visual record of a particular sound phenomenon, some sort of photographic or equivalent recording is made; the analysis of complex sounds has already been mentioned. Other instances are found in sound-ranging records in air and water (*see* section V.5 below).

3. Architectural Acoustics.—Architectural acoustics is essentially concerned with the production and reception of sound in closed spaces such as lecture halls, theatres, church auditoriums, school classrooms and living rooms in homes. For many years, even though the fundamental principles of acoustics were well understood and correctly applied in many ways, the subject was almost entirely neglected or handled in an unsatisfactory manner. The first attempt to apply the concepts of the science of acoustics was made around 1895 by W. C. Sabine of Harvard university in his study of the reverberation time, that is, roughly speaking, the time required for a given sound in a room, when its source is stopped, to die away or, more strictly, to be reduced to one-millionth part of its initial intensity. By careful experiments Sabine discovered the relation between this quantity and the volume of the room and the sound absorbing power of its surfaces. All modern architectural acoustics dates from Sabine's work.

Requirements for Good Room Acoustics.—The optimum acoustical qualities of a room depend on the acoustical use to be made of it. Thus the primary desideratum in a living room is that it be free of unnecessary noise which might impede normal conversation. In a lecture hall the requirement is that speech be heard articulately in every place where an auditor might sit, while keeping the intensity level at a reasonable figure. In a concert hall devoted to instrumental music the listener wishes sufficient intensity or loudness and at the same time desires to hear the contribution of the various individual instruments. In the case of a singer both articulation and intensity are desired. In both cases the sound should be rather evenly distributed or diffused through the room, so that there are no marked inequalities in reception from one position to another. Each room, depending on its acoustic use, has an optimum reverberation time. Moreover, each must be acoustically treated so as to control the normal modes of sound vibration associated with its geometrical shape (*see* section II.15 above). Methods based on the principles of acoustics set forth earlier in this article are available for meeting all these requirements and the acoustical design of buildings for all purposes can now be completely and reliably carried out before construction. (*See* also Acoustics of Buildings.)

Sound Distribution in a Room.—Mathematically speaking, the problem of the distribution of sound in a closed space arising from a source within that space is extremely complicated. It involves the solution of the acoustic wave equation (section II.2 above) subject to boundary conditions. This is impossible in closed form for any but very simple geometrical shapes. However, it is fairly easy to see physically what goes on. The source will excite the normal modes of the geometrical space in question just as blowing an organ pipe excites the characteristic frequencies or modes of the pipe.

For a room with its three dimensions these are naturally more numerous than those for a linear system like a pipe or finite string or even for a two-dimensional vibrator like a membrane. The simplest case is provided by a rectangular enclosure with sides equal to l_1, l_2, l_3 respectively. The characteristic frequencies or normal modes are then given by

$$\nu_n = \frac{c}{2}\sqrt{(n_1/l_1)^2+(n_2/l_2)^2+(n_3/l_3)^2} \qquad \text{(V.3-1)}$$

where c is the velocity of sound as usual and n_1, n_2, n_3 are any set of integers. The modes of lowest frequency correspond to small values of n_1, n_2, n_3. There are no harmonics in the sense of the vibrating string or organ pipe. Between any two given frequencies the number of modes increases with the average frequency, the dependence being as the square if the frequency is high enough.

Recalling the discussion of the forced oscillations of a vibrating system in section III.2, we see that after the sound source in the room is turned on, all the normal modes will be excited, but ultimately only those with frequencies equal to those of the source will continue to exist. The process of building up is complicated since there is interference between the sound reflected from the walls and other room surfaces (producing the normal modes) and the sound emitted directly from the source. Moreover, there is always some absorption at the room surfaces. This in fact is critical for the growth of sound. Obviously, if the reflection from the surfaces were complete, the sound energy in the resonant modes would increase indefinitely. To make the analytical problem simpler geometrical acoustics (*i.e.*, sound rays) is usually employed, and it is assumed that all ray directions are equally probable, so that the diffusion of sound in the room is perfect. Then it can be shown that the root-mean-square excess acoustic pressure in the room in dynes/cm.^{2} as a function of the time t from the turning on of the source is

$$p_{\text{ave}} = 1{,}300\ (W/a)^{1/2}(1-e^{-cat/4V})^{1/2} \qquad \text{(V.3-2)}$$

where W is the acoustical power output of the source in watts, V is the volume of the room in cubic feet, c is the velocity of sound in feet per second and a is the total amount of surface absorption in square foot units or sabins. We must now examine the absorption. If when sound impinges on a surface the fractional amount α is absorbed per unit of time, then α is called the absorption coefficient of the surface in question. Thus if 30% of the incident sound is reflected back into the room, the absorption coefficient of the surface material is 0.7. Each square foot thus acts as the equivalent of 0.7 sq.ft. of a perfect absorber. It is customary to use the "sabin" (after W. C. Sabine) as the unit of sound absorption at a surface. One sabin is the equivalent of one square foot of perfectly absorbing material. When α is known for each element dS of surface of the room, the total absorption of the whole room surface is given by

$$a = \int \alpha\, dS \qquad \text{(V.3-3)}$$

in sabins or square foot units, where the integral is taken over all the surfaces of the room. Actually it is customary in practical calculations to replace the integral by the sum of products, thus

$$a = \sum \alpha_i S_i \qquad \text{(V.3-4)}$$

where α_i is the average absorption coefficient over the surface S_i, and the sum is taken over all different surfaces of the room.

According to equation (2), the sound pressure in the room continues to grow for an infinite time before attaining the steady state value

$$p = 1{,}300\ (W/a)^{1/2} \qquad \text{(V.3-5)}$$

However for most rooms encountered in practice $ca/4V$ is considerably greater than unity (*e.g.*, for a cubical room of side l, $ca/4V = 3\bar{\alpha}c/l$, where $\bar{\alpha}$ is the average absorption coefficient) and hence the steady state is effectively attained in a few seconds, unless indeed the room is very large and the absorption very small as, for example, in a cathedral. The pressure in (5) is normally expressed in terms of the number of decibels for the equivalent intensity above the intensity corresponding to the standard base pressure (average minimum audible) 2×10^{-4} dynes/cm.^{2} This level then becomes in decibels above minimum audible

$$L = 10 \log_{10} (W/a) + 136 \qquad \text{(V.3-6)}$$

The same kind of analysis which leads to equation (2) indicates that if the sound source is stopped when the effective steady state has been reached, the pressure will decay in accordance with the equation

$$p = 1{,}300\ (W/a)^{1/2} e^{-cat/8V} \qquad \text{(V.3-7)}$$

The time taken for the pressure to drop to $\frac{1}{1{,}000}$ of its initial value (or 60 db. equivalent drop in intensity) is defined to be the reverberation time of the room. From equation (7) this becomes

$$T = 0.161\ V/a \qquad \text{(V.3-8)}$$

if V is expressed in cubic metres and a in equivalent square metres of absorption. If V is expressed in cubic feet and a in sabins the corresponding result is

$$T = 0.049\ V/a \qquad \text{(V.3-8')}$$

The two preceding formulas contain Sabine's law of reverberation time, first established by him on purely experimental grounds. Actually this law is not completely accurate, particularly for rooms with abnormally high absorption (what are called "dead" rooms). By considering the multiple sound reflections at the walls more carefully, C. F. Eyring was able to show that a more correct expression for the reverberation time is (using feet for dimensions)

$$T = \frac{0.05\ V}{-2.30\ S \log_{10}(1-\bar{\alpha})} \qquad \text{(V.3-9)}$$

where S is the total surface area of the room and $\bar{\alpha}$ is the average absorption coefficient. If $\bar{\alpha} << 1$, the expansion of the logarithm at once leads to equation (8') as the first approximation. However, if $\bar{\alpha} > 0.5$, as in a dead room (*e.g.*, a radio broadcasting studio), formula (9) is a better representation of the facts and has a definite economic value since it implies that much less absorbing material is needed to make a dead room of specified reverberation time than would be calculated by Sabine's formula.

It is important to emphasize that the reverberation formulas of both Sabine and Eyring depend on the assumption of uniform diffusion of sound in the room. Qualitatively what this means is clear, but it is desirable to have a concrete way of assessing it, particularly since complete diffusion is never realized in any actual room. Since if sound were perfectly diffused there would be no average fluctuation of intensity with frequency at any point, it was thought desirable to seek a measure of diffusion in the quantity

$$\psi = \frac{\Sigma p_{\max} - \Sigma p_{\min}}{n \Delta \nu} \qquad \text{(V.3-10)}$$

where $\Sigma p_{\max}$ denotes the sum of the pressure maxima in the frequency interval $\Delta\nu$, presumably covering the frequency range of greatest interest in the room in question, while similarly $\Sigma\, p_{\min}$ denotes the sum of the corresponding pressure minima. The number n is the number of pressure maxima in the frequency interval.

It is found experimentally that when diffusing material in the form of wall irregularities (like cylindrical bosses or straight wall sections meeting at obtuse angles near to 180°) is introduced into the room, leading subjectively to a feeling that the sound is better diffused, the quantity ψ becomes smaller than in the room without the diffusing treatment. It therefore seems likely that ψ may be used as a criterion of diffusion. Some authors refer to it as the diffusity of the room.

Optimum Acoustical Conditions.—The optimum acoustical conditions in a room depend considerably on its use. The reverberation time is almost always of primary importance. If a room is used mainly for speaking a relatively short reverberation time is indicated in order to ensure good articulation. With long reverberation the successive speech sounds overlap each other unless the speech is abnormally slow.

A standard articulation test is based on the understanding by an observer of meaningless monosyllabic sounds produced by a speaker. The percentage of sounds written correctly by the observer (suitably averaged over the room and preferably over several observers) is called the percentage articulation. Though simple it has proved very useful.

Experience indicates that if the articulation is 85%, about 97% of actually comprehensible speech will on the average be understood. If the articulation falls to 75%, 6% of ordinary speech will be misunderstood, and the hearing may be said to be good, though attentive listening is required. If the articulation falls to 65%, hearing conditions are bad, with nearly 10% of all speech being missed by the auditor.

Experiments in auditoriums which have a volume of the order of 2×10^5 to 5×10^5 cu.ft. indicate that the maximum reverberation time for satisfactory articulation is about 2.5 sec. Best listening conditions would indeed demand more nearly the figure 1 sec. It must be kept in mind, however, that these figures are somewhat idealized since they are based on the existence of a low general noise level (*i.e.*, about 30 db. above minimum audible) and assume good sound diffusion. In very large rooms (10^6 cu.ft. and up) the unaided voice of no speaker can be heard well no matter how low the reverberation time is. A public-address system is clearly required in such cases. Moreover, the situation alters if the room is used primarily for music. Generally the good hearing of music requires a longer reverberation time than the articulate hearing of speech. Thus for an auditorium of volume in the range between 2×10^5 and 5×10^5 cu.ft., music usually requires an optimum reverberation time of about 1.7 sec., with organ music running nearer to 2 sec. These figures are for a frequency of about 500 cycles per second. From the Sabine law it is known that since the absorption a generally increases with the frequency at any rate up to about 1,000 cycles per second (the precise dependence is complicated), the reverberation time varies inversely. Most reverberation times are quoted at 500 cycles per second. In general for lower frequencies they will be higher and for higher frequencies slightly lower (though often not appreciably so).

Although the reverberation time is usually treated as having primary importance for the acoustics of a room, the direct reflections of sound from various surfaces in the room also turn out to be of significance. Experiments indicate that such reflections reaching an auditor within the first 0.05 second after the sound coming directly from the source serve to prevent the sound from appearing too "directional" and not sufficiently "mixed." They can also serve to give the impression of greater loudness. Places in a room where such reflections reach an auditor seem to be acoustically preferred, though the complete reason for this was not understood in the mid-1950s. Of course reflected sound arriving too long after the direct sound produces unpleasant echoes and should be avoided.

Acoustic Absorbing Materials.—Since for a room of given size the total sound absorption is the primary factor controlling the reverberation time, numerous types of sound-absorbing materials have been manufactured. In general these depend for their absorbing character on their natural porosity (as in the case of cork products and textiles, for example) or on artificially made perforations or fissures. From the theoretical studies on sound absorption referred to in section II.13 above, it is evident that a hard smooth surface is a poor absorber compared with a rough one with many indentations and air pores. The problem is one of acoustical engineering and no attempt will be made here to present figures.

Sound-Amplification Systems.—Even when a room has been designed as carefully as possible for good acoustics, if it is very large (or if, in addition, good articulation must be available in every portion, as in a theatre), artificial amplification may be necessary to achieve satisfactory results. Sound-amplification systems are available for this purpose. Briefly, they consist of one or more microphones, appropriate electronic amplifiers and an array of loud-speakers. In a single-channel system all the microphones, though spread over the whole sound source area (*e.g.*, stage of a theatre) feed into the same amplifier which in turn operates all the loud-speakers, the latter being strategically placed so that the amplified sound appears to originate at the actual source. For rooms with a volume ranging from 5×10^4 to 5×10^5 cu.ft. the total acoustical power output in watts necessary for satisfactory reproduction of speech is approximately $V \times 10^{-7}$, where V is the

volume. This provides an idea of the amplification required in any specific case. The efficiency of the amplifier-loud-speaker combination must be known before the actual electrical power needed for the system can be computed.

More natural sound reception is achieved by a multichannel or stereophonic system in which each microphone feeds through its own separate amplifier into its own speaker. When properly installed this can ensure, for example, that sound originating from the left side of a stage will appear to the listener to be coming from that direction.

Noise Control in Rooms.—A vital factor in good room acoustics is the noise level, noise being here defined as unwanted sound whether intelligible or unintelligible. In order to use a room effectively for the intended acoustical purpose it is necessary to control the noise level, since noise tends to mask other sounds and render hearing fatiguing. If sufficiently intense and prolonged, it can impair hearing permanently and even perhaps have an effect on the nervous system.

In most rooms noise originates in three principal ways: outside sources, machinery and people. In a business office the last-named source can account for a large part of the noise, although office machinery, unless specially silenced, can be annoying. In warm weather much noise may come in through the open windows. Noise can also enter a room from other parts of the building and from the outside through the walls and partitions. Measurements show that the average noise level in the living room of an ordinary residence is about 43 db. without a radio turned on and 50 db. with the radio (higher in some cases). In a large office with more than ten desks the figure may run to 65 db. and in a factory 80 db. Outdoor noise usually originates in motor traffic, with aircraft becoming a serious problem near airports.

Noise originating inside a room, particularly that caused by people, can usually be controlled by appropriate sound-absorbing material. This is, to be sure, not necessarily true of machinery. Here special devices are necessary to reduce the vibration which results in sound radiation both to the air and to the floor. Noise coming from outside the room can be controlled only by proper construction of the partitions. Heavy walls transfer sound largely by vibrating as diaphragms. If massive enough they can act as adequate sound insulators. Actually, it is often just as effective to employ layered partitions of lighter material with air spaces in between, which act like acoustic filters (*see* section V. 7).

4. Musical Acoustics.—The scientific aspects of music make it a part of acoustics, though until relatively recent times, acoustical theory as developed by physicists had relatively little bearing on the production of music. As an illustration the problem of standard pitch may be cited. Though a conference of physicists in Stuttgart, Ger., in 1834 proposed that the musical note A should be standardized at 440 cycles per second, great variety in practice followed, and it was not until 1938 that this standard was adopted internationally. Undoubtedly the advent of large-scale radio broadcasting ultimately hastened this move.

Musical instruments are special sources of sound, and their behaviour can be analyzed by the methods set forth in section III above. They are rather complicated examples of coupled vibrating systems or, to put it in another way, vibrating systems of many degrees of freedom. They can also be considered as acoustical filters of greater or lesser complexity since when excited they in general emit only certain frequencies or combinations thereof.

Organ pipes are a special case of vibrating air columns, discussed in section II.15 above. The real nature of edge tones which excite the ordinary pipe is the subject of close study in physical acoustics (section III.11 above). The horn has been extensively studied for its physical characteristics; its behaviour is largely the result of its resonance frequencies plus the fact that the flare or increase in diameter in going from the throat end produces a better acoustic impedance match with the surrounding air and hence leads to more radiation from a source of given strength than a tube of uniform cross section and of the same length.

There has been much interest in the musical possibilities of electrophonic instruments. The simplest type is illustrated by the Theremin, which, is essentially a beat frequency oscillator connected to a loud-speaker through an amplifier. The capacity controlling the frequency is alterable by the motion of the hand (or a baton) relative to a metal loop. A continuous range of single frequencies is thus available. More elaborate is the Novachord or Hammond organ, a keyboard instrument in which each key controls a rotating wheel which in turn induces electromagnetically a current pulse in a coil. The various pulses are mixed and amplified and ultimately fed to a loud-speaker. A somewhat similar type of instrument employs variation in electrostatic capacity produced by a series of rotating disks tied in with the grid circuits of a set of vacuum tubes. Still another example is the photoelectric organ, in which light from a source of controllable intensity falls on the edge of a rotating disk bearing a sinusoidal pattern painted in black paint. After being chopped by the disk the light falls on a photocell and the resulting current after amplification is fed to a loud-speaker.

5. Military Acoustics.—Most military operations involve the production of sound, and with the increasing use of technology in the prosecution of warfare it is not surprising that many acoustical applications have been developed, particularly after World War I. The detailed weaponry which resulted, particularly during and after World War II, remained subject to security classification, but the basic ideas are available in published literature. Technical problems arising in these applications stimulated new research into fundamental properties of sound transmission. One example was the inquiry into the cause of abnormal absorption in liquids.

The principal use of acoustics in warfare has been in the detection of military objects at a distance. For practical purposes this is best divided into atmospheric and underwater applications.

Sound Ranging in Air.—If an impulsive sound from a source at O (*see* fig. 22) is received at microphones placed at A and B, knowledge of the velocity of sound and the differences in time of arrival at the two microphones is sufficient to establish the direction of the source. For if C is midway between A and B on the base line of the microphones, it follows that

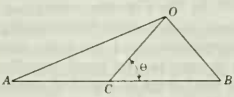


FIG. 22

$$\cos\theta = c(t_A - t_B)/l \qquad \text{(V.5-1)}$$

if l is the distance between A and B, t_A and t_B are the arrival times and c is the velocity of sound. It is here assumed that l is very short compared with either ct_A or ct_B. If another microphone is placed along the same base line another bearing on the source can be obtained and the intersection of the two yields the position of the source relative to that of the microphones; this serves to range the source and if the latter is the discharge of a gun, provides in principle a method of gun ranging. Various difficulties are inherent in the scheme, however. The projection of a shell from a gun involves ultimately three impulsive sounds: (1) the actual discharge of the gun; (2) the bow wave from the projectile which comes as a sharp crack; and (3) the sound from the bursting of the shell. Successful ranging of the gun demands the ability of the microphone to distinguish these sounds; hot-wire and condenser-type microphones with resonance in the region of low frequencies are employed for this purpose. Another difficulty is the uncertainty about the velocity of sound, which is dependent on temperature and humidity of the air and also on wind, leading to the necessity for concomitant meteorological measurements. Errors in location are reduced if more than three microphones are employed on a longer base line.

Acoustic detection of airplanes by means of large horns using the binaural effect (*see* section IV.2 above) was practised with some success in World War I, but abandoned in favour of short electromagnetic wave methods (radar) in World War II. Airplane noise, particularly from jet planes, is indeed a serious problem both from military and civilian points of view and has stimulated much research on high-amplitude sound (macrosonics—*see* sections II.2 above and V.6 below).

Underwater Sound.—Acoustics has been extensively applied to the detection of underwater objects like submarines. Here electro-

magnetic radiation is too rapidly damped (sea water being a good conductor of electricity) to be of much use, whereas the sea is a rather good transmitter of sound radiation. Moreover, since its characteristic impedance is much closer to that of solid sound sources than that of air, it is relatively easier to radiate sound into water than into air.

Two general methods of underwater detection may be employed. The first, like sound ranging in air, simply listens passively to the sound being produced by the object in question. If the latter is a ship or a submarine the sound comes mainly from the noise associated with the cavitation produced in the water by the propellers. Some engine noise will of course escape into the water via the hull and the latter will itself vibrate as a diaphragm or solid shell. The latter sound is mainly of low frequency, while the cavitation noise will have both sonic and ultrasonic components. For localization in passive listening the binaural effect is employed, usually with an array of underwater microphones termed hydrophones. The latter are usually of the magnetostrictive or piezoelectric variety. They are connected to an elaborate electronic circuit which permits phase differences of the sound at the various hydrophones to be compensated electrically so that the whole array can be "steered" in the direction of the sound source without actually rotating.

For actual location of an underwater object, which may itself be producing little or no sound, an echo method was devised, similar in principle to the echo-sounding scheme in which an ultrasonic pulse is emitted straight down into the sea from near the surface. A certain portion of it is reflected from the sea bottom and returns to the surface to be picked up by a receiver there. The latter may indeed be the same transducer as the sender. From the time difference between the direct and reflected pulse and a knowledge of the velocity of sound in sea water, the depth may be calculated, or read off a specially designed recorder. When applied to the detection of a distant object this method is known as sonar (or asdic in Great Britain). If the emitted pulse is of high frequency, it travels out in a more or less narrow beam. By rotation of the transducer this beam sweeps the surrounding medium and when it encounters an obstacle, a reflected pulse returns to the sender, who can locate the object from bearing and distance. This method is the acoustic analogue of radar. Since the absorption coefficient of sound in water increases more or less as the square of the frequency, the method cannot secure long ranges at high frequency. On the other hand, with very low frequency the beamlike quality of the emitted sound is lost unless the source itself is large. Hence a compromise must be worked out. As in the case of air-borne sound discussed above, the emitted sound is subject to disturbing effects if the water is nonhomogeneous because of variations in temperature from place to place as well as currents, changes in salinity, etc. The sound beam then suffers refraction, diffraction and scattering as well as absorption. Corrections must be made if the echo method is to give accurate results.

6. Ultrasonics.—The transmission of sound through a medium, characterized by its velocity, attenuation and scattering, depends on the physical properties of the medium. This is already obvious, for example, in the dependence of velocity on the elastic moduli and the density. In section II.13 it was seen that the attenuation or absorption of sound in a fluid can depend not merely on the viscosity and thermal conductivity but also on internal energy changes in the molecules of the fluid (thermal relaxation). Scattering of sound by a suspension depends on the size and distribution of the suspended particles. These illustrations show that acoustical measurements are competent to shed light on the macroscopic physical properties of matter and in certain cases to give information about its microscopic structure as well. Further research was made possible by the development of means for producing and detecting sounds of high frequency over a wide range extending far beyond the upper limit of audibility. This branch of acoustics is known as ultrasonics and has been principally responsible for the use of acoustics as a tool in the study of matter. Just as X-rays (very short or ultrahigh-frequency electromagnetic radiation) can tell more about the structure of crystals than ordinary visible light (of relatively lower frequency), so ultrasonic radiation can provide more information about many details of structure than can ordinary audible acoustical radiation. Moreover, ultrasonic radiation, because of its high frequency, possesses valuable practical applications in technology, chemistry and medicine (*see* ULTRASONICS).

7. Noise and its Control.—Noise or unwanted sound had become in the 1950s one of the most vital aspects of practical acoustics. The point made at the beginning of this article, that man lives immersed in a world of sound, was no understatement. More and more of this is man-made and its reaction on human beings is a problem of great consequence. By the middle of the 20th century noise control in every part of man's environment had developed into a major branch of acoustical engineering.

Every advance in the transformation of heat energy into mechanical energy has involved a noise problem, and in general it increases with the power production. The jet airplane is a good example: the large-scale turbulence of the exhaust gases in the jet forms an unusually intense source of sound the control of which is more difficult than that in the exhaust of an automobile engine in which a relatively simple acoustic filter or muffler reduces the noise level to endurable proportions.

All noise control problems involve the three standard acoustical elements: source, transmission path and receiver. Examples of noise sources are so common as to need no enumeration, but we may mention machinery of all kinds, some so-called musical instruments, jets, and finally living things, including people. Transmission may be through the open air, as in street noises or airplane racket; through the sea, as with fish sounds that interfere with underwater signaling; through the walls of buildings, etc. The most vulnerable receiver is doubtless the human ear, though noise in the form of unwanted vibration may be communicated from one part of a machine to another or to its support, with possible deleterious consequences. In each case the first important problem is to decide how much noise reduction is required for the comfort or safety of the given receiver, and the second is to try to devise ways of achieving it.

As an example, consider the reduction of factory noise, which, if sufficiently intense and sustained for long intervals of time, can permanently damage the hearing of a worker. Various methods are possible (as shown by the work of L. L. Beranek); in one, a partial enclosure may be built around the noise machine and lined with acoustic absorbing material. This can provide an attenuation of from 5 to 10 db. for a worker on the outside. Full enclosure of the machine can provide perhaps 40 or 50 db. attenuation. But if the worker must conduct his business right at the machine, ear defenders (small acoustic filters inserted in the ear canal) may be the only answer.

See also NOISE AND ITS CONTROL.

BIBLIOGRAPHY.—L. L. Beranek, *Acoustic Measurements* (1949), *Acoustics* (1954); L. Bergmann, *Der Ultraschall*, 6th ed. (1954; *Ultrasonics*, English trans. of 1st ed. by H. S. Hatfield, 1938); W. H. Bragg, *The World of Sound* (1920); C. A. Culver, *Musical Acoustics*, 4th ed. (1956); H. Fletcher, *Speech and Hearing in Communication*, 2nd ed. (1953); T. F. Hueter and R. H. Bolt, *Sonics* (1955); F. V. Hunt, *Electroacoustics* (1954); L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics* (1950); V. O. Knudsen and C. M. Harris, *Acoustical Designing in Architecture* (1950); H. Lamb, *Dynamical Theory of Sound*, 2nd ed. (1925); W. P. Mason, *Electromechanical Transducers and Wave Filters* (1942); P. M. Morse, *Vibration and Sound*, 2nd ed. (1948); H. F. Olson, *Acoustical Engineering* (1957); Lord Rayleigh, *Theory of Sound*, 2 vol. rev. (1926), 1 vol. ed. (1945); E. G. Richardson, *Sound*, 5th ed. (1953), *Ultrasonic Physics* (1952), (ed.) *Technical Aspects of Sound*, vol. i (1953), vol. ii (1956); G. W. Stewart and R. B. Lindsay, *Acoustics* (1930); J. Tyndall, *On Sound* (1867); A. B. Wood, *Textbook of Sound*, 3rd ed. (1955).

Acoustical journals of importance are the *Journal of the Acoustical Society of America*, *Revue d'Acoustique* and *Acustica* (International Journal of Acoustics). *Soviet Physics—Acoustics* is now being translated into English and is available through the American Institute of Physics (New York). Useful reviews of progress in acoustics will be found in *Reports on Progress in Physics*, published annually by the Physical Society of London, vol. i appearing in 1934. *See* also the *Proceedings* of the congresses sponsored by the International Commission on Acoustics (*Proceedings of the First ICA Congress* [Netherlands, 1954], *Proceedings of the Second ICA Congress* [U.S., 1957]).

(R. B. L.)

SOUND, THE, the easternmost of the straits giving entrance to the Baltic sea from the Cattegat, between the Danish island of Zealand and Sweden (Danish Oresund). Its extreme length from

the promontory of Kullen to that of Falsterbo is 70 mi. Its narrowest point is between Helsingör in Denmark and Helsingborg in Sweden, which are 3 mi. apart. Its extreme width, 30 mi., is at Kjöge bay. Three islands lie in it—Hven, belonging to Sweden, and Saltholm and Amager, belonging to Denmark.

SOUND BARRIER (also called SONIC BARRIER) is a sudden increase in the air resistance that occurs when an airplane or other flying object approaches the speed of sound. If the aircraft is not properly designed to fly at supersonic speeds, it might be crippled or destroyed by the strain resulting from the increased resistance. An associated phenomenon is the sonic boom, an explosive sound somewhat like a thunderclap that is caused by a shock wave which forms along the wing surfaces as they surpass the speed of sound; on some larger aircraft under certain conditions a second wave forms at the tail surfaces producing a double wave that is heard by ground observers as two sharp reports.

See AERODYNAMICS: *Supersonic Aerodynamics.*

SOUNDING, the act of determining the depth of the ocean or any body of water. The number representing the depth of water at any position is "a sounding"; and a ship is said to be "on soundings" when she is in a portion of the ocean, as over the continental shelves, where bottom is easily reached, and "off soundings" when in deeper water.

Sounding by Pole.—The first Stone Age man who shoved off his crude dugout into a muddy estuary no doubt quickly discovered that he needed a way to determine the depth of water, and he doubtless soon made use of a marked pole which he poked vertically toward the bottom. A sounding pole is shown in an Egyptian carving of 1500 B.C. Although they are still employed in some parts of the world, long poles are dangerous and tiring devices to use from a boat in any but calm waters, and the development of the lead line probably came quite early in the history of navigation.

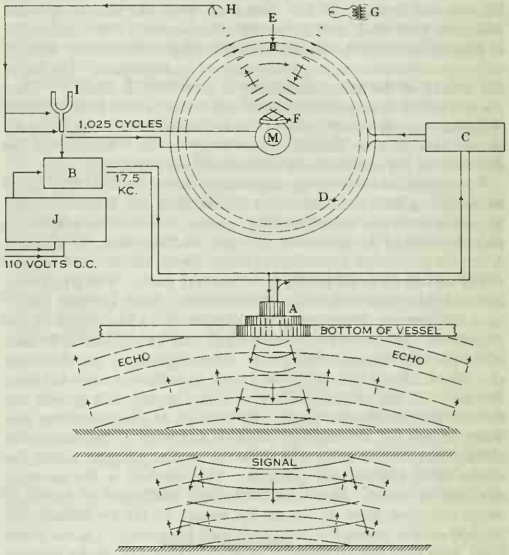
Sounding by Line.—The lead line essentially is a length of fishing line equipped with a sinker (the sounding lead) and marked at appropriate intervals with knots, coloured tags or similar indicators. Its use is mentioned by Herodotus (430 B.C.) and in Acts xxvii, 28 ("And they sounded, and found twenty fathoms . . ."), and it is still employed under suitable conditions. Formerly the sounding lead was provided with a hollow at the bottom, into which tallow was placed. A lead so fitted was described as "armed," and with it a sample of the bottom could be obtained at the same time as the depth was determined. Since in many areas there is more variation in the type and colour of bottom than there is in depth, the navigator had an additional bit of information to help in ascertaining his position.

To obtain an accurate sounding the lead line must be vertical; hence the ship must be nearly motionless or else in very shallow water. In order to permit a ship to obtain a sounding while underway, various expedients were tried that measured the depth of water independently of the length of line paid out. One device was a vertically oriented propeller, which rotated as the lead sank to the bottom and thereby actuated a counter. It was difficult to haul this device back in at high speeds, and in unskilled hands it could give erroneous results; nevertheless it was fairly widely used in the 19th century. As steamers came into wide use and the speed and size of vessels increased, investigation of causes of shipwrecks showed more and more cases where soundings had been neglected, the captains preferring to take chances rather than risk delays through slowing or stopping the ship to obtain soundings.

The problem was presented to Sir William Thomson (later Lord Kelvin; *q.v.*), professor of natural philosophy at Glasgow university, in 1878, and his solution received wide use as the Kelvin-White sounding machine. This device consisted of a small hand winch holding a reel of piano wire, which was as strong as the hemp lines previously used and much easier to retrieve. To the sounding lead was attached a glass tube, coated on the inside with silver chromate and sealed at the upper end but open at the lower. As the tube descended in the water, hydrostatic pressure compressed the air and permitted the water to rise. A chemical reaction with the sea water changed red silver chromate to white silver chloride, so that after the tube was hauled back aboard the depth could readily be determined. Thus, if the sea water had come halfway up the tube, the hydrostatic pressure had been two atm., corresponding to a depth of 33 ft.; two-thirds of the way up was three atm. or 66 ft.; and so on. As long as the lead reached the bottom, a good record of the depth was obtained regardless of the amount of wire that had been paid out.

The Kelvin-White system, with improvements such as the use of power on the winch and stranded wires, was standard on most vessels until after 1930. The echo sounder, which measures the time required for sound waves to return to the ship after being reflected from the bottom, gradually replaced it.

Echo Sounders.—A typical echo sounder, the Dorsey Fathometer, is shown schematically in the figure.



BY COURTESY OF U.S. COAST AND GEODETIC SURVEY

SCHEMATIC DIAGRAM OF DORSEY FATHOMETER (*see* TEXT)

The transceiver A is a combined transmitting and receiving unit, mounted in a hole cut in the bottom of the vessel. It receives power from a 17.5-kc. oscillator and amplifier B. The received echo signals in turn are fed to the echo amplifier C, which amplifies the weak voltage and flashes the neon tube D. An aluminum disk rotated at 1,230 r.p.m. by a synchronous motor M hides the neon tube except for a narrow slot E. The motor also rotates a mirror F. Once each revolution the mirror reflects the beam from a light G onto a photocell H, which keys the oscillator B into vibrating for 0.7 milliseconds. Thus, each time the index slot E is vertical, a signal is generated in the oscillator-amplifier B and transmitted into the water by the transceiver A. Assuming that the water is 120 ft. deep, the signal, traveling at 4,800 ft. a second, reaches the bottom in $\frac{1}{40}$ sec., and the echo returns in another $\frac{1}{40}$ sec., or a total of $\frac{1}{20}$ sec. Since the disk rotates at 1,230 r.p.m. or 20.5 r.p.s., the index slot will have completed almost one complete revolution, and the flash of the neon tube will be seen just to the left of the zero mark. In this configuration, the echo sounder has a scale range of 120 ft. (20 fathoms), but deeper ranges can be provided by driving index disks through suitable reduction gears. Since the accuracy of the depth determination depends entirely on accuracy of timing the revolutions, a tuning fork I is provided to regulate the frequency of the power supply. A converter J produces alternating current from the ship's direct current supply.

Numerous modifications of the echo sounder have been made. The Fathometer in the figure uses ultrasonic frequencies; others

use sound in the audible range, which is capable of being transmitted for greater distances in sea water. This Fathometer has a magnetostrictive transmitter, but piezoelectric, electromagnetic or percussion sound sources can be used. For survey purposes it is desirable to have a permanent record of the soundings, and echo sounders are used that draw a continuous profile of the bottom on a roll of paper.

Deep-Sea Sounding.—The sounding methods described above are adequate for the ordinary purposes of navigation "on soundings." Commencing about 1850, with the promotion of submarine cables, there arose an interest in the exact determination of deep oceanic depths, which could be met only by the use of much more elaborate equipment. Devices were developed that released the lead on contact with the bottom, so that the line could be hauled in more readily, while accurate meters measured the amount of wire paid out. Even so, a sounding might require several hours to complete, and although the ship was maneuvered to permit the line to descend vertically as it entered the water, there was the possibility that different currents at depth might cause the line to depart from the vertical. Echo sounding has proved as feasible in deep water as in shoal, when signals of suitable power are sent out. *See* also Index references under "Sounding" in the Index volume.

BIBLIOGRAPHY.—U.S. Coast and Geodetic Survey, *Hydrographic Manual* (1942); Henry Galway, *Echo Sounding at Sea* (1951); Austin Melvin Knight, *Modern Seamanship*, 12th ed., (1953); *Admiralty Manual of Seamanship*, 2 vol. (H.M.S.O., 1951–52). (J. L.)

SOUP. Soup is a liquid food made by slowly cooking in water either meat, poultry, fish or vegetables or a combination of such ingredients. The cooking tenderizes the foods, extracts desirable nutrients and develops a pleasing aroma and flavour. The liquid portion is often thickened by the addition of cereals, cream or eggs. Seasonings, spices and herbs are added to suit individual taste or to impart a characteristic flavour.

There are three main classes of soup: (1) Clear soups include consommé, bouillon and the various broths. These are the clear water extracts of meats, poultry, fish or vegetables and may be served with or without the addition of garnishes. (2) Thick soups are purées, *veloutés* and creams. These are prepared by straining the main ingredient through a fine sieve to thicken the liquid (purée) or by the addition of a cereal flour (*velouté*), milk, cream or eggs to give the liquid portion a thick consistency. (3) The third class might be considered a subdivision of the second because it also is thickened slightly, but usually this characteristic is obtained by adding pieces of meat, fish, vegetables or whole grain cereals. Vegetable soup, Scotch broth and the various chowders are typical of this class.

Soup through the ages has played a role similar to bread as a basic food. Every country, every civilization has its soup—France its *pot-au-feu* and *petite marmite;* Italy, *minestrone;* Russia, *borsch;* Spain, *puchero;* Holland, *erewensoep;* Japan, *osumashi* (clear soup); China, *yen-wo-tang* (bird's-nest); the United States, tomato.

Soup has been celebrated in literature through the centuries, from the Bible (Jacob's "pottage of lentils") to Lewis Carroll ("Soup of the evening, beautiful soup!"). It has inspired gourmets to lyrical descriptions: "There is nothing like a plate of hot soup, its wisp of aromatic steam making the nostrils quiver with anticipation . . . one whiff of a savory, aromatic soup and appetites come to attention" (Louis P. De Gouy, *The Soup Book*, Greenberg, New York, 1949). Queen Elizabeth I and Queen Victoria were fond of sipping mutton broth to wake up their morning appetites; and breakfast soups have been popular with such musically famous personages as Ignace Paderewski and Leopold Stokowski. The Italian composer Giuseppe Verdi attributed much of his inspiration to the warming and sustaining effects of large bowls of noodle soup.

From time immemorial, many countries have observed the ritual of the old-fashioned soup kettle simmering all day on the kitchen stove to utilize edible scraps and bones. During periods of disaster or stress, soup kitchens have served as an excellent expedient in feeding large masses of people.

Soup is often served as an appetizer before a meal, although some of the hearty soups may be used as the main course. They are healthful and stimulating foods. Properly prepared from high quality ingredients, they have good nutritional value. Because many of them contain ingredients of both plant and animal origin they have a good nutritional balance of protein, carbohydrate, fat, vitamins and minerals.

Although some soups, such as consommé or broths, are relatively low in caloric value, certain others such as pea, bean and meat soups may supply a high percentage of the human dietary requirements and are often good sources of one or more important nutritional factors. Because of their digestive and nutritious qualities, soups play a very important role in diets for infants, older people and convalescents.

Until the latter half of the 19th century practically all soups were homemade or were prepared in restaurants or institutions for consumption on the premises. Canning of soup was started in England shortly after 1800 by Bryan Donkin and John Hall, but commercial production did not reach substantial proportions until the creation of condensed soup in 1897 by John T. Dorrance of Camden, N.J. He believed that if he made a canned soup double strength and let the housewife add the water, important savings could be made from smaller cans, cheaper labels, lower freight costs and lower warehousing charges. These savings would make it possible for the soup to retail for as little as ten cents a can, a price which would stimulate volume which in turn would lead to still greater production economies.

The earliest "condensed" soup was made in two 50-gal. kettles located in a mincemeat room of what became the Campbell Soup company at Camden. Condensed soups proved popular and by mid-20th century constituted the great bulk of all soup consumed in the U.S., although there was still a small market in commercially made "ready-to-serve" soups which contained the full amount of liquid; these were especially suited for restaurant or vending machine use. Condensed soups are one of the earliest and best examples of the modern trend toward "convenience" foods—foods which require little preparation and effort for the housewife to serve. The timesaving element of such foods is well illustrated by the fact that it used to take the housewife an average of five hours to prepare and cook vegetable soup in the kitchen. Condensed vegetable soup may be heated and served within five to ten minutes.

There are three modern methods of making and preserving soup commercially: (1) heat processing (or sterilizing); (2) freezing; and (3) dehydration. In heat processing, the cans, after being filled with the soup and sealed, are heated by steam in a closed-pressure vessel called a retort. In the most modern plants they are introduced into a long cylindrical pressure cooker through which they move on a spiral reel. While passing through the cooker the cans are rotated slowly to permit rapid penetration of heat and to obtain efficient sterilizing effect of steam under pressure. The length of time the cans remain in the cooker depends on the type of soup being processed, but generally varies from 10 to 25 minutes.

Heat-processed soup will remain in good condition for years if stored in a cool, dry place. Caches of canned foods left by explorers in the Antarctic were found to be fresh and edible 46 years later.

While it is believed that homemade soup had been frozen for years, this method of preserving did not assume real commercial importance until about 1953. The advance preparation is similar to the preparation for heat processing, up to and including the filling and closing of the cans. The sealed cans are then frozen in a cold room at a temperature of −20° to −30° F. or by immersion in a low-temperature bath maintained at about −35° F. by a mechanical refrigerating system. The frozen cans are then brushed clean and put immediately into freezer storage at −10° F. or lower.

Freezing has been found superior to other methods for preserving soups containing heat sensitive ingredients such as milk or cream, as in oyster stew, New England style clam chowder and many cream soups.

One of the earliest methods of preserving foods was drying. Meat, fruits and vegetables were preserved in this manner long

before anyone understood the causes of spoilage, in fact, long before anyone knew that microorganisms such as bacteria, yeasts and certain molds existed. The principle is simply the removal or reduction of water to the point where spoilage organisms cannot grow. Salt often is added to retard the growth of organisms. Soups also are preserved by removing water. The term dehydrated is a modern word indicating that the material has been dried under somewhat better-controlled conditions and generally in a shorter time than in earlier days. (*See* DEHYDRATION OF FOODS.)

Some commercial dehydrated soups are prepared by actually dehydrating a soup that has been prepared in the usual manner. Most of them, however, are made by combining ingredients that normally contain no appreciable amount of water or those that have been previously dehydrated.

World consumption figures were not available in the late 1950s, but in the U.S. it was estimated that more than 17,500,000,000 servings of soup were consumed annually, of which approximately 40% were made in the home and 60% commercially.

BIBLIOGRAPHY.—Louis P. De Gouy, *The Soup Book* (1949); H. D. Renner, *The Origin of Food Habits* (1944); Cora Louise Brown *et al.*, *Soups, Sauces and Gravies* (1939); Henry Sarson, "Introducing Soup," *Chemistry and Industry* (Jan. 13, 1951); Mary D. Chambers, *A Book of Unusual Soups* (1923). (O. G. W.)

SOURIS, an important tributary river flowing into the right bank of the Assiniboine river, about 20 mi. S.E. of Brandon, Man., Can. The river is 500 mi. in length and drains an area of about 23,600 sq.mi. in the provinces of Manitoba and Saskatchewan and the state of North Dakota. Rising in the southern part of the Regina plains near Weyburn, Sask., it receives drainage from Moose mountain to the east, and the Coteau du Missouri on the west. Flowing southeasterly, it crosses the international boundary into North Dakota, making a long loop before recrossing into Manitoba and turning northeasterly to join the Assiniboine river. A considerable part of the drainage basin is floored by silt and clay deposited on the bottom of an ancient glacial lake known to geologists as Lake Souris. The river follows the valley cut by the outfall from the old lake for some distance near Souris, Man., but suddenly turns northeastward, cutting a deep, narrow valley through the Tiger hills to join the Assiniboine. The downstream extension of the old valley is drained by the Pembina river. The grassy plains drained by the Souris river were once the grazing grounds of countless buffalo herds. They form an important part of the hard red spring wheat belt of central North America. Near Estevan and Bienfait, Sask., there is extensive strip mining of lignite coal. Productive oil wells are found near Midale, Sask., Virden, Man., and other points. There are coal mines also near Minot, N.D., the largest city on the Souris river. (D. F. PU.)

SOURSOP, also called guanabana (botanical name *Annona muricata;* family Annonaceae, *q.v.*), a small evergreen tree, about 20 ft. high, native to tropical America and the West Indies. It bears large, dark-green, ovoid or heart-shaped fruit, six to eight inches or more in length and one to five pounds or more in weight, covered with short fleshy spines. The white cottony-textured pulp, juicy, slightly acid and aromatic, has a flavour suggesting that of mango and pineapple; it makes excellent jelly and preserves and is used to prepare a delicious sherbet and refreshing beverage. The soursop is cultivated in the West Indies, tropical South America, Mexico, India, Laos, parts of Polynesia and the west coast of Africa. With some frost protection it can be grown along the southern coast of Florida. It is propagated by seed and by shield budding.

See also CHERIMOYA; CUSTARD APPLE; SWEETSOP.

SOUSA, JOHN PHILIP (1854–1932), U.S. composer and bandmaster, was born at Washington, D.C., on Nov. 6, 1854. At 19 he conducted for Milton Nobles; he played in Offenbach's orchestra at the Centennial exhibition, Philadelphia, in 1876, and conducted *H.M.S. Pinafore* in 1879 for the New York production. In 1880 he became bandmaster of the U.S. marine band, which acquired a high reputation under his control. The title of the "march king" was bestowed upon him by an English brass band journal. In 1892 he organized Sousa's band, which won great renown. It visited Europe (1900, 1901 and 1903) and toured the world (1910–12). His compositions include the comic operas, *The Smugglers, Désirée, Queen of Hearts, El Capitán, The Bride Elect, The Charlatan, Chris and the Wonderful Lamp, The Free Lance, The Glass Blowers* and *The American Maid;* many songs; 15 suites; more than 100 marches, of which the most popular were "The Washington Post," "High School Cadets," "Liberty Bell," "Semper Fidelis" and "The Stars and Stripes Forever"; *The Chariot Race* (Ben Hur), a symphonic poem. He wrote *The Fifth String, Pipetown Sandy, Through the Year With Sousa* and an autobiography. He died March 6, 1932.

SOUSA, LUIZ DE (MANOEL DE SOUSA COUTINHO) (1555–1632), Portuguese historian, was born at Santarem, a member of the noble family of Sousa Coutinho. In 1576 he broke off his studies at Coimbra university to join the order of Malta and shortly afterward was captured at sea by Moorish pirates and taken prisoner to Argel, where he met Cervantes. A year later Coutinho was ransomed and, landing on the coast of Aragon, passed through Valencia, where he made the acquaintance of the poet Jaime Falcão, who seems to have inspired him with a taste for study and a quiet life. Between 1584 and 1586 he married. Settling at Almada, he divided his time between domestic affairs, literary studies and his military duties as colonel of a regiment.

In 1599, to escape the pestilence that devastated Lisbon, the governors of the kingdom for Philip II decided to move their quarters to Coutinho's residence. Finding his protest against this arbitrary resolution unheeded, he set fire to his house, and to escape the consequences of his courageous act had to leave Portugal. Going to Madrid, he was able to pursue his literary studies in peace and to publish the works of his friend Falcão (1600). Nothing is known of how he passed the next 13 years, although there is a tradition that he spent some time in America. In 1613 he and his wife agreed to a separation and he took the Dominican habit in the convent of Bemfica, while his wife entered the convent of Sacramento at Alcantara.

In 1616 the task of writing a history of the order and a life of the famous archbishop of Braga, D. Frei Bartholomew of the Martyrs, was confided to Frei Luiz. The *Life of the Archbishop* appeared in 1619, and the first part of the *Chronicle of St. Dominic* in 1623, while the second and third parts appeared posthumously in 1662 and 1678; in addition he wrote, by order of the government, the *Annals of D. John III,* published by Herculano in 1846. After a life of about 19 years spent in religion, he died in 1632, leaving behind him a memory of strict observance and personal holiness.

SOUSSE, a city of Tunisia, in the governate of the same name, on the gulf of Hammamet, in 35° 49′ N. and 10° 39′ E., 36 mi. N.E. of Kairouan, of which it is the port, and 93 mi. S.E. of Tunis. Pop. (1956) 48,172.

Sousse (formerly Susa) occupies part of the site of the ancient Hadrumetum, an important Phoenician city which served Hannibal as his base of operations at the end of the second Punic War. Trajan made it a colony, and about A.D. 300 it became the capital of the new province of Byzacena. Later it received the name of Justinianopolis. In the 9th century it was fortified by the Aghlabite rulers of Kairouan. It shared the general fortunes of Tunisia and became a noted haunt of pirates who raided the coast of Italy. In 1537, it was unsuccessfully besieged by the marquis of Terra Nova, in the service of Charles V, but in 1539 was captured for the emperor by Andrea Doria. As soon as the imperial forces were withdrawn it became again the seat of Turkish piracy. The town was attacked by the French and the Knights of St. John in 1770 and by the Venetians in 1764. It remained, however, in the possession of the bey of Tunis. In 1881 it was captured by the French, and in April 1943 by the British 8th army during the north African campaign in World War II, after advancing 80 mi. from Sfax in two days.

The town is built on the side of a hill sloping seaward and is surrounded by a crenelated wall, strengthened by towers. Recesses in the inner side of the wall are used as shops and warehouses. The kasbah, or citadel, built on the highest point within the town, was thoroughly restored by the French after their occupation of the country. Interesting buildings in the old town are

the Kasr-er-Ribat, a square fortress with a high tower and seven bastions, built not later than the 9th century A.D., and the Kahwat-el-Kubba (Café of the Dome), a curious house, square at the base, then cylindrical, and surmounted by a fluted dome. It was probably a church during the Byzantine period. Southwest of the town are vast Christian catacombs. The modern port was completed in 1901.

Exports are chiefly phosphates and other minerals, olive oil, esparto and cereals.

SOUTH, ROBERT (1634–1716), English priest, was born at Hackney, Middlesex, in Sept. 1634. He was educated at Westminster school and at Christ Church, Oxford. Before taking orders in 1658 he was in the habit of preaching as the champion of Calvinism against Socinianism and Arminianism. He also at this time showed a leaning to Presbyterianism, but on the approach of the Restoration his views on church government underwent a change. On Aug. 10, 1660, he was chosen public orator of the university, and in 1661 domestic chaplain to Lord Clarendon. In March 1663 he was made prebendary of Westminster. In 1667 he became chaplain to the duke of York. He was a zealous advocate of the doctrine of passive obedience, and strongly opposed the Toleration act. In 1676 he was appointed chaplain to Lawrence Hyde (afterward earl of Rochester), ambassador-extraordinary to the king of Poland. In 1678 he was presented to the rectory of Islip, Oxfordshire. In 1693 South published with transparent anonymity *Animadversions on Dr. Sherlock's Book, Entitled a Vindication of the Holy and Ever Blessed Trinity,* in which the views of William Sherlock were attacked with much sarcastic bitterness.

Sherlock, in answer, published *A Defence* in 1694, to which South replied in *Tritheism Charged Upon Dr. Sherlock's New Notion of the Trinity, and the Charge Made Good.* The controversy was finally stopped by the king. In 1710 he ranked himself among the partisans of Henry Sacheverell. He declined the see of Rochester and the deanery of Westminster in 1713. He died on July 8, 1716, and was buried in Westminster abbey.

His *Sermons,* six volumes (1692) were reprinted many times; there is a modern edition in Bohn's classics, two volumes (1845), and a complete edition of his *Works,* five volumes (1842).

See also W. C. Lake, *Classic Preachers of the English Church,* 1st series (1877).

SOUTH, THE, an extensive region covering approximately the southeastern quarter of the United States and including the states of Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Tennessee and Arkansas, as well as parts of the states of Texas, Oklahoma, Missouri, Kentucky, West Virginia, Maryland and Delaware.

Geography.—In its physical geography, the south does not possess any natural unity. In fact, it includes a number of quite diverse geographical districts. Along the coast, from Maryland to the Rio Grande, there is a low coastal plain, marked by a series of alluvial rivers flowing eastward to the Atlantic or southward to the Gulf of Mexico. On the Atlantic shore, this coastal plain rises gradually to the "fall line" where the streams cease to be navigable. There begins an area known as the piedmont, where the land is rolling or hilly, erosion has taken place, and reddish-yellow clay soils predominate in contrast to the sandy or alluvial soils of the coast. At its western edge, the piedmont rises into the Appalachian mountain system, the southern part of which extends from Maryland to northern Georgia and Alabama. The easternmost part of this system consists of the Blue Ridge mountains, behind which lies the main Appalachian ridge. Beyond the Appalachian ridge is the high and mountainous country of the Cumberland plateau. This mountainous terrain dominates the western parts of Maryland, Virginia, North Carolina, the northern part of Georgia, the eastern parts of Tennessee and Kentucky and most of West Virginia.

Just as the Appalachian mountain system is the key to the geography of the eastern or Atlantic coastal part of the south, so the Mississippi river valley is the key to the western or Gulf coastal part. Although portions of Alabama, Mississippi, Louisiana and Texas drain directly into the Gulf of Mexico, other parts of these states and all of Kentucky, Tennessee, Missouri and Arkansas drain into the Mississippi. On its west bank, such major rivers as the Missouri, the White, the Arkansas, the Ouachita and the Red flow eastward through fertile alluvial valleys directly into the Mississippi. On the eastern side of the valley, the principal southern rivers are the Cumberland and the Tennessee, but instead of flowing westward from the Appalachian ridge into the Mississippi, these streams describe great semicircular arcs, first swinging southward and then northward across Tennessee and Kentucky into the Ohio river and thence into the Mississippi. Their courses run through rich limestone valleys, which have given great and distinctive beauty to the Nashville basin in Tennessee and the bluegrass region of Kentucky.

Population.—This geographical variety has produced a diversity rather than a uniformity of southern folk—a diversity ranging from the isolated mountaineers of Appalachian valleys to the semi-amphibious bayou dwellers of southern Louisiana, and including the "red necks" of the southern hill country, the poor whites of the piney woods, the cattle-herding plainsmen of Texas, the peasantlike Negroes who form more than 80% of the population of some counties along the Mississippi and the cotton-planting farmers, both great and small, who, more than any others, have constituted a "southern" type.

In terms of the origins of its population, the south presents a pattern unlike that of the rest of the country, but which, again, shows lack of uniformity within the region itself. The contrast with the nonsouth lies in the fact that the south received only a small proportion of the immigrants who came to the United States after 1800, and hence it has a relatively small number of people of southern, central and eastern European antecedents. Most of the population is descended from the settlers who arrived during the colonial period. In the 17th century, English settlers established themselves in the coastal areas of Virginia, Maryland and the Carolinas. In the 18th century, a heavy influx of Scots-Irish (Presbyterians of lowland Scotland who had migrated first to northern Ireland and then to America) and German settlers moved southward from Pennsylvania into the southern piedmont. During the same century, the French settled in Louisiana, where their language, religion and culture are still in evidence, and at Mobile, in southern Alabama. The Spanish at the same time occupied Florida and Texas but did not settle in sufficient force to leave an important Hispanic tradition. The most extensive influx of the 18th century was the involuntary immigration of Negroes brought as chattel slaves directly or indirectly from the west coast of Africa. The importation of slaves had begun as early as 1619, but did not attain large numbers until after 1700. By the year 1790, however, Negroes formed more than one-third of the population in the states of Maryland, Virginia, South Carolina and Georgia, and in 1960 they continued to form more than 20% of the population in Virginia, North Carolina, South Carolina, Georgia, Alabama, Mississippi, Louisiana and Arkansas. Florida, Maryland, Tennessee, Delaware and Texas had Negro populations of between 10% and 20%. Illinois, with 10.3%, was the only state outside the south with as much as 10% Negro population.

Climate and Soil.—In terms of climate and soil, the south shows features of uniformity, somewhat more than it does in geography and in the ethnic origins of its population. Almost the entire region lies in the same latitudes as north Africa, and the climate is, in general, mild to hot. In the summers, temperatures rise, day after day, above 90° F. (32° C.) in the shade. But the maximum heat is less significant than the long growing season—a period of 200 to 290 frost-free days during which slow-growing crops such as cotton, rice and sugar may be brought to maturity. Also, there is abundant rainfall, averaging 40 to 50 in. per annum in the upper south and 50 to 60 in. in the Gulf coast region.

Because of the warmth of the region, it was not glaciated during the Ice Age, and this lack of glaciation gives a distinctive quality to the soil. Except in the limestone basins of Kentucky and Tennessee, the black belt across parts of central Georgia, Alabama and Mississippi and the river valleys with their rich alluvial deposit, the soil is predominantly of coarse texture and of mediocre fertility, with high proportions of clay in the piedmont region,

with its "red hills," and of sand in the coastal districts. Soil exhaustion has occurred widely in these areas, both because of their natural lack of fertility and because of the soil-depleting effects of the cotton and tobacco cultures. As a result, the south regularly consumes more than half of the nation's output of commercial fertilizer. Also, erosion has occurred very widely in the piedmont area because of the torrential rains, coupled with the exposure of the soil in plowed hills and furrows which are used in the row cropping of cotton and tobacco. This erosion had been so severe that by 1930 the south suffered a total loss of fertility in 22,500,000 ac. of its land—an area larger than the entire state of South Carolina.

Crops.—The one feature above all others that has made the south historically a distinctive and somewhat homogeneous region is the fact that, for more than three centuries, it relied almost wholly upon an agricultural economy, centred upon staple crops which were cultivated for sale on the world market. Even as late as 1948, the south still derived twice as much of its income from agriculture (16%) as did the nonsouth (8%). The first of these crops was tobacco, first cultivated in Virginia about 1612, just five years after the founding of the colony. The tobacco culture spread into Maryland and northern North Carolina, and for nearly 200 years it provided the chief export product of North America. Meanwhile, rice about 1698 and indigo about 1744 took hold in the Carolinas and quickly grew to dominate the economy of the coastal part of that colony. The culture of sugar was successfully launched in French Louisiana in 1794, and it also expanded to become the chief economic activity of the region. But none of these crops rivaled cotton, which created southern wealth, shaped southern institutions and, as "king cotton," grew to symbolize the whole south at the height of its power.

The long-fibred, or black-seed cotton was profitably cultivated on the coastal islands of Georgia and Carolina before the American Revolution, but the only strain that would grow in most areas on the mainland was the short-fibred, green-seed plant. The latter seemed commercially worthless because the lint clung so firmly to the seed that the separation of the fibres was impracticable. This situation was suddenly changed in 1793, however, when Eli Whitney, a Connecticut Yankee teaching school in Georgia, invented a device, known as a cotton gin, to separate the fibres. Whitney's gin made it possible for the south to supply cotton for a world textile market in which the new technology of mechanical spinning and weaving had created an insatiable demand for raw material. The production of southern cotton doubled every decade between 1800 and 1861, until it reached 4,400,000 bales in 1861. As this growth took place, the staple-crop economy spread into the Gulf coast region and within a single generation the south expanded from the Oconee river in Georgia to the banks of the Brazos in Texas.

The Plantation System.—The staple-crop economy of tobacco, rice, indigo (declining sharply after the Revolution), sugar and cotton shaped the character of the south. Since production on a big scale was profitable, large agricultural units known as plantations grew up; although vast numbers of small, owner-cultivated farms remained, the plantation became a dominant factor. Since the crops were used for commercial purposes rather than for subsistence (such as prevailed in the backwoods), it paid the proprietor to employ labour and thus increase production. The slow-maturing crops needed attention most of the year and thus required a permanent labour force. Because the agricultural tasks were simple enough to be performed by unskilled labour working in gangs, owners resorted to the use of Negro slaves. Thus the one-crop system, the plantation unit, and the presence of a biracial population, with the Negroes in a subordinate position, were all developed in the south early in the 18th century, and their grip was tightened by the overpowering rise of the cotton economy. By 1850 there were 3,204,000 slaves in the south and it was estimated that 1,815,000 of these were connected with the cultivation of cotton.

Ultimately, the complex of staple crops, plantations and Negro cultivators (whether slave or free) gave to the south qualities that made it a distinctive region and set it in an adverse sectional relationship to other parts of the United States. But until the 19th century, the sectional alignment was far from clear-cut. Historians now reject a once-prevalent interpretation which explained the differences between north and south in terms of the dissimilarities between Puritans in colonial New England and cavaliers in colonial Virginia. The British colonists who settled the two regions were not so sharply differentiated as the tradition has claimed. Moreover, in the colonial south, the plantation economy and slavery were rather narrowly confined to limited areas around Chesapeake bay and in the coastal district of the Carolinas and Georgia. The democratic, Presbyterian, Scots-Irish who populated the vast back country of the piedmont did not share in this regime. They practised a subsistence economy, cultivated grain crops, had no use for slaves and were frequently at odds with the aristocratic, Anglican planters. These subsistence farmers of the interior seemed to have more in common with the similarly circumstanced settlers in the interior of Massachusetts and the backwoods of Pennsylvania than either group had with the grandees of the coastal districts—whether merchant princes in Boston, Mass., or planter-aristocrats in Charleston, S.C., and along the James river. Correspondingly, the latter groups showed a greater affinity with one another than with their neighbours to the west. The antagonism between the coastal and the interior districts was chronic, and repeatedly manifested itself in incidents of friction such as Shays's rebellion in Massachusetts and the uprising of the Regulators in the Carolinas. At the time of the American Revolution this east-west division seemed likely to remain the primary line of sectional cleavage.

North-South Differences.—At that time, the divisions between north and south still remained somewhat tenuous. The leaders of both sections agreed in regarding slavery as an evil. At the same time, both sections practised it, and it was sanctioned by law in every state. The actual concentration of slaves was overwhelmingly in the south—94% of all slaves in 1790 were held south of Pennsylvania. But many districts in the south had almost no slaves, and slavery was widely regarded as a localized and declining institution.

Between 1775 and 1830, however, both north and south experienced transformations which heightened their dissimilarities and generated deep antagonisms between them. The northern transformation came as the region slowly began to industrialize. This process committed it politically to the idea of economic self-sufficiency within a national market protected by tariff walls, where agricultural areas and manufacturing areas would complement one another by producing to meet each other's needs and serving as markets for each other's products. Industrialization generated rapid economic change in the north, and this gave sanction to the idea of progress and to the idea of a free society in which the individual can alter his occupation and status to meet changing conditions. Moreover, the use of machines freed the worker from many brute tasks and, by increasing productivity, raised the standard of living. Both of these developments tended to enhance the dignity of the worker.

Meanwhile, in the south the advent of the cotton economy made it possible to convert the whole region to staple-crop agriculture, as had never been possible with tobacco, rice and sugar. Cotton would grow almost everywhere in the south and it was more flexible than other staple crops in lending itself to production on any scale, large or small. The cotton economy moved quickly into the upcountry, thus obliterating the former divisions between the coastal and the piedmont areas. By carrying the plantation and slavery with it, it increased the value of slaves and reinvigorated the declining slave system.

Economically, the south looked to the British textile industry for its market, and it therefore opposed the growing economic nationalism of the north and west. Socially, the plantation system discounted the commercial values of thrift, prudence, enterprise and progress; it exalted qualities of magnanimity, command, manly prowess and physical courage; and it adopted a cult of chivalry, with a code duello, to enshrine these virtues. The southern social philosophy, holding to a country-gentry ideal, presented a sharp contrast with that of the north, for it stressed the con-

servative values of status in a fixed social order rather than of freedom, of stability rather than of progress, and of a way of life rather than accumulation of wealth. Such an emphasis was almost inevitable in a region where one-third of the population occupied a position of fixed legal subordination.

The Slavery Controversy.—The growing divergence between north and south, as well as the dissimilarities in their values, found a focus in the question of slavery. In the north, the antislavery sentiment of the Revolutionary era continued to grow. Between 1777 and 1804, every state north of Maryland either abolished slavery outright or provided for gradual abolition. The early antislavery movement was moderate in tone, and it sought to accomplish emancipation gradually by persuading slaveowners to free their slaves voluntarily. But about 1830 it entered a new and more militant phase, in which immediate abolition was demanded and slaveowners were denounced in most abusive terms. The militant abolitionists were always a minority, and historians disagree as to the extent to which the north became "abolitionized," but it is clear that antislavery sentiment became very widespread. The immense popularity of Harriet Beecher Stowe's *Uncle Tom's Cabin* (1852) and the election to the presidency of Abraham Lincoln, who had said that "if slavery is not wrong, then nothing is wrong," serve as indexes to this feeling.

While the north passed from mild opposition to strong condemnation of slavery, the south passed from mild opposition to an unqualified defense of the "peculiar institution." Various southern spokesmen defended slavery on the ground that it was sanctioned by the Bible, that the economic exploitation of wage earners was worse than the exploitation of slaves, that the Negro was biologically inferior and that in a well-ordered society it was better for one class to be set apart for the menial duties. Belief that slavery was a positive good became a test of southern orthodoxy.

With all the cultural dissimilarities and economic conflicts of interest between north and south converted into a dispute on the slavery question, sectional antagonisms were chronically acute between 1830 and 1860. The south perceived that it was not growing as rapidly as the industrial north in wealth, power or population. A minority psychology became fixed upon the whole region, and a number of its political leaders, of whom John C. Calhoun was foremost, developed elaborate constitutional arguments, proclaiming the ultimate sovereignty of the states and their right to secede from the union. The full explosive effect of the slavery question was to some extent kept out of politics by the fact that no one contested the right of individual states to permit or prohibit slavery within their own borders. But whenever new lands were acquired, either by the Louisiana Purchase (1803) or by the War with Mexico (1846–48), bitter contests took place over the status of slavery in these areas prior to statehood. Thus there was a series of crises: in 1820 over slavery in the Louisiana territory; in 1832–33 over South Carolina's effort to nullify, or prevent enforcement, of the tariff act; and in 1848–50 over slavery in the territory acquired from Mexico. In each of these crises a clash was averted by the adoption of measures of compromise, which have been compared to treaties between two nations. But despite these efforts to save the union, relations between the sections became increasingly strained and they were brought almost to the breaking point by a series of violent episodes in the 1850s—the repeal of the Missouri Compromise by the Kansas-Nebraska act in 1854, open combat between antislavery and proslavery factions in "bleeding Kansas," the decision of the supreme court in 1857 in the Dred Scott case that congress could not prohibit slavery in the territories, and John Brown's raid at Harpers Ferry, in 1859. Finally, in 1860, Abraham Lincoln was elected president on a platform pledged to oppose slavery in the territories, and by a purely sectional majority. The states of the lower south took his election as a signal to act. Seven of them (South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Texas) adopted ordinances of secession, and proceeded (1861) to form a new union, the Confederate States of America.

Lincoln denied the legality of their act and sought to maintain the authority of the union. As a result, fighting began at Ft. Sumter, in Charleston harbour (April 1861). When it did so, four other states (Virginia, North Carolina, Tennessee, Arkansas) also seceded and joined the Confederacy. For four years, the United States and the Confederate states fought each other fiercely in the American Civil War (1861–65).

Southerners in the heat of secession and war believed that the "southland" or "Dixie," as they now called it, had achieved a full and separate nationality, not only in the political but also in the cultural sense. Historians since then, with an eye to features of cultural distinctiveness, have also recognized an important degree of "southern nationalism." It is a serious question, however, whether southern separatism resulted primarily from deep cultural differentiation, or whether it was the reaction of a group who were still an integral part of the American people, but who had been psychologically (rather than culturally) alienated from their fellow Americans by a long period of minority status, conflicting interests and defensive guardianship of the institution of slavery. The readiness with which the south returned to the union after the war, with a minimum of "irredentism," suggests the latter.

The Reconstruction Era.—At any event, the south lost the war, and this loss settled two questions. First, it killed the idea of state sovereignty and secession. Second, it ended the institution of slavery, for Lincoln's Emancipation Proclamation freed all slaves in areas which were still resisting federal authority on Jan. 1, 1863, and the 13th amendment, almost three years later, freed the rest. But there were other problems which the war did not solve, and which continued to make for southern distinctiveness for another century. For one thing, the south continued to rely on a one-crop economy, and it continued to cultivate this crop with the labour of Negroes, who discovered that as long as a man went on making his living by hoeing cotton, the transition from slavery to freedom did not altogether revolutionize his life. For another, the south continued to insist upon the inferiority and the subordination of the Negro. Slavery had been partly an economic system for the ownership of labour and partly a social system of racial control, and though emancipation ended the one, it did not end the other.

During the so-called Reconstruction period (1865–77), the victorious Republican party made more or less earnest efforts to assure Negro equality, but those who wanted the party to help the Negroes were probably never as numerous as those who wanted the Negroes to help the party. After the latter perceived that the party might fare better without Negro help, the crusade for equality was generally abandoned, and the south was left to work out its own arrangements. As a consequence, the institution of slavery was replaced by three institutions: the economic system of sharecropping, the political system of one-party politics and the social system of segregation, supported both by law and by custom.

Sharecropping.—The system of sharecropping or tenancy resulted from need for a new link between the soil and the cultivator. The former slaves lacked funds to buy land; the landowners lacked funds to pay wages. Hence, an agreement was made that the landlord would furnish the land, the freedman the labour and each would receive a share of the proceeds for his contribution. The cultivator, however, necessarily received the necessities of life on credit during the annual periods between harvests, and the cotton economy was so unprofitable that the tenant's share seldom left a decent surplus after his debts had been settled. Thus southern tenant farmers were chronically in debt; their indebtedness limited freedom to move or change occupation; and they had a lower standard of living than any other large group in the American population. Moreover, the chronic overproduction of cotton, with its attendant low prices, forced more and more small landowning farmers into tenancy. Between 1880 and 1930 the proportion of southern farms operated by the tenants increased from 36% to 55%. During the last 30 years of this span, white tenants increased by 400,000 and Negro tenants by 147,000, so that what had begun as an arrangement for the freedmen became more and more a biracial institution.

One-Party System.—The one-party system was an arrangement to neutralize the political power of the Negroes, who had legally been enfranchised by the 15th amendment (1870). After

a brief period of southern support for Populism in the 1890s, southerners conducted their political activity solely within the Democratic party. Contending that the Democratic party was a private and not a public instrumentality, they excluded Negroes from participation in the primary elections in which party nominees for office were chosen. Moreover, between 1890 and 1910, most southern states placed in their constitutions literacy and property qualifications from which large groups of whites, at least in some states, were exempted by clauses allowing them to qualify if their forebears had voted before 1860 ("grandfather clauses"). By these devices, by inequitable enforcement procedures and by a relatively costly poll tax which most Negroes found difficult to pay, Negroes were effectively excluded from political life. At the same time, the white "solid south" remained Democratic: 11 states from Virginia to Texas elected only four Republican governors between 1876 and 1960. Because of this concentration in one party, southerners gained a disproportionate share of committee chairmanships in Democratic congresses and tended to dominate congress during Democratic administrations.

Segregation.—The third of the institutions succeeding to slavery was legalized segregation. The practice of segregation, it may be noted, was far older than the segregation laws and it extended far beyond the south. Throughout the United States, Negroes were generally denied access to places of public resort. They were excluded from trade unions, forced to live in segregated residential areas, prohibited from marrying whites and prevented from mingling freely with the rest of the population. In most areas, this was done informally or by extralegal devices such as restrictive housing covenants, but in the south it was done by law. Among southern segregation laws, the statutes requiring segregated education were important. These laws were upheld as constitutional by the supreme court in the case of *Plessy* v. *Ferguson* (1896).

Economic Trends.—Until 1932 the south remained an impoverished and undiversified region. The growth of a textile industry in the Carolinas, and the movement to develop a "new south" after the Civil War had not seriously qualified the region's commitment to cotton, to agriculture and to a rural way of life. The Negroes remained a kind of peasantry and the income of the south stood at only $372 per capita in 1929, while income outside the south was $797 per capita. But after 1933 the south experienced an economic revolution, resulting partly from the New Deal and the collapse of cotton tenancy and partly from rapid economic diversification and industrialization.

During the depression that began in 1929, cotton prices sank to a point that bankrupted the cotton economy. The plight of the sharecropper was desperate and cultivation continued only because of the lack of alternatives. When this situation was at its nadir, the government stepped in with two devices that exercised an effect more revolutionary than was intended. Agricultural benefit payments for taking acreage out of production led landlords to curtail sharecropping operations. Partly because of this and partly because of the nation-wide shift of population to the cities, the number of tenant farmers in the south declined from 1,552,000 in 1930 to 592,000 in 1954, while the number of southern Negro tenants declined from 600,900 to 254,000. The second New Deal device was the system of unemployment relief, which drew Negroes temporarily to southern towns where relief payments were issued. During and after World War II, however, there was heavy Negro migration to northern industrial centres. By 1950 the proportion of Negroes who lived in the south was only 62%, whereas in 1920 it had been 85%. By 1950 the average Negro was no longer a rural-dwelling, sharecropping, cotton cultivator but was a city-dwelling, wage-earning industrial worker.

With World War II and its aftermath, industrial developments and diversification revolutionized the economy within the south. Military installations were heavily concentrated in the south during the war, and after the war several new industries enjoyed spectacular growth. In 1948 the value of cotton produced was $2,189,000,000, but this amount was surpassed both by the value of livestock ($3,072,000,000) and of textiles ($2,832,000,000), and was rivaled by the value of chemical products ($1,082,000,000), lumber products ($889,000,000), petroleum products ($642,000,000) and paper products ($539,000,000), all of which had increased by more than 247% in the preceding eight years. Cotton was no longer king.

As a result of this economic change, the percentage of workers in the south employed in agriculture declined from 50.9% in 1920 to 21.6% in 1950 (compared with a national decline from 17% to 9%). Also, the proportion of southerners living in cities rose from 15% in 1900 to 43% in 1950 (compared with a national increase from 39% to 59%). But perhaps the most drastic change of all was reflected in the greater increase of per capita income in the south. While such income was rising from $703 to $1,847 for the U.S. as a whole between 1929 and 1955, it rose from $368 to $1,291 for the southern states (excepting Texas).

Desegregation.—At the middle of the 20th century, the steady growth of industrialism, urbanization and wealth were constantly diminishing those factors of poverty, agrarianism and ruralism that had operated to preserve the south's historical traditions and its distinctive identity. But although the distinctive elements in the southern region had been much reduced, the southern adherence to a biracial system remained strong enough to assert itself with powerful force when challenged in 1954.

From about 1905 (when a protest was framed by Negro leaders at Niagara, N.Y.), there had been a rising demand in liberal circles for the elimination of discriminatory practices toward Negroes, and after about 1927 (when the white primary was ruled invalid), the supreme court began to interpret the guarantees of the 14th amendment as prohibiting various forms of discrimination. For many years it did not question the legality of segregation in cases where facilities were "separate and equal," but in 1954 in the case of *Brown* v. *Board of Education*, it reversed the *Plessy* v. *Ferguson* decision and declared that segregation in publicly supported schools was unconstitutional.

Sectional antagonisms flared again as the southern states denounced "judicial usurpation," and invoked the principle of local self-government, while liberals criticized the south for racial prejudice and for the defiance of federal authority. Eleven states (Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Texas, Arkansas and Tennessee) adopted laws to avoid, directly or indirectly, the integration of the schools, and "citizens' councils" were organized to resist it. In the border states, where resistance was not so strong, desegregation achieved several relatively quick and easy triumphs, as in Washington, Baltimore, St. Louis and notably Louisville. But in the lower south, resistance was so determined that seven years after the Brown decision there were still eight states where less than one Negro child in 1,000 was attending a desegregated school. Federal court orders were resisted and some states adopted policies of "massive resistance." In 1957 Pres. Dwight D. Eisenhower sent troops to support desegregation in Little Rock, Ark., and in 1962 Pres. John F. Kennedy sent troops to the University of Mississippi. By 1963, partly because of continued southern segregation and partly because of generalized race discrimination throughout the country, Negro unrest had become very intense and mass demonstrations were taking place on a wide scale. A century after the south had made its supreme assertion of a separate identity, and after many decades of steady decline in the significance of the differentials which distinguish the south, the historical heritage of the south's biracial system was still strong enough to create a major sectional issue and to induce a high degree of southern solidarity in asserting local autonomy as opposed to uniform national policy. In the presence of this immediate issue, the less tangible aspects of southernism—its emphasis upon the virtues of the agrarian way of life, its conservative concept of the social order, its ideal of social standards set by the leaders rather than by the masses—were so much overshadowed that it is difficult to estimate the extent to which these features, by themselves, would have continued to make the south a significant and distinctive entity.

See also references under "South, The" in the Index volume.

BIBLIOGRAPHY.—W. H. Stephenson and E. M. Coulter (eds.), *A History of the South,* 10 vol. (1947 *et seq.*); Almon E. Parkins, *The South: Its Economic-Geographic Development* (1938); Rupert P.

Vance, *Human Geography of the South* (1932); Howard W. Odum *et al., Southern Regions of the United States* (1936); Calvin B. Hoover and Benjamin U. Ratchford, *Economic Resources and Policies of the South* (1951); V. O. Key and A. Heard, *Southern Politics in State and Nation* (1949); Jay B. Hubbell, *The South in American Literature, 1607–1900* (1954); Lewis C. Gray, *History of Agriculture in the Southern United States to 1860*, 2 vol. (1933); Ulrich B. Phillips, *Life and Labor in the Old South* (1929) and *American Negro Slavery* (1918); Kenneth Stampp, *The Peculiar Institution* (1956); C. Vann Woodward, *Origins of the New South, 1877–1913* (1951); Wilbur J. Cash, *The Mind of the South* (1941); Twelve Southerners, *I'll Take My Stand: the South and the Agrarian Tradition* (1930); Southern Education Reporting Service, *Southern Schools: Progress and Problems* ed. by P. McCauley and E. D. Ball (1959); Harry S. Ashmore, *An Epitaph for Dixie* (1958); William B. Hesseltine and David L. Smiley, *The South in American History* (1960). (D. Pr.)

SOUTH AFRICA, REPUBLIC OF (Republiek van Suid-Afrika), a country occupying the southernmost part of the African continent between the Atlantic and Indian oceans. From 1910 to 1961 it was known as the Union of South Africa, a constituent part of the Commonwealth of Nations, comprising the former British colonies of Cape of Good Hope, Natal, Transvaal and Orange River colony, which became the four provinces of the Union, the last named being known since union as Orange Free State. The southern and western limits of the republic are formed by the coast line; the eastern limit is the coast line and the southwestern boundary of Portuguese East Africa (Mozambique). The northern boundary is formed mainly by the Limpopo river in the east, the Molopo and Nossob rivers (which separate South Africa from Bechuanaland and Southern Rhodesia) and the meridian of 20° E. in the middle, and the lower Orange river in the west. The area of the republic is 472,359 sq.mi., comprising Cape of Good Hope province (Cape Province) 278,465 sq.mi., Natal 33,578 sq.mi., Transvaal 110,450 sq.mi. and Orange Free State 49,866 sq.mi. The administrative capital is Pretoria; Cape Town is the seat of the legislature. South West Africa (317,887 sq.mi.), formerly a German possession, was mandated to the Union after World War I and is administered now as part of the republic. Within the confines of the republic, but not part of it, lie the British High Commission territories of Basutoland (11,716 sq.mi.) and Swaziland (6,704 sq.mi.).

In Dec. 1947 South Africa formally annexed two small subantarctic islands, Prince Edward and Marion, which lie in the Indian ocean about 1,200 mi. S.E. of Cape Town.

This article is divided into the following sections:

I. Physical Geography
1. Geology and Structure
2. Physiography
3. Climate
4. Vegetation
5. Animal Life
II. Geographical Regions
1. The Interior Plateau
2. The Marginal Areas
III. The People
1. Racial Groups
2. Religions
3. Languages
IV. History
A. Dutch Foundations
B. South Africa Divides (1795–1870)
C. Through War to Union (1870–1910)
D. The Union and the Republic
V. Population
VI. Administration and Social Conditions
1. Constitution
2. Government
3. Taxation
4. Living Conditions
5. Welfare Services
6. Justice
7. Education
8. Defense
VII. The Economy
A. Production
1. Agriculture
2. Forestry
3. Fisheries
4. Minerals
5. Power
6. Industries
B. Trade and Finance
1. Foreign Trade
2. Banking and Currency
3. National Finance
C. Transport and Communications
1. Roads
2. Railways
3. Shipping
4. Air Transport
5. Postal Services and Telecommunications

I. PHYSICAL GEOGRAPHY

1. Geology and Structure.—The geological formations range from some of the oldest rocks of the earth's crust to the most recent formations. The oldest rocks include metamorphosed sediments and lavas of the Swaziland or Kheis system which, together with other more restricted formations, form the Archean complex. Gold and other minerals are associated with these rocks. Next in age are the Proterozoic formations, the oldest of which is the Dominion Reef system of lavas and sediments. This is followed by the Witwatersrand system, 2,500 ft. thick, consisting mainly of quartzites and shales with thin reefs of "banket" (quartzitic conglomerates) from which most of the gold of Transvaal and Orange Free State is derived. Next come the Ventersdorp system of lavas and sediments, and the Transvaal system consisting of Black Reef quartzite at the base followed by dolomite and a succession of quartzites, shales and lavas of the Pretoria series, the system attaining a thickness of about 16,000 ft. Toward the close of the deposition of the Pretoria beds vast intrusions of basic and acid magma in central Transvaal formed a lopolith, *i.e.*, a saucer-shaped mass of norite and allied basic plutonic rocks, and the red granite of the Bushveld Igneous complex. Now exposed by the removal of the overlying beds, this lopolith is revealed as the largest feature of its kind in the world. In these rocks are many minerals, including vast quantities of chrome, platinum and iron. After the intrusion of the Bushveld magmatic rocks the deposition of the sandstones and conglomerates of the Loskop, Matsap and Waterberg systems took place in the late Pre-Cambrians.

In the Paleozoic, deposition continued with the formation of the Nama system, of quartzite, limestone and shale, now found in

BY COURTESY OF SOUTH AFRICAN TOURIST CORP.

COAST LINE AT CLIFTON-ON-SEA, A SUBURB OF CAPE TOWN

western Cape Province and in southern South West Africa. The Cape system followed; at its base the Table Mountain Sandstone series, 5,000 ft. thick, forms most of the prominent mountain features of the western Cape, including Table mountain itself. The Bokkeveld series of sandstones and shales, 5,000 ft. thick, was next deposited and the third series, the Witteberg, of whitish

sandstone, terminated the deposition of the Cape system. In Permian times the whole system was strongly folded. In the Carboniferous the deposition of the Karroo system, which now covers a greater area of the republic than any other formation, began. Having a maximum thickness of about 25,000 ft., the system is divided into the Dwyka, Ecca, Beaufort and Stormberg series. At the base of the Dwyka series is the Dwyka tillite, which was followed by sandstones and shales with numerous doleritic intrusions and crowned in Basutoland with lava sheets now about 4,000 ft. thick—the Stormberg (or Drakensberg) lavas, probably of early Jurassic age. Except in the Cape folded zone and the eastern area affected by the Natal monocline the Karroo sediments have remained almost horizontal.

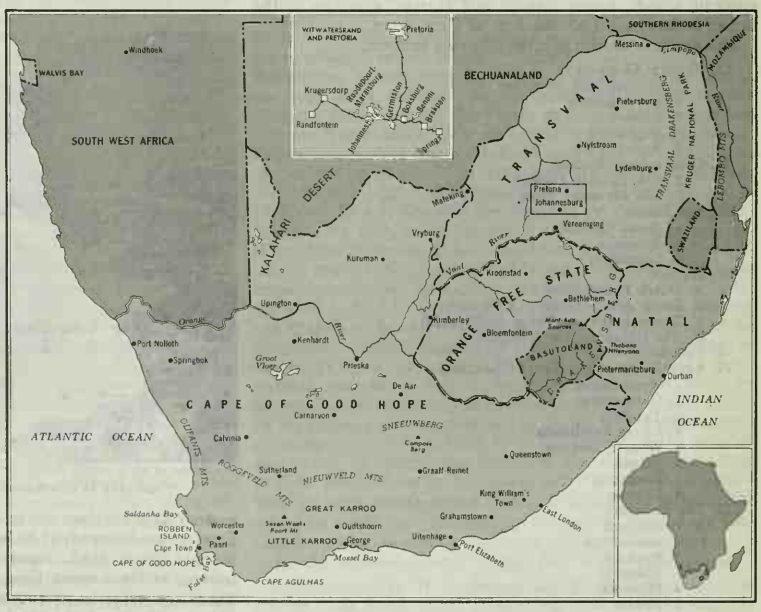
During the Cretaceous there was intense erosion and deposition in the valleys of the Cape folded zone and marine deposition along the south and east coasts; these movements brought into being the valley conglomerates and the estuarine sandstones and clays of the Uitenhage series and the coastal limestones of Zululand. Deposition continued into the Tertiary in the coastal areas, forming the limestones of the Alexandria and Bredasdorp formations in the south and southwest and in the east the Zululand Tertiary limestones.

In the plateau basin of the interior (the Kalahari basin) to the north of the Orange river, the Kalahari system has been deposited, at first apparently under fluviatile conditions and later by aeolian action. At the base of the system beds of calcareous grits, sandstones and conglomerates vary in thickness from about 30 to 300 ft., overlain by a mantle of Kalahari sand varying in thickness from a few to as much as 200 ft. This sand mantle extends continuously northward into the southern part of the Republic of the Congo.

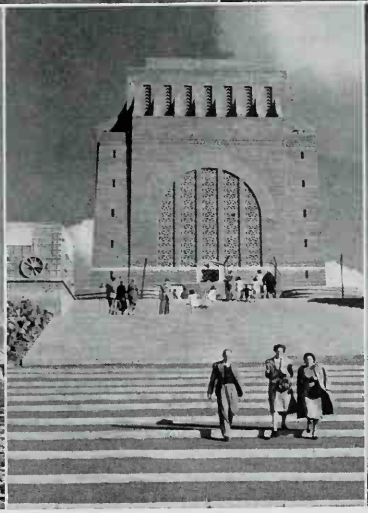
2. Physiography.—The surface of South Africa may be considered as having two generally distinct physiographic elements: the plateau surface and the surface between the plateau edge and the coast, which may be called the marginal areas (*see* also *Geographical Regions* below). Separating these two divisions is the plateau edge, or the Great Escarpment, which is the feature formed by the headward erosion of the drainage of the marginal area. This escarpment is sharply defined or indistinct dependent on whether the plateau rocks are hard-overlying-soft or generally of undifferentiated hardness. Thus the boldest part of the escarpment is along the border of Basutoland and Natal where the Stormberg lavas overlie the softer Stormberg sandstones. There is formed the Natal Drakensberg (Dragon's Mountain) with spectacular peaks such as the Mont-aux-Sources (10,822 ft.) in the north, Champagne Castle (11,072 ft.) and Giant's Castle (10,868 ft.) farther south, the highest point so far surveyed near the plateau edge, and the highest point in South Africa, being Thabana-Ntlenyana (11,425 ft.), about 15 mi. S.W. of Giant's Castle. The height of the scarp above the comparatively flat surface of the Natal high veld is about 6,000 to 7,000 ft. To the north of the Natal Drakensberg the Stormberg lava does not cap the Karroo sediments and the plateau edge is less well defined, but in eastern Transvaal another structural element, the Black Reef quartzite of the Transvaal system, is responsible for the boldness of the escarpment which is there known as the Transvaal Drakensberg, its highest point, Mt. Anderson, a little behind the Black Reef escarpment itself, being 7,498 ft. To the north of the Black Reef quartzite the escarpment is cut in granite and is generally not well defined.

The southern edge of the plateau is given prominence by the presence of strong dolerite sheets which sharply define the Sneeuwberge (Snow mountains) to the north of Graaff-Reinet, where the dolerite Compass Berg has an altitude of 8,215 ft. The Nieuwveld escarpment to the north and west of Beaufort West is of similar formation. The western edge of the plateau is cut in the south in Karroo beds (the Roggeveld, *i.e.*, Rye Veld, mountains), in the middle in the Table Mountain sandstone (the Bokkeveld, *i.e.*, Goat Veld, mountains), and in the north in granite, the Nama beds and the rocks of the Richtersveld complex. It is continued in South West Africa (*q.v.*).

Within the Great Escarpment the plateau surface ranges in altitude from about 3,000 ft. in the west to nearly 11,500 ft. in the east. Two types of plateau surface are distinguishable: the Karroo surface and the pre-Karroo surface, now being exposed by erosion. The Karroo surface, on account of the predominantly horizontal lie of the sediments and their association with dikes and sills of hard dolerite, is tabular in character. Great expanses of flat surface are broken by table mountains and *kopjes* ("hillocks"), generally formed by dolerite capping soft shales or sandstones. In the eastern part of the plateau the Stormberg lavas and the vigorous erosion of the Orange river headwaters have given the Basuto highlands a rugged, dissected character. The pre-Karroo surface has a more diversified topography chiefly because of the greater variety of the rocks. Where the rocks are hard and disturbed the surface is generally rugged, with prominent ridges of quartzite or ironstone and deep valleys formed by the softer rocks. The quartzite ridges of the Witwatersrand are of this character, as are also the Magaliesberg group of ridges which form much of the margin of the Transvaal bushveld basin. In northern Transvaal the Soutpansberg ridges are of a similar type, the quartzites dipping toward the Limpopo trough. Where, however, the pre-Karroo rocks are more nearly horizontal, local plateau surfaces are formed such as the Waterberg plateau in western Transvaal and the Kaap plateau in northwestern Cape



PROVINCES, MAJOR CITIES AND TOWNS IN THE REPUBLIC OF SOUTH AFRICA



PHOTOGRAPHS, (CENTRE LEFT) ACE WILLIAMS FROM BLACK STAR, (ALL OTHERS) BURTON HOLMES FROM EWING GALLOWAY

VIEWS OF SOUTH AFRICAN CITIES

Top: Union buildings, Pretoria, administrative capital of the Republic. Cape Town, however, is the legislative capital and Bloemfontein the judicial

Centre left: The harbour at Cape Town with Table mountain in the background

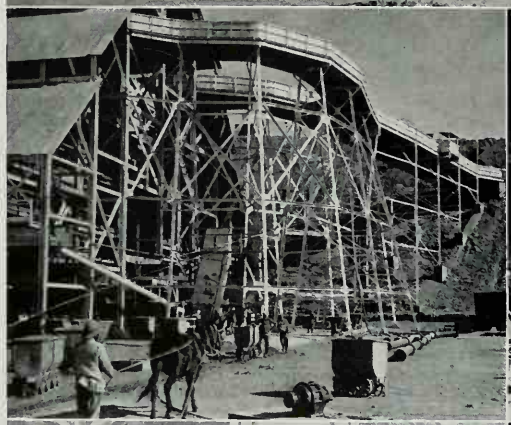
Centre right: Street in Durban showing influence of Indian architecture

Bottom left: Department store of Cape Town

Bottom centre: Voortrekker memorial, Pretoria, commemorating Boer pioneers of 1835–43 who moved into the northern provinces to escape British rule

Bottom right: View of downtown Johannesburg

SOUTH AFRICA, REPUBLIC OF



PHOTOGRAPHS, (TOP, CENTRE RIGHT, BOTTOM LEFT) CONSTANCE STUART FROM BLACK STAR, (CENTRE LEFT) BURTON HOLMES FROM EWING GALLOWAY, (BOTTOM RIGHT) ABC—VAN SPRANG FROM BLACK STAR

INDUSTRY AND NATIVE LIFE

Top: Bushman spear hunters of the Kalahari desert
Centre left: Sifting plant of the diamond mines at Kimberley
Centre right: General store in the native village of Ixopo, near Durban
Bottom left: Matabele woman of Transvaal making a broom
Bottom right: Workmen on a sheep farm shearing wool

Province. At the northern boundary of Transvaal the low floor of the middle Limpopo basin is the result of large-scale trough faulting.

In the marginal zone below the plateau edge prominent features are largely caused by the vigorous erosion of steeply graded streams. Thus hard rocks, such as the quartzites and ironstones of the Swaziland system, form the prominent ridges of northern Swaziland and adjacent Transvaal. The Lebombo mountains, along the eastern border of Transvaal, are low ridges, about 1,500 to 2,500 ft. above sea level, formed of Karroo lavas dipping steeply eastward. In the midlands of Natal there has been monoclinal folding which has given the Karroo beds a steep eastward dip. The main ridges in this region are gigantic spurs jutting out from the Drakensberg between the marginal stream valleys. In the south the marginal zone contains the "Cape folds" of Permian age. In the western part of the folded zone the anticlinal ridges of Table Mountain sandstone run nearly north-south, slightly arcuate and concave to the coast. Two main lines of ranges are distinguishable: to the west the Hottentots Holland mountains, stretching northward to the Olifants River mountains, and to the east the Cedarberg line trending from just south of Vanrhynsdorp toward Ceres. The highest point in this western zone is the Great Winterhoek mountain, 6,815 ft. In the eastern part of the folded zone two main ranges trend eastward, slightly concave to the south: the northern range, known by various names, may be called the Swartberge; the southern range is the Langeberg line. Between these two is a third rather fragmented series of ridges which may be called the Kammanassie line. The highest peak in this eastern zone is the Seven Weeks Poort mountain, 7,632 ft. Between the folded zone and the coast is the coastal foreland and between the Langeberg line and the Swartberge lies the Little Karroo basin. Between the folded belt and the Great Escarpment is the erosional basin known as the Great Karroo (*see* also KARROO).

Hydrography.—The main physiographic division between the interior plateau surface and the marginal area is reflected in the drainage systems of South Africa. The plateau river systems are much larger than the marginal systems; the two most important rivers draining the plateau are the Orange (*q.v.*) (with its affluent the Vaal [*q.v.*]), which flows into the Atlantic ocean, and the Limpopo (*q.v.*), which discharges into the Indian ocean through Portuguese East Africa.

There are very few lakes in South Africa, and although there are natural hollows on the surface of the land, most of these are water filled for only part of the year. These accumulations of water are called pans (Afrikaans, *panne*) and are found especially on the flatter outcrops of the Dwyka and Ecca beds of the Karroo system where the surface has been weathered into shallow hollows. The sandveld of Orange Free State has many pans (of sizes ranging up to about 12 sq.mi.) and these flooded flats occur because percolation is hindered by the presence of hardpan which has formed two or three feet below the surface. The water in pans may be saline, brackish or almost fresh.

In contrast to these flat areas, where occasional heavy rain causes sheets of standing water, are sloping surfaces which are vulnerable to erosion by swift-moving water. If the protective vegetation cover is removed by burning or close grazing, soil is exposed to the attack of heavy rain and inches of topsoil may be washed downhill in a single storm. Runoff may concentrate in tracks made by animals going regularly to waterholes, and as these bare paths offer no check to running water they are quickly attacked and dongas (gullies) are formed. Cultivated land in parts of the eastern heavy summer-rainfall region has been gashed by dongas and lesser runnels called rills.

Coasts.—The comparative straightness of the northern part of the west coast is probably due to the granite character of this area and the lack of large rivers. The southern part has the large inlet of Saldanha bay. Lack of fresh water in the vicinity led to its abandonment as a harbour in favour of the more exposed Table bay where abundant fresh water was available. Table bay is backed by the spectacular Table mountain (*q.v.*) and by the Cape peninsula which ends southward in Cape point, the "Cape of Good Hope." On the south coast the chief features are False bay, Cape Agulhas (Portuguese signifying "needles"), the southernmost point of Africa, and several other capes. At the eastern end Algoa bay is typical of bays formed by erosion of the softer rocks between Table Mountain sandstone outcrops. The east coast is remarkable for its straightness, which is thought to be caused by crustal bending and faulting. Earthquakes off the coast of Zululand indicate present crustal instability.

Off the south coast lies the Agulhas bank extending about 130 mi. S.S.E. of Cape Agulhas, where its depth is about 130 fathoms. Certain parts of the bank are favourable for fishing. Off the west coast, about 7 mi. from Cape Town, lies Robben Island, formerly noted for its seals (Dutch, *robben*). The island, about $2\frac{1}{4}$ mi. long and $1\frac{1}{4}$ mi. wide, has been the site of a prison settlement, a leper asylum and a defense base; it is now designated again as a prison station. The seals have entirely disappeared.

3. Climate.—South Africa is dominated by the tropical belt of high atmospheric pressure which encircles the globe between the latitudes of about 25° and 30° S. In the winter months this belt is continuous over land and sea; in summer it is broken by the occurrence of low pressure systems over the land surfaces. In summer, therefore, a monsoon effect is produced by the occurrence of the "low" over South Africa and the "highs" which move eastward and northeastward along the south and east coasts. The interplay of the wind systems resulting from this general relationship affects both the temperature and the rainfall.

Actual temperatures are strongly affected by altitude which offsets to some extent the normal increase of temperature resulting from decrease of latitude. From the south coast northward the temperatures increase little except in the lowland areas. At the coast Port Elizabeth has a mean annual temperature of 17.8° C. (64° F.); Graaff-Reinet, at an altitude of 2,440 ft. in the Great Karroo basin, 63° F.; Johannesburg (5,750 ft., on the high veld) 61° F.; Pietersburg (4,244 ft., northern high veld) 63° F.; Messina, at 1,800 ft. in the Limpopo trough, 23° C. (73° F.). The monthly range of temperature is least at the coast and greatest in the interior. Thus the range between the mean monthly temperatures of January and July are for Port Elizabeth 6.7° C. (12° F.), Graaff-Reinet 11.7° C. (21° F.); Johannesburg 11.1° C. (20° F.); Pietersburg 10.6° C. (19° F.); Messina 9.4° C. (17° F.). There is also a marked effect of the coastal currents; the cold Benguela current flowing northward along the west coast gives low temperatures to the west coastal areas, while the warm southwestward-flowing Mozambique current on the eastern side of the subcontinent is responsible for the higher temperatures of the eastern coastal areas. Thus on the west coast Port Nolloth has a mean annual temperature of 13.9° C. (57° F.); the corresponding temperature of Durban on the east coast is 21.1° C. (70° F.).

Frost occurs over most of the interior plateau and in the higher parts of the marginal zone. The mean duration of the frost period is highest in the Basuto highlands (more than 180 days) and lowest in the low veld and coastal areas where frost rarely occurs except occasionally in the valley bottoms. The duration of sunshine varies from more than 80% of the possible duration in northern Cape Province to 70% over most of the remainder of the interior, and to less than 60% in the coastal areas.

Three fairly distinct rainfall regions can be recognized according to seasonal occurrence: the summer, the winter and the all-season regions. In the centre and east of the republic the rainfall has a predominantly summer incidence, being mainly monsoon in character and often associated with heat thunderstorms. The moisture is brought into the interior by winds with an easterly component blowing toward the continental low pressure area developed over the southern Kalahari. There is thus a gradual progression of the summer rainfall from the east coast toward the northwestern part of Cape Province from August to January, and a recession eastward from February to May. Except along the east coastal zone little rain falls in this area in the winter. The highest rainfall in this eastern area occurs on the Great Escarpment, where the average annual precipitation is in places more than 60 in., on other high mountain features and on parts of the eastern coastal belt. The summer rainfall is mostly of high in-

tensity, which, coupled with high evaporation, reduces its value for agriculture and increases its soil-eroding tendencies.

In the southwest of the republic is the winter rainfall area in which the rain occurs mainly from May to October. This rainfall is mainly cyclonic and orographic in type, occurring most heavily on the mountain ranges where annual falls of 200 in. are not uncommon. In the coastal lowlands the highest annual averages are about 25 in.; to leeward of the mountain ranges the averages fall sharply to about 10 in. or less. In the third region rain may fall in any season. Thus in the south coastal area rainfall is almost evenly distributed through the year and varies between 20–30 in., while in the narrow corridor in the interior between the margins of the summer and winter rainfall areas average annual rainfall is generally less than 10 in. Taking the republic as a whole it is estimated that about 30% of the surface receives on an average less than 10 in. of rain a year; about 34% receives between 10 and 20 in., 25% between 20 and 30 in. and the remainder more than 30 in. The reliability of the annual rainfall varies from 70% in northwestern Cape Province to about 85% in the southern and eastern coastal areas and more than 90% along the western part of the south coast. It is estimated that about 51% of South Africa is in drought once every five years, 38% once in three years and 24% once in two years, drought being defined here as the occurrence of less than 85% of the average rainfall for the year.

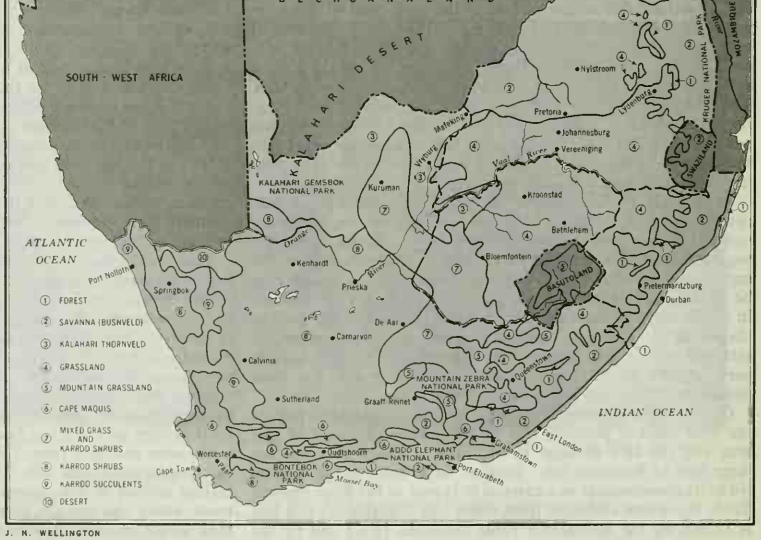
4. Vegetation.—The vegetation of South Africa may be divided into five main types: forest, savanna, grasslands, semidesert and desert, and sclerophyll or Cape maquis.

Forest.—Two general types of forest can be recognized, the temperate and montane, and the eastern subtropical and tropical. The largest area of temperate forest is situated in the eastern part of the south coastal zone where mean monthly temperatures range from about 12° C. (53° F.) to about 22° C. (71° F.), and the mean monthly rainfall from about 1½ to 4½ in. The best-known part of this association is the Knysna forest in which the main trees include the tall yellowwoods (*Podocarpus* species), growing to a height of 150 ft., the black ironwood (*Olea laurifolia*), the Cape beech (*Myrsine melanophloeos*), the stinkwood (*Ocotea bullata*), the sneezewood (*Ptaeroxylon obliquum*) and many others. Similar associations occur in smaller areas on the south-facing slopes of the Cape mountains, especially on the Hogsback and Amatola escarpments and in the Natal midlands where, however, only small remnants of the former great forests are left. In eastern Transvaal large forests formerly covered the escarpment of the Transvaal Drakensberg, but these have been exploited for mining operations and the largest natural forest in this area is now conserved in the Woodbush mountains. The subtropical and tropical forest formerly stretched along a narrow zone behind the east coast where the rainfall is 30–50 in. a year. The dense bush and evergreen forest includes trees of the genera *Sclerocarya, Protorhus, Albizzia, Trichilia, Ficus* and many others. Formerly there were valuable trees such as the Cape mahogany (*Trichilia emetica*), the Cape ebony (*Heywoodia lucens*) and the Cape box (*Notobuxus macowani*), but these have almost all been exploited. (*See* also *Forestry* below.)

Savanna.—The bushveld, as it is generally called in South Africa, is generally the dominant association where the rainfall is between about 15 and 30 in. a year and frosts and cold are not severe enough to kill the young trees. The bush may be open, with sparse tree growth, or closed, with dense tree growth. Where no species is dominant the bush is termed mixed. In the republic mixed bush is the most common type, occupying most of the middle veld (between about 2,000 and 4,000 ft.) in eastern Cape Province and Natal, and in Transvaal covering the whole of the plateau area below the altitude of about 4,000 ft. In the drier areas the most common trees are acacias, with aloes and euphorbias, and with the baobab (*Adansonia digitata*) forming a distinctive feature in northern Transvaal. In the moister parts of Transvaal deciduous broad-leaved trees dominate, among which are the maroola (*Sclerocarya caffra*), various sterculias, the South African mahogany (*Afzelia cuanzenzis*) and in the lower valleys various palms. The commonest grasses are of the genera *Themeda, Digitaria, Panicum,* with many others.

Grasslands.—These occupy the higher, cooler areas where the annual rainfall exceeds about 15 in. It is usual in South Africa to distinguish between "sweet" veld grasses, which occur mainly in the western, drier part of the grassland area, and which are nutritious and palatable to stock throughout the year, and "sour" veld, consisting of grass types which during the dry season lose their nutritious character and become woody and indigestible. On the plateau the climax type over most of the area is the red grass (*Themeda triandra*). Retrogression to the more xerophytic *Eragrostis* species follows overgrazing and soil erosion and, later, to *Aristida,* including spear grass, or *Cynodon* (couch grass). In the higher parts of Basutoland the veld is alpine in type, including the *glauca* variety of the red grass, the fescues *Festuca rubra* and *F. caprina,* and *Danthonia disticha.*

Semidesert.—Vegetation is comprised mainly of scattered Karroo shrubs and succulents, which occupy most of the area having an annual rainfall of between 5 and 10 in. The shrubs are mostly perennial woody bushes about 3 to 12 in. high and spaced from about one foot to several feet apart. The leaves are generally very small and the root systems large. The commonest include various species of *Pentzia,* the bitter karroo (*Chrysocoma*



J. N. WELLINGTON

NATURAL VEGETATION REGIONS OF SOUTH AFRICA

tenuifolia) and, in brack soil, various gannas (*Salsola* species) and the saltbushes (*Atriplex* species). Some of these bushes can survive a drought of nine or ten months and they form excellent subsistence feed for small stock. The succulents are of two types: low bushes, rarely more than eight inches high, with ericoid shrublets and many species of *Mesembryanthemum*, and a taller type consisting mainly of euphorbias, aloes and *Portulacaria*. In the lower Orange river valley the rocky slopes harbour many desert types such as the remarkable *Pachypodium namaquanum* and other succulents such as *Ceraria namaquensis, Aloe dichotoma, Euphorbia virosa, Commiphora sexicola* and many sarcocaulons.

Cape Maquis.—The sclerophyll vegetation, or Cape maquis, is confined to southwestern and southern Cape Province where the plants are protected against the hot, dry summers. The shrubs are found from sea level to about 3,000 ft. With annual rainfalls of about 20–30 in. they grow to a height of about 15 ft.; near the lower rainfall limit of about 15 in. the shrubs are 3–5 ft. high, the leaves often very small and leathery. The families typical of the region are Proteaceae, Restiaceae, Ericaceae, Peneaceae, Grubbiaceae, Bruniaceae, Rutaceae and Iridaceae. A remarkable difference between this region and most other regions of Mediterranean vegetation is the general absence of trees in the Cape area.

5. Animal Life.—Zoologically the republic is recognized as forming the southern part of the southern African subregion of the Ethiopian faunal region.

Mammals.—South African mammals are all placental mammals. The insectivores include little elephant shrews and golden moles. There are more than 20 genera of bats.

The primates are represented by the bush baby or night ape (*Galago*) and by the Cercopithecidae, which include the vervet monkey, the samango monkey and the chacma baboon. The order Pholidota is represented by the Cape pangolin or scaly anteater.

The carnivores include Canidae (foxes and jackals), Mustelidae (polecat, weasel, ratel or honey badger, otter), Viverridae (civet, genet, mongoose, meerkat), Protelidae (aardwolf, hyena) and Felidae (wildcat, serval, lynx, leopard, lion, cheetah). The Pinnipedia are represented by the Cape sea lion, or fur seal, and the elephant seal, which occasionally visits the south coast. The Tubulidentata has one representative, the aardvark or antbear; the Proboscidea also one, the African elephant. Of the Hyracoidea there are two genera, the hyrax and the dassie, the former being one of the commonest mammals in the republic.

Among the odd-toed hooved mammals are two species of the genus *Diceros*, the black rhinoceros and the white rhinoceros, and two species of the genus *Equus*, the mountain zebra and Burchell's zebra. The even-toed ungulates include the bush pig, the wart hog, the hippopotamus, the giraffe and about 20 genera of antelope (or buck, as they are generally called in South Africa), including duiker, grysbok, oribi, suni, klipspringer, vaal rhebok, reedbuck, waterbuck, impala, springbok (or springbuck), gemsbok, roan and sable antelopes, tsessebe (or sassaby), bontebok, blesbok, hartebeest, black and blue wildebeest, bushbuck, inyala, kudu, eland and African buffalo. The order Lagomorpha has two genera of hares.

In point of numbers the largest order is probably that of the rodents. Six families are represented, the animals included in them being the mole rat, cane rat, squirrel, spring hare and dormouse, but the largest family is that of the Muridae, which includes rats, mice and gerbils. Five genera of murids are peculiar and endemic. Of the Cetacea in the South African seas about 20 genera of whales and dolphins are found in greater or lesser numbers.

The present state of the fauna in the republic may be summarized by saying that most of the small mammals are still found wild. The larger animals, especially the carnivores and antelopes, survive only in protected areas. Lions and elephants are found now only in or near the game reserves. The larger buck are preserved on some farms and in the reserves. Smaller buck still roam wild, and the leopard and baboon, the jackal and hyena are still found in the less frequented areas. The blue antelope (bloubok) and the quagga of the Cape have become extinct. The elephants of the Knysna forest—the tallest of all African elephants—in the early 1960s existed in a herd of fewer than ten beasts and also the Addo elephant herd had no adult bull.

Birds.—Most of the bird families are of the Ethiopian type but of the groups of families one is tropical in character, another antarctic or southern in its origin and yet another consists of migratory types. The tropical group consists of well-known types from the central African forests such as the parrots and lovebirds, the touracos, the rollers and many others. These are found mainly in the eastern part of the republic, having come southward along the east African savanna corridor. The southern group includes most of the sea birds: the jackass penguin (*Spheniscus demersus*), the albatrosses and petrels, including the Cape pigeon (*Daption capensis*), the Cape hen (*Procellaria sequinoctialis*) and several shearwaters now also included in the genus *Procellaria*.

Of the large group of migratory birds from the Palearctic region are such well-known species as the European swallow, the martin, swift, cuckoo, willow warbler, garden warbler, yellow wagtail, tree pipit, golden oriole and many others. Of the endemic Ethiopian families only the Promeropidae, the long-tailed sugarbirds, are peculiar to South Africa, the Cape sugarbird and Gurney's sugarbird being confined respectively to the Cape and the eastern areas. Among the other interesting species are the rock jumper (*Chaetops*), the ground woodpecker (*Geocolaptes*), the bald ibis (*Geronticus*) of the eastern mountains, the blue crane (*Tetrapteryx*) and the blue korhaan (*Eupodotis*), all of which are peculiar to South Africa. The largest number of families are in the orders Passeriformes (including, among others, the larks, wagtails, babblers, bulbuls, thrushes, warblers, shrikes, sunbirds, weavers, sparrows and canaries); the Charadriiformes (including the plovers, phalaropes, skuas, gulls and terns); and the Gruiformes (including the cranes, bustards, curlews, jaçanas, crakes and coots). The family Struthionidae is represented by the southern ostrich, which is still wild in certain areas; the Sagittariidae by the snake-killing secretary bird, which is widely dispersed over the savanna and grassveld. An interesting recent newcomer is the Indian myna, which was introduced at first into Natal and now is found in southern Transvaal.

Reptiles and Amphibia.—Turtles and terrapins are represented, but tortoises are of special importance, six genera being recognized. The lizards, chameleons, leguans and geckos include numerous species and subspecies native to the subcontinent. The representative of the Crocodilia is the Nile crocodile, whose southern limit on the west side of the subcontinent is the Cunene river and on the east the Tugela river.

Of the more than 100 species of snakes about 25% are poisonous. The nonpoisonous snakes include the useful mole snake and the curious egg-eaters (*Dasypeltis scabra*). The back-fanged snakes, which are moderately poisonous, include such well-known species as the skaapstekers or "sheep-strikers" and the boomslang or tree snake. Among the highly poisonous front-fanged snakes are the deadly mamba, three species of cobra and the closely allied ringhals, *i.e.*, "ringed-neck," a spitting snake peculiar to South Africa. The vipers are represented by the puff adder and the horned adder. South Africa also possesses one species of python, which may attain a length of 20 ft., and also one venomous sea snake (*Hydrus platurus*).

The amphibians of South Africa are chiefly frogs and toads. There are no tailed amphibians, and there is only one genus of limbless amphibians. Of frogs and toads there are 5 families and about 80 species, the size varying from the minute form *Arthrolepis*, which may be less than half an inch, to the great bull frog *Rana adspersa*, or *Pyxicephalus adspersus*, which sometimes attains a length of more than a foot. Among the toads one of the largest is the leopard toad, *Bufo regularis*, which is four to five inches long and is widespread in South Africa. The smallest is the tiny striped mountain toad, *Bufo rosei*, which is about an inch long. From the commercial point of view the most important amphibian is the aquatic, tongueless, clawed frog *Xenopus laevis*, the Platanna, which provides a dependable test for human pregnancy in the early stages.

Fresh-Water Fishes.—Of about 200 species of fresh-water fishes nearly half belong to the carp family, of which the best known is the barbel. Kurpers comprise about one-quarter, and of the remaining quarter the most numerous are the catfish and tiger fish. Trout have been successfully introduced.

Insects.—All the important orders of insects are represented in South Africa. The orthopterans are numerous and diverse, simulating leaves, flowers, sticks and other forms. All are carnivorous. The brown locusts which formerly swarmed over South Africa are now destroyed in the immature stages. Termites or "white ants" are abundant in species and numbers. Many families of bees, wasps and ants also are represented. True bumblebees are absent but the carpenter bee (*Xylocopa*) is common and there is an immense variety of solitary wasps. There are many endemic beetles; especially abundant are the scarabs, or dung beetles, and the chafers. A few types of butterflies and moths are found in the southwest but many occur in the east, some of the genera being peculiar to South Africa. The Diptera include the culicine and anopheline mosquitoes, and the tsetse fly which was formerly widespread in Zululand.

South African arachnids are abundantly represented by scorpions and solifuges, or hunting spiders, which flourish especially in the hot, arid conditions of the Karroo and Namaqualand. Ground-living spiders are abundant. *Peripatus,* the wormlike "missing link" in the ancestry of insects, is endemic but is fast disappearing as the natural forest is destroyed.

Game Reserves.—In South Africa these serve the important purpose of conserving the game, many of which are in danger of becoming extinct. There are five national parks and a large number of provincial reserves. The Kruger National park, the largest, extends for about 200 mi. along the Transvaal-Mozambique border following the Crocodile-Limpopo river, with an average width of about 40 mi. Nearly all the larger game animals (the notable exceptions being the gemshok and the springbok) and the carnivora are found in the park, as are also the elephant, hippopotamus and giraffe, but not the rhinoceros. The eland and the inyala frequent only the northern parts of the park. The Kalahari National park is situated between the Nossob and Auob rivers in the extreme north of Cape Province. This is also known as the Gemsbok reserve on account of the large numbers of these animals in the park, but springbok also are very numerous and there are also hartebeest, eland and kudu, and a fair number of carnivores including lion, leopard, cheetah and wild dog. The Addo Elephant National park, about 35 mi. N. of Port Elizabeth, is notable for its small herd of elephants which, with the Knysna herd, are the southernmost of the African elephants. The Cape buffalo also thrives there, the only place now in Cape Province where it is found.

The Bontebok National park (about 1,800 ac.), about 17 mi. S. of Bredasdorp in Cape Province, was formed to preserve the last remaining bontebok, of which a herd of about 150 has been established. The Mountain Zebra National park, near Cradock in Cape Province, was formed to preserve the few remaining mountain zebras. In the early 1960s these were increasing slowly. Of the provincial reserves the most notable are those of Natal. The Hluhluwe and Umfolozi reserves (in Zululand) have preserved the white and the black rhinoceros.

II. GEOGRAPHICAL REGIONS

In South Africa physiography offers the best single basis for regional division. As described above (see *Physiography*) the two primary divisions are the interior plateau and the marginal areas.

1. The Interior Plateau.—Five major regions may be recognized: the high veld, the middle veld, the Transvaal plateau basin, the Limpopo trough and the Kalahari basin.

The High Veld.—Lying between the altitudes of about 4,000 and 11,500 ft. above sea level, the southern part of the surface is formed by the beds of the Karroo system (see *Geology* above). Eroded by the Orange-Vaal headwaters, this structure gives rise to a tabular type of physiography consisting of flat plains and scattered relict mountains and *kopjes,* generally capped by a sill of hard dolerite. The surface is ideal for farming of a pastoral type, being covered with grass suitable for cattle in the eastern part, and with grass and karroo shrubs in the western part suitable for sheep and goats. Soils, however, are rather infertile except on dolerite outcrops. The summer rainfall in the eastern part (25–35 in.) is adequate for maize (corn), but not in the western part (15–20 in.), where kaffir corn is more suited to the drier conditions. In the east of the region, the Basuto highlands, the high plateau surfaces provide summer pastures only; in the valleys are narrow alluvial terraces which form sheltered habitable areas in winter and provide limited cultivation of maize and kaffir corn. The northern part of the high veld region has a different type of surface. The Karroo beds have been removed by long-continued erosion exposing the pre-Karroo surface which was planed by the Permo-Carboniferous ice sheet. This surface, now being eroded by the headwaters of the Limpopo and Vaal rivers, has become more uneven than the Karroo surface as the Proterozoic quartzites, shales and lavas have been thrown into relief. The great economic factor is the mineral content of the Witwatersrand beds, the "banket" or quartz-conglomerates of which contain the gold particles that have made the Rand the world's greatest gold field. The development of the Witwatersrand has been facilitated by the abundant coal supplies from the neighbouring Karroo beds.

The Cape Middle Veld.—This lies to the west of the high veld at an altitude of about 2,500–4,000 ft. There is no natural boundary between these two regions, but the lower altitude of the middle veld is responsible for its higher summer temperatures, and its greater distance from the east coast for its lower rainfall (5–15 in.). The region is thus arid to semiarid and where there is no irrigation is suitable only for small stock farming of an extensive type. Intensive irrigation is practised along the Orange river in this region.

The Transvaal Plateau Basin.—This basin, to the north of the high veld, comprises the great Transvaal bushveld basin and its highland rims. The floor of the basin is formed mainly of norite and red granite. The norite soils are fairly fertile and repay cultivation, but the granite soils are sandy and infertile. Overlying the plutonic rocks in the northern part of the basin is the Stormberg lava which has decomposed into very fertile black clay soil; the rainfall, however, is low (25 in.) and rather unreliable. The Springbok flats, as this part of the basin floor is called, is thus an area of mixed farming rather than of intensive cultivation. The highland rim of the basin is formed by the high veld in the southeast, by the *bankeveld* (or cuesta) topography in the southwest and east, and mainly by the Waterberg plateau in the north. The Pietersburg plain and the Soutpansberg ranges form the extreme northern part of the region. The highest economic potentialities of the region are associated with the irrigation areas along the Crocodile, Olifants and Marico rivers and with the smaller schemes in the valleys of the *bankeveld* marginal areas. Minerals such as chromite and platinum are present in large quantities in the plutonic complex of the basin floor and low-grade iron ore in almost inexhaustible quantities in the marginal quartzites.

The Limpopo Trough.—This is the middle portion of the Limpopo basin. Formed mainly by large-scale faulting, the valley has an altitude of between about 1,500 ft. at its eastern end (Pafuri) and 3,500 ft. at its southwestern end (Lohatsi). With low rainfall (15–25 in.) and high temperatures this area is of little agricultural importance, being mainly suited to extensive ranching. Copper in the Messina area is the only important mineral.

The Kalahari Basin.—Only the extreme southern part of the Kalahari falls within the republic (see *Geology* above). The sand surface is more or less covered with grass and sparse tree growth and water is generally absent except in scattered pans and along the dry watercourses. The area is thus scarcely inhabited away from the river courses such as the Molopo in the east and the Nossob and Auob in the west.

2. The Marginal Areas.—In the area marginal to the plateau the eastern, southern and western areas form fairly distinct types of country.

The Eastern Marginal Area.—The fall from the plateau edge to

the coast is in many parts taken in three steps or surfaces, forming a high veld surface above about 4,000 ft., a midland surface between about 2,000 and 4,000 ft. and a coastal zone below about 2,000 ft. In Natal and the Transkei these are all areas of fairly high summer rainfall (25–40 in.) so that cultivation is generally possible, the crops being determined mainly by the temperature. On the eastern high veld surface the dominant occupation is pastoral farming, in the midlands mixed farming and in the coastal zone of Natal the cultivation of sugar cane.

The Southern Marginal Area.—The surface is dominated by the Cape folded ranges (see *Physical Geography* above). Agriculture is confined mainly to the well-watered coastal belt and to the irrigable valleys of the Little and Great Karroos. The Karroo vegetation is well suited to sheep and goat farming and the irrigated lands to the cultivation of lucerne and, in some areas, of fruit.

The Western Marginal Area.—The surface falls rather irregularly and the climate is too dry for any sustained agriculture to the north of about Vanrhynsdorp. Southward the area between the folded mountains and the coast is well suited to winter cereals, especially wheat, and to summer fruits such as grapes, peaches, plums and apricots. (J. H. Wn.)

III. THE PEOPLE

1. Racial Groups.—South Africa contains peoples of many diverse ethnic origins and differing cultures, languages and forms of social organization. This complex situation may best be described historically, in terms of immigrations into the region.

Bushmen and Hottentots.—These, the surviving representatives of the original inhabitants of South Africa, are of non-Negro physical type, with languages which, though related to each other, are unrelated to any other language families. They are together known as Khoisan (*q.v.*). The Bushmen are, and would seem always to have been, hunters and gatherers, living a wandering life and divided into small isolated bands lacking any form of centralized chieftainship. Traces of their inhabitance, mainly in the form of rock engravings, show that they were once widespread throughout southern and central Africa; but they are now virtually confined to the more arid areas of South West Africa and Bechuanaland. The Bushmen of what is now the Republic of South Africa were mainly killed, driven out or absorbed by Negroes and whites. The Hottentots were pastoralists, keeping cattle and sheep, and occupied most of the southwestern corner of Africa at the time of the Dutch arrival at the Cape. Only the Nama Hottentots of South West Africa now form a distinct Hottentot-speaking group. Others have become part of various mixed-blood communities, described below. In the early 1960s Bushmen numbered about 55,000 and the Nama 25,000. (*See* Bushman; Hottentot.)

Nguni.—The main element of the South African population consists of various Negro peoples speaking southern Bantu languages. They are, indeed, conventionally classified according to their linguistic affiliation and collectively referred to as Bantu for statistical purposes. The first Negro group to have entered South Africa would seem to have been the cluster of tribes speaking Nguni dialects. They occupied what is now Natal and had covered much of the eastern Cape before the arrival of Vasco da Gama, and held the same areas when met by the Dutch and English in the 18th century. Nguni languages contain three of the so-called "clicks" (*c*, *q* and *x*) of the Bushmen and Hottentots, and the Nguni tribes show some Hottentot physical characteristics; presumably they drove the Hottentots before them in their southern migrations and absorbed many of the Hottentot tribes. The Nguni are usually divided into the southern Nguni of the Cape and the northern Nguni of Natal. The southern or Cape Nguni consist of several tribes, mostly living in the Transkei area: the Pondo, Thembu (Tembu), Xhosa, Bomvana, Mpondomise and others, in all numbering well over 2,500,000. The northern Nguni include the Zulu, Swazi, Bhaca, Fingo and others, and also total over 2,500,000, in Natal and the surrounding areas. As a consequence of the political upheavals of the early 19th century associated with the rise of the Zulu king Shaka (Chaka), several offshoots of northern Nguni moved northward: the Ndebele (Matabele) of the Transvaal, the Ndebele of Southern Rhodesia, and the Ngoni of Nyasaland, Northern Rhodesia and Tanganyika. All the Nguni of South Africa were traditionally organized into chiefdoms, reckoning descent patrilineally and observing strict patriarchy, and with economies of mixed agriculture and cattle keeping. Some of them, particularly the Zulu and Swazi, established kingships and became important military nations in the 19th century. (*See* Ngoni; Nguni; Swazi.)

Sotho.—Behind the Nguni came the various Negro peoples speaking Sotho dialects, which are closely related to Nguni but lack the Hottentot-derived clicks. The Sotho-speakers consist of the southern Sotho, or Suto, of Basutoland; the western Sotho, or Tswana, of Bechuanaland; and the northern Sotho of the Transvaal (the Pedi, Lovedu and others). In the northern Transvaal are also the Venda, sometimes classified with the Sotho but usually counted as a distinct group. In their economy the Sotho peoples resemble the Nguni tribes, but they differ in certain other respects: some of them live in large "towns" rather than in the dispersed settlements of the Nguni, and they approve of marriage between cousins, whereas the Nguni prohibit it. They reckon descent patrilineally, like the Nguni, and are also organized into chiefdoms, often with rulers of marked sacred or even divine attributes. Some rulers, such as those of the Lovedu in Transvaal, are women. They play a male role and "marry" "wives," whose children may inherit whereas any children of the ruler are counted as illegitimate. The rulers of the Western and Southern Sotho chiefdoms have greater secular and military power and resemble those of the Zulu and Swazi. (*See* Suto; Tswana.)

European Intrusion.—The Europeans have lived in South Africa since the founding of Cape Town in 1652. Originally Dutch, the white population was first concentrated in the western Cape. The Dutch made early contact with the Hottentots, and seem at first to have treated them with kindness and to have regarded them as of more or less equal standing (there was occasional intermarriage and much concubinage). With the extension of white farms along the Cape seaboard, whites came into direct competition with Hottentots over water and grazing and drove some of them into the more arid regions to the north and made others into little more than serfs. The Dutch imported slaves, mainly from the East Indies and Madagascar. There was much miscegenation between whites (both Dutch and English) and slaves, Hottentots and Bushmen, and the descendants of these various unions later became the Cape Coloured people. The Dutch, to show their allegiance to their new country, took the name Afrikaner.

The whites came into contact, and also into conflict, with Negroes on the eastern borders of the Cape during the 18th century. The ensuing "Kaffir Wars," fought between white and Negro farmers who desired the same lands and water supplies, since both groups were mixed agriculturalists and cattle keepers, led to considerable tribal disruption and movement. This resulted in the founding of the Basuto nation largely of refugees from elsewhere, the appearance of the Fingo of the Ciskei from the remnants of broken Nguni tribes, and later of many small racially mixed communities of white-Hottentot-Negro origin (the Griquas, Bastaards and others).

With the growth of the diamond and gold industries and the development of cities such as Johannesburg and Kimberley there developed the system of migrant Bantu labour to the towns, mines and white farming areas outside the Cape (in the Cape itself Coloured labour has always been predominant). This has led to an increasing Negro population in the towns, both purely migrant (*i.e.*, temporary) with its roots in the tribal reserves, and also fully urbanized and detribalized, living permanently in the locations of the cities and towns. A good proportion of the temporary labour, especially that employed in mines, comes from outside South Africa, from Basutoland, Bechuanaland, Portuguese East Africa and the Rhodesias. The permanent African urban population has tended to drop tribal customs and even tribal languages and to become a new amalgam with an urban western way of life, although economically an extremely depressed one.

Asians.—The last main element in the South African population

is Indian. Indians were brought to Natal in 1860 mainly as indentured labourers on the sugar estates in the humid coastal zone. A century later they numbered over 300,000 in Natal, two-thirds of whom live in Durban, with smaller communities elsewhere, mainly as traders. There are also a few hundred Chinese and smaller numbers of Mauritians, St. Helenians and others, almost all in the towns.

2. Religions.—The traditional religious beliefs and practices of the peoples of South Africa vary widely. Those of the Bushmen and Hottentots were based upon worship of the moon and other heavenly bodies, although both peoples believed in an afterlife also. Among the Bantu-speaking groups the principal indigenous cults are those of the ancestors, sacrifice to them being made by the heads of lineages and clans. A high god was recognized but sacrifice was not made to him. Beliefs in witches and sorcerers were and are prevalent, witches being thought to be women throughout this region. In most of the states and large chiefdoms the ruler has important ritual duties at first-fruit ceremonies and other rites, which have both religious and political aspects. In some states the ruler has great sanctity: the king of the Venda is regarded as divine before his death, for example, and the queen of the Lovedu is recognized widely for her near divinity and power to control rain and fertility.

Christianity has become widespread among all the Bantu groups. In most areas there are, besides orthodox mission churches, separatist churches led by Bantu priests, bishops and archbishops (the latter usually self-consecrated), who resent white control of most mission churches and regard themselves as prophets or messiahs. The separatist churches consist mainly of various Zionist churches, in which speaking with tongues and "rolling" are practised, and Ethiopian churches in which the antiwhite, tribalistic and nationalistic emphasis is strong.

Most Indians follow Hindu or Muslim rites. The majority of the Cape Coloureds are Christian, only the Cape Malays being Muslims. Among Europeans the most powerful churches are the various Calvinistic sects which form the Dutch Reformed churches. These churches, which have a large Coloured membership, accept racial segregation as divinely ordained, and as a result they have come under much criticism from Protestant churches outside South Africa. (*See* REFORMED CHURCHES.)

3. Languages.—The two principal official languages are Afrikaans and English. At the 1960 census Afrikaans, a Netherlandic language, was given as first among 58% of the white population and English among 37.3%, but with instruction in both languages in schools, an increasing number of people, particularly the younger generations, are bilingual. (*See* also SOUTH AFRICAN LITERATURE: *Afrikaans*).

Africans speak various Bantu dialects; some, such as Xhosa (for Transkei Bantustan), are accepted as official languages within the homelands and are used as vernaculars in education (*see* BANTU LANGUAGES). A large number of Africans also speak English or Afrikaans, and a lingua franca known as Fanagolo is also used in the mining areas. (J. F. M. M.)

IV. HISTORY

Whether or not the Phoenicians sailed around Africa as claimed by Herodotus, southern Africa was unknown to the Greeks and the Romans and to Europeans before the 15th century. Then, largely as a result of the patronage of Prince Henry the Navigator (1394–1460), Portuguese mariners steadily pushed down the west African coast until in 1488 Bartolomeu Diaz de Novaes rounded the Cape of Good Hope and in 1498 Vasco da Gama, in a flagship of 120 tons, 80 ft. long, reached the objective, India. Thereafter the Portuguese gained maritime supremacy in the Indian ocean and used it to exploit the Asiatic spice trade and, less effectively, the east African gold trade. But, though they colonized Angola and Mozambique, they made no settlements as far south as what is now the Republic of South Africa, regarding its resources as negligible and its inhabitants as too primitive for commercial or evangelical effort. Throughout the 16th century, therefore, South African waters were often traversed by Portuguese fleets, but the only civilized people who spent much time ashore were the survivors of the many ships that were wrecked on the coast.

A. Dutch Foundations

Trading Companies.—By the end of the 16th century Dutch and English seamen had also begun to trade with Asia by the Cape route. They found it desirable to pause for refreshment, the death rate on the long voyage being often as high as 50%; so during the first half of the 17th century they usually stopped at St. Helena, the Cape or Mauritius. In 1615 the English tried vainly to form a settlement at the Cape with ten felons reprieved from the hangman's noose, and five years later two English captains made the empty gesture of annexing all Africa for James I; but it was St. Helena that eventually became the main port of call for the English company.

Meanwhile the Dutch East India company outstripped all rivals and obtained a virtual monopoly of the East Indian spice trade. For several decades Dutch ships made a practice of putting into Table bay to take in fresh water, to barter meat from the Hottentots and to find the latest information about the affairs of Europe or Asia in dispatches they left one another under inscribed stones. Then in 1647 Leendert Janssen was shipwrecked in Table bay with the crew of the "Noord Haerlem," and after his return to the Netherlands he presented a report advising the company to found a permanent refreshment station at the Cape. Largely because the Anglo-Dutch War of 1652–54 was imminent, the directors eventually agreed; three ships were to take out materials for building, seeds and implements, and 90 men were to build a fort and develop a vegetable garden.

Jan van Riebeeck.—To command the expedition the directors chose Jan van Riebeeck, a doctor by training, who had already had six years of commercial experience with the company. He landed at the Cape on April 7, 1652, and remained ten years, during which time droughts, floods and pests, indiscipline among his subordinates and thefts by the Hottentots often brought the settlement to the brink of disaster. But it survived; a timber-and-sod fort was built, the coast line was charted, the interior was explored, and a belt of land extending about eight miles from Table bay to Wynberg was brought under cultivation.

However, the Cape did not become the simple outpost which the directors had had in mind. At that time Hottentots, pastoral nomads loosely organized in small tribes, occupied most of the coastal regions of South Africa as far east as Natal. The Cape peninsula had been used by three such tribes, who regarded the Dutch as intruders in their domain, especially when it became apparent that they were not birds of passage but were building fortifications and extending their area of cultivation. The Dutch bartered with the Hottentots for sheep and cattle and from time to time the Hottentots stole them back again. Van Riebeeck calculated whether he was strong enough to use force, but when the directors cautioned him to do no such thing he fell back on the idea of confining the settlement within a palisade. Yet even at this early stage in South African history it was not possible to keep the races completely apart, because the Dutch continued to barter with the natives.

Van Riebeeck also initiated two experiments which had far-reaching consequences. At first all the work of the station was done by the company's servants, but in 1657 a few slaves were imported and more came in 1658 and thereafter. The second experiment started in 1657 when the company, hoping it would save money if corn were grown and stock bred by unpaid freemen instead of by paid employees, freed nine married men from their contracts and gave them 30-ac. farms along the Liesbeeck valley at Rondebosch. These people, who were known as burghers, were to produce corn and stock and sell them to the company at agreed prices. So Van Riebeeck left behind him an embryonic colony of settlement which used slave labour and was confronted with a "native problem."

Immigration and Expansion.—In 1662 the Cape settlement numbered about 250 white persons, of whom nearly half were company servants, and in the next few years there was little increase. But during the governorships of Simon van der Stel (1679–99) and his son W. A. van der Stel (1699–1707) the company made its

one serious effort to encourage immigration. In 1707 a census showed a burgher population of 1,779 men, women and children, owning 1,107 slaves. This was a slow rate of progress compared with that of the English colonies in North America. Nevertheless, the modern Afrikaner people are descended for the most part from those who were enumerated in 1707. These were company servants who had taken their discharge at the Cape and people who were sent out to settle. Most of them were of Dutch and German stock. There was also a small but significant group of rather fewer than 200 French Huguenots who, after the revocation of the Edict of Nantes in 1685, went to the Netherlands and then were sent to the Cape. As a matter of policy the company scattered the Huguenots among the other settlers, and within little more than a generation they became fully assimilated, discarding the French language and even pronouncing their names in the Dutch manner. The area of settlement expanded with the population—to Stellenbosch beyond the sandy Cape flats (1679) and then to the Berg river valley. To meet the needs of local administration a general purposes official (landdrost) was appointed to Stellenbosch in 1685; sitting with four local burghers (heemraden), he was given jurisdiction in minor civil cases. Nearly all the company servants, however, lived in Cape Town, where the governor and his chief subordinates formed a council of policy which issued regulations for the colony, subject to the overriding authority of the governor general and council of the Indies in Batavia and the directors in the Netherlands.

From the first the burghers had fretted at many of the regulations, which imposed restrictions and taxes on internal trade; for example, the rights to deal in much local produce, including cattle, were auctioned annually to contractors, as well as the rights to retail imported goods. Furthermore, by the end of the century the Cape was producing more wine, brandy and corn than was consumed locally and by passing ships, and there was little opportunity for export.

What was particularly galling to the burghers was the tendency of the officials to farm and trade on their own account, using their authority to make sure that their own produce did not remain unsold. This led to an outcry against the governor, W. A. van der Stel, who embarked on large-scale farming operations and cornered the market so successfully that some malcontents, including the Huguenot diarist Adam Tas, sent a petition to the directors, who recalled Van der Stel in 1707. But the underlying cause of the discontent—the commercial policy of the company—was not remedied.

Evolution of the Afrikaner People.—The company, probably convinced that white settlement led to political agitation, had no vision of a New Netherlands in South Africa and sponsored no immigration schemes after the time of the Van der Stels. Nevertheless, during the 18th century a New Netherlands of a sort was arising from the materials present. The white population grew, mainly by natural increase, to about 15,000 in 1795, by which time they regarded South Africa as their only home and spoke a language, Afrikaans, which deviated from the language of the Netherlands. Those who lived in Cape Town—the only port and the only real town in the colony—were mostly traders and innkeepers. In the southwestern part of the colony, near Cape Town, were the corn and wine farmers, some of them proprietors of large estates run by slave and Hottentot labour. Beyond them were the trekboers, about half the total white population. The trekboers were seminomadic, many of them completely nomadic, pastoral farmers. Most of the land they occupied was unsuitable for cultivation; moreover, the market for agricultural produce was glutted and slaves did the skilled work of the colony. There was, therefore, nothing to keep the increase of the white population in the southwestern Cape and much to encourage them to disperse and become self-sufficient pastoral farmers. The way was paved by W. A. van der Stel, who, unlike his father, allowed burghers to concentrate on stock farming. Within a few years it became possible for any burgher to take occupation of 6,000 ac. or more of land and treat it as his property, paying the company a small annual licence. The result was a continuous expansion of the frontiers of white settlement. In the north the trekboers were slowed up by aridity and finally stopped near the Buffels river about 1760. In the east they advanced along either side of the arid Great Karroo until in the 1770s the two streams of pioneers began to mingle beyond it in the vicinity of what became Graaff-Reinet, about 400 mi. from Cape Town.

Thus a few thousand white persons scattered over an area about twice as large as England. Their simple needs were met by their stock and by hunting and by traveling traders who bought their sheep and cattle for the Cape Town market and sold them a few imported goods. The company did little for them beyond establishing courts of landdrosts and heemraden at Swellendam (1745) and Graaff-Reinet (1785). From time to time officials expressed concern that the isolation of the trekboers was causing them to degenerate. But although all the trekboers were extreme individualists and some were ruffians, most of them contrived to preserve a sound domestic morality, a smattering of literacy and a respect for their religion, a rather strait Calvinism.

Slaves, Hottentots and Bushmen.—During the 18th century the slaves, the Hottentots and some of the Bushmen were beginning to interbreed with one another and with Europeans to produce the community now known as the Cape Coloured people. Although few slaves were imported, the slave population outstripped the white to number nearly 17,000 in 1795. Negroid types from the west and east coasts of Africa predominated, with Malays from the East Indies forming an important minority. There was a considerable amount of extramarital miscegenation between white men and slave women. Most of the slaves were the property of the agricultural farmers, for whom they formed the main labour supply. Others were owned by the company and the Cape Town burghers, a few by the trekboers. The treatment of the slaves, which had at first been fairly mild, tended to deteriorate, and the 1754 slave code was a stiff one.

The nearest approach to a Hottentot war had taken place in the 1670s. But it was smallpox that took the heaviest toll of the Hottentots, especially a shattering epidemic in 1713. After that, as the trekboers advanced into their territories, the surviving Hottentots lost their land and their cattle—their only means of an independent existence. Many were reduced to a status not unlike serfdom, forming the main labour supply of the trekboers, whose simple demands they were able to satisfy. Others withdrew beyond the frontiers of white settlement to maintain a more or less independent existence until the next century.

There was a certain amount of miscegenation between white men and Hottentot women, although intermarriages had been unlawful since 1685. The offspring were usually disowned by their fathers and became known, bluntly, as Bastaards; some entered into service, others retreated before the advancing trekboers, leading trekboer lives themselves.

The small Bushmen were still more primitive than the Hottentots. Before white settlement began they had been driven into the more arid and mountainous parts of South Africa by the Hottentot and African tribes. But, though every man's hand was against them, they contrived to make hit-and-run raids to get sheep and cattle for food, and, if cornered, they were able to kill human beings with their poisoned arrows. As the trekboers in turn began to enter Bushman country and to suffer from their raids they organized Boer commandos to hunt them. Toward the end of the 18th century several Bushmen bands made a prolonged stand, operating from mountain fastnesses along the northern fringes of the area of white settlement; but the commandos gradually cleared them out, killing 2,500 Bushmen and capturing 650 in the decade ending in 1795. Those captured were children, who were taken into service.

Their experience with slaves, Hottentots and Bushmen had important and enduring effects on the racial attitudes of the Afrikaner people. Most of the nonwhites they came into contact with were at so low a cultural level that the Afrikaners tended to think of themselves as innately superior. Their response was not merely hostile, however, because they all—Capetonians, agriculturalists and trekboers—came to depend on the use of nonwhite servants.

African Opposition and Economic Discontent.—Before the last quarter of the 18th century the trekboers had not been

seriously impeded by human obstacles. Then, however, they came up against formidable opposition. There were two main groups of African (Bantu-speaking) tribes in southeastern Africa: the Nguni group, who inhabited the well-watered coastal strip between the mountains and the Indian ocean; and the Sotho group, on the plateau. These Africans were more numerous and more advanced than the Hottentots; they were organized into larger tribes and, though they too were primarily pastoralists and hunters, they also practised a rudimentary type of agriculture. While the trekboers were moving eastward in search of new pastures, some of the Nguni tribes were moving westward toward them from Natal for the same purpose. By 1700 the Xhosa tribe of the southern Nguni in the van had reached the Great Kei river; by 1750, the Keiskama river. Boer hunters made contact with the Xhosas and bartered ivory and cattle from them. The company vainly tried to prevent such contacts and delay the convergence of the Boers and the Xhosas by issuing proclamations prohibiting barter and defining limits beyond which the Boers were not to go. When the governor, J. van Plettenburg, toured the frontier in 1778 he found that Boers were already intermingled with Africans in the vicinity of the Fish river.

A long series of frontier "wars" followed—basically struggles for possession of the land. The first fighting took place in 1779–81 and was sufficient to show the Boers that they faced tough opposition. They therefore appealed to the company for aid; although a landdrost, M. Woeke, was appointed to Graaff-Reinet in 1785, no troops were sent. Woeke's life was soon made intolerable by the Boers, and in 1793 he was superseded by H. C. D. Maynier. Further fighting broke out in that year and again the result was inconclusive. The Boers were disgruntled. For nearly a century they had been left by the company to make their own terms with man and nature, and when at length they had asked for help all they received was a government official who not only failed to defeat the Xhosas but also allowed Hottentots to come to his court with complaints against them. In 1795, on the eve of the British conquest of the Cape, they drove out Maynier and declared Graaff-Reinet to be a republic. But the name "republic" was a euphemism; in fact there was anarchy at Graaff-Reinet. Similar events took place at Swellendam, where another "republic" came into being.

Meanwhile there was trouble in the southwestern part of the colony also. The company had maintained rigorous commercial restrictions, together with a system of government that gave the colonists practically no voice in the determination of policy. The trekboers had managed to escape most of the effects, but the restrictions had always borne heavily on the agricultural farmer who had produce to sell and on the Cape Town man who lived by trade. In the 1770s, which were lean years, the discontent came to a head with the organization of the Cape Patriots movement, which had some connections with the anti-Orange Patriot party in the Netherlands. A petition was drawn up criticizing many of the officials and asking for full burgher representation on the council of policy and the right to trade freely throughout the company's possessions. When it was laid before the directors in Amsterdam, their only response was to dismiss the official who was most roundly criticized—the fiscal, W. Boers.

In 1784 a second petition was taken above the heads of the directors to the states-general of the Netherlands. This time the burghers were given equal representation with officials on the court of justice (though not on the council of policy), and the right to trade with the company's possessions—a right that was nullified by the conditions that were attached. By then, however, the disaffection had subsided because the American Revolutionary War had brought a large French garrison to Cape Town to help the Dutch to defend it against the British, and the garrison provided an abnormally large market for Cape produce and brisk business for the traders and innkeepers. In the early 1790s bad times came again as the company, trying desperately to stave off bankruptcy, made a final effort to make the colony pay its way. The garrison was reduced, public works were suspended and—the last straw—new taxes were imposed. So, when the British invading force arrived in 1795, it found depression in the west and rebellion in the east. The company had never done much for the white colonists and still less for the nonwhite races.

B. South Africa Divides (1795–1870)

British Conquest.—The Cape was captured by a British naval and military force in Sept. 1795, nominally on behalf of the prince of Orange, who had taken refuge in England from the Dutch republicans. In their struggle with France the British considered it essential not to let the Cape fall into enemy hands, and until the peace of Amiens a hold was kept on it for strategic reasons. The British tried to conciliate the Afrikaners, and by abolishing the restrictions on internal trade and maintaining a large garrison they caused an economic revival in the southwestern part of the colony. On the eastern frontier, on the other hand, the British were no more successful than their predecessors. Although the republics of Swellendam and Graaff-Reinet soon submitted to the new government, anarchy developed in Graaff-Reinet again in 1799, after the arrest of a prominent trekboer, A. van Jaarsveld, on a charge of forgery. Not only the Xhosas but also many of the Hottentot servants of the Boers made the most of the opportunity created by the divisions among the whites and caused havoc as far west as Knysna. The British general F. Dundas patched up a truce; but H. C. D. Maynier, appointed resident commissioner of the two eastern districts, failed to satisfy the trekboers, who rebelled again in 1801, when once again their Hottentot servants joined with Xhosa tribesmen to harry the farms over a wide area.

The eastern districts were therefore in turmoil when in 1803 the Cape was handed back under the treaty of Amiens. By this time the company was defunct and the Dutch government—the Batavian republic organized by the French in 1795 after their conquest of the country—succeeded to its charge. J. A. de Mist, the commissioner general, and Gen. J. W. Janssens, the governor, tried to restore the morale of the Boers and showed a considerable understanding of the problems of the country; but these men did not stay long enough to consolidate their reforms, for in 1806 the British sent out another force to recapture the Cape from Napoleon's Dutch allies. Janssens lacked the means to put up much resistance and once more the colony passed into the hands of the British, whose title was confirmed in 1814 as part of the general peace settlement.

British Policy, 1806–36.—Until 1823 British policy was cautious and conservative. Although the governor had autocratic powers, the Afrikaners were used to that. Moreover, their Roman-Dutch law was retained and local administration remained as of old in the hands of landdrosts and heemraden, most of them Afrikaners. These were comparatively prosperous years. The southwestern farmers profited from their access to British markets, where their wine sold well thanks to a substantial tariff preference. Elsewhere, however, extensive pastoral farming remained the norm. Nevertheless, the British did introduce some changes in these years. They appointed more landdrosts. In 1809 they promulgated a code for the treatment of Hottentot servants, who were to be employed only under written contracts registered at a landdrost's office and were given legal protection against ill treatment; on the other hand, they were not to be allowed to leave their employer's farm without a pass signed by the employer. From 1811 onward high court judges, who had previously sat only in Cape Town, made annual circuits to hear cases in all the districts.

Mild though these reforms were, they caused trouble. There was much grumbling in 1812 when the judges on circuit in the eastern districts tried a number of Afrikaners on charges of ill treating their Hottentot servants. Three years later a trekboer, F. C. Bezuidenhout, opened fire on a party that had come to arrest him for repeatedly ignoring his landdrost's orders to answer a charge of cruelty; in the skirmish Bezuidenhout was killed, whereupon his brother proceeded to organize opposition to the government; but the Slagtersnek rebellion soon ended when about 60 rebels were rounded up and tried, and five ringleaders were hanged.

The British also brought superior force to bear against the

Xhosas on the eastern frontier and thus started the process of the conquest of the African tribes. In 1811 British regular troops and Boer commandos drove the Xhosas back to the eastern side of the Fish river, and in 1819 to the eastern side of the Keiskama.

In 1820, 5,000 British settlers were brought out and placed on 100-ac. lots in the Zuurveld, on the western side of the Fish river. The 1820 settlers did not, however, form the human barrier that was intended, because the land was not suited to agriculture; within a few years most of them had abandoned their lots and become townsmen in Port Elizabeth and Grahamstown. Others became traders among the African tribes to the east. Those who stayed on the land eventually made good by producing wool for export to England, especially when large farms came onto the market at the time of the Great Trek. This was the first subsidized immigration scheme since the 17th century, and the 1820 settlers were the first white immigrants who were not assimilated by the Afrikaner people.

When they had got their bearings the 1820 settlers were dismayed to find that the legal system of the colony was alien to them, the judges were officials without any particular legal training and Lord Charles Somerset, the governor from 1814 to 1826, was an autocrat who made it impossible for them to publish a newspaper. Their agitation for British liberties and institutions evoked some response. The Charter of Justice of 1827 created a supreme court of qualified judges with security of tenure, abolished the heemraden and introduced English rules of procedure and evidence, including the jury system, while it left untouched the Roman-Dutch substance of the law. In 1834 an executive council of officials and a legislative council of officials and nominated non-officials were established. By that time the freedom of the press had been won, English had become the official language of the courts, and British teachers had been imported to start village schools in which English was the medium of instruction.

During these years South Africa was attracting the attention of the British humanitarians or evangelicals who were striving to free the nonwhite races in the colonies from their legal disabilities. They were supplied with severe criticisms of South African conditions by John Philip, South African superintendent of the London Missionary society, which had established a number of mission stations up and down the country. Philip reached the conclusion that the Hottentot code of 1809, with its stringent pass regulations, was unjust; in 1826 he returned to England to appeal to the government and the British public, and in 1828 the house of commons passed a resolution for the emancipation of the Hottentots. Seeing the trend of affairs, Gen. (afterward Sir) Richard Bourke, the acting governor, had already passed an ordinance, no. 50 of 1828, which abolished the pass system and gave "free persons of colour" complete liberty of movement. Meanwhile, with their eyes focussed primarily on the British West Indies, the British government had been insisting on improvements in the treatment of slaves; and in 1833 the reformed parliament passed the Emancipation act under which all slaves in the colonies became free after a period of apprenticeship, and £20,000,000 was voted as compensation to their owners. When their period of apprenticeship ended in 1838 the former slaves in the Cape colony stepped into the position of other "free persons of colour." Thereafter the emancipated slaves and Hottentots—the Cape Coloured people—became a rural and urban working class in the Cape colony.

Meanwhile frontier trouble had flared up again. Unrest had developed among the border African tribes for various reasons, including the frequency of military raids on their villages made in reprisal for the theft of stock from white farmers. In 1834 Gen. Sir Benjamin D'Urban (*q.v.*) arrived as governor with instructions to devise a more equitable policy, but he did nothing about it until large bodies of Africans invaded the colony and did serious damage. D'Urban then reasserted British military superiority and annexed the land up to the Great Kei river, from which most of the Africans were to be expelled (May 1835). But he found it impossible to enforce the expulsion, and the colonial secretary, Lord Glenelg, disapproved; consequently, in 1836 the territory was disannexed and an attempt was made to pacify the tribes by making treaties with their chiefs. By that time, however, the frontier was no longer a clear line of division between the races; white traders and missionaries were working among the tribes to the east of it, and Africans were working on white farms to the west.

The Great Trek.—For several generations the trekboers had been left more or less to their own devices. Now, under an alien government actuated by liberal ideals which they did not comprehend, many of their institutions were being transformed. They were short of labour and they found it difficult to comply with the regulations and get the compensation money for their slaves; their property was being pilfered by wandering Hottentots and they suffered further losses during the African invasion in 1834. Behind all these things they saw the hand of missionaries, who, they considered, misrepresented them in England. The reversal of D'Urban's frontier policy was the last straw that made many of them decide to throw off British dust and find a new home where they could, in the words of Piet Retief, their ablest leader, "preserve proper relations between master and servant." Later, Anna Steenkamp recorded that the British had placed their slaves "on an equal footing with Christians, contrary to the laws of God and the natural distinctions of race and religion, so that it was intolerable for any decent Christian to bow down beneath such a yoke; wherefore we withdrew in order thus to preserve our doctrines in purity." Imbued with this spirit the voortrekkers, about 12,000 Afrikaners, men, women and children, left the colony with their sheep, their cattle, their ox wagons and their Coloured servants between 1835 and 1843.

In the previous two decades there had been widespread destruction and dislocation among the African tribes, caused by the Zulus under Shaka, and their offshoot, the Ndebele or Matabele, under Mzilikazi (Moselekatse). By the time of the Great Trek two areas seemed suitable for white settlement—Natal, south of the Tugela river, and the high veld, on either side of the Vaal. But appearances were deceptive, for these two areas were kept denuded by Zulu and Ndebele impis or warriors. The first voortrekkers made for the high veld, where they were attacked by the Ndebele, whom they routed in 1837. Most of them then decided to settle in Natal. Piet Retief and a pioneer group went to the kraal of Dingaan, Shaka's successor, to negotiate for a cession of land; and there, in Feb. 1838, he and 70 Boers and 30 Coloured servants were treacherously massacred, while another 300 Boers, mainly women and children, and 200 Coloured servants, were killed in upper Natal. On Dec. 16, 1838, however, Andries Pretorius gained a decisive victory over the Zulus at Blood river. Most of the voortrekkers then proceeded to stake out farms in Natal. Border troubles ensued, and when in 1840 a Boer commando expedition made a sortie to the south against the kraals of a chief named Ncapaai, his neighbour Faku, who had a treaty with the British, appealed for aid. Previously the British had been undecided what to do about the voortrekkers; the Cape of Good Hope Punishment act of 1836 had made them liable, as British subjects, to trial in British courts for crimes committed south of latitude 25° S., but little had been done to make it effective. Then, however, anxious lest the activities of Natal commandos would lead to disturbances on the Cape frontier, the governor, Sir George Napier, sent a small force to Natal which, after some fighting, became British territory in 1843. Most of the voortrekkers who had settled in Natal then inspanned their oxen again and trekked back to the high veld, north of the Vaal river, where they proved to be immune from further British attention for a generation.

Between the Orange and the Vaal, however, the British were drawn in for a while. There was a racial medley in that region: Griquas (Hottentot-European half-castes), who had been living on the northern side of the Orange river for several decades; Africans, survivors of the Ndebele devastation of the high veld, who had been rallied by Moshesh, the Basuto chief, near the Caledon river; and Afrikaners. The British had made a treaty with the West Griqua chief, Andries Waterboer, in 1834, and they made treaties with the East Griqua chief, Adam Kok III, and with Moshesh in 1843, hoping to use them to keep the peace along the Orange river. Trouble followed when Kok, in accordance with his treaty, tried to hand an Afrikaner over to the Cape authorities for trial under the Punishment act. The Afrikaners

took up arms and were dispersed by troops from the Cape at Zwartkoppies (1845). Two years later Gen. Sir Harry Smith became governor of the Cape colony and British high commissioner in South Africa. An optimist, he believed that the South African imbroglio would be pacified if the area of British authority was extended. Accordingly, after overcoming further African resistance on the eastern frontier of the Cape, he annexed the land up to the Great Kei river as the province of British Kaffraria (1847) and the land between the Orange and the Vaal as the Orange River sovereignty (1848). The colonial secretary, Earl Grey, accepted these annexations most grudgingly.

Humanitarian influence was waning in Britain and the main concern of the colonial office was to cut down expenses. Therefore, when further fighting broke out on the eastern frontier and similar difficulties developed in the sovereignty with Moshesh, the British government decided to reduce its commitments in South Africa by recognizing the independence of the voortrekkers. This was done in two stages. In 1852 the commissioners W. Hogge and C. M. Owen granted the Transvaalers "the right to manage their own affairs without any interference on the part of the British government" in a convention signed at the Sand river; and in 1854 Sir George Clerk made similar arrangements with the Afrikaners of the Vaal-Orange area at Bloemfontein. South Africa was thus divided into two camps: Afrikaner republics and British colonies.

Afrikaner Republics.—At the time of the Bloemfontein convention there were about 15,000 white persons engaged in pastoral farming in what then became the Orange Free State. They soon agreed on a rigid constitution, which vested legislative power in an elected council (*volksraad*), executive power in an elected president—the voters being the white male adults—and local administration in landdrosts and heemraden on the old Cape model. The East Griqua question was settled in 1861 when Adam Kok sold out his remaining land rights in the Free State and trekked with his people to found a new East Griqualand nearer the coast (*see* GRIQUALAND, EAST AND GRIQUALAND, WEST). The Basuto question was more difficult. The British had abandoned the sovereignty because they shirked the expense of controlling the brisk rivalry between the Afrikaners and the Basuto for the Caledon river valley, which included some of the best arable land in southern Africa.

When fighting broke out in 1858 the Basuto outmaneuvered the Free Staters, who were glad to accept the mediation of the high commissioner Sir George Grey (first convention of Aliwal North). The Free Staters, conscious of their weakness, then tried to unite with the Transvaal but were stopped by Grey, who threatened to withdraw the recognition of their independence and advised them to federate with the Cape colony; yet when they expressed a desire to do so, the British government refused. Grey considered that a federation, under the crown, of all the colonies and republics in South Africa would be in the general interest. The British government, however, was averse to any extension of its area of responsibility. Further fighting started in 1864; this time the Free Staters were more successful and were able to dictate stringent terms, by which Moshesh (*q.v.*) lost most of his arable land (treaty of Thaba Bosiu, 1866). But the peace was not kept and Moshesh asked Britain to annex Basutoland. This was done in 1868 and the next year the frontier line between the Free State and Basutoland was agreed upon (second convention of Aliwal North).

The intervention of the British preserved for the Basuto people more land than had been left to them in 1866; but it gave serious offense to the Free Staters, depriving them of the fruits of their victory. Thereafter the Free state, its frontiers pacified and absolved from the responsibility of administering a large number of Africans, became what James (afterward Viscount) Bryce once described as a model republic, under the wise leadership of Johannes (afterward Sir Johannes) H. Brand, president from 1864 to 1888. The Cape colony took over Basutoland in 1871, but later got into difficulties in trying to force the Basuto to forfeit their arms and was glad to hand it back to Great Britain in 1884.

Meanwhile the voortrekkers north of the Vaal river, numbering perhaps 22,000 at the time of the Sand river convention, preserved their old trekboer mode of life. It was not until 1857 that most of them overcame their factional disputes and united to form the South African republic under the presidency of M. W. Pretorius, son of the victor of Blood river (*see* PRETORIUS). Their constitution, which created institutions similar to those in the Free State, frankly stated that "The people desire to permit no equality between coloured people and the white inhabitants, either in church or state." Otherwise it was an ambiguous document and from the first the *volksraad* ignored any restraint but that of force. In 1863 there was further factional fighting; and Boer expansion frequently led to disputes with African tribes, the weaker of which were broken up, the remnants becoming squatters and servants under white landowners.

British Colonies After the Great Trek.—Not more than a quarter of the Afrikaners had left the Cape colony during the Great Trek, though others went north in later years. Those who remained gradually became accustomed to the more equal order and in the course of time a liberal tradition developed in the colony, in contrast with the illiberal tradition of the republics. When the colony received representative government in 1853, the only franchise qualifications were low economic ones, irrespective of race or colour; but most of the Coloured people were unable to qualify and the vast majority of the electorate were white men. By that time the loss of population had been offset by further British immigration; and the decline in wine farming caused by the removal of British tariff preferences was remedied by an increase in the production of wool for export. In British Kaffraria the high commissioner, Sir George Grey, tried to achieve quick results from a policy of civilizing the Africans by encouraging trade, mission schools, medical services and public works, by bringing in white settlers and by using white magistrates to undermine the authority of the chiefs.

There was a reaction in 1857, when the Africans destroyed their own stock and crops in the superstitious belief that the white man would then vanish from the country. The result was catastrophic: thousands starved to death and emaciated survivors poured into the colony in search of work. In 1865 British Kaffraria was incorporated in the Cape colony. Soon afterward the British government urged the colony to assume full responsibility for its own affairs, and in 1872 a responsible government bill was carried and J. C. Molteno became first prime minister.

British immigrants had settled in Natal, so that its white population became predominantly British and crept up to 9,000 in 1856, when representative institutions were established, and to 50,000 in 1893, when responsible government was obtained. Mixed farming was practised inland and sugar was grown in the subtropical coastal belt. One result of the defeat of Dingaan was an influx of Africans from Zululand into Natal, many of them returning to the region from which they had formerly been evicted by Shaka. Thus the African population rose steeply from about 20,000 in 1840 to nearly 500,000 in 1893. They were handled by Sir Theophilus Shepstone (*q.v.*), the son of a missionary who had come to South Africa with the settlers in 1820. He placed most of the Africans in a number of reserves or locations, comprising about one-seventh of the total area of the colony. Their tribal laws and customs were maintained and they were able to preserve their tribal economy for another generation, when they began to suffer from land shortage. Africans subject to tribal law were not eligible for the franchise, but in 1864–65 laws were passed which enabled Africans to apply to be exempted from tribal law and, seven years later, for the franchise. However, the regulations were such that few became exempted and hardly any became voters.

The colonists were critical of Shepstone's policy, mainly because it did not make the Africans work for them. The coastal sugar planters, in particular, were short of labour and it was to help them that Natal began to import Indian labourers in 1860. By 1866, 6,000 had arrived, and when their five-year contracts expired most of them chose to stay in Natal, as they were entitled to do. In 1872 the Indian government prohibited further emigration because some of the Indians had complained of ill-treatment; but the migration was resumed on a large scale, with a

heavy government subsidy, in 1874, and by the end of the century the Indians outnumbered the whites in Natal.

C. Through War to Union (1870–1910).

Diamonds and Gold.—After 1870 the South African economy was rapidly transformed. Previously South Africa had been an economic backwater, with most of its inhabitants, whites and nonwhites, engaged in inefficient near-subsistence farming; and it had lacked the means to attract the capital and the skilled personnel to create the facilities of a modern country. In 1867 alluvial diamonds were found along the Orange and the Vaal rivers; in 1870 the dry diamond diggings began to be worked at Kimberley; and in 1886 the gold rush to the Witwatersrand began to eclipse all previous gold rushes. The effects were remarkable. The sum of South Africa's imports and exports rose from £49,000,000 in the decade 1861–70 to £128,000,000 in 1871–80, £179,000,000 in 1881–90, £357,000,000 in 1891–1900 and £700,000,000 in 1901–10. In the last of these decades gold (59%) and diamonds (19%) accounted for 78% of South Africa's exports; and by 1910 £121,000,000 had been invested in the Witwatersrand gold-mining industry and dividends totaling £77,000,000 had been paid by it. Immigration raised the white population from about 328,000 in 1875 to 1,117,000 in 1904. Railway construction was even more striking: in 1870 there were only 69 mi. of railway track in all South Africa; by 1886 there were 1,800 mi. and the line from Cape Town had reached Kimberley; by 1895 there were 3,600 mi. and Johannesburg was connected with five ports—Cape Town, Port Elizabeth, East London, Durban and Lourenço Marques (Portuguese East Africa). Thus South Africa at last entered the world economy and acquired a special place in it as the source of about a third of the world's annual supply of gold and more than half its diamonds.

The control of these great industries became concentrated in a few hands. The digger phase on the diamond fields was short-lived; companies forced out individuals and then the companies amalgamated until in 1891 De Beers Consolidated Mines controlled the entire South African diamond industry. The digger phase on the gold fields was even shorter, but there the process of concentration never reached quite the same finality; companies became organized into groups or corporations, and the groups became associated in the chamber of mines for certain purposes only, notably for the recruitment of African labour. Thus great financial power came to be exercised by a few men, above all by Cecil Rhodes (1853–1902), who went to South Africa for his health in 1870 and became the most successful of all the mining financiers, controlling both De Beers and the Consolidated Gold Fields, one of the strongest gold corporations.

The skilled work in the mining industries was done by whites, mainly immigrants at first, who drew high wages. The unskilled work was done by low-paid Africans who came to the mines from all over southern Africa, especially from Portuguese East Africa, to work for limited periods during which they lived in closed compounds, separated from the ordinary life of the community. In 1898, for example, the Witwatersrand gold-mining industry employed 9,476 whites at an average monthly wage of £26 and 67,797 Africans at an average monthly wage of £2 9*s*. Previously many Africans had been employed as farm labourers in return for squatting rights or low wages. Then the agrarian pattern of race relationships was being adapted to the mining industries, which, having a voracious appetite for labour, were giving a distinct impetus to the long-drawn-out process of the disintegration of tribalism.

The discovery of diamonds led to a serious controversy. The Transvaal claimed part of the diamondiferous area and the Orange Free State claimed the rest of it, while David Arnot, a white attorney, claimed it all on behalf of his client Nicholas Waterboer, the chief of the West Griquas. Arnot managed to persuade the British colonial office that the Griquas had the best case; therefore when a new high commissioner, Sir Henry Barkly, reached South Africa in Jan. 1871 he had instructions to take a strong line. Barkly persuaded M. W. Pretorius, the Transvaal president, to submit his case to arbitration by R. W. Keate, the governor of Natal. When Keate discovered that Pretorius had unwittingly placed great reliance on a forged document, he decided against him. Although this award did not directly affect the Free State claims, which were the strongest, Barkly, failing to understand their merits, annexed to the crown the entire area claimed for Waterboer in the same month (Oct. 1871). Five years later a land court decision exposed the fallacies in Waterboer's claim to sovereignty over the diamond fields; whereupon J. H. Brand, the Free State president, went to London and exacted £90,000 compensation from the imperial government. In 1880 Griqualand West was incorporated in the Cape colony.

Annexation and Retrocession of the Transvaal.—Lord Carnarvon, colonial secretary in Disraeli's 1874 ministry, hoped that the South African states and colonies might unite in a self-governing federation under the British crown. The following year he summoned a conference to discuss the project in London, but it was a failure: the Free State, smarting under the recent annexations, would have nothing to do with the project; nor would the Cape, whose ministers' dignity had been ruffled by the conduct of the historian J. A. Froude, who had toured the colony on behalf of Carnarvon urging federation. Federation from the south having failed, Carnarvon turned his attention to the north and found that the South African republic was ripe for plucking.

That republic, unlike the Free State, had never acquired stability. Pretorius was bundled out of the presidency after the annexation of Griqualand West, and the Transvaalers chose as his successor Thomas François Burgers (*q.v.*), a Cape colonial preacher of the Dutch or Netherlands' Reformed Church. But in Burgers they had more than they had reckoned for; he was critical of Transvaal backwardness and proposed to remedy it by social, educational and religious reforms. He also borrowed money, intending to build a railway to Delagoa bay to free the Transvaal from its dependence upon British trade routes; but not a single track of the railway was laid, whereas the loan strained the meagre resources of the republic and alarmed the Cape and Natal merchants whose interests it threatened. When in 1876 Burgers personally waged an unsuccessful campaign against the Pedi (Bapedi) tribe in the northern Transvaal, he was totally discredited. The Afrikaners, including Paul Kruger, were plotting against him, and the traders were conniving with the Cape and Natal merchants to bring in the British. Carnarvon heard their cries and commissioned Shepstone to annex the Transvaal. Entering the republic in Jan. 1877 with 25 police and a staff of seven, Shepstone annexed it in April in a proclamation which promised "the fullest legislative privileges compatible with the circumstances of the country and the intelligence of the people."

But British rule in the Transvaal was neither efficient nor tactful and the elected legislature did not materialize. Then the Zulu War revealed the weakness of British arms: the Zulus, their martial spirit revived by Cetywayo (*q.v.*), routed a large British force at Isandhlwana in 1879 before they were vanquished in the following year. By that time the Transvaalers, who had passively acquiesced in the annexation, were moving toward open rebellion under the leadership of Paul Kruger, Petrus Jacobus Joubert (*qq.v.*) and M. W. Pretorius, and in 1881 they, too, wiped out a British force at Majuba. Gladstone's ministry, which assumed office in Britain in 1880 but failed to reform the Transvaal administration in time to avert the rebellion, then gave the Transvaalers a qualified independence (Pretoria convention, 1881). Three years later they allowed the republic full internal autonomy, while maintaining British control over its relations with foreign states, other than the Free State (London convention, 1884). Carnarvon's project had completely miscarried, and his permissive act for the unification of South Africa (1877) was stillborn.

The Rhodes-Hofmeyr Alliance.—In the Cape, which remained the most civilized of the South African states, the advent of responsible government and the series of British annexations prompted a group of Afrikaners to form a political organization, the Afrikaner Bond, in 1879. At first the Bond's program was crudely sectional, but three years later Jan Hendrik Hofmeyr (*q.v.*) (1845–1909) gained control and turned it in the direction of Anglo-Afrikaner co-operation, with a British South African federa-

tion, which should include the republics, as the ultimate goal. Meanwhile Cecil Rhodes had entered the Cape parliament and, as his vision of British expansion in Africa was reconcilable with Hofmeyr's, he was able to form a ministry with Bond support in 1890.

Moreover, although the Cape had mishandled Basutoland, a more promising native policy was developing in the Transkeian territories up to the Natal border, which were gradually annexed between 1872 and 1894. Nearly all the land was left in African possession, African tribal law was recognized, and a corps of able and sympathetic administrators was created. But these annexations raised serious political questions because they caused the African population of the colony to increase threefold and far outstrip the white population. The Cape parliament accordingly granted the territories a modicum of political representation, while making it clear that occupation of land on tribal or communal tenure did not constitute a qualification for the franchise (1887), raised the economic qualifications slightly and added a simple educational test (1892) and encouraged the Africans to adopt individual land tenure and to elect local administrative councils (1894). The native policy of the Cape was thus based on the principle that the Africans should gradually change from the tribal to the western mode of life, and that those who had made the change should have the same civil and political rights as white men. By 1895, therefore, the Rhodes-Hofmeyr alliance was bidding fair to heal Anglo-Afrikaner quarrels and to maintain a promising basis for relations between the whites and the nonwhites in the Cape colony.

British Clash With the Transvaal.—Meanwhile the scramble for Africa was taking place, and in 1884 Germany took over South West Africa and Transvaalers began to encroach westward toward it, setting up the republics of Stellaland and Goshen. This expansion, which seemed likely to cut off the British from access to central Africa via the Cape, was checked, largely on the insistence of Rhodes, in 1885, when Great Britain, after a display of force, proclaimed a protectorate over northern Bechuanaland and annexed southern Bechuanaland as a crown colony. Transvaal efforts to keep their north clear of the British were also foiled by Rhodes, who secured a prospecting concession from the Ndebele chief Lobengula in 1888, obtained a royal charter incorporating the British South Africa company to exploit the concession in 1889, and dispatched a pioneer expedition to occupy what became known as Rhodesia in 1890 (*see* RHODESIA, SOUTHERN: *History*). Finally, the Transvaalers tried to push eastward to get a port on the Indian ocean, but, although they incorporated fresh territory including Swaziland, they did not reach the coast, as Britain annexed Zululand in 1887 and Tongaland, the last gap, in 1895. The Afrikaner republics were thus encircled by the colonies of Great Britain and its oldest ally, Portugal—hemmed in a kraal, as Paul Kruger complained. Southern Bechuanaland was taken in by the Cape colony in 1895; Zululand and Tongaland by Natal in 1897; but northern Bechuanaland remained a British protectorate.

In 1883 Paul Kruger (1825–1904) was elected president of the Transvaal. His character had been molded in the hard school of the Great Trek and of commando fighting against Afrikaner factions and African tribes; and recently he had led the opposition to President Burgers and the British. His policy was to regain complete independence for the republic and to preserve Afrikaner control of it. In 1885 he was constrained to suggest a customs union with the Cape, whose government rejected the offer. Soon afterward his financial difficulties were solved by the gold discoveries and he proceeded to use the mining industry as a milch cow, imposing heavy taxation and granting monopolies of essential materials, such as dynamite. Nevertheless, the growth of such an industry in the heart of his republic posed serious problems. The Afrikaners became outnumbered by the newcomers who poured in from all parts of the world, most of them British subjects. To give these *uitlanders* ("outlanders"), as they were called, an effective vote seemed like suicide; to withhold such a vote was bound to cause trouble with the British. Kruger and the *volksraad* never hesitated; they progressively raised the franchise qualifications until by 1894 no *uitlander* could vote in presidential elections, and only those who were 40 years old and had lived 14 years in the Transvaal, during 12 of which they had been subject to an oath of allegiance, could vote in *volksraad* elections. In short, on the one hand the government was corrupt and fleeced the *uitlanders* and denied them political rights; on the other hand the *uitlanders* continued to swarm into the republic where many of them prospered exceedingly. Such a situation required patience. Although Kruger was re-elected for a third term in 1893, his majority over the more moderate P. J. Joubert was small and a change seemed likely in 1898.

But Rhodes was not patient. Finding that Kruger stood between him and a British South African federation, he planned to overthrow him. Kruger nearly played into Rhodes's hands in 1895 when, contrary to the London convention, he tried to injure Cape trade with the Transvaal and the Cape railways that were competing with the Delagoa bay line by making the rates over the Transvaal sector of the Cape line prohibitive and by closing the Vaal drifts to prevent the carriage of goods by wagon. But he gave way when the colonial secretary, Joseph Chamberlain, agreed to support the Cape ministry's protests, by force if necessary.

The Jameson Raid.—Rhodes then pushed ahead with his conspiracy: the *uitlanders* were to rise, an armed force under Leander (afterward Sir Leander) Starr Jameson (*q.v.*) was to come to their assistance, and the high commissioner was to hurry to Pretoria to "restore order." Chamberlain knew of Rhodes's plans when he provided a suitable jumping-off place for Jameson's force by ceding to the British South Africa company a strip of land along the Transvaal border of the Bechuanaland protectorate, but he later withdrew his support when he learned that the *uitlander* rising would not take place to give the invasion an air of respectability. Therefore, when Jameson, contrary to Rhodes's last-minute instructions, rode into the Transvaal from Pitsani near Mafeking on Dec. 29, 1895, he was disowned by the colonial secretary and the high commissioner. Four days later he surrendered to Transvaal commandos at Doornkop.

Outbreak of the South African War.—The raid fiasco cleared the decks for the South African War. Rhodes had to resign the premiership of the Cape colony, where, in 1898, W. P. Schreiner became head of a ministry which relied upon the support of the Bondsmen, who had been made deeply suspicious of British designs. In the Orange Free State, which had opposed the Transvaal over the drifts crisis, M. T. Steyn was elected president and formed a military alliance with the Transvaal. While Steyn and the Bond were not uncritical of the Kruger regime, Kruger's Anglophobia seemed to have been vindicated in the Transvaal, where he won the 1898 presidential election by a large majority. In England, on the other hand, a parliamentary committee failed to probe Chamberlain's dealings with the raiders, the kaiser's telegram congratulating Kruger turned Jameson and Rhodes into popular heroes, and the press paid much attention to *uitlander* grievances.

As the tension mounted, Chamberlain, taking the high ground of British paramountcy in South Africa, claimed to be competent not merely to control the foreign relations of the Transvaal (a right expressly granted by the London convention) but also to intervene on behalf of the *uitlanders* (which was not). If the right to intervene existed, there were grounds enough for exercising it. The tendency of the *volksraad* to give sweeping powers to the president was exemplified in the judges' crisis: in 1897 Chief Justice John (afterward Sir John) Gilbert Kotze delivered a judgment which meant that the greater part of the laws of the Transvaal had been enacted unconstitutionally and were null and void, whereupon the *volksraad* deprived the courts of the testing power and gave Kruger the power to dismiss any judge who disagreed; and the next year Kruger dismissed Kotze. In 1898 an Anglo-German agreement removed the likelihood of German intervention in the event of war in South Africa, and Chamberlain and the high commissioner, Sir Alfred (afterward Viscount) Milner, proceeded to apply the screw. In so doing they brushed aside several attempts by the Cape government and the Afrikaner Bond to avert a catastrophe. The central issue was the Transvaal fran-

chise. In June 1899, in conference at Bloemfontein, Milner proposed a simple five years' franchise, Kruger refused and the negotiations collapsed. The last chance of peace was lost in August when Jan Christian Smuts (*q.v.;* 1870–1950), Kruger's young state attorney, offered a five years' franchise and Chamberlain rejected the conditions that were attached to the offer. Milner had long since made up his mind that war was "inevitable," and in September Chamberlain arranged for 10,000 British troops to be sent to augment the meagre British forces in South Africa. The republicans, who had been arming furiously for several years, replied by issuing an ultimatum which expired on Oct. 12, 1899. (For events of the war, *see* SOUTH AFRICAN WAR.)

Reconstruction and Union.—After the peace of Vereeniging of May 31, 1902, Milner, high commissioner and governor of the crown colonies—the Transvaal and the Orange River colony—concentrated on their material reconstruction from the ravages of war. The gold mines resumed large-scale operations, and the Afrikaners were returned to their land from the prison and concentration camps and supplied with food, stock, seed and implements, no less than £10,000,000 being spent by Great Britain on their rehabilitation. Milner also prepared the ground for political union by bringing all the colonies into a South African customs union and by amalgamating the railways of the Transvaal and the Orange River colony. All this work was well done by Milner and his band of able young men, who were dubbed the "kindergarten." But the opportunity was not used to emancipate the nonwhite inhabitants of the crown colonies from their legal disabilities, which remained much as they had been during the republican regime. Nor did Milner gain the confidence of the Afrikaners: they disliked him as "the man who made the war," who openly supported an agitation for the suspension of the Cape constitution in 1902, who imposed an English educational system on the crown colonies and who was responsible for bringing Chinese labourers to the Witwatersrand gold mines in 1904. For the importation of Chinese labourers there were, indeed, strong economic reasons, because the supply of African mine labour had temporarily fallen off; but in Britain the decision played a large part in the fall of A. J. Balfour's Conservative government in Dec. 1905 and the change of government was followed by a radical change in the distribution of power in South Africa. Sir Henry Campbell-Bannerman's Liberal ministry, anxious to make amends to the Afrikaners for a war which had been brought about by their predecessors, decided to give the new colonies responsible government, with exclusively white electorates. The Liberals defended this decision by referring to the treaty of Vereeniging, but it did not, in fact, debar them from allowing Coloured and Asian people to vote. Consequently in March 1907 Louis Botha (*q.v.*) (1862–1919), former commandant general of the forces of the South African republic, became premier of the Transvaal, with Smuts as his right-hand man, and in Dec. 1907 A. Fischer became premier of the Orange River colony. The swing of the pendulum was completed in Feb. 1908, when J. X. Merriman's South African party, consisting largely of the Afrikaner Bond, ousted Jameson's Progressives from office in the Cape colony.

The high commissioner, Lord Selborne, who succeeded Milner in 1905, argued the case for union in a memorandum published in July 1907, and, once in power, the three Afrikaner parties took it up enthusiastically, with the support of their oppositions. White public opinion in South Africa moved steadily in favour of union, partly out of an idealistic feeling that it was best to bury the hatchet and make a new start on a basis of white equality, and partly because an African rebellion in Natal in 1906 led many to think union essential for white security; some, moreover, saw in union the best way of guarding against any further British interference in South African affairs. There were also compelling economic reasons for union: the four colonies were interdependent and yet, without political union, their material interests were so divergent that the customs union seemed bound to collapse.

Accordingly a national convention, comprising 30 members appointed by the four colonial parliaments and 3 nonvoting members from Rhodesia, met in 1908 and 1909 under the chairmanship of Sir (afterward Baron) Henry de Villiers, chief justice of the Cape. A constitution was drafted and unanimously approved by the convention and it was carried with scarcely any dissentients in the parliaments of the Cape colony, the Transvaal and the Orange River colony, and by a three-to-one majority of the electorate in a referendum in Natal. It was enacted by the British parliament in Aug. 1909, substantially as it had been submitted to the British government by delegates from South Africa. The South Africa act came into force on May 31, 1910, and the four colonies became the provinces of the Union of South Africa (*see* also *Constitution,* below).

The convention, except for the Natal members, wanted a close union and complete flexibility, and they were largely successful; the principal feature of the South African constitution was the grant of power to parliament to legislate on practically every subject by simple majorities in each house. Nevertheless, on two subjects they did not grant parliament such extensive powers. First, the Afrikaner members wanted special protection for the section which gave Dutch (to which Afrikaans was added in 1925) equal status with English as an official language of the Union. Second, the convention could not agree on a uniform franchise for the Union, the Cape delegates favouring a colour-blind franchise and the others a rigid colour bar. Eventually a compromise was reached whereby the franchise qualifications were to remain as they were in each province and the established franchise rights of the Cape nonwhites were specially protected. Accordingly, the sections on language equality and the Cape nonwhite franchise were to be amended only with the approval of two-thirds of the members of both houses of parliament in a joint sitting. The act also made it possible for Rhodesia to join the Union on terms to be approved by the privy council and for the high commission territories of Basutoland, Bechuanaland protectorate and Swaziland to join on terms laid down in a schedule, but only with the consent of the imperial government, which had special commitments to their African inhabitants.

D. The Union and the Republic

The Race Issue.—The dominating factor in the history of South Africa has been the exceptionally complex character of the population. In 1910 there were nearly 6,000,000 inhabitants, of whom 21½% were whites, 67% Africans, 9% Coloured people and 2½% Asians. By 1960 there were nearly 16,000,000, of whom 19½% were whites, 68% Africans, 9½% Coloured people and 3% Asians. In 1910 the whites included an Afrikaner majority, who were mainly rural, and a British minority, who were mainly urban. Between them they owned most of the land and the capital, did most of the skilled work and possessed 93% of the votes. The Africans were still predominantly tribal rather than western in culture, but, as a result of their conquest and of white missionary activity and economic enterprise, there was already a distinct trend toward the disintegration of tribalism. Although most Africans still had homes of sorts in the reserves, those were scattered lands which amounted to only one-fourteenth of the area of the country and by no means provided them all with a livelihood; consequently, many Africans went out to earn wages on white farms and in the towns. Moreover about 1,000,000 Africans had long been established as squatters on white farms and a few were already completely urbanized. The Africans were subject to pass laws which restricted their movements outside the reserves, to Masters and Servants laws which made breach of contract a criminal offense in many types of employment, to special taxation and to other special laws. These laws varied from province to province, and in the Cape there were about 7,000 African voters who were exempted from some of them. The Coloured people were western in culture. Most of them lived in the Cape province, where their status was that of the whites, and about 14,000 were voters, whereas in the other provinces they were subject to discrimination. The Asians were mainly the product of the indentured Indian immigration scheme to Natal, where they outnumbered the whites and were subject to special laws, including a £3 tax. They were excluded from the Orange Free State, but about 11,000 had settled in the Transvaal, where, too, they came under special laws.

The policies of South African governments differed on many important issues; but all, being responsible to a predominantly white electorate, stood more or less explicitly for "the maintenance of white supremacy," a task which became increasingly difficult as a result of the rapid tempo of economic development in the Union and the change in the balance of power and the climate of public opinion elsewhere.

The Botha-Smuts Regime, 1910–24.—In 1911 the Afrikaner parties of the former colonies merged to form the South African party, which, under Botha and Smuts, governed the country until 1924. Botha and Smuts had been zealous fighters for republican independence so long as there was a chance of success, but they were realistic enough to accept the inevitable at Vereeniging and imaginative enough to see great prospects for South Africa when the hot mood of British imperialism had spent itself. By 1910 they hoped for the coalescing of the British and Afrikaner elements into a white South African nation which would freely cooperate with the British Commonwealth in peace and in war. The ideal went too far, however, for the more conservative Afrikaners, who were concerned with preserving their group identity, and not far enough for the more self-conscious nonwhites, for whom it implied a permanently inferior status.

The government wished to keep most of the Africans in the reserves to prevent the whites from being swamped, and to use African manpower as the unskilled base of all forms of economic enterprise—objectives which involved certain contradictions. As the Africans became more accustomed to a money economy, some of them were liable, unless checked by law, to compete successfully with the less efficient whites. This was already happening in the gold-mining industry, and laws were passed in 1911 to preserve the racial hierarchy in that industry. Such competition was also happening in some rural areas, where Africans were pooling their resources to buy more land; a Natives Land act was passed in 1913 to limit the areas within which such purchases could be made. Neither of these acts was wholly effective. In 1916 a commission reported that if the policy of territorial segregation was to be carried out the reserves should be substantially increased in size, but little was done about it because the whites were not prepared to make the necessary sacrifices.

The Asian question had reached a more crucial phase. Already before union Mohandas Karamchand Gandhi (*q.v.*) had organized passive resistance against the Transvaal government, and after union he resumed the struggle against the £3 tax in Natal. As a result the Indian government stopped the importation of indentured Indian labourers in 1911 and the £3 tax was removed in 1914. By that time the pattern of later disputes had developed: the Union government regarded the Indians as temporary visitors and tried to persuade them to return to India; and most of the Indians regarded themselves as permanent residents and demanded full rights of citizenship.

The conservative Afrikaners had found their leader in J. B. M. Hertzog (1866–1942). Although he had joined the Botha cabinet in 1910, Hertzog regarded his colleagues' policy as liable to cause the Anglicization of the Afrikaner people, and he publicly advocated a different, "twin-stream" policy, which led to his exclusion from a reconstituted cabinet in 1912 and to the foundation of the Nationalist party. When in 1914 the government unhesitatingly took the part of Great Britain in World War I and parliament voted funds for the conquest of German South West Africa, a number of former republican generals, some of whom held appointments in the South African defense force, came out in rebellion to avenge Vereeniging. After mastering the revolt, Botha took command of the South-West African expedition and forced the Germans to capitulate on July 9, 1915. South African volunteer contingents also fought in east Africa and on the western front in Europe and many individuals joined British units. Smuts, after serving as commander in chief in east Africa, did notable work in the British war cabinet. In July 1919 he and Botha, on behalf of South Africa, signed the treaty of Versailles and the covenant of the League of Nations, under which South West Africa became a Union mandate. Back in South Africa Botha died before August was out, and Smuts, succeeding to the premiership, faced the discontents caused by thwarted Afrikaner nationalism, a steep rise in the cost of living and industrial troubles.

As early as 1907 Smuts had intervened on the side of the employers in a strike of white workers on the Witwatersrand, and in 1914 he had ended another strike by declaring martial law and illegally deporting nine strike leaders. After the war there was a serious crisis. In Dec. 1921 the Chamber of Mines, faced with rising costs and a fall in the price of gold, decided to organize the industry more rationally by using Africans for semiskilled work. There was a violent reaction by white labour, which Smuts suppressed at a cost of 230 lives, with the result that, although the threat to relax the colour bar was not carried out, the Labour party, representing the aggrieved white workers, made an electoral pact with the Nationalists. In the hope of redressing the political balance, Smuts wooed the Southern Rhodesian electorate to accept incorporation in the Union, but when a referendum was held on the issue in 1922 they preferred to remain separate. Two years later the South African party was heavily defeated at the polls and Hertzog became premier of a Nationalist-Labour coalition.

Hertzog's Nationalist Policy, 1924–33.—Hertzog's main objectives were to complete the emancipation of South Africa from imperial control and to provide greater protection for the whites from the Africans and for the Afrikaners from the British. He played a notable part in the events leading to the Balfour report (1926) and the Statute of Westminster (1931), which gave statutory definition to the established convention that the British government could not exert authority over a dominion. South Africa's sovereign status was also asserted by the adoption, after a long and bitter controversy, of a distinctive national flag (1927), by the appointment of ambassadors to Rome, Washington and The Hague (1929), and by the separation of the office of governor general, the head of the South African government, from that of high commissioner, the representative of the British government in South Africa (1931).

Economic nationalism was fostered by the foundation of a state-controlled iron and steel works at Pretoria, by increased protection for agriculture and industry, by a reduction in imperial preferences and by a commercial treaty with Germany. White supremacy was bolstered by the provision of sheltered employment for "poor whites" in state enterprises, by a Mines and Works Amendment act (1926), which was more effective than its predecessor in shutting Africans out from skilled mining trades, by a Native Administration act (1927) and a Riotous Assemblies act (1930), which gave the executive wide powers over individuals, and by franchise acts (1930–31), which extended the vote to all white men and women, while they left the Cape nonwhite vote as before, restricted to men who possessed property and educational qualifications. Hertzog's legislative program was still incomplete when the 1930 world depression undermined the prosperity of the country and the popularity of the government, which continued to cling to the gold standard after Great Britain had abandoned it in 1931. After the government, yielding to pressure from some of its own supporters, left the gold standard in Dec. 1932, there was a fairly rapid economic recovery, but political confusion continued until, in 1933, Hertzog and Smuts formed a coalition government which secured overwhelming support from the electorate.

Hertzog-Smuts Regime, 1933–39.—The coalition was based on a great deal of common ground between Hertzog and Smuts. Hertzog no longer feared political co-operation with Smuts's British supporters, believing that the Afrikaners had been raised to the point where they were immune from the danger of Anglicization, and Smuts was reconciled to most of Hertzog's achievements and was prepared to collaborate with him in rounding them off. Consequently, in 1934 the two major parties fused to form the United party; but Hertzog failed to carry with him a small group of Afrikaner irreconcilables, led by D. F. Malan, who formed the Purified Nationalist party, while C. F. Stallard at the other extreme dissociated himself from Smuts and formed the Dominion party. By then the constitutional settlement had been completed by the Status act and the Seals act (1934) which secured, so far as words could do, that South Africa was master in its own

house. The government then proceeded to complete Hertzog's segregation legislation. A Representation of Natives act (1936) removed the Cape province African voters from the common roll and gave them the right to elect three white members to represent them in the lower house, gave the Africans throughout the Union the right to elect four white senators, and created a Natives' Representative council with advisory powers. A Native Trust and Land act (1936) provided for a considerable increase in the size of the reserves. A Native Laws Amendment act (1937) authorized the executive to prevent more Africans from entering the towns and to compel municipalities to segregate African from white residents. It was Hertzog's hope that the enlarged reserves would become capable of maintaining almost all the African people, so that those who worked for whites could be regarded as temporary visitors from the reserves. But it is notable that he saw the need for consultation with an African council, and that he did not try to discriminate against the Coloured people. Moreover, his government expanded the social services not only for the whites but also for the nonwhites; there was, for example, a considerable increase in government grants for education.

Nevertheless, Africans, Asians and Coloured people were becoming disturbed by the great contrast between their living standards and those of the whites and by an accumulation of laws which caused 500,000 Africans to be convicted of statutory and municipal offenses in a year and many more to come into unpleasant contact with a police force recruited from among the poorer Afrikaners, who were imbued with the strongest race prejudices. As tension began to mount the Nationalists claimed that "white South Africa" would not be safe unless the restrictions on Africans were increased and the Coloured people were also segregated from the whites. Though Hertzog strongly disagreed, the Nationalists were able to profit from the celebrations held to mark the centenary of the Great Trek in 1938, when they appealed to all Afrikaners to remain true to the principles of the voortrekkers and of Paul Kruger, placing special emphasis on their race attitudes.

World War II.—The outbreak of World War II caused a crisis in South Africa. Although Hertzog and Smuts had co-operated successfully on many issues, they differed widely on foreign affairs. Hertzog took an indulgent view of Nazi Germany, believed that its expansion was no concern of South Africa, and moved in the house of assembly, on Sept. 4, 1939, that South Africa should remain neutral. Smuts, keenly aware of the wider implications of Naziism, maintained that it was the interest and the duty of South Africa to support Great Britain, and he won the crucial division by 80 votes to 67. On Sept. 5 Smuts formed a ministry with the support of the Labour party, the Dominion party and the majority of the United party, and war was declared on Germany.

The Smuts government concentrated its energies on the war. About 200,000 white men (more than half of them Afrikaners) and 125,000 nonwhites (mainly Africans and Coloured men) joined the forces, and many of them served with distinction in the Ethiopian, Mediterranean and Madagascan theatres. The nonwhites were not allowed to bear arms but were distributed among the combatant units, for whom they performed vital work as stretcher-bearers, labourers, etc. Industry was efficiently switched to the production of munitions and clothing for military purposes. Smuts himself remained in close contact with Winston Churchill and watched carefully the strategy of the war, often leaving his able lieutenant, J. H. Hofmeyr, Jr. (1894–1948), to carry a lion's share of the burden of administration. South Africa's achievements were impressive, considering the strength of the opposition to the war. Fortunately for the government, the opposition splintered into fragments during the most critical period: J. F. van Rensburg's Ossewabrandwag ("Guard of the Oxen Cart") and Oswald Pirow's New Order, accepting the racial doctrines of the Nazis and their contempt for parliamentary government, patently hoped to profit from a Nazi victory; the Malanites expounded the ideal of an Afrikaner republic in which, as in Kruger's republic, the British would not necessarily have political rights; while the Hertzogites stood firm by their principle of equality between the whites. In the first flush of his disappointment, indeed, Hertzog led his defeated minority from the United party to join hands with Malan and form the Reunited Nationalist party, but the reunion was not a happy one and before the end of 1940 Hertzog had been driven out for his tenderness toward the British. His colleague, N. C. Havenga, subsequently founded the Afrikaner party to maintain his ideals. The general election of 1943 was a victory for Smuts in that his prowar coalition secured a majority of 67 seats in a lower house of 153; but it was also in a sense a victory for Malan's Nationalists, who won all the 43 opposition seats. Thereafter, as the prospect of a Nazi victory faded, the Nationalists consolidated their position as the political instrument of self-conscious Afrikanerdom, absorbing elements from the Ossewabrandwag, the New Order and the Afrikaner party, and drawing support from a variety of cultural and economic organizations, including the Broederbond, a secret group of Afrikaner elite.

As the war drew to a close Smuts drafted the preamble to the United Nations charter signed at San Francisco, Calif., in 1945 and returned to mold postwar South Africa along the lines of a generous demobilization scheme, an expansion in the social services for all races, a planned development of agricultural, mineral and industrial resources, and large-scale white immigration.

Industrial Expansion in South Africa.—Subsequent events can be understood only in relation to the fact that from about 1938 to about 1951 South Africa was experiencing an industrial expansion as intense as the British industrial revolution of the early 19th century, with the added complication of racial differences between the unskilled workers on the one hand and the skilled workers, the capitalists and the majority of the electorate on the other. The way had already been paved by the rise of the mining industries and by the foundation of many manufacturing industries during World War I and the 1930s; the rate of expansion increased during and after World War II. In the 15 years between 1936 and 1951 the total population increased by 32%. In the 13 years between 1938 and 1951 the output of coal, pig iron, iron ore, copper and many other mineral products doubled. In the same period the net national income increased threefold, the revenue fourfold, and the value of imports and exports fivefold; while the net value of the output of the manufacturing industries increased nearly sixfold. Even when a correction is applied for the change in the value of money, those increases were exceptional.

The gold-mining industry continued to play a central role in the Union's economy, and fears that the supply would soon be exhausted were removed by the opening up of new mines in the Orange Free State, where production began in 1951. Nevertheless, gold production declined between 1941 and 1954, and that decline, together with the expansion of the manufacturing industries, reduced the relative importance of the gold-mining industry, which produced only 9% of the national income and 22% of the exports in 1951 (compared with 18% and 75% respectively in 1936). The mining industries continued to be organized on racial lines and the manufacturing industries adopted a similar basis. Skilled work in South Africa remained virtually a white monopoly and was well paid; most of the unskilled work continued to be performed by nonwhites for low wages.

As a result of this expansion nearly three-quarters of the white population were townsmen by 1951, the "poor whites" were absorbed by industry and the earlier economic distinctions between the Afrikaners and the British inhabitants became blurred. Industry also attracted many nonwhites. More than 2,300,000 Africans (over a quarter of the total African population) were in the towns at the time of the 1951 census. Into some of these towns the influx was so rapid that housing and other amenities were grossly inadequate and there was much crime. But in spite of low wages and squalid living conditions, the urban Africans earned far higher incomes than the rural Africans.

The 1948 Election.—As the 1948 election approached, disturbing facts were becoming known in South Africa. Although the reserves had been enlarged under the 1936 act, only 40% of the African population was to be found in them at the time of the 1946 census. The social and economic planning council reported that the quality of the reserve lands was deteriorating; and early in 1948 a commission revealed that many Africans were becoming

permanent inhabitants of the towns. These facts did not square with the argument that the Africans had adequate homes in the reserves and could therefore be treated as inferiors elsewhere. Moreover the Natives' Representative council was frustrated because the government ignored its advice and the African National congress was demanding the removal of discriminatory laws in terms of the ideals of the United Nations. This raised the question whether white supremacy, which had developed in the simple pastoral economy of the 18th century, could be maintained in the industrial economy of the 20th century.

The Smuts government continued to profess adherence to the segregation policy but tried to deal with practical grievances in a conciliatory spirit, while Hofmeyr went farther than his colleagues on occasion, feeling toward a policy which might be acceptable to all races. But Asians were protesting vigorously against legislation that prevented their buying or occupying new premises in Durban, and the Indian government attacked South Africa in the United Nations general assembly. Consequently, when Smuts tried to persuade the assembly in 1946 to allow the Union to incorporate South-West Africa, he met with a rebuff, and the assembly passed the first of a series of resolutions condemning racial discrimination in South Africa.

The Nationalist leaders wished to reduce South Africa's links with Britain, the commonwealth and the United Nations, to advance the power of the Afrikaner people, and, above all, to preserve white supremacy by every possible means. They expressed horror at the "liberalism" of the United party, declared it would cause the "suicide" of white South Africa, and proclaimed a policy of apartheid (*q.v.*), according to which whites, Coloureds, Asians and Africans would be separated from each other and each race would be able to "develop along its own lines in its own area." Though as far as the Africans were concerned this was little more than Hertzog's segregation policy in a new dress, it was presented to the electorate as something new and dynamic. The Nationalists also exploited the discontents caused by wartime controls and were assisted by the fact that the electoral system favoured the rural voters. The result was that in May 1948 Malan, with the assistance of Havenga's small Afrikaner party, was able to form a government, with a majority of five in the house of assembly. This result, which was a surprise to Malan as well as to Smuts, was decisive for South Africa.

Nationalist Party Policy From 1948.—A spate of legislation followed. Mixed marriages and interracial unions out of wedlock were made illegal; South Africa began to be divided into a vast number of uniracial zones for residence and for business; African tribal authority was bolstered up, under chiefs who were strictly subordinated to the government; African schools were brought under government control; and the Natives' Representative council was abolished. The discretionary powers of the executive, already considerable, were increased by the sweeping Suppression of Communism act and others, and a Public Safety act gave it vast powers in an emergency. Smuts's scheme to bring in large numbers of white immigrants was suspended. A Citizenship act extended the period of residence of British immigrants before they could apply for citizenship and the franchise. The white people of South West Africa were given six seats in the house of assembly. But an act to remove the 50,000 Coloured voters from the common roll in Cape Province and give them, like the Africans, the right to elect three whites to represent them in the house of assembly was declared invalid by the supreme court, as being in conflict with the South Africa act, since it had not been passed by a two-thirds majority at a joint sitting; and although under a further act the Nationalist members of parliament sitting as a "high court" reversed that decision, the supreme court held that the high court of parliament was itself invalid.

As the 1953 election approached, the Nationalist party swallowed the Afrikaner party and called for a further mandate "to save white South Africa." The Nationalists made several gains in the election, but they still lacked the two-thirds majority they needed to remove the Coloured voters from the common roll. However, the United party, led by J. G. N. Strauss since the death of Smuts in 1950, was weakened by internal differences and in 1954 some right-wing members broke away to form a short-lived Conservative party.

In Nov. 1954 Malan retired from politics at the age of 80. J. G. Strijdom, the head of the Nationalist party in Transvaal, was then elected leader of the party by the parliamentary caucus and became prime minister, while the more moderate Havenga, who had aspired to the premiership, retired from politics. Under Strijdom the party moved to a more extreme position. The civil service and the judiciary were increasingly afrikanerized; an Industrial Conciliation act gave the minister of labour the power to reserve any class of work for members of any race, and this power was used mainly to reserve skilled jobs for white people; and in Dec. 1956 more than 150 men and women from all racial groups were arrested on charges of high treason. A way was also found to overcome the constitutional safeguard for the Cape Coloured franchise. The senate was enlarged and its composition was changed in 1955 to give the Nationalists 77 seats out of 89 in that house and a two-thirds majority in a joint sitting of both houses. The Coloured voters were removed from the common roll by the necessary two-thirds majority in 1956. In the 1958 election the Nationalist party made further gains.

Strijdom died in Sept. 1958 and was succeeded by H. F. Verwoerd, another Transvaaler, who had been minister for native affairs since 1950 and was a principal architect of the policy of apartheid. His government proceeded to put the coping stones on the work of its predecessors. After 1959 it became unlawful for nonwhite students to enter the established universities, though the universities of Cape Town, the Witwatersrand and Natal had previously admitted them and wished to continue to do so. Instead the government created—and closely controlled—new "university colleges" for Africans in Zululand and the Transvaal, for Coloured people in Cape Town, and for Indians in Natal; and it also took over the African college at Fort Hare, which had previously been affiliated to Rhodes university, Grahamstown. The Promotion of Bantu Self-Government act, 1959, removed the Natives' Representatives from both houses of parliament and from the Cape provincial council, and made the tribal authorities (composed mainly of chiefs and councilors appointed by and dismissable by the government, and urban location boards which had no powers) the only African institutions which had official recognition. The house of assembly then contained 156 whites elected by the white voters of the Union and South West Africa (*i.e.*, white citizens aged 18 and over) and four whites elected by the Cape Coloured voters. The senate, as again reconstituted by an act of 1960, contained 43 whites elected by white electoral colleges and 11 whites nominated by the government, one of whom was nominated to "represent" the Coloured people. Thus the compromise embodied in Hertzog's 1936 legislation was obliterated. The only nonwhites who had any say in the composition of the parliament which was supreme over South Africa were the Coloured people of the Cape province, and their say was meagre.

Opposition to Nationalist Rule.—Nevertheless, Nationalist ascendancy was fundamentally insecure. South Africa remained, in essence, a single society. The Asian and Coloured communities continued to increase in numbers and to form parts of the general economy and the 1960 census revealed great increases in the African town populations; so that no progress had been made toward the declared objective of separating the races into distinct and potentially autonomous regions, and no real justification for the discriminatory laws existed. Furthermore, economic development had slowed down since 1951. While the net value of the output of the manufacturing industries increased nearly sixfold between 1937 and 1951, it increased by only 37% between 1951 and 1957; and while the net national income increased threefold between 1938 and 1951, it increased by only 57% between 1951 and 1959.

The government encountered considerable opposition from white South Africans. People of British descent in Natal and elsewhere were estranged by the afrikanerization of the civil service and the judiciary, and by republicanism. Businessmen reached the conclusion that the colour bar was a serious impediment to economic stability. Intellectuals and churchmen denounced apart-

heid, in practice, as immoral and unchristian; and even within the Afrikaner universities and the Dutch Reformed churches there were many who expressed doubts about the moral basis of apartheid. Though the United party, led by Sir D. P. de Villiers Graaff after the retirement of Strauss in 1956, remained equivocal in its attitude to apartheid, a Liberal party was founded in 1953 and a Progressive party in 1959, on the basis of the rejection of discrimination on racial grounds.

Opposition from nonwhites became increasingly vigorous. After 1948 nearly every nonwhite leader claimed full citizenship rights for his people and the African National congress gained a considerable following. Unable to obtain redress of grievances by ordinary constitutional means, it adopted a policy of "passive resistance to unjust laws." Several hundred people, mainly Africans, deliberately courted arrest and imprisonment in 1952–53. About 45,000 Africans walked from their locations to their places of work in Johannesburg and Pretoria for ten weeks in 1957, rather than pay increased bus fares. There were also intermittent riots and clashes between Africans and police in many parts of the country, both urban and rural. These reached a high point in 1960. In January a party of police was attacked by an African mob in Cato Manor, a location near Durban, and four white and five nonwhite policemen were killed. The Pan-African congress, an offshoot of the African National congress, organized demonstrations against the pass laws in March, when police opened fire upon an African crowd at Sharpeville, near Johannesburg, killing 67 and wounding 186, and blood was also shed in Cape Town, after which most of the Africans in Johannesburg and Cape Town abstained from work and there were riots elsewhere. The government declared a state of emergency under the Public Safety act (1953), banned the African National congress and the Pan-African congress, mobilized the citizen force and arrested about 2,000 white, Coloured, Asian and African leaders, as well as 18,000 "vagrants and undesirables." Then in April the prime minister narrowly escaped assassination, when David Pratt, a white farmer, fired two revolver shots at point-blank range during the opening of the Rand Jubilee exposition. Before the year was out there were further clashes between police and Africans in Johannesburg, while in rural Pondoland there was a rebellion against the system of government-controlled tribal authorities.

There was also strong and concerted opposition to the South African government from abroad. Apartheid was continually denounced by the UN general assembly. In 1959, by 62 votes to 3 with 7 abstentions, the assembly expressed regret that the Union government had not responded to previous appeals to reconsider its policies. After the Sharpeville shootings the UN security council met in special session and passed a resolution by nine votes to nil, with Britain and France abstaining, deploring the loss of life, blaming the government's policy for it, stating that the situation might endanger international peace and calling on South Africa to abandon apartheid. Severe criticisms also came, not only from Afro-Asian countries, but also from the United States, in a declaration from the state department, and from Britain, in a resolution of the house of commons.

South Africa's Withdrawal From the Commonwealth.— In Oct. 1960 the white voters of the Union were asked to indicate in a referendum whether they wished South Africa to become a republic. A small majority expressed a preference for a republic and the government then decided that the change should take place on May 31, 1961. During the commonwealth prime ministers' conference in London in March 1961 Verwoerd initially requested that South Africa might remain a member of the Commonwealth of Nations after the change took place; but the criticisms of South Africa's racial policies, initiated by the Afro-Asian members and supported by Canada, were so vehement that Verwoerd withdrew his request. Consequently on May 31, 1961, South Africa not only became a republic, with C. R. Swart, elected by the members of parliament, as president, but also left the commonwealth. A general election was held in Oct. 1961. From the white electorate, the Nationalists increased their seats from 102 to 105; and although the United party increased from 42 to 49 seats, this was at the expense of the Progressive party, which lost all but 1 of its 11 seats; while the National union also won one seat—J. D. du Plessis Basson, a former Nationalist (who later joined the United party). Most of the Coloured voters boycotted the election and those who did vote returned one United party member and 3 Independents. The election showed that the anxieties which existed among some Afrikaner intellectuals and churchmen were not politically significant. Verwoerd had established a complete mastery over the Nationalist party and the party's control over the political apparatus was more secure than ever. Moreover, since the United party had purged itself of its liberal wing, the only member of the new house of assembly who was opposed to the principle of racial discrimination was the Progressive party member, Helen Suzman.

In 1962 the new parliament voted a substantial increase in the military budget and enacted a drastic law in restraint of sabotage. The government also announced plans to create an African "parliament" in the Transkei, to consist for the most part of chiefs and chiefs' nominees, with a minority of elected members; but these plans did not allay African discontent and acts of sabotage continued to take place. During 1962 external criticisms of South Africa became more violent; and in November, at the instigation of the Afro-Asian bloc, the UN general assembly, by 67 votes to 16, with 23 abstentions, called upon member states to apply economic sanctions against South Africa and to break off diplomatic relations with the republic. For the effects of apartheid on the commonwealth *see* COMMONWEALTH OF NATIONS: *Changing Commonwealth.* (L. M. T.)

V. POPULATION

At the first census (1911) after union the population totaled 5,972,757 but it had nearly trebled by 1960, reaching 15,841,128. The increase is due only slightly to immigration and overwhelmingly to natural increase which in 1960 was 16.4 per 1,000 for whites, 28.5 for Asians and 32.9 for Coloureds. No reliable statistics are available for the Bantu (a term that has officially replaced "native" or "African").

From Tables I and II the numerical preponderance of the Bantu in the population is evident. Since 1911 they have maintained their rather more than three to one lead over the whites. About 30% of the Bantu population is urban, the majority living in municipal townships with a considerably smaller number in white areas as servants. Industrial compounds house about 146,000 and about 474,000 live in mine compounds. Of the Bantu rural population (70% of their total numbers) more than half live in the

TABLE I.—*Population of South Africa, 1911–60*
(in 000)

Census year	All races		Whites		Bantu		Asians		Coloureds	
	Total	Urban	Total	Urban	Total	Urban	Total	Urban	Total	Urban
1911	5,973	—*	1,276	—*	4,019	—*	152	—*	525	—*
1936	9,588	3,204	2,003	1,361	6,596	1,246	220	157	769	440
1946	11,416	4,393	2,372	1,769	7,831	1,856	285	203	928	565
1951	12,671	5,397	2,642	2,071	8,560	2,329	367	284	1,103	713
1960	15,841	—*	3,068	2,565	10,808	—*	477	—*	1,488	—*

*Not issued.

TABLE II.—*Provincial Areas and Populations, 1960 (Preliminary Census)*

Provinces and capitals	Area (sq.mi.)	All races	Whites	Bantu	Asians	Coloureds	Density per sq.mi.
Cape Province (Cape Town)	278,465	5,308,839	997,377	2,976,827	20,243	1,314,392	19.1
Natal (Pietermaritzburg)	33,578	2,933,447	340,293	2,155,824	394,237	43,093	87.4
Transvaal (Pretoria)	110,450	6,225,052	1,455,372	4,601,545	62,918	105,217	56.4
Orange Free State (Bloemfontein)	49,866	1,373,790	274,596	1,073,613	16	25,565	27.5
Total	472,359*	15,841,128	3,067,638	10,807,809	477,414	1,488,267	33.5

*Data exclude Walvis Bay (374 sq.mi.)

Bantu homelands; most of the remainder are labourers or servants on white farms. Table II shows that the Bantu are most numerous in Transvaal, which is partly because of their extensive employment in the industries of the southern Transvaal industrial and mining area and partly because of the numerous homelands in the province. The second highest number is in Cape Province because of the population of about 1,500,000 in the homelands of the Transkei and Ciskei. Of the total number in Natal (including Zululand) about 900,000 are in the homelands, the rest being in the white area. The white population has increased by more than 1,000,000 since 1936. From Table I it is clear that this section of the population is predominantly urban. From Table III it is seen that the majority of urban whites is concentrated in the southern Transvaal industrial area, which includes the Witwatersrand (notably Johannesburg, Germiston, Benoni and Springs), Pretoria and Vereeniging-Vanderbijlpark. The four ports of Cape Town, Durban, Port Elizabeth and East London together contain about 617,000 whites. With the provincial administrative capitals

TABLE III.—*Metropolitan Areas, 1960*

Metropolitan area	All races	Whites	Bantu	Asians	Coloured
Witwatersrand	2,088,399	734,505	1,240,807	37,772	75,315
Johannesburg	1,096,541	389,690	622,831	27,069	56,951
Cape Town	731,484	279,404	65,133	9,369	377,578
Durban	655,370	194,276	204,071	231,385	25,638
Pretoria	415,989	202,664	198,295	7,805	7,225
Port Elizabeth	270,815	94,085	111,651	4,165	60,914
Germiston	204,605	84,419	113,620	2,310	4,256
Bloemfontein	140,924	61,213	73,446	3	6,242
Benoni	135,467	41,305	85,774	2,989	5,399
Springs	135,231	36,445	95,911	1,364	1,511
Vereeniging-Vanderbijlpark	115,892	46,295	67,542	903	1,152
East London	114,584	48,725	56,056	1,753	8,050
Pietermaritzburg	95,124	39,472	29,305	21,741	4,606
Kimberley	77,180	23,987	35,796	1,076	16,321

Bloemfontein and Pietermaritzburg the white population in these "conurbations" is rather more than two-thirds of the total white population. Of the rural white population about 400,000 are on the 112,000 white farms, which cover an area of about 216,000,000 ac.

The Coloured population, deriving from a mixture of various racial groups, is concentrated mainly in Cape Province, where they supply most of the labour for the farming and manufacturing industries. Included for statistical purposes in the Coloured total are 63,000 Malays, of whom 55,000 are urban in Cape Province.

The Asians have increased more rapidly than any other racial group. They consist almost entirely of Indians who were brought over from India in the second half of the 19th century for the Natal sugar plantations. Of the total 477,414 in 1960, 394,237 were in Natal, about 307,000 being urban, and of the urban Asian population 231,385 were in the Durban metropolitan area, exceeding the whites in this area by nearly 40,000. (J. H. WN.)

VI. ADMINISTRATION AND SOCIAL CONDITIONS

1. Constitution.—Following the passage of the South Africa act, 1909, by the British parliament, the Union of South Africa was formed on May 31, 1910, out of the colonies of the Cape of Good Hope, Natal, Orange River Colony (Orange Free State) and Transvaal, which became the subordinate provinces of the new state. The Union was a member of the British empire, and its sovereign independence was later guaranteed by the Statute of Westminster (1931) and the Status of the Union act, 1934. On May 31, 1961, after a popular referendum held among white voters in Oct. 1960, which favoured by a narrow margin the establishment of a republic, the Union became the Republic of South Africa under a constitution adopted by the Union parliament. On becoming a republic South Africa ceased to be a member of the Commonwealth of Nations (see *History: South Africa's Withdrawal From the Commonwealth*, above).

The new constitution (further amended in 1963) is virtually the same as that of the Union enacted in 1910, with certain changes such as effected in the article dealing with the office of president and substitution of "the State" for "Crown" or "the King or Queen." It provides for a state president (who is head of the republic), an executive council of ministers of state appointed by him, a parliament comprising a senate and a house of assembly and, in each province, an administrator appointed by the state president-in-council for five years, an executive committee of four members under the administrator's presidency, and an elected provincial council which elects the executive committee and, like it, serves for five years. In 1920 the former German territory of South West Africa was assigned to South Africa under mandate. The people of this territory are regarded as South African subjects and are represented in the parliament of the republic (*see* SOUTH WEST AFRICA).

The Executive.—The state president is elected for a term of seven years by secret ballot of an electoral college consisting of members of both houses of parliament. He is a constitutional ruler acting on the advice of his ministers, whom he appoints from among sitting and prospective members of parliament having the confidence of the majority in the house of assembly, and they together form a cabinet which in practice is the executive council. Ministers in charge of departments may not exceed 18 in number but the state president may also appoint not more than six deputy ministers.

The Legislature.—Parliament has unrestricted legislative power "to make laws for the peace, order and good government of the republic." It consists of an upper house, the senate, and a lower house, the house of assembly. Ordinarily, it legislates by means of bare majorities in both houses sitting separately, but in the case of the entrenched clause of the Republic of South Africa Constitution act providing for the equality between the two official languages, English and Afrikaans, it must legislate by means of a two-thirds majority of all members of both houses sitting together.

Although both houses are original legislative bodies, in practice the senate is in the main a revising chamber with delaying powers in respect of bills introduced in the lower house. In 1960 the senate's constitution was changed to form a chamber consisting of 54 members, of whom 14 were elected in the Transvaal, 11 in Cape Province, 8 each in Natal and the Orange Free State, and 2 in South West Africa. In addition, eight were nominated for South Africa (four of them for their knowledge of the Coloured peoples) and two for South West Africa, one for similar knowledge. The Coloured population in the Cape had one representative. Qualifications for senators were the same as for members of the lower house, except for a minimum age limit of 30 years. Four seats formerly held by senators elected by Bantu voters were abolished in 1960.

In 1961 the house of assembly comprised 150 members directly elected by white voters in the republic, 6 members directly elected by white voters in South West Africa and 4 members elected by Coloured voters in Cape Province. Three representatives previously elected by Bantu in the Cape ceased to be members of the house in 1960. Members must be adult South African citizens of white descent, and the representatives of white voters are elected from single-member constituencies delimited by a judicial commission at intervals from five to ten years. Seats are allotted to the provinces (and provincial councils) in proportion to their adult white population. Under the 1958 delimitation the Cape returned 52 members, Natal 16, the Orange Free State 14 and the Transvaal 68. Electoral divisions within each province have an approximately equal number of voters, but may exceed or fall short of the average by 15% under a so-called loading system.

With the abolition of Bantu representation and the lowering of the age qualification for white voters in 1959, the parliamentary electorate comprises white South African citizens of both sexes who are 18 years and over and a separate roll of Coloured males in the Cape possessing prescribed literacy and property qualifications.

2. Government.—There are 38 departments of state, staffed by about 241,000 public or civil servants. Departmental responsibility is vested in individual members of the cabinet, which is presided over by the prime minister. Statutory bodies have been created to supervise broadcasting, electricity supply and the development of residential group areas according to race; to produce

iron, steel phosphates and other fertilizers; to extract oil from coal; to finance the development of industries and fisheries; and to promote the economic development of the Bantu and Coloured peoples in separate areas. The railways, harbours and airways are operated by an administration responsible to the minister of transport. About 80 major public bodies have been set up for regulatory and advisory purposes, including an economic advisory council, a national education advisory council and an immigration council.

Provincial Administration.—The legislative powers of the provincial councils are defined and limited by the Republic of South Africa Constitution act of 1963 and the Financial Relations Consolidation and Amendment act, 1945, and their ordinances may not be repugnant to an act of parliament. The constitution guarantees the provincial boundaries. The provincial administrations comprise the administrator; the executive committee, which under an act of 1962 is to be elected by the majority parties in the councils instead of by proportional representation as in the past; a provincial secretary; and various departments and branches. They deal with education, hospitals, roads, shop hours, public entertainments, horse racing, libraries, new townships, town planning and motor vehicle licensing. They also control local government authorities. Under the constitution, parliament may not abolish a provincial council or abridge its powers except on a petition by the council. The South West Africa administration is similarly constituted.

Local Government.—Each province has its own system of local government. In the Cape there are municipalities with city and town councils, village management boards, periurban (suburban) local boards and divisional councils (rural). They are wholly elected except for local boards. In Natal there are elected city and town councils and town boards, and elected or appointed health committees. Periurban development is controlled by a local health commission appointed by the administrator. In the Orange Free State there are municipalities with city and town councils and village management boards, all elected. Elected small holdings committees of management in periurban fringes or settlements are controlled by an appointed small holdings area control board. In the Transvaal, there are elected city and town councils and village councils, and elected and appointed health committees. Periurban development is controlled by a special health board appointed by the administrator. South West Africa has institutions similar to those of the Cape.

Nonwhite Administration.—Local authorities control Bantu populations within their borders and establish elected Bantu advisory boards in their Bantu townships. In 1961 provision was made for the establishment of partially elected Bantu urban councils in the townships. In the homelands, a reorganization of local government was undertaken into tribal, regional and territorial authorities under tribal chiefs and the supervision of the department of Bantu administration and development. In 1963 legislation was passed conferring limited autonomy (*i.e.*, excluding foreign affairs, defense, etc.) on the Transkei homeland, the first of the projected "Bantustans." Local councils for Coloured people are being created in the Cape under the supervision of the department of Coloured affairs and the Coloured advisory council. Similar bodies were planned for the Indian population by the department of Indian affairs.

Political Parties.—There are two main political parties in South Africa and one other with representatives in parliament. They are the Nationalist party (105 seats in the house of assembly in 1962), the United party (51 seats) and the Progressive party (1 seat). Smaller groups unrepresented in parliament are the Liberal party and the Conservative Workers' party; Coloured voters are represented by four independents. The Nationalist party came into office in May 1948, since when the United party has been the official opposition.

3. Taxation.—Three-quarters of national taxation comprises income tax, customs and excise duties and death duties. There is a growing reliance on the income tax (and super tax) levied progressively, which accounts for more than 50% of total revenue, but incidence is not excessive and is reduced by family and sickness allowances. In 1963 the pay-as-you-earn system of income taxation was introduced. The provinces rely mainly on personal and income taxes (18% of their income), motor vehicle licence revenue, and racing and entertainment taxation; but they are state aided to the extent of at least 50% of their income. In contrast, local authorities receive only 7% in grants, and 40% of their income derives from rates on property, generally considered regressive in incidence. Charges for goods and services account for more than 30% and trading profits 10%.

While such taxation facilitates central budgetary control in the interests of economic stability, increasing urbanization throws a growing financial burden on local authorities, and an inquiry into the redistribution of revenue resources began in 1956. By the early 1960s approximately 80% of all taxation was levied centrally, 13% provincially and 7% locally.

4. Living Conditions.—The standard of living is high for the white population; for most Bantu it remains low but is nevertheless higher than in most African countries.

Employment and Wages.—A maximum 46-hour week was introduced in 1941 and the tendency has been to reduce working hours, in some cases to a 40-hour week. The Industrial Conciliation act, 1956, provides for the registration of employers' organizations and trade unions, the voluntary settlement of disputes by industrial councils or conciliation boards and, finally, arbitration (which is compulsory in essential services where strikes and lockouts are prohibited). Africans are excluded from this act and cannot become members of registered trade unions, but the Native Labour (Settlement of Disputes) act, 1953, created a central native labour board and subordinate regional committees and labour officers. It also prohibited strikes by, and lockouts of, Africans. Outside this legislation, wage determinations are made by a wage board established under the Wage act, 1937, which also determines African wages where the minister refers disputes to it from the central native labour board.

Other legislation includes the Registration for Employment act, 1945, which created a national employment service (including employment exchanges and extending to sheltered employment and unemployment relief), and the Workmen's Compensation and Unemployment Insurance acts, 1941. African labour bureaus and unemployment are dealt with by the department of Bantu administration and development, either regionally or through local authorities.

Trade Unions.—Only a few trade unions, properly so-called, survive; *i.e.*, unions that represent the workers in a specific trade. Most unions have become industrial labour organizations, representing all workers in a particular industry. Under the Industrial Conciliation acts, as amended in 1956, trade unions are prohibited from affiliation with political parties. In the early 1960s the total number of industrial organizations and trade unions exceeded 180, over half of which were for white workers, less than one-quarter were Coloured and more than one-quarter were racially mixed. The major organizations include the South African Trade Union council, the South African Federation of Trade Unions, the Koordinerende Raad van Suid-Afrikaanse Vakverenigings, the National Union of Liquor and Catering Trades and the Mechanics Unions' Joint Executives.

Bantu unions do not exist, but the rights of Bantu workers are safeguarded by the Native Labour (Settlement of Disputes) act of 1953. This act provides for a Central Bantu Labour board, Bantu labour committees and works committees and for Bantu labour officers. Disputes which cannot be settled by the labour officers in collaboration with the regional labour committees and the divisional inspector of labour are referred to the central board. Under the Industrial Conciliation act (1956) certain occupations are reserved for certain racial groups in order to prevent the racial groups from unfair competition.

Housing.—At the 1951 census 564,277 non-African residential units were enumerated, of which 443,616 were occupied by whites. At the time of the 1960 census new dwellings completed in 18 principal municipalities and 45 smaller towns averaged more than 12,600 a year. Residences for whites are usually built privately, but public housing projects were stimulated by wartime measures

that set up a national housing and planning commission with the power to erect houses, to assist local authorities and utility companies and to make loans to individuals. Most projects are economic or "no-loss," but assisted housing is provided for lower income groups in which the central and local authorities share the loss. A ministry of community development and housing was established in 1961 to advance housing programs for all races.

Bantu housing is assisted by a Bantu housing board and undertaken in Bantu townships by urban local authorities with public and private funds. A massive project to rehouse urban Bantu was undertaken from 1951 onward. The "squatters' camps," or slum shacks, which had accumulated around towns to which Bantu had been attracted by industrialization, were removed and new townships built. During ten years more than R.200,000,000 (£100,000,000 or $280,000,000) was spent in providing these families with new houses, services and transportation. The townships are planned on modern lines and built by modern methods; each house has its own garden, and there are community amenities, schools, halls, playgrounds, churches, etc. Ethnic groups are housed in their own separate quarter. An example is Daveyton, outside Benoni. Houses are leased, and rents are government-subsidized where incomes fall below R.360 or £180 ($524) a year (or R.300 in some cases), though some Bantu have bought their houses. The cost of services to urban native townships is defrayed from funds accruing under the Native Services Levy act, 1952, which obliges employers to pay a maximum of *2s. 6d.* (35 cents) a week for each Bantu employed, part of which is reserved for transport to the townships. Outside these areas Bantu (and also Coloured and Asian) housing is provided by the group areas development board. Very little nonwhite housing is privately built.

5. Welfare Services.—Welfare services are run by central, provincial and local authorities and by private organizations. The department of social welfare administers legislation dealing with the protection of children, adoptions, family allowances, probation services, juvenile delinquency, the registration of welfare organizations, assistance to the blind, pensions and poor relief. It subsidizes the voluntary work of about 1,800 registered organizations, including national bodies for cripple care, child welfare, the blind, the deaf, first aid, mental health, and maternal and family welfare. Church organizations also undertake relief and family welfare work.

In the provinces and municipalities, other services are financially assisted or provided in conjunction with hospitals, public health, education and recreation (all races). Co-ordination of welfare services is effected by a statutory nonofficial national board and local boards.

Health.—Public health and medical services fall under a central ministry of health, the four provincial councils and more than 700 local authorities. The ministry of health, created under the Public Health act of 1919, has advisory and executive functions, chiefly in regard to the prevention and control of infectious and communicable diseases and the promotion of health. It has in its employ numerous full-time or part-time district surgeons in urban and rural areas. Health centres and clinics serve mainly the indigent. Larger municipalities have health departments and salaried medical officers. The provinces maintain public general hospitals, including chronic sick homes, for whites and nonwhites. These hospitals are financed from taxation, and in necessitous cases treatment is free or partly free. In addition there are many private, mission, mining and industrial companies' hospitals. Medical-aid societies operate among sections of the public in a number of areas. In 1962 a central authority to co-ordinate this movement on a nation-wide voluntary basis was established. By the early 1960s South Africa had about 8,000 medical practitioners, mostly white but including a number of Bantu, Coloured and Asian doctors.

6. Justice.—The common law of the republic is based on Roman-Dutch law (*q.v.*), the uncodified law of Holland that was retained after the Cape's cession to Britain in 1806. The judiciary comprises a supreme court with local, provincial and appellate divisions, the high court and circuit courts of South West Africa, and inferior courts. The judges, headed by a chief justice, are appointed by the state president-in-council; they cannot be removed from office except by an address of both houses of parliament, and their salaries cannot be reduced. Appeals from the divisional courts and the high and circuit courts of South West Africa lie to the appellate division of 11 judges, with a quorum of 5 except where the validity of an act of parliament is in question, when the quorum is 11.

The inferior courts comprise magistrates' courts, Bantu courts and water courts (dealing with the use of water resources). Most civil and criminal litigation is a matter for the magistrates' courts, each responsible for a magisterial district. Regional magistrates are appointed in populous areas with criminal jurisdiction over a number of districts. The Bantu courts do not have jurisdiction over Bantu in criminal cases or where whites are involved.

Police.—The police are organized nationally into 11 divisions. The force comprises about 11,000 whites and a similar number of nonwhites, and it extends to South West Africa. Each division contains a number of districts and police-station areas. Outside this machinery there are the security police, directly responsible to head office, and the traffic police of provincial and local authorities.

7. Education.—The department of education, arts and science is responsible for primary and secondary education for non-African children who are physically handicapped or in need of care and for vocational schools, including technical high schools and colleges. It is also responsible for the autonomous universities of Cape Town (founded 1873), Natal (1909), Rhodes (1904) and the Witwatersrand (1922) (all English speaking), the Orange Free State (1855), Potchefstroom (1869), Pretoria (1930; from 1908 it was the University College of the Transvaal), Stellenbosch (1918) (Afrikaans speaking) and South Africa (a correspondence university). In addition there are three university colleges for Bantu, at Fort Hare (1916), Ngoye (1959) and Turfloop (1959), and one for Coloureds—the University College of the Western Cape—at Bellville (1959). A university college for Indians was opened at Durban in 1961. In 1959 the department of Bantu education assumed responsibility for all education of Africans, whether primary, secondary, technical, teaching or university, other than education given by mission schools (which are subsidized and partly controlled except in the case of Roman Catholic schools).

The provinces administer all other primary and secondary education and teachers' training colleges (private schools excepted). Natal apart, local school committees are elected by parents, and elected and nominated district boards undertake general administration and (with the committees) select teachers. In Natal education is directly administered by the province. In all provinces schools are generally single medium (*i.e.*, either English speaking or Afrikaans speaking), and school selection depends on the language in which the child is more fluent. Attendance is compulsory and free for white children between 7 and 16 years of age, and there are separate government schools for Coloureds and Asians, where attendance is not compulsory but free up to certain secondary standards, varying with the province. There are extensive bursary facilities for higher education. In 1963 the ministry of education established a national advisory education council to foster a uniform national education policy.

8. Defense.—The South African Defense (Further Amendment) act, 1961, provides that every white citizen attaining the age of 18 years is subject to nine months' compulsory military training for one year and another period of three weeks during the following two years.

The South African defense force comprises the permanent force, active citizen force, rifle commandos, reserves, school cadets and auxiliary services (nonmilitary and nonwhite personnel). The permanent force consists of the army, air force and navy organized in 11 army commands, 3 air force groups and a naval base. It is staffed by regular personnel primarily concerned with training and administration. The active citizen force comprises personnel undergoing peace-time training, and the school cadets consist of boys between 12 and 17 years of age. (L. P. Gr.; L. H.; X.)

VII. THE ECONOMY

South Africa has passed through three phases in economic development. The first phase was the pioneer period of subsistence agriculture which lasted until about 1870, when the diamond fields were opened up. Gold discoveries followed, culminating in the development of the Witwatersrand gold field and giving a stimulus to agriculture and foreign exchange for the purchase of manufactured goods. With the outbreak of World War I and the inability to get supplies from overseas came the stimulus to manufacturing, ushering in the third phase, in which secondary industry became a greater producer of wealth than either agriculture or mining. This is reflected in the national income; between 1914 and 1953 manufacturing industries passed from eighth to first place among contributors to the national income.

The government's policy is to support private enterprise, with the minimum of interference. Governmental assistance is given in the fields of development, finance and export promotion. The National Resources Development council, set up in 1947, advises the minister of economic affairs in order to promote general development along sound lines. There are a large number of regional development associations formed to guide local development. Foreign investments by British, other European and U.S. companies have been heavy, especially since World War II, although in 1960 and 1961, because of racial unrest (see *History*, above), the South African economy suffered from the outflow of foreign capital, which was later stemmed.

A. Production

1. Agriculture.—The two forms of agriculture, cultivation and animal husbandry, have developed largely on a natural basis of climate, topography and soil. Cultivation has flourished where the ground is moderately flat, the rainfall adequate or irrigation practicable and the soils sufficiently fertile. Two major parts of the country fulfill these requirements—the lowland surface of southwestern Cape Province, with a winter rainfall exceeding 10 in., and the eastern part of the country, with a summer rainfall of more than 25 in. In southwestern Cape wheat and the vine, with deciduous fruit trees, are of special importance; in the eastern part maize (corn), sugar and subtropical fruits are leading products.

Stock farming, while common in all crop-growing areas, is found mainly in the moister parts too high or too steep for cultivation or in areas of grassland too dry for cultivation. Many of these stock farms rely on underground sources of water for their animals and through boreholes tap supplies in a variety of water-bearing rocks. Cattle are largely confined to the eastern summer-rainfall areas with long grass, sheep and goats to the drier areas where Karroo shrubs can withstand long droughts and provide adequate sustenance for small stock. Field crops and horticultural products are together of greater value than livestock products, though the value and volume of both sides has been steadily increasing. In the early 1960s about 1,500,000 people were engaged in farming, including 150,000 whites.

Crops.—Maize (corn), the staple diet of the African, is the most valuable of the republic's crops. Two kinds are generally grown: white dents and yellow flints. About 4,000,000 morgen (8,400,000 ac.) are sown to maize, the production fluctuating between about 30,000,000 bags (of 200 lb.) and 60,000,000. Excess of production over local requirements has necessitated exports involving considerable annual losses. This is one of the republic's problems since production is likely to increase as more hybrid seed becomes available. Wheat is grown mainly in southwestern Cape, and also in eastern Orange Free State and, under irrigation, in southern and central Transvaal. Annual production has varied from about 5,000,000 to 9,000,000 bags. As the Bantu are buying more wheaten bread, an annual import of wheat is generally necessary. Sugar cane growing is limited mainly to the east coastal area of Natal (including Zululand). The subtropical climate necessitates a 16- to 20-month reaping cycle. About 600,000 ac. are under cane, of which about 290,000 ac. are harvested each season. Production is about 1,000,000 tons of sugar from 10,000,000 tons of cane, about one-third of which is exported, but production was cut by about 20% after 1960 owing to limited export possibilities. Other important crops include tobacco, peanuts, potatoes and kafir (sorghum).

Fruit.—The gross value of the annual production of fresh and dried fruit is about R.70,000,000 (R.2 = £1 sterling). Deciduous tree fruits and grapes flourish best in southwestern Cape; citrus and subtropical fruits in the eastern part of the republic. The main deciduous tree fruits grown are (in order of quantity exported) apples, pears, plums, peaches and apricots. Exported grapes total approximately 25,000 short tons, and 3,000,000 gal. of wine and a smaller quantity of brandy are also exported. More than 80% of the citrus grown consists of Valencia and navel oranges; grapefruit, lemons and naartjies (tangerines) are grown in much smaller quantities.

Livestock.—The livestock population comprises mainly cattle, sheep and goats, with comparatively small numbers of pigs and donkeys and many fowls. Cattle number about 12,000,000; approximately 4,000,000 are in Cape Province (mainly in the eastern part), 3,500,000 in Transvaal, 2,500,000 in Natal and nearly 2,000,000 in Orange Free State. Of these about 7,000,000 belong to whites, 1,200,000 to Bantu on white farms and 3,800,000 to Bantu in the homelands. Sheep number approximately 38,000,000, of which nearly 25,000,000 are in Cape Province, nearly 9,000,000 in the Free State, over 3,000,000 in Transvaal and over 1,000,000 in Natal. About 28,000,000 of these are merinos, 1,200,000 karakuls and 4,000,000 nonwooled sheep. About 3,500,000 of the nonmerino sheep belong to Bantu. Goats number nearly 5,500,000, more than 3,500,000 being in Cape Province, 900,000 in Natal, 767,000 in Transvaal and 77,000 in the Free State. Of these about 870,000 are angora goats, 125,000 belonging to Bantu.

Cattle products include meat, fresh milk, butterfat, cheese milk and milk for condensing. Products from sheep and goats include not only meat and milk (from goats) but wool and mohair. Miscellaneous livestock products include ostrich feathers.

2. Forestry.—Out of the republic's total land area of slightly over 302,000,000 ac., natural forest covers less than 925,000 ac. There are more than 3,000 timber and wattle plantations varying from 25 to over 10,000 ac. Of the large plantations (5,000 ac. or more) most are in eastern Transvaal, Natal, southern Cape Province and Zululand. In the natural forests the most valued hard-

AUTHENTICATED NEWS

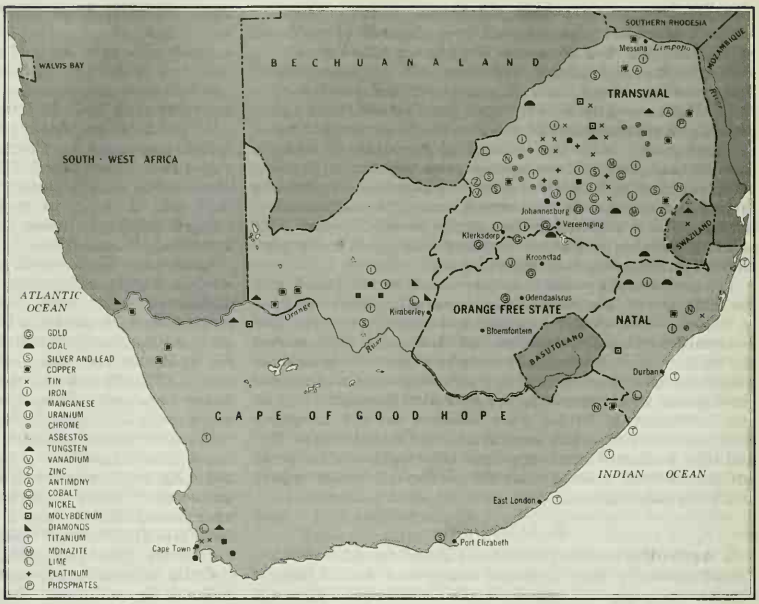
MINEWORKERS EXCAVATING A PUMP CHAMBER AT THE WESTERN DEEP LEVELS GOLD MINE, ONE OF THE DEEPEST WORKINGS ON THE CENTRAL RAND

woods are stinkwood and sneezewood. The softwoods include the yellowwoods, the Cape box and many others. The natural forest is carefully preserved by the state forestry department, and only small quantities of timber are cut each year. The chief broadleaf species include Eucalyptus trees, matchwood poplars and the tan wattles. An important factor in the development of exotic softwood plantations is the high natural wood increment which is one of the highest in the world (the average annual increment is of the order of 150 cu.ft. per acre, which may be compared with European and North American areas where it is between 17 and

35 cu.ft.). In some of the wetter afforested areas in South Africa the average increment is 250 cu.ft. per acre. South Africa's afforestation potential is estimated to be about 2,000,000 ac.

The national consumption of softwoods is about 130,000,000 cu.ft. of logs per annum. Of this about two-thirds is imported while one-third is produced locally. Since, however, pines take 30–40 years to mature, and well over half of the plantations have been planted since 1950, 75% of the logs delivered to the sawmills are immature. The local woods are therefore generally marketed at prices 15%–25% below those of imported timber, but mature South African timber compares favourably with the imported wood. The export of forest products includes wattle bark and bark extract. (*See* also *Vegetation*, above.)

3. Fisheries.—Marine fishing is carried on mainly off the west and south coasts. The west coastal waters with their low temperatures and abundant marine plankton abound in fish life and provide the largest catches. Off the west coast the most profitable types of fishing are inshore shoal fishing and offshore trawling. The two principal shoal fish are the pilchard and the maasbanker, with mackerel varying considerably in quantity. Trawling is the main source of fresh fish consumed locally. Stockfish, or cape hake, is by far the most important item in the trawled catch, though much smaller catches are taken of the 14 or 15 other varieties, of which the most valuable are kingklip, the Agulhas sole, panga and kabeljou. Most of the catch of line fisheries is of snoek, or sea pike, the remainder consisting mainly of geelbek (Cape Salmon), kabeljou, steenbras, redfish and sharks. Vitamin A is derived from the shark livers and there is an export of dried and salted shark to the far east. An important item in the inshore catch is the crawfish, commercially termed rock lobster. This is found mainly off the west coast and is exported largely to the United States in the form of "frozen tails" and canned lobster. There is a whaling industry which is based on Durban and Saldanha bay. Fin, sperm, sei, humpback and blue whales are caught for whale oil, sperm oil, whale meat and meat extract, and vitamin oil.



DEPARTMENT OF INFORMATION, SOUTH AFRICAN EMBASSY, LONDON

PRINCIPAL MINERAL DEPOSITS IN THE REPUBLIC OF SOUTH AFRICA

4. Minerals.—Mining provided in the early 1960s the second largest contribution (after manufactures) to the national income. Production of gold bullion exceeded 20,000,000 fine oz. annually and provided a revenue of more than R.500,000,000 (£250,000,000 or $700,000,000). Most of the output was produced by the mines of the Transvaal and Orange Free State Chamber of Mines which controls 55 producing mines in the five main gold fields: (1) the Witwatersrand; (2) the Far West Rand (sometimes known as the "West Wits line"); (3) the Klerksdorp field; (4) the Free State field; (5) the Kinross field, east of the Witwatersrand. All of these mine the "reefs" or gold-bearing conglomerates of the Witwatersrand geological system, which has surface outcrops on the Witwatersrand but is found only at depth in the other gold fields. On the Rand (*i.e.*, Witwatersrand) the main reef dips gently southward and is mined to a depth of over 11,000 ft.

Since 1952 uranium, which in the form of uraninite is associated with the gold in the reefs, has been produced in ever-increasing quantities. The republic's uranium ore reserves are estimated at 1,100,000,000 tons, of which the uranium content appears to be greater than the combined reserves of Canada and the United States. Nineteen of the older mines, though producing gold at a loss, make a substantial profit on uranium.

Employment on the gold mines includes about 51,000 whites and 400,000 Bantu drawn from many parts of southern Africa. The place of gold in the economy of the republic is well established, but perhaps its most important effect is as a stabilizer in the fluctuations of foreign trade (see *Foreign Trade*, below).

Of other precious minerals platinum is produced as the market demands; the annual output is valued at about R.16,000,000. Small quantities of silver and osmiridium are found in the gold ore. Diamonds to the value of about R.33,000,000 are derived from mines and alluvial deposits. In the base mineral group coal from the Ecca beds of the Karroo system is mined in southern Transvaal and Natal; in the early 1960s the output exceeded 40,000,000 tons, of which nearly 1,000,000 tons were exported. Iron ore production, mainly from Thabazimbi in Transvaal, Postmasburg in northern Cape Province and from northern Natal, fluctuated between 3,000,000 and 4,000,000 tons, the bulk of the ore being a hard, high-grade hematite with low phosphorus and sulfur content. Other minerals include asbestos, copper, manganese, chrome ore and tin. (*See* map of mineral resources.)

5. Power.—In 1922 the South African government established the Electricity Supply commission (E.S.C.O.M.), a nonprofit-earning body, to control and supply power for the industrial development of South Africa. The commission operates in seven licensed areas of supply: Rand and Orange Free State, Cape Western, Natal Southern, Natal Central, Border (*i.e.*, eastern Cape Province), the Eastern Transvaal and the Cape Northern areas. The total power output of the commission in the early 1960s was in excess of 18,000,000,000 kw.hr., representing about 70% of the country's output, the rest being provided by municipalities and private undertakings (mines and industries). The supply to rural settlements and farms has steadily increased. The power stations, with two small exceptions, are all thermal, coal-burning types.

6. Industries.—Government-sponsored industrial activities include the South African Coal, Oil and Gas corporation (S.A.S.O.L.), brought into being for the extraction of oil from

coal; the Iron and Steel Industrial Corporation, Ltd. (I.S.C.O.R.), which owns furnaces and mills and provides about two-thirds of the country's steel requirements; Klipfontein Organic Products (K.O.P.), which produces chemicals and insecticides; and the Phosphate Development corporation, producing fertilizers. Industrial research is sponsored by the South African Council for Scientific and Industrial Research (C.S.I.R.), a corporate body set up to advise the minister of economic affairs on all questions of scientific and technological methods affecting the utilization of the natural resources of the republic. The C.S.I.R. controls the bureau of standards and a number of national laboratories including those for research in chemistry, physics, telecommunications, building, mechanical engineering, road construction, personnel administration, nutrition and industries. An important function of the council is the promotion of co-operative research by the formation of industrial research associations which are independent bodies subsidized by the council.

Manufacturing is by far the largest item (about 85%) of all private secondary industry, while building construction, electricity, gas and steam, and services make up the remainder. In the manufacturing sector four areas are outstanding in importance: southern Transvaal (including Pretoria and Vereeniging-Vanderbijlpark), western Cape, Durban-Pinetown, and Port Elizabeth-Uitenhage.

Of the total gross output of private manufacturing, the groups producing most were: food, metal products, clothing and footwear, basic metal industries, chemicals and chemical products, and nonmetallic mineral products. The total number of employees in manufacturing industries in the early 1960s exceeded 690,000, comprising about 30% whites, 50% Bantu and 20% Asians and Coloureds.

B. Trade and Finance

1. Foreign Trade.—A characteristic feature of South Africa's trade has been a constant disequilibrium between the import and export of merchandise, excluding gold. In 1925, for example, the excess of imports over exports, excluding gold, was £12,800,000; in 1930, £17,100,000; in 1935 £44,300,000. In each year the deficit was changed into a surplus by exports of gold bullion amounting respectively to £34,300,000, £35,900,000 and £71,400,000. Gold is therefore of increasing importance in the balance of payments.

Although South Africa is no longer a member of the commonwealth, preferential tariff rates are maintained with Great Britain under a bilateral agreement and in the framework of the General Agreement on Tariffs and Trade (G.A.T.T.).

Imports.—In the early 1960s by far the most valuable item was the class "metals, metal manufactures, machinery and vehicles." In this class the chief commodities were automobiles and trucks, railway locomotives and rolling stock, tractors and parts, and airplanes. These goods were supplied by the U.K. and other commonwealth countries; of the other suppliers the U.S. and West Germany far exceeded the other trade partners (Japan, Italy, the Netherlands, Rhodesia and Nyasaland, France, Belgium and Ceylon). Second in value was the class "fibres, yarns, textiles and apparel." Third in value came "oils, waxes, resins, paints and varnishes," with crude oil, paraffin and motor spirit (gasoline) from Iran as the main items. Next in value were "drugs, chemicals and fertilizers," "foodstuffs," "minerals, earthenware, glassware and cement," "wood, cane and wicker manufactures."

Exports.—In the early 1960s the class with the highest value (excluding gold) was "metals, metal manufactures, etc." Atomic energy material covered almost half of this amount. Other minerals were of comparatively small value: copper, lead ore concentrates, manganese ore, chrome ore. Manufactured metal goods included mining machinery, ferroalloys, wrought iron and manufactured iron and steel. Small items deriving from gold mining were silver and osmiridium. Foodstuffs came second, the chief items being fresh and preserved fruits (citrus exports go to more than 35 countries), maizé (corn) and maize meal, fish (fresh, dried, frozen and fish meal) and sugar. The class "animals, agricultural and pastoral products" was next in value. The most important commodities in this class were wool, hides and skins, fish meal, wattle and wattle bark extract and mohair. Minerals, earthenware, and glassware came next (diamonds, asbestos, etc.), followed by jewelry (cut and polished diamonds), then by oils, waxes, etc., and other commodities of small value.

The chief trading partners receiving exports and re-exports were the United Kingdom, Rhodesia and Nyasaland, the United States, Japan, West Germany, Italy, Israel, Belgium, France and the Netherlands.

2. Banking and Currency.—In South Africa the ministry of finance determines the country's general financial policy, which is administered by the central bank and operated mainly by the commercial banks and other specialized financial institutions. The central bank is the South African Reserve bank, which came into existence in 1921 during the monetary chaos that followed World War I. The bank is a privately owned joint-stock company with limited liability; its governor and deputy governor and 3 other directors on the board of 11 are appointed by the government. The bank has the sole right of note issue in the republic, but its notes are not redeemable in gold.

Under the terms of the South African Reserve Bank act of 1944 the bank was required to maintain in gold coin and bullion a reserve of at least 30% of the aggregate amount of its note issue and of its other liabilities to the public, but by an amendment in 1948 the minimum reserve ratio of the bank was reduced to 25%.

The commercial banks are required to hold on deposit with the Reserve bank balances amounting to not less than 10% of their demand liabilities payable in the republic and 3% of their time liabilities. The Reserve bank is, however, empowered to require the commercial banks to maintain, when the national interest demands it, supplementary reserves not exceeding 10% of their total liabilities to the public. The Reserve bank found it necessary in the national interest to exercise these powers in 1958. The Reserve bank may alter the supplementary reserves to 6% to ease a strain on the banking system. Registered banking institutions include commercial banks, people's banks, loan banks and deposit-receiving institutions.

To promote agricultural development the Land and Agricultural Bank of South Africa was established in 1912 as a nonprofit-making institution. To promote an active money market the National Finance Corporation of South Africa was established in 1949. A close co-operation with the Reserve bank is provided for, and the corporation may not be placed in liquidation except by an act of parliament.

Currency.—The South African coinage was decimalized in Feb. 1961 when the rand (issued as two gold coins of R.1 and R.2) became the new monetary unit equal to the former half-pound (10 shillings or $1.40). The new silver coins were 50 cents (crown), 20 cents (florin), 10 cents (shilling), 5 cents (sixpence), 2½ cents (threepence); and the bronze coins comprising the cent (twelve-tenths of a penny), half-cent and quarter-cent. Although no longer a member of the commonwealth South Africa remains within the sterling area.

3. National Finance.—Exchequer receipts fluctuated in South Africa from £147,400,000 (equivalent to R.294,800,000) in 1950 to more than R.730,000,000 by the early 1960s; in the same period exchequer issues increased from R.293,600,000 to more than R.720,000,000. On the revenue account (in excess of R.600,000,000) the receipts were derived from customs and excise, post office and inland revenue. Of these the inland revenue amount was by far the largest, with income tax the highest source of revenue, followed by departmental receipts, interest and dividends, transfer duty and nonresident shareholders' tax.

Parallel expenditure from the revenue account included, as its main items, provincial administration, social welfare and pensions, posts and telegraphs, defense, public debt and agriculture.

In the decade following 1953 the net national income increased from R.2,854,800,000 to around R.4,600,000,000. The main sources were manufacturing, mining, agriculture, forestry and fishing, and transportation.

The Bantu homelands five-year plan (1961–66) called for an

estimated expenditure of R.114,000,000 a year.

C. Transport and Communications

1. Roads.—In South Africa the making and upkeep of roads is the responsibility of provincial administrations. In each of the provinces four classes of roads are recognized: national, main, district and minor roads. Provincial main roads are 80 ft. wide and, in general, are of the all-weather type, with gravel or bituminous surface. District roads, 60 ft. wide, are usually of the fair-weather type, with gravel surface. The minor roads vary from almost veld tracks to ordinary farm roads. The extent of such road types is of the order of 3,000 mi. of blacktop (bituminous, asphalt, tarmac, etc.), 10,000 mi. of gravel-surfaced and 80,000 mi. of earth roads.

In 1935 the National Roads act created a National Road board for the administration of national roads under the responsibility of the central government. The board's program, announced in 1936, provided for the construction, reconstruction and repair and maintenance of about 5,120 mi. of roads, declared "national." These were specified as bituminous surfaced with a minimum width of 100 ft. Fifteen routes were designated, their function being to link together the larger centres and to facilitate communication with adjoining territories. Thus route no. 1 links Beitbridge (over the Limpopo river on the Rhodesian border) with George on the south coast, passing through Pretoria, Johannesburg and Bloemfontein; route no. 2 links Cape Town with Durban, passing through Mossel Bay, Port Elizabeth and East London; route no. 3 links Johannesburg with Durban; route no. 4 links Pretoria with Komatipoort on the Mozambique border. The remaining routes interconnect these four routes or provide branch roads or extensions to other important centres.

2. Railways.—With few exceptions the railways and harbours of South Africa and South West Africa are the property of the government, the control and management being under the direction of the South African Railways and Harbours administration, responsible to the minister of transport. Of the republic's 12,000 mi. of railway lines 5,000 mi. are main lines of 3-ft.-6-in. gauge, and nearly 7,000 mi. are branch lines, of which less than 500 mi. are of 2-ft. gauge. Connected with the railways are the South African Railways' auxiliary road motor services, which cover almost 30,000 route miles.

Electrification of South Africa's railways has taken place mainly in industrial areas as, for example, southern Transvaal, where, in addition to electrification on the Witwatersrand and between Johannesburg and Pretoria, the lines from the coal mines of the Witbank area were, in the 1960s, connected by electrified lines with Kroonstad and the Free State gold field. In Natal the main line has been completely electrified from Durban to the Transvaal border, and in southwestern Cape the line to the interior has been electrified for 256 route mi. to Touws River. The 179-mi. stretch between Touws River and Beaufort West is partly electrified. Diesel locomotives are replacing steam chiefly for traction in the drier parts of the country and in South West Africa.

3. Shipping.—There were about 150 ships registered in the mercantile marine in the early 1960s with a registered gross tonnage of about 250,000. The total number of ships arriving at South and South West African ports in a year at that time was nearly 14,000, including coasters. In respect of cargo Durban is the first port in the republic and in Africa, and is followed in South Africa by Cape Town, Port Elizabeth and East London. Ship arrival and passenger transits, however, are greater at Cape Town than at Durban.

A marked feature of the shipping in South Africa is the large number of foreign cargo- and passenger-carrying liners calling regularly at ports either as destination or in transit to other parts of the world. There is thus generally ample transport for South Africa's overseas imports and exports. Since, however, precooling is essential for the country's important fruit exports the large precooling stores at Cape Town and Port Elizabeth have special significance. An important feature of Cape Town harbour is the Sturrock graving dock, which is able to accommodate even the largest oil tankers.

4. Air Transport.—South African airways is included in the South African Railways and Harbours organization and is the only operator with regular services. There are more than 270 licensed civil aerodromes in South Africa. Of these, however, Jan Smuts airport (Johannesburg) is the only one of international standard; three others are of national standard: Louis Botha airport (Durban), D. F. Malan airport (Cape Town) and J. B. M. Hertzog airport (Bloemfontein). Services are classified as domestic, regional (Johannesburg-Rhodesia-Mozambique) and international (Johannesburg-London, Johannesburg-Australia).

5. Postal Services and Telecommunications.—These are maintained at about 3,300 post offices, 1,500 of which are also money-order offices. The post office has a savings bank and also issues listeners' radio licences.

Included in telephone services are more than 400,000 farm connections. Besides internal telephones there are radiotelephone services to foreign countries and ships at sea. A radiotelegraph and telephone system between South Africa and South West Africa was introduced in 1960. There were more than 650 departmental teleprinters in use, and country offices equipped for teleprinter working totaled over 150. Nine radioteleprinter subscriber circuits were in operation to overseas countries, while international telex service is available via London to 30 overseas countries.

Domestic radio transmission is controlled by the South African Broadcasting corporation, a public utility established in 1936, which is responsible to the minister of posts and telegraphs and derives its revenue from listeners' licences and from its commercial service (Springbok radio). There are 12 broadcasting stations transmitting three national programs, English, Afrikaans and commercial. Bantu services are broadcast on the first two of these.

See Cape of Good Hope; Natal; Orange Free State; Transvaal; *see* also references under "South Africa, Republic of" in the Index volume. (J. H. Wn.; X.)

Bibliography.—*Physical Geography:* Weather Bureau, *Climate of South Africa,* parts i–iv (1957); *South African Geographical Journal* (annual, 1917 *et seq.*); publications of the Republic Geological Survey; L. C. King, *South African Scenery,* 2nd ed. (1951); J. H. Wellington, *Southern Africa,* vol. i (1955); R. S. Adamson, *The Vegetation of South Africa* (1938); E. Palmer and N. Pitman, *Trees of South Africa* (1961); J. R. Ellerman *et al., South African Mammals* (1953); A. Roberts, *Birds of South Africa,* rev. ed. (1957); W. Rose, *Reptiles and Amphibians of Southern Africa* (1950); J. L. B. Smith, *The Sea Fishes of Southern Africa* (1950); S. H. Skaife, *African Insect Life.*

Economy: Government publications: *Monthly Bulletin of Statistics; Official Yearbook of the Republic of South Africa; Farming in South Africa* (monthly); *Commerce and Industry* (monthly); *Statistics of Production: Agriculture, Industries; Monthly Abstract of Trade Statistics; Trade and Shipping of South Africa and the Territory of South-West Africa* (annual); N. N. Franklin, *Economics in South Africa,* 2nd ed. (1954); M. Cole, *South Africa* (1961); J. H. Wellington, *Southern Africa,* vol. ii (1960); L. P. Green and T. J. D. Fair, *Development in Africa: a Study in Regional Analysis* (1962). (J. H. Wn.)

The People: I. Schapera, *Western Civilization and the Natives of South Africa* (1934), *Government and Politics in Tribal Societies* (1956), *The Khoisan Peoples of South Africa* (1930), *The Tswana* (1953), (ed.) *The Bantu-Speaking Tribes of South Africa* (1937); L. Marquand, *The Peoples and Policies of South Africa,* 3rd ed. (1962); E. M. Thomas, *The Harmless People* (1959); N. J. van Warmelo, *Preliminary Survey of the Bantu Tribes of South Africa* (1935); E. H. Ashton, *The Basuto* (1952); A. T. Bryant, *Olden Times in Zululand and Natal* (1929); M. Hunter, *Reaction to Conquest* (1936); E. J. Krige, *The Social System of the Zulus* (1936); E. J. and J. D. Krige, *The Realm of a Rain-Queen* (1943); H. Kuper, *An African Aristocracy* (1947), *The Swazi* (1953), *Indian People in Natal* (1960); V. G. J. Sheddick, *The Southern Sotho* (1953); J. H. Soga, *The Ama-Xosa* (1932); H. A. Stayt, *The BaVenda* (1931); B. G. M. Sundkler, *Bantu Prophets in South Africa* (1948); J. S. Marais, *The Cape Coloured People 1652–1937* (1939); S. C. Patterson, *Colour and Culture in South Africa* (1953); W. D. Hammond—Tooke, *Bhaca Society* (1962). (J. F. M. M.)

General Historical Surveys: C. W. de Kiewiet, *History of South Africa, Social and Economic* (1941); *The Cambridge History of the British Empire,* vol. viii (1936); A. M. Keppel-Jones, *South Africa* (1949); A. J. H. van der Walt *et al.* (eds.), *Geskiedenis van Suid Afrika,* 2 vol. (1951); J. H. Hofmeyr, *South Africa,* 2nd ed. by J. P. Cope (1952); L. Marquard, *The Story of South Africa* (1955); E. A. Walker, *History of Southern Africa,* 3rd ed. (1957), *Historical Atlas of South Africa* (1922).

Special Subjects: J. du Plessis, *A History of Christian Missions in South Africa* (1911); W. M. Macmillan, *The Cape Colour Question*

(1927), *Bantu, Boer and Briton* (1929); C. W. de Kiewiet, *British Colonial Policy and the South African Republics, 1848–72* (1929), *The Imperial Factor in South Africa* (1937); R. I. Lovell, *The Struggle for South Africa, 1875–99* (1934); I. D. MacCrone, *Race Attitudes in South Africa* (1937); S. H. Frankel, *Capital Investment in Africa* (1938); P. J. van der Merwe, *Die Trekboer in die Geskiedenis van die Kaapkolonie* (1938); J. S. Marais, *Maynier and the First Boer Republic* (1944), *The Fall of Kruger's Republic* (1961); S. T. van der Horst, *Native Labour in South Africa* (1942); M. Roberts and A. E. G. Trollip, *The South African Opposition, 1939–45* (1947); E. A. Walker, *The Great Trek*, 3rd ed. (1948); M. W. Spilbaus, *The First South Africans* (1949); J. van der Poel, *The Jameson Raid* (1952); G. B. Pyrah, *Imperial Policy and South Africa, 1902–10* (1955); S. C. Patterson, *The Last Trek: a Study of the Boer People and the Afrikaner Nation* (1957); Gwendolen M. Carter, *The Politics of Inequality* (1958); L. M. Thompson, *The Unification of South Africa, 1902–10* (1960); F. A. van Jaarsveld, *The Awakening of Afrikaner Nationalism* (1961). Monographs and articles are published in the *Archives Yearbook for South African History* (1938–), the *South African Journal of Economics* (1927) and *The Round Table* (1910 *et seq.*) and in the publications of the South African Institute of Race Relations and the South African Bureau of Racial Affairs.

Biographies: J. H. Hofmeyr and F. W. Reitz, *The Life of Jan Hendrik Hofmeyr* (1913); B. Williams, *Cecil Rhodes*, new ed. (1938); I. D. Colvin, *The Life of Jameson*, 2 vol. (1922); E. A. Walker, *Lord de Villiers and His Times* (1925), *W. P. Schreiner* (1937); F. V. Engelenburg, *General Louis Botha* (1929); C. M. van den Heever, *General J. B. M. Hertzog* (1946); J. Rose Innes, *Autobiography* (1949); E. W. Smith, *The Life and Times of Daniel Lindley* (1949); J. C. Smuts, *Jan Christiaan Smuts* (1952); W. K. Hancock, *Smuts: the Sanguine Years, 1870–1919* (1962).

Collected Historical Records: J. Bird, *The Annals of Natal, 1495–1845*, 2 vol. (1888); G. M. Theal, *Basutoland Records, 1833–68*, 3 vol. (1883), *Records of the Cape Colony, 1793–1831*, 36 vol. (1897–1905), *Records of South-Eastern Africa*, 9 vol. (1898–1903); G. W. Eybers, *Select Constitutional Documents Illustrating South African History, 1795–1910* (1918); A. P. Newton, *Select Documents Relating to the Unification of South Africa*, 2 vol. (1924); C. Headlam, *The Milner Papers, South Africa, 1897–1905*, 2 vol. (1931–33). The publications of the Van Riebeeck Society (1918 *et seq.*) comprise original journals, reports, etc.

Official publications and reports of government departments on special topics are published at Pretoria by the government. *See* also the government publication *Union Statistics for Fifty Years, 1910–1960.*

Bibliographies: The most accessible detailed bibliography is in the *Cambridge History of the British Empire*, vol. viii (1936). *See* also S. Mendelssohn, *South African Bibliography*, 2 vol. (1910); G. M. Theal, *Catalogue of Books and Pamphlets Relating to Africa South of the Zambesi* (1912); I. Schapera, *Select Bibliography of South African Native Life and Problems* (1941). For recent publications, *see* the *Quarterly Bulletin of the South African Public Library* (1946–).

(L. M. T.)

SOUTH AFRICAN LITERATURE. The earliest writings about South Africa date back several centuries, but South African literature is of more recent origin. After a difficult start in its development from the Dutch language, Afrikaans developed a strong literature, and a review of South African literature must be divided into that written in English and that in Afrikaans.

ENGLISH

Various reviewers have held that English literature in South Africa has followed the main lines of development in the rest of the English-speaking countries. This view must be accepted with reserve; it is true that many early writers were British and that most younger writers left the country to work as part of other traditions, but South African literature proper has, none the less, individual characteristics.

In poetry this individuality is revealed by an indigenous subject matter and the use of imagery and diction that are characteristically South African. The first poet of note was Thomas Pringle (*Ephemerides*, 1828; *African Sketches*, 1834). After a considerable lapse there were William Charles Scully (*Wreck of the Grosvenor*, 1886; *Poems*, 1892; *By Veldt and Kopje*, 1907; *Voices of Africa*, 1944), Francis Carey Slater (*Footpaths Thro' the Veld*, 1905; *The Sunburnt South*, 1908; *Settlers and Sunbirds*, 1919; *The Karroo*, 1924; *Veld Patriarch*, 1949), Arthur Shearly Cripps (*Lyra Evangelistica*, 1909; *Pilgrimage of Grace*, 1912; *Lake and War*, 1917) and Charles Murray (*Hamewith*, 1909; *A Sough of War*, 1917; *In the Country Places*, 1920). These are poets of the older school and their verse shows no particular talent or individuality other than this "South African" quality. Cullen Gouldsbury (*Rhodesian Rhymes*, 1909; *Songs out of Exile*, 1912; *From the Outposts*, 1914), a poet and novelist of great promise, died prematurely.

Roy Campbell wrote notable South African poetry in *The Flaming Terrapin* (1924), *The Wayzgoose* (1928), *Adamastor* (1930), *Flowering Reeds* (1933) and many others. William Plomer (*Family Tree*, 1928, and others) also became prominent. Later poets include Frank Prince, R. M. Titlestad, Anthony Delius and Guy Butler.

Anthologies include *Treasury of South African Poetry* (edited by E. H. Crouch, 1907 and 1909), *Centenary Book of South African verse* (edited by Francis Carey Slater, 1925 and 1945) and *South African Poetry* (edited by Roy MacNab and Charles Gulston, 1948).

The South African novel also has characteristics which differentiate it from other novels in English. Few of the important novels are free from a topic peculiar to South Africa. Notable exceptions are Olive Schreiner's *Story of an African Farm* (1883); Sarah Gertrude Millin's *Mary Glenn* (1925), *Sons of Mrs. Aab* (1931) and *What Hath a Man?* (1938); Ethelreda Lewis' *Harp* (1924); and Elizabeth Charlotte Webster's *Expiring Frog* (1946). Olive Schreiner's *Story of an African Farm* has the distinction of being the first novel of note to come from Africa and, despite some failings, is a work of merit. Her later works, *Trooper Peter Halket of Mashonaland*, *From Man to Man* and *Undine*, fell below its standard.

The most popular subject for novels is the trials of the "mixed" or coloured people, who have been developed by centuries of miscegenation and are isolated from both Africans and Europeans. Sarah Gertrude Millin (*God's Stepchildren*, 1924; *King of the Bastards*, 1950; *Burning Man*, 1952) became the major writer on this theme, although William Plomer in his short stories *I Speak of Africa* (1927), Laurens van der Post in *In a Province* (1934) and Perceval Gibbon in *Souls in Bondage* (1904) wrote with insight. Peter Abrahams, a coloured writer, produced some impressive work (*Dark Testament*, 1942; *Song of the City*, 1945; *Mine Boy*, 1946; *Return to Goli*, 1953).

The native formed the subject of hundreds of adventure stories and "ethnographical" novels, but the best work was done in books showing the difficulties of the native people in society. Cullen Gouldsbury (*God's Outpost*, 1907; *Tree of Bitter Fruit*, 1910), William Plomer (*Turbott Wolfe*, 1925), Sarah Gertrude Millin (*Coming of the Lord*, 1928; *Herr Wilchdoctor*, 1941), William Charles Scully (*Daniel Vananda*, 1923), Alan Paton (*Cry, the Beloved Country*, 1949) and J. Grenfell Williams and Henry May (*I am Black*, 1936) produced a unique and sometimes powerful body of work.

Other novels include historical works by Stuart Cloete (*Turning Wheels*, 1937; *Watch for the Dawn*, 1939; *Hill of Doves*, 1941) and Dorothea Fairbridge (*That Which Hath Been*, 1910); A. A. Smith's *Trader Horn* (1927–28); Percy Fitzpatrick's charming dog story *Jock of the Bushveld* (1907); Horace Rose's *Golden Glory* (1915); and Daphne Muir's *Virtuous Woman* (1929).

Pauline Smith's *Little Karroo* (1925) must head any list of South African short stories. William Westrup's *Old McBein* (1937) and *Trek Chain* (1929) and Ernest Glanville's *Kloof Yarns* (1896), *Tales From the Veld* (1897) and *Claw and Fang* (1923) are outstanding for their description of animal and veld. Among later books, Herman Charles Bosman's *Mafeking Road* (1947) and Nadine Gordimer's *Face to Face* (1949) revealed remarkable gifts.

Travel books were the earliest books written about South Africa, and the works of Thomas Baines, John Barrow, William John Burchell and David Livingstone became well known. In the popular travel book by Laurens van der Post, *Venture to the Interior* (1952), the scene is Nyasaland, but it seems that the writer's understanding of the country has its origin in his native soil. In biography an extensive literature gathered round the names of Cecil Rhodes, L. S. Jameson, Louis Botha and J. C. Smuts, the best work being done by Sir Thomas Fuller, Sarah Gertrude Millin, Ian Colvin, Earl Buxton, J. Crafford and H. C. Armstrong. The letters of Lady Anne Barnard (*South Africa a Century Ago*, 1901)

give a vivid account of the years 1797–1801 at the Cape.

No significant drama in English had been produced in South Africa by mid-20th century. (J. P. L. S.)

AFRIKAANS

Afrikaans, one of the youngest members of the Teutonic group, is the language used by the descendants of Dutch, German and French colonists who settled in South Africa after 1652 and before the British occupation in 1806. The 17th-century dialects of the Netherlandish province (South) Holland form the linguistic basis of Afrikaans; but, although the Afrikaans vocabulary is even today almost identical with the original stock, the grammatical structure has been much simplified. The Afrikaans noun has no grammatical gender, has lost all inflectional forms but the singular and plural, uses one form of the verb for all persons, forms its tenses with modal words and has obliterated the difference between the weak and strong declension. Afrikaans structure is, therefore, even more analytical than that of English. This development is mainly a result of an early lack of the conserving influence of schools and literature, new surroundings and occupations and the incorporation of German and French immigrants. There is every reason to believe that Afrikaans reached its modern form not much later than 1750.

Until the second decade of the 20th century Dutch was still used in South Africa for literary and official purposes. Toward the middle of the 19th century Afrikaans was used in letters in some newspapers and in a few propaganda pamphlets. In 1875 a society, Die Genootskap vir Regte Afrikaners, was founded to translate the Bible into Afrikaans and to foster the use of the language officially. The society founded a magazine and newspaper and started a printing press, but of the literary production of the period 1860–1900 only J. L. Cachet's didactic stories *Die Sewe Duiwels* and S. J. du Toit's historical novel *Die Koningin van Skeba* are still read.

The South African War saw the end of the society, but it also saw the birth of an entirely new spirit in the movement for the recognition of Afrikaans. The old phonetic orthography was discarded in favour of a spelling more closely related to Dutch; new writers and scholars, better equipped for the task, took the lead; new journals were founded. Die Afrikaanse Akademie was founded in 1909. As a result of all these activities Afrikaans was adopted as a medium of instruction in schools (1914), university professorships in Afrikaans were instituted, and by 1924 all the Afrikaans churches had adopted Afrikaans as their official language. In 1925 Afrikaans was officially recognized by the entire country. The complete Afrikaans version of the Bible was published in 1933. The first volume of the Afrikaans dictionary, begun in 1925, was completed in 1950. After 1935 an ever-increasing number of Afrikaans journals, magazines and newspapers was established.

A literature distinctly Afrikaans in colour and sentiment grew out of the experience of the South African War. The first group of poets of real merit whose poems arose from this experience were Eugène N. Marais, Jan F. Celliers (*Die Vlakte*, 1908; *Martjie*, 1911), J. D. du Toit (*Totius—By die Monument*, 1908) and C. L. Leipoldt (*Oom Gert Vertel E. A. Gedigte*, 1911). Largely because of the national spirit of its poetry Afrikaans came to be recognized as the cultural language of the country within an extremely short space of time.

Advances in the field of prose were soon made. G. S. Preller and C. J. Langenhoven, both keen supporters of Afrikaans, D. F. Malherbe and Jochem van Bruggen were still largely concentrating on the national experience during the war, but they soon turned their attention to the realistic novel and made a sound contribution toward the beginning of an Afrikaans prose style (Malherbe, *Die Meulenaar*, 1916; Bruggen, *Ampie*, a trilogy, 1924–42). A. A. Pienaar's masterly novelettes on wild life in the African jungle, *Uit Oerwoud en Vlakte* (1921), belong to this period.

Meanwhile a new generation of poets had arisen who were mainly concerned with the impact of modern science and technology on a mainly agrarian and patriarchal culture and religion. Consequently the stress was shifted from national to personal experience. Toon van den Heever especially, in his *Gedigte* (1919), succeeded in embodying his doubt and his rebellion against the biblical tenets of his people in his poems. Although it is characterized by differences of outlook, style and interests, the poetry of the following is actually part of this period: T. J. Haarhoff, *Die Liefde van Catullus* (1933); S. J. du Toit, *Tussen die Dae* (1947); A. G. Visser, *Gedigte* (1925) and *Die Purper Iris* (1930).

About 1925 a new group emerged to dominate the literary scene for the next decade. The most marked advance was made in the field of the novel and the short story. The realistic and the social novel was much favoured. The dominating figure from that time was C. M. van den Heever (*Op die Plaas*, 1927; *Droogte*, 1929; *Somer*, 1935; *Laat Vrugte*, 1939; *Die Held*, 1948; and a biography, *Generaal J. B. M. Hertzog*, 1943). Other prominent authors of the period were C. H. Kühn (*Toiings*, a trilogy, 1934–44; *Klaradynstraat*, 1947), J. van Melle (*En Ek is Nog Hy*, 1942), G. C. and S. B. Hobson (*Kees van die Kalaharie*, 1929; *Buks*, 1933), Holmer Johanssen (*Die Onterfdes*, 1944), P. J. Schoeman (*Op die Groot Spore*, 1942) and G. H. Franz (*Moeder Poulin*, 1946). Those members of this group who were also poets—Van den Heever (*Deining*, 1932; *Aardse Vlam*, 1938) and I. D. du Plessis (*Stryd*, 1935; *Kwatryne*, 1941)—might be regarded as the transition to the next generation, the poets of 1935.

In the poetry of this latter group national experience is at last entirely ignored and the poems centre on the feelings of the individual. The old rhyming verse is discarded in favour of modern blank verse. The new form of poetry passed its zenith soon after 1950, but until then it dominated the field. The outstanding members of the group were N. P. van Wyk Louw (*Alleenspraak*, 1935; *Die Dieper Reg*, 1939; *Raka*, 1940), D. J. Opperman (*Heilige Beeste*, 1945; *Negester Oor Nineve*, 1947; *Engel uit die Klip*, 1951) and W. E. G. Louw (*Die Ryke Dwaas*, 1934). The younger poets differed from the above in their lyrical excellence. They include Uys Krige (*Oorlogsgedigte*, 1942; *Hart Sonder Hawe*, 1949), Elizabeth Eybers (*Belydenis in die Skemering*, 1936; *Die Stil Avontuur*, 1939), S. J. Pretoria (*Vonke*, 1943; *Inkeer*, 1940) and Olga Kirsch (*Mure van die Hart*, 1948).

There are few Afrikaans plays, but dramatists of a high literary standard include C. L. Leipoldt (*Die Heks*, 1923), J. F. W. Grosskopf (*Drie Eenbedrywe*, 1926; *Die Klipdolk E. A. Kort–Spele*, 1941), H. A. Fagan (*Die Ouderling*, 1934) and Uys Krige. A younger dramatist than these, W. A. de Klerk, came to the fore with his play *Die Jaar van die Vuuros;* it received the national prize for drama during the Riebeeck centenary of 1952 as well as the prize of the Suid-Afrikaanse Akademie, the award in both cases being shared with G. Beukes's *As Ons Eers Getroud Is.*

See C. M. van den Heever and P. de V. Pienaar (eds.), *Kultuurgeskiedenis van die Afrikaner*, 3 vol., a symposium with exhaustive lists of books (Cape Town, 1945–50). (A. J. Ce.)

SOUTH AFRICAN WAR, also known as the ANGLO-BOER WAR or BOER WAR, was fought between Great Britain and the two Boer republics, the South African Republic (Transvaal) and the Orange Free State, between October 1899 and May 1902.

Causes.—The war had its origins in the relations that had developed between Britain and the South African Republic. Britain, especially after 1895, when Joseph Chamberlain became colonial secretary, aimed at establishing its supremacy over southern Africa and at preventing the encroachment of its colonial rivals, France and Germany, while Paul Kruger (*q.v.*), president of the South African Republic, was determined to maintain the political and economic independence of his Boer republic. Several factors helped build up the tension that ultimately led to war. The first of these was the issue of British suzerainty. In 1877 Britain annexed the Transvaal, but in consequence of a Boer rebellion (December 1880) and the defeat of British troops at Majuba (February 1881), by the Pretoria Convention (August 1881) the Transvaal was granted self-government subject to the suzerainty of Queen Victoria; three years later, by the London Convention (February 1884), it received a fuller measure of independence (*see* TRANSVAAL: *History*). But Britain's retention of control over the republic's treaties, other than with the Orange Free State, and its claim at a later stage that it had not abandoned suzerainty were a

source of resentment to Kruger and his burghers (enfranchised citizens).

The second factor was the influx of large numbers of immigrants (known as *uitlanders*), who came into the republic from Britain and elsewhere following the discovery of gold on the Witwatersrand in 1886. Although they paid the bulk of the republic's taxation, they were denied any effective share in the government; the English language was banned in the law courts and schools; and juries consisted only of burghers. The *uitlanders* also complained of the government's grant of monopolies in dynamite and other commodities vital to the mining industry and of corruption on the part of the officials. But Kruger, fearing that if given political rights the *uitlanders* might gain control of the Boer republic, rejected their demands for the franchise and for redress of their grievances.

The third factor was the conflict between Kruger's ideal of independence and isolation and the attempts of Cecil John Rhodes (*q.v.*), head of the British South Africa Company (which had control over the country north of the Limpopo River) and prime minister of the Cape Colony, to bring about an economic federation of the whole of southern Africa that would lead to a political federation under the British flag. The British occupation between 1885 and 1895 of territory to the north, west, and east of the republic, which encircled it and cut it off from access to the sea, aroused the fears of Kruger. He resisted Rhodes's attempts to include the republic in a customs union, and to ensure economic independence from the British colonies to the south he gave Hollander interests a concession to build a railway line from Pretoria to Delagoa Bay in Portuguese East Africa. A fierce rivalry in railway rates developed between the Netherlands Railway Company, which was responsible for this line, and the Cape authorities. War seemed imminent when in October 1895 Kruger closed the Vaal River drifts (fords) that the Cape authorities had been using to send goods into the Transvaal by wagon to avoid the high rates imposed on the Transvaal section of their Cape to Johannesburg line; but under pressure from Chamberlain, Kruger reopened the drifts on Nov. 7.

The Jameson Raid.—Hostility between the Boers and the British was further increased by the fiasco of the Jameson Raid (December 1895). Rhodes arranged that L. S. Jameson should lead an armed force into the republic from Pitsani, in Bechuanaland, to support a rising of the *uitlanders* against the Kruger government. The rising failed to materialize, and Jameson's raid ended in his capture at Doornkop (Jan. 2, 1896), but it aroused among the Boers deep suspicion of Britain's intentions. (*See* JAMESON, SIR LEANDER STARR.) Kruger, convinced that war with Britain could be only a matter of time and, by receiving a congratulatory telegram from the German emperor, William II, encouraged to hope for German assistance in the event of war, began a systematic program of rearmament. Moreover, Kruger's prestige was increased not only among his own people but also among the Cape Dutch and the Free Staters, and in 1897 he formed an alliance with the Orange Free State.

Relations between Britain and the South African Republic steadily deteriorated after the raid. Sir Alfred (later Viscount) Milner, who was appointed high commissioner in 1897, was determined to uphold British supremacy in South Africa and was prepared to "work up a crisis" by demanding the franchise and other civic rights for the *uitlanders*. Kruger, on the other hand, regarded firmness on the franchise question as essential for the maintenance of Boer independence. At the Bloemfontein Conference (June 1899) both the high commissioner and Kruger adopted an attitude of intransigence, and the negotiations broke down over the question of the *uitlander* franchise. Thereafter Chamberlain openly adopted the *uitlanders'* cause; for him, too, it was part of the more important issue of British paramountcy. In August, J. C. Smuts, the state attorney of the South African Republic, offered a franchise for the *uitlanders* after five years' residence, but Chamberlain rejected the conditions attached, which included abandonment of the British claim to suzerainty over the republic. Both sides then drifted rapidly into war. On Sept. 22 the British government drew up a drastic program of demands and decided to mobilize an army corps. Mobilization also proceeded both in the South African Republic and in the Orange Free State, which, although having endeavoured throughout to avert war, had resolved to support its ally. On Oct. 9 the republics presented an ultimatum, demanding the withdrawal of British troops from the frontiers, which only narrowly forestalled the British ultimatum.

THE COURSE OF THE WAR

On the expiry of the Boer ultimatum (Oct. 11, 1899) there were only 27,000 British troops in South Africa, while the republics had about 48,000 men under arms. The British Army Corps that was later dispatched to South Africa included highly trained field artillery and cavalry, as well as some mounted infantry, which were at that time unique in Europe. But it was hampered by incompetence in the high command and by the difficulty of the terrain—vast expanses of open country with poor communications. The Boers had an initial advantage in their mobility and their experience of the country. They were mounted riflemen, loosely organized in district commandos (*see* COMMANDO) under elected officers, each army being supported by a small corps of state artillery. But in the early stages of the war the lack of discipline and the weakness in the command of the Boer forces offset these advantages.

The war can be divided into two stages. The first was the period of regular warfare, which ended with the annexation of the Transvaal in September 1900; the second was the period of prolonged skirmishing, which continued until May 1902, when peace was signed.

The war opened with the investment of Mafeking by a Transvaal force under Gen. P. A. Cronje and the envelopment of Kimberley by Free State commandos under C. J. Wessels (October 1899). Meanwhile, a large Boer army under Gen. P. J. Joubert, commander of the Transvaal forces, had invaded the northern triangle of Natal, where Lieut. Gen. Sir George White was commander in chief of 16,000 British forces, most of which were divided between Ladysmith and Dundee. The Boers advanced on these centres from the north, east, and west. Although one Boer force under Lukas Meyer was repulsed at Talana Hill, outside Dundee, on Oct. 20, and another under Gen. J. H. M. Kock was defeated at Elandslaagte on the following day, the Boers forced the British to fall back on Ladysmith. White's attempt to break the Boer lines resulted in the Battle of Ladysmith (Oct. 30), in which the British losses totaled more than 1,200 (69 killed, 249 wounded, 954 missing) against the Boers' 200. Thereupon the Boers started the long siege of Ladysmith and seemed about to overrun the whole of Natal.

Buller's Offensive.—Gen. Sir Redvers Buller, who had been appointed to the supreme command in South Africa, arrived in Cape Town (Oct. 31, 1899) followed by his army corps. The situation with which he had to deal was serious. On Nov. 1, M. T. Steyn, president of the Orange Free State, sent his commandos into the Cape Colony, where they found support among the colonists, despite the imposition of martial law by the Cape government in the invaded districts. On the western front Mafeking and Kimberley were in a state of siege, and in Natal the British position was critical. Buller split his army corps into three unequal parts. One, under the command of Lord Methuen, was to relieve Kimberley and Mafeking in turn; the second, under the command of Lieut. Gen. Sir William Gatacre, was to contain the commandos invading the Cape Colony; the third, under Buller's own command, was to relieve Ladysmith. But Buller's whole scheme broke down in the series of disasters that the British sustained during "Black Week" in the middle of December. On Dec. 10 Gatacre essayed a night march and attack on the Boer position near Stormberg Junction, but, having been misled by his guides, he was himself surprised and forced to return to Molteno with a total loss of about 700 men. On the following day Methuen, who had succeeded in crossing the Modder River on his march to the relief of Kimberley, was defeated in the Magersfontein Hills, just south of the besieged town, by the forces of the Boer generals J. H. de la Rey and Cronje. The British lost about 1,000 men, including Maj. Gen. A. J. Wauchope. The Boers' victory was

largely due to De la Rey's original and effective measure of digging and defending long narrow trenches along the foot of the hills. Then on Dec. 15 Buller was decisively defeated at Colenso, on the Tugela River, by the able Boer general Louis Botha. Botha routed the British forces and lost 29 men against the British loss of 1,139. Buller immediately sent orders to White, who was still besieged in Ladysmith, to destroy his guns and surrender the town. But White refused to do so, and the siege continued.

In England "Black Week" brought a reaction of deep shock. The Cabinet appointed (Dec. 16) Field Marshal Lord Roberts to the supreme command, with Maj. Gen. Lord Kitchener as his chief of staff. Volunteer forces, both mounted and infantry, were raised, and the militia were invited to come forward. The empire—Australia, New Zealand, Canada, and others—contributed contingents (which amounted in the end to a total of 40,000 volunteers), and at the same time the Boer forces were augmented by volunteers from many countries of Europe. In the early days of 1900 the struggle swayed to and fro. On the western front, in the far north Cape, the settlement of Kuruman surrendered (January 1900) to the Boers' besieging force; in the Cape midlands Maj. Gen. John French (later 1st earl of Ypres) was pressing on Colesberg; while in Natal the fate of Ladysmith hung in the balance. On Jan. 6 the Boers made an attempt to overwhelm the besieged town but without success. This was followed by two attempts on the part of Buller to relieve it. In the first of these he was forced to withdraw with very heavy losses in the Battle of Spion Kop (Jan. 24). On Feb. 5 he again crossed the Tugela River but failed through indecision to press the advantage he had attained and was defeated at Vaal Krantz.

Roberts' Advance.—Roberts and Kitchener arrived in Cape Town on Jan. 10. Roberts' plan was to assemble his army on the Modder River in the west and to carry out an eastward march to strike at Bloemfontein, the capital of the Orange Free State. He kept his plan secret and created the impression that he was concentrating his forces in the Cape midlands area, near Colesberg. On Feb. 11 Roberts began his march from between the Modder and Orange rivers. He sent French from Ramdam (near Graspan) on a flanking movement northeast toward Kimberley, where Rhodes's vociferous demands and the serious health situation in the town, whose populace was near starvation, created an urgent need of relief. Sweeping round Cronje's main force in the Magersfontein Hills, French put an end to the siege of Kimberley on Feb. 15. Thereupon Cronje abandoned his Magersfontein defenses and escaped eastward through the British lines, which were also moving east. But, hampered by his long train of wagons, he was surrounded at Paardeberg (Perdeberg), southeast of Kimberley, and after a long and stubborn resistance was forced to surrender to Roberts with about 4,000 men on Feb. 27. The next day Buller in Natal at last succeeded in forcing the Tugela defense line and relieved Ladysmith, which by this time was in a desperate state. The Boers' besieging force retreated in disorder. The turning point in the first stage of the war had been reached. Kimberley and Ladysmith had been relieved, the Boer Army in the west had been captured almost entire, that in Natal was retreating in confusion, and in the Cape the invaders De la Rey and H. Schoeman had been forced to flee to Bloemfontein. On March 5 Presidents Kruger and Steyn offered peace proposals to the British government on the basis of continued independence for the republics, but after the fall of Bloemfontein these were rejected by the British prime minister, the marquess of Salisbury.

Capture of Bloemfontein.—Roberts, meeting with little resistance from the Boers, continued his advance on Bloemfontein, which he entered and occupied on March 13. He considered that the war in the Free State was over and on March 15 issued a proclamation allowing those burghers who laid down their arms to return to their homes. But the Boers gained a valuable breathing space by the delay of Roberts' army in Bloemfontein, where an outbreak of enteric fever took a heavy toll of lives. At the same time the death of Joubert (March 27) brought the more vigorous Louis Botha into command of the Transvaal forces. Further strengthened by a decision to reduce all transport to a minimum and to impose stricter discipline on their forces, the Boer leaders again took the initiative. Christiaan de Wet captured the waterworks at Sannah's Post outside Bloemfontein after inflicting a defeat on Gen. R. G. Broadwood (March 31) and followed this up by securing the surrender of a British detachment near Dewetsdorp, 40 mi. SE of Bloemfontein, on April 4.

Relief of Mafeking.—Early in May, Roberts set out from Bloemfontein on his march to the Boer capital of Pretoria. On May 17 a flying column, from near Kimberley, relieved Mafeking, where Col. R. Baden-Powell had maintained morale in increasingly difficult circumstances. Roberts, having reached the Vaal River and announced the annexation of the Orange Free State as the Orange River Colony (May 24), continued his successful advance, invaded the Transvaal, and on May 31 entered Johannesburg. With practically no resistance he was able on June 5 to occupy Pretoria, from which Kruger had already fled east along the Delagoa Bay railway line to Middelburg. The occupation of Pretoria was secured by Roberts' displacement of Botha's force at Diamond Hill (June 11–12), east of the capital. But the end of hostilities was not yet in sight. As Roberts concentrated on his eastward offensive against Botha, his rear was threatened by De la Rey, who harried the British forces in the western Transvaal. At the same time De Wet was engaged in disrupting railway communications in the Orange Free State. However, on July 30, a party of more than 4,000 Boers under Marthinus Prinsloo was forced to surrender after being trapped in the Brandwater Basin, south of Bethlehem, in the Orange Free State, but De Wet himself had already escaped and made his way northward.

Roberts' main army was joined (Aug. 14) by the forces of Buller, who had marched through northern Natal. On Aug. 27 they defeated the Boer commandos at Bergendal, near Dalmanutha, 45 mi. E of Middelburg, and caused them to disperse. Kruger, who had been fleeing east in advance of the British forces, left the republic (Sept. 11) and entered Portuguese East Africa. On Oct. 19 he set sail for Europe, where it was felt he would be safe and might obtain aid for the Boer cause. S. W. Burger was appointed to act as president in his absence.

The Guerrilla War.—Roberts had already proclaimed the annexation of the Transvaal (Sept. 1). Then, believing that the war was over, he laid down his command in South Africa in favour of Kitchener (November 1900) and in January returned to England. But a new war had already begun, in which the Boers adopted guerrilla tactics. They concentrated on disrupting railway communications, capturing isolated posts, and avoiding contact with the columns sent in pursuit of them. In the Orange Free State, De Wet roused the burghers to the new tactics and raised local commandos. He was unable at the beginning of December 1900 to cross the swollen Orange River and carry out his proposed invasion of the Cape, but two of his generals, P. H. Kritzinger and J. B. M. Hertzog, entered the colony (Dec. 16) and made their way southward, evading the British columns sent in pursuit of them. The new tactics were also adopted in the Transvaal, De la Rey operating in the west and Botha in the east.

Concentration Camps.—Kitchener, on taking over command in November 1900, was faced with the problem of dealing with the new war. Having appealed to Britain for more manpower, he extended the "scorched earth" policy that Roberts had already begun. Boer farms were devastated, and the women and children of the men on commando were interned in concentration camps, where for a time the death rate from disease was very high. From the British point of view Kitchener's internment policy was not a success; it added an extra burden to the fighting forces, and it further stimulated the spirit of the Boers. At the end of January 1901 Kitchener launched two large-scale offensives, the first against De Wet, who on Feb. 10 crossed the Orange River to invade the Cape Colony. But closely pursued by Col. H. C. O. (later Viscount) Plumer, he gave up his attempt and retired (Feb. 28) across the river. A second offensive, under French, was launched (Jan. 28) against Botha in the eastern Transvaal, from where he was planning an invasion of Natal. Botha managed to escape with his forces from Ermelo (Feb. 6), whereupon he was asked by Kitchener to discuss peace terms. A meeting took place at Middelburg on Feb. 28, 1901. Kitchener proposed peace terms, which included a general amnesty for Boers and for Cape and Natal rebels and the grant of self-government to the northern colonies as soon as possible. But, largely because the British government was not in favour of an amnesty for the rebels, the negotiations broke down (March 16).

Blockhouses.—As the Boers continued their skirmishing tactics, Kitchener adopted a policy of building corrugated iron blockhouses along the railway lines. When these proved effective, he extended them away from the railways to form fortification lines across the Free State and later across the Transvaal, so that eventually the whole country was split up into small areas, within which the Boer forces could be localized and overwhelmed by British troops. Meanwhile, he continued his policy of interning women and children and on Aug. 7 issued a proclamation to the effect that all Boers who had not surrendered by Sept. 15 would be punished and the leaders permanently banished from South Africa. But the Boers, who had been exhorted by the exiled President Kruger to continue fighting, were not daunted. Botha made an unsuccessful attempt at invading Natal early in September. On Sept. 3 Smuts invaded the Cape Colony and by the end of October was within 50 mi. of Cape Town. During the next six months Kitchener concentrated his efforts on the three able Boer generals who were still harassing the British forces: Botha in the eastern Transvaal, De la Rey in the western Transvaal, and De Wet in the eastern Free State. The last notable Boer victory was gained by De Wet at Tweefontein, 50 mi. W of Harrismith, on Christmas Day. Thereafter Kitchener brought his blockhouse system to completion and by a series of raids forced Botha to retire from the eastern Transvaal in February 1902. In the same month he launched intensive drives across the Free State, carefully planned to comb the countryside, and in this way he succeeded in rounding up large numbers of De Wet's men. In the western Transvaal, De la Rey's forces were gradually worn down during March, while in the Cape Smuts continued his skirmishing. The Boers' resistance steadily weakened. On April 12 their leaders met Kitchener in Pretoria and agreed to discuss peace proposals. The Treaty of Vereeniging, which was signed on May 31, 1902, provided for the surrender of the Boers and their submission to the sovereignty of the British crown. The war, which had lasted two and a half years, had cost Britain about 22,000 men and the Boers not far short of 6,000, as well as 4,000 women and 16,000 children who had died in the concentration camps. At the end of the war there were almost 26,000 Boers in prison camps and more than 116,000 in the concentration camps.

RESULTS OF THE WAR

As a result of the war the two Boer republics ceased to exist, and crown colony rule was established in the Transvaal and in the Orange River Colony. The whole of South Africa was brought under the British crown, and thus the subsequent political union of the former republics with the existing colonies of the Cape and Natal was made possible. The Treaty of Vereeniging contained a promise of self-government for the new colonies, and this was granted to the Transvaal in December 1906 and to the Orange River Colony in June 1907. For both political and economic reasons a movement for the closer union of the four self-governing colonies developed, and this led to the establishment of the Union of South Africa in 1910.

The war had the effect of arousing a new national consciousness among the Afrikaner people throughout South Africa. This gave rise first to a cultural and then to a political revival. In the crown colonies, in opposition to Lord Milner's policy of anglicization, a "Christian National Education" movement developed. Private schools were established, and these fostered Afrikaner national sentiment among the rising generation. National consciousness expressed itself also in a language movement. The Taalbond (language union), revived in 1903, encouraged the use of High Dutch to combat English. Attempts were also made to promote Afrikaans, the spoken language of the Afrikaner people, as a written language and to obtain official recognition for it. The Afrikaans poetry and prose that the language campaign produced gave rise to a virile national literature. At the same time the Afrikaners in the new colonies were beginning to form political organizations. In the Transvaal the Het Volk ("The People") Party was formed in January 1905, followed in May 1906 in the Orange River Colony by the Orangia Unie ("Orange Union"), both pledged to the ideal of self-government.

For Britain the war brought to an end the period of late-19th-century imperialism. Moreover, it revealed to Britain many defects in its military organization and taught it tactical lessons, such as the importance of accurate rifle fire and of taking cover, which were to prove valuable in World War I.

BIBLIOGRAPHY.—L. S. Amery (ed.), *The Times History of the War in South Africa, 1899–1900*, 7 vol. (1900–09); Sir J. F. Maurice and staff, *History of the War in South Africa* (comp. by direction of His Majesty's Government), 4 vol. (1906–10); C. Headlam (ed.), *The Milner Papers: South Africa, 1897–1899, 1899–1905*, 2 vol. (1931–33); D. Reitz, *Commando* (1929); C. R. de Wet, *The Three Years' War* (1902); A. C. Martin, *The Concentration Camps, 1900–1902* (1957); R. Kruger, *Good-bye Dolly Gray* (1959). For detailed bibliography (official publications, contemporary works and later works) *see* bibliography section xii in *Cambridge History of the British Empire*, vol. viii, 2nd ed. (1963). (C. E. M. O'D.)

SOUTHALL, a municipal borough (1936) of Middlesex, Eng., which under the London Government Act 1963 was amalgamated on April 1, 1965, with Acton and Ealing to form the new London borough of Ealing (*see* LONDON). Southall Municipal Borough, which includes Norwood, is 12 mi. W of Charing Cross, by London Airport. Pop. (1961) 52,983. Area 4.1 sq.mi. Southall is a parliamentary constituency. In 830 the manor of Norwood was bequeathed to the archbishop of Canterbury; the manor house (1581) is now municipal offices. In 1698 William III granted a weekly market, now held on Wednesdays (cattle) and Saturdays (shopping). Martinware pottery was made in Southall but present industries are chiefly the making of motor vehicles and chemicals, food processing, and engineering. (E. G. MA.)

SOUTH AMERICA, the fourth largest of the continents, is located between longitudes 34° and 82° W and latitudes 13° N and 55° S. It has an approximate area of 6,879,000 sq.mi., about 13% of the land surface of the earth. Its estimated population (1963–64) was 160,000,000, equal to approximately 23 persons per sq.mi. The continent, with adjacent islands, is made up of ten sovereign republics (Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela) and three European colonies (British Guiana, French Guiana, and Surinam, or Dutch Guiana). South America is joined to Central America by the Isthmus of Panama.

Geographically, the continent is dominated by the Andes Mountain chain, which extends in a generally north-south direction for 5,000 mi. The Magdalena (Colombia), Orinoco (Venezuela), São Francisco (Brazil), and Río de la Plata (Argentina, Brazil, Paraguay, and Uruguay) are major river systems, but each is dwarfed by the Amazon, the largest river in the world in terms of area drained and volume of flow. The barren, saline Atacama Desert of the west coast stands in marked contrast with the tropical jungles of Bolivia, Brazil, Ecuador, Peru, and the Guianas on the one hand and temperate Valdivia, Chile, with an average annual rainfall in excess of 100 in., on the other.

This article is organized according to the following outline:

I. Physical Geography
  A. Geology
  B. Physiography
    1. The Coasts
    2. Landforms
    3. Drainage
  C. Climate
  D. Vegetation
  E. Animal Life
II. Natural Resources
    1. Water
    2. Soils
    3. Minerals
    4. Land Use
III. Anthropology
    1. Racial Type
    2. Linguistic Groups
    3. Marginal Culture
    4. Tropical Forest Culture
    5. Circum-Caribbean Culture
    6. Central Andean Culture
IV. Population
V. History

I. PHYSICAL GEOGRAPHY

A. Geology

The most ancient core areas of South America lie in the Brazilian Shield and the Guiana Highlands. Analogous to the Laurentian Shield of North America, they consist largely of a basement complex of pre-Cambrian gneisses and schists, exposed in some areas and overlain in others by sedimentaries mostly of Paleozoic Age. Many of the older formations have their counterparts in age, composition, and attitude in southern and western Africa, and the belief is widely held, although not universally accepted, that the two continents were joined until Jurassic time. Much of the rest of the South American continent must ultimately owe its existence to deposition in marginal seas of the products of erosion from the core areas. Extensive and repeated periods of uplift and subsidence brought about major transgressions and retreats of the sea over these marginal areas, with consequent deposition of massive beds of marine sediments and a gradual increase in the size of the continent.

A vast geosyncline on the western and northern margins of the continent became the repository in the Mesozoic Era of accumulations of sediments many thousands of feet thick. In the Upper

Cretaceous the great weight of these deposits gave rise to crustal instability with crumpling and uplift of the strata, a process that continued throughout the Tertiary and resulted in the formation of the Andes. Intrusions of magma accompanied this activity, and volcanism was widespread. Volcanic activity is still a feature of the Andes, although it has entirely ceased in such older lands to the east as the Paraná Basin of southern Brazil and Paraguay, where successive and widespread outflows of basaltic lava engulfed the land in the Jurassic period.

Extensive glaciation has occurred in South America in at least two distinct geologic epochs, the Carboniferous and the Pleistocene. In the former case the main centre appears to have been in southern Brazil; in the latter, in the southern Andes. A few small relics of a formerly extensive continental ice sheet still survive in southern Patagonia.

B. Physiography

1. The Coasts.—The South American continent rises abruptly from the ocean floor along much of its coast, but the steepness of the continental margin is more marked on the western side than on the eastern. The west coast between latitudes 5° and 35° S displays little development of a continental shelf and drops off sharply to a series of deep submarine trenches parallel to the shore. In the neighbourhood of 26° S the Richards Deep exceeds 25,000 ft. in depth within 50 mi. of the coast. In general, the Pacific coast north of the Patagonian archipelago is but little articulated and is completely lacking in good natural harbours.

From 41° S to Cape Horn the Pacific coastal zone is a series of islands, channels, and fjords, results of subsidence and glaciation. North of 41° S the land between the Andes and the sea averages 40–50 mi. in width. In southern and central Chile a longitudinal interior valley is bordered on the west by a coast range whose elevation increases toward the north. Northern Chile has a high cliff coast situated in front of a series of interior basins. In Peru, Ecuador, and Colombia the coast range is largely absent.

The east coast is in strong contrast with the west. Between Cape São Roque and Rio Grande do Sul the coast generally follows the sinuous eastern edge of the Brazilian Plateau. Along this coast lie some of the world's most beautiful harbours, notably those of Recife, Salvador, Vitória, Rio de Janeiro, and Florianópolis. Farther south the estuary of the Río de la Plata is the most marked indentation of the entire South American coast. It is the outlet for the combined waters of the Paraná and Uruguay river systems. The Patagonian coast is poorly endowed with natural harbours. Although many embayments occur, their utility is reduced by shallow water, high tidal ranges, and coastal cliffs.

The east coast of South America has few islands, and these, except Trinidad, off the coast of Venezuela, and Marajó, at the mouth of the Amazon, are mostly small. Trinidad (area 1,980 sq.mi.) is separated from the continent by the Gulf of Paria. Along the island's northern edge is a range of mountains with a maximum height of 3,085 ft., which is geologically a continuation of the Cumaná Range of Venezuela. On the south side of this island is the famous asphalt lake—the largest known deposit of its kind. West and north of Trinidad and lying farther off the coast are several small islands of historical interest and considerable commercial importance: Tobago, Margarita, Tortuga, and the Curaçao group. The main island of the Fernando de Noronha group of volcanic islands, 230 mi. off Cape São Roque, has an area of only 7 sq.mi. Although separated from the mainland by a channel 13,000 ft. deep, it really stands upon the submerged corner of the continent. The Falkland Islands in latitude 51° S also stand upon the submerged edge of the continent. Their fauna and flora indicate that they were once a part of the mainland, from which they are now separated by shallow water. On the west coast north of the Patagonian archipelago are the Juan Fernández Islands west of Valparaíso, the famous guano islands close to the mainland of Peru, the Galápagos Islands on the Equator, and the islands of the Guayas River delta near Guayaquil.

2. Landforms.—The continent of South America is composed of the following major surface forms: the cordilleras of the Andes in the west and north; the Guiana, Brazilian, and Patagonian uplands in the east and south; and the lowlands that divide the highlands from one another, namely, from north to south, the Orinoco depression, the Amazon Basin, the Paraguay Basin, and the pampas.

The Andes, one of the greatest mountain systems in the world, are more than 4,500 mi. long and are exceeded in average height only by the Himalayas. The southern Andes are a single chain, but north of latitude 28° S they form a complex system. Widening to a broad plateau in Bolivia, they narrow to a series of parallel ranges in Peru, continue in two parallel ranges in Ecuador, and then fan out in Colombia into three distinct ranges, two of which end near the Caribbean shore while the easternmost continues through northern Venezuela into Trinidad. Volcanic cones are present in three separate areas (for further details, *see* Andes.)

The Guiana Highlands and the Brazilian Shield together represent the ancient core area of the South American continent, divided into two unequal sections by the Amazon geosyncline. Waterfalls and rapids occur along all the streams passing over the edge of these uplands. Angel Fall in the Guiana Highlands is the highest uninterrupted waterfall in the world (3,212 ft.). The Paulo Afonso Falls on the São Francisco River in northeastern Brazil were harnessed in the late 1940s for a major hydroelectric installation. The crystalline rocks, in areas of high rainfall, weather into characteristic domelike shapes, as exemplified by the celebrated Sugar Loaf overlooking the bay at Rio de Janeiro. The highest point in the Guiana Highlands is Mt. Roraima (9,094 ft.), and on the Brazilian Shield, Pico da Bandeira (9,495 ft.). Patagonia is mostly a succession of arid plateaus, characterized by horizontal beds of sandstone and widespread lava caps.

In South America the development of lowlands is far greater in the interior of the continent than near the periphery. Only on the Atlantic coast, around the mouth of the Amazon and near the Río de la Plata, do extensive lowlands occur next to the sea. Typically, the lowland areas are pouch shaped, with their greatest width inland. The Llanos del Orinoco occupy a structural depression between the northern Andes and the Guiana Highlands. The Orinoco River does not flow along the central axis of the llanos but at the base of the Guiana Highlands, in part a consequence of the great deposition of alluvium by tributary streams issuing from the rapidly eroding Andes. The Amazon Basin has an exceedingly low gradient, Manaus, 800 air miles from the Atlantic, being only 130 ft. above sea level, and Iquitos, 1,700 air miles inland, only 350 ft. In spite of this low gradient, far less of the Amazon lowland is subject to inundation than is commonly supposed, the great bulk of the land lying above the level of highest floodwater on the major streams. In its widest section the upper Amazon Basin affords access over low divides (about 1,000 ft. elevation) into the Orinoco depression to the north and the Paraguay Basin to the south. The Paraguay-Paraná geosyncline begins in the north with the great region of swamp (Pantanais grandes) in western Mato Grosso and with the Gran Chaco of Bolivia, Paraguay, and Argentina. Slopes are so gentle that drainage is slow and imperfect. Toward the south the Chaco grades into the pampas, while the swamps of Mato Grosso find their general counterpart in the adjoining Mesopotamia section of Argentina.

3. Drainage.—The most important river systems of South America—the Amazon, Orinoco, and the Paraná-Río de la Plata—have the greater part of their drainage basins in the lowland belt. The chief headwater tributaries of the Amazon and the Orinoco have their sources in the Andean Highlands, where they have cut deep valleys. The principal sources of the Paraná-Río de la Plata system are in the Brazilian Plateau. These three river systems together drain an area of about 3,700,000 sq.mi. The Amazon is the largest river in the world, though not the longest. At the narrows of Óbidos, where it is constricted by spurs of highland projecting from north and south, it has a width of 1¼ mi., a depth of 270 ft., and an average current of 2.7 mph. During flood stages it rises about 25 ft., and the current in the narrows increases to 4.2 mph. The main stream is navigable for ocean steamers as far as Iquitos, about 2,300 mi. from the sea. In time of high water the main stream of the Orinoco is navigable for 1,000 mi. or more.

Other important rivers are the Magdalena in Colombia, the Es-

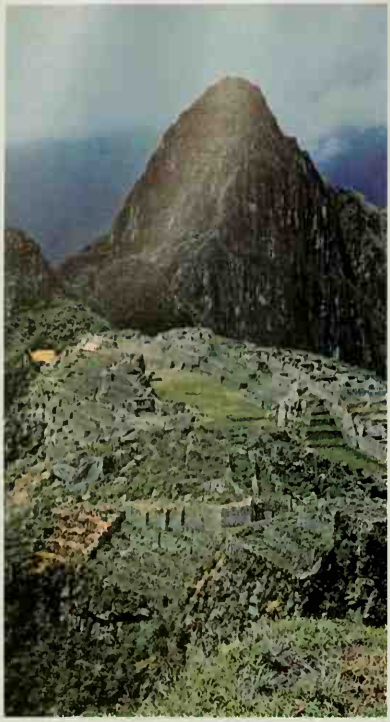
Lake Maracaibo, Venezuela, centre of the petroleum industry of Latin America. The lake receives the drainage of the slopes of the eastern cordillera of the Andes Mountains. Although it is called a lake, it is, in fact, a 133-mi.-long bay of the Gulf of Venezuela

Magdalena River, Colombia. A major commercial artery for several centuries, the Magdalena drains the central and eastern slopes of the Colombian Andes and empties into the Caribbean near Barranquilla

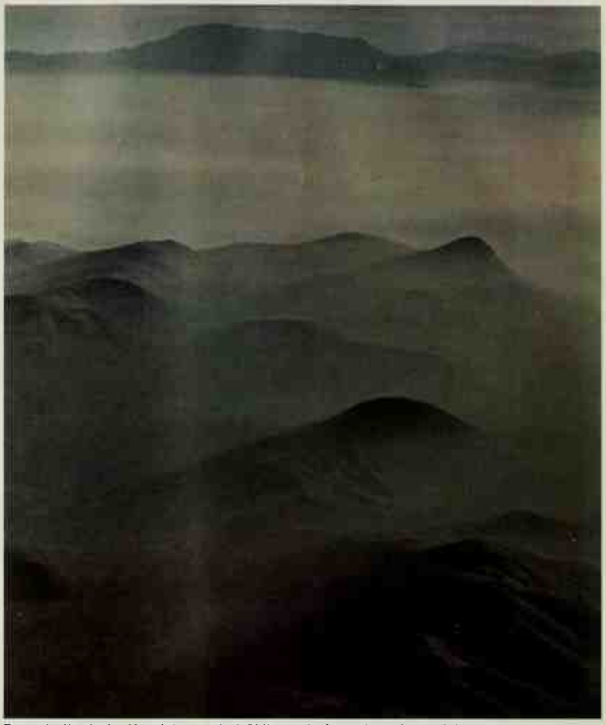
Plains near Esmeralda, Venezuela, part of the Orinoco watershed which lies between the northern Andes and the Guiana Highlands (in background)

PHOTOGRAPHS, (TOP) J. D. BARNELL—IMAGE INTERNATIONAL, (BOTTOM LEFT) THOMAS B. HOLLYMAN—PHOTO RESEARCHERS, INC., (BOTTOM RIGHT) HECTOR R. ACEBES—PHOTO RESEARCHERS, INC.

SOUTH AMERICA



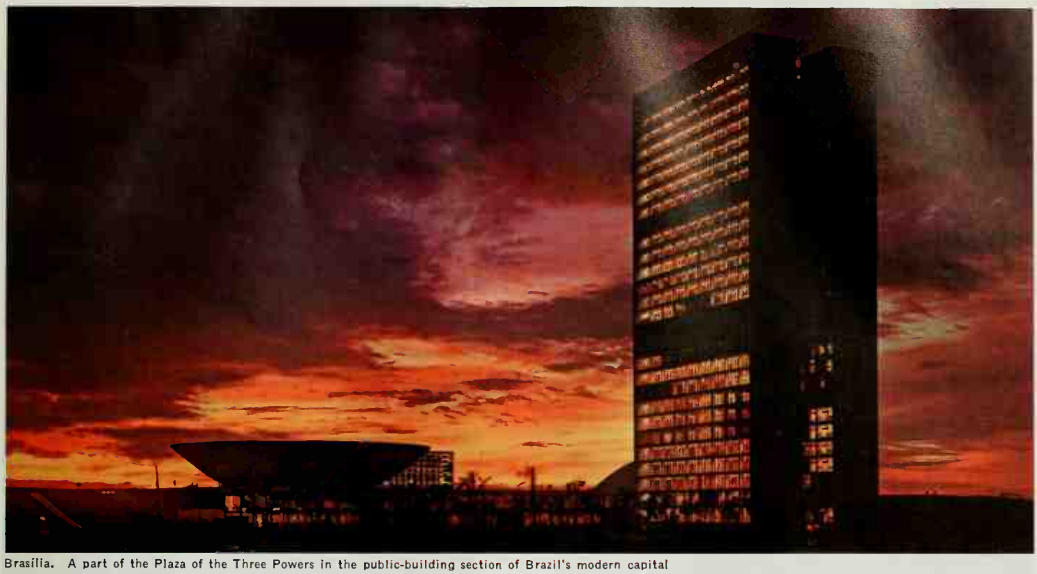
Machu Picchu ruins, Peru. Ancient Inca village perched high in the Andes overlooking Urubamba canyon 70 mi. from Cuzco, capital of the Inca empire. Behind the ruins is the peak of Machu Picchu from which the village takes its name



Range in the Andes Mountains, central Chile, part of a system of mountain ranges stretching the entire length of the Americas and attaining its highest elevation along the Chile-Argentina border

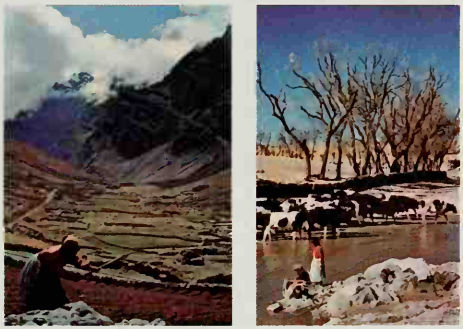
Volcanic peaks, Imbabura province, northern Ecuador. The eastern cordillera of the Andes in Ecuador is one of the world's major areas of volcanic activity

PHOTOGRAPHS, (TOP LEFT) MARC AND EVELYNE BERNHEIM—RAPHO GUILLUMETTE, (TOP RIGHT) RONNY JAQUES—PHOTO RESEARCHERS, INC., (BOTTOM) THOMAS B. HOLLYMAN—PHOTO RESEARCHERS, INC.



Brasília. A part of the Plaza of the Three Powers in the public-building section of Brazil's modern capital

Tropical vegetation, Amazon basin, Brazil. The wet, warm climate of the lowland equatorial portion of South America fosters luxuriant plant and animal growth. The tribal Indian culture to which the hunter belongs is undergoing modification

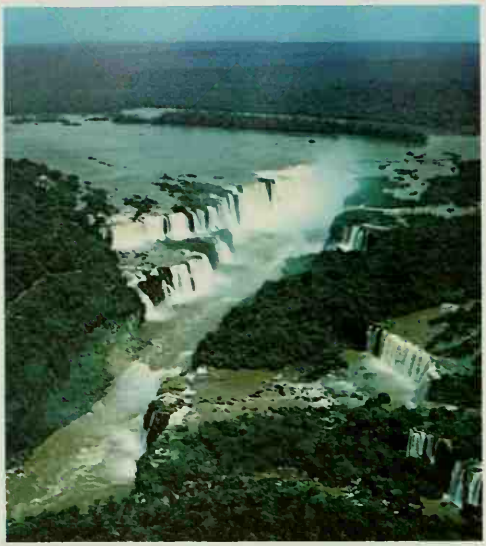


Some aspects of land usage and topography in South America: (left) a fertile mountain valley under cultivation near La Paz, Bolivia; (right) cattle in Santa Cruz district of eastern Bolivia, a developing farming area; (below) Marajó Island at the mount of the Amazon River, Brazil

PHOTOGRAPHS, (TOP) J. D. BARNELL—IMAGE INTERNATIONAL, (LEFT) KARL WEIDMANN FROM NATIONAL AUDUBON SOCIETY, (CENTRE LEFT) THOMAS B. HOLLYMAN—PHOTO RESEARCHERS, INC., (CENTRE RIGHT) RENE BURRI—MAGNUM, (BOTTOM RIGHT) SIMON—PHOTO RESEARCHERS, INC.

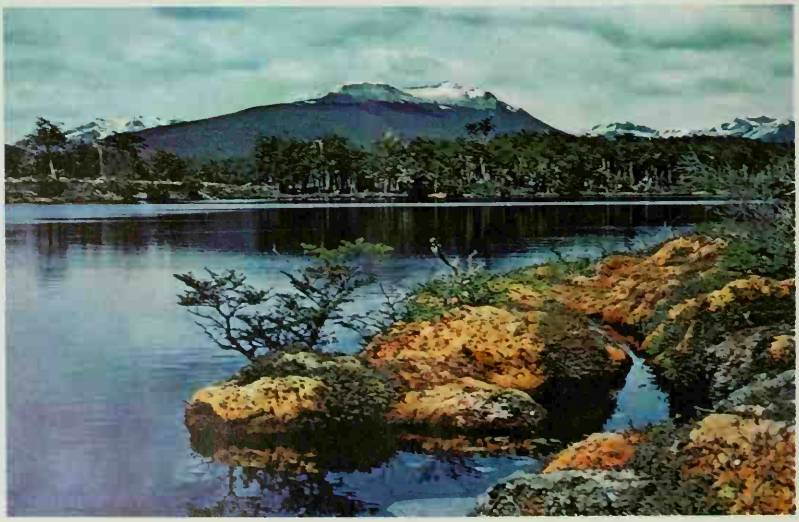
SOUTH AMERICA

Boat dock on the Paraná River north of Buenos Aires. Larger of the two river systems that empty into the Rio de la Plata, the Paraná rises in Brazil and drains large areas of Brazil, the Paraguay basin, and the pampas



Iguaçu Falls. Along part of Brazil-Argentina border, the 2½-mi.-wide Iguaçu River drops about 270 ft. over a series of about 200 falls separated by islands and grouped roughly in a horseshoe pattern

Copper mine, Chuquicamata, Chile. One of several mines situated in the Atacama Desert, it works the largest proven copper deposits in the world



Tierra del Fuego. Glacial lakes, deciduous forests, and rugged, and, in places almost barren terrain are typical of this southernmost part of South America, separated from the mainland by the Strait of Magellan

PHOTOGRAPHS (TOP LEFT) ED DREW—PHOTO RESEARCHERS, INC. (TOP RIGHT) ALPHA (CENTRE RIGHT) FPG (BOTTOM) ROGER TORY PETERSON—PHOTO RESEARCHERS, INC.

sequibo in British Guiana, and the São Francisco in Brazil. The Magdalena is navigable in two sections separated by the rapids at Honda for river steamers as far as Girardot, the river port of Bogotá. The São Francisco, which rises in the highlands of Minas Gerais, is navigable nearly to the falls of Paulo Afonso, 140 mi. from its mouth, and for 850 mi. above the falls.

Most of the lakes of South America are mountain lakes in the Andes or along their base. Of the Andean lakes, Titicaca (*q.v.*) and Poopó are the largest. The glacial lakes along the eastern base of the Andes in southern Argentina and those of the Llanquihue District of southern Chile are described in the articles ANDES, ARGENTINA, CHILE, and PATAGONIA. Along the east coast there are occasional lakes of brackish water, such as the Lagoa dos Patos of southern Brazil, produced by depression of the coast and the closing of the mouths of estuaries formed thereby or by barrier beaches thrown up by the sea. Lake Maracaibo in northern Venezuela is a large, narrow-necked bay of the Caribbean rather than a true lake, but its waters are brackish.

C. Climate

South America, the only southern continent to extend far into temperate latitudes, tapers toward the South Pole and hence does not experience the extremes of temperature of the northern continents. The winters in temperate latitudes are warmer and the summers cooler than in North America or Asia. The continent is broadest in the equatorial section and has larger areas of hot tropical lowlands than any other continent.

Average annual temperatures in the Amazon Basin are about 80° F. (27° C.), and seasonal variations are slight. Quito, near the Equator but at an elevation of 9,350 ft., averages about 55° F. (13° C.) in every month. Seasonal temperature ranges increase with latitude and reach their maximum in northwestern Patagonia, where July averages 32° F. colder than January. Winter freezing at low elevations is common in Tierra del Fuego and Patagonia, is by no means rare on the Argentine pampas or in south-central Chile, and occurs in southern Brazil at elevations above 1,000–2,000 ft.

The northward-flowing Humboldt (Peru) Current and associated upwelling water keep temperatures on the Pacific coast cool as far north as the Equator. The coast of Argentina is cooled by the Falkland Current. Elsewhere around the continent warm water washes the shores: the South Atlantic equatorial current flowing west encounters the bulge of Brazil and divides into the south-flowing Brazil Current and a coastal current that flows northwest into the Caribbean; on the northern Pacific coast the equatorial countercurrent sends a warm-water eddy to wash the shores of western Colombia and Ecuador.

The rainfall pattern of the continent is closely connected with the distribution of winds and air masses. Most of tropical South America east of the Andes has prevailing winds from the northeast, east, or southeast. They transport moisture from the warm Atlantic Ocean and give copious rainfall by convection. Rainfall totals average 90–120 in. per year along the lower Amazon and the coast of the Guianas. The middle Amazon Basin has somewhat less rain (60–70 in.), and the upper basin has more than 90 in., possibly exceeding 150 in. in exposed portions of the Andean foothills. The interior of the Brazilian bulge is generally dry and periodically subject to extreme drought. The Llanos del Orinoco and the Guiana Highlands have moderate to high rainfall, with a strong maximum in summer. Much of northwestern South America has two rainfall maxima associated with the two periods of high sun. The Pacific shores of Colombia and northern Ecuador are extremely rainy, while the Gulf of Guayaquil marks the transition to the dry west-coast conditions of Peru and northern Chile. The Atacama Desert along this part of the coast is one of the driest areas in the world. South of 33° S rainfall is adequate west of the Andes, and south of 38° S it becomes excessive. Cyclonic precipitation is characteristic of middle and southern Chile, associated with the westward movement of storms from the Pacific. Valdivia (latitude 40° S) averages 105 in. of rain per year, and the fjorded coast farther south receives even more. Argentine Patagonia lies in the rain shadow of the Andes and is mostly steppe and desert. The grasslands of Argentina, Uruguay, and southern Brazil have moderate rainfall (25–50 in.) well distributed throughout the year. Eastern Paraguay is moist (50–75 in.), while the Chaco is moderate to dry and has a dry winter.

D. Vegetation

The pattern of vegetation for the continent is complex, especially in the highlands and in areas of critical rainfall differences. The actual vegetation cover deviates most from the natural forms in areas of dense human settlement and in regions where the land has been converted into permanent farms or artificial pasture. A question still remains as to how much influence the pre-Columbian Indian exerted on the natural vegetation cover. It is possible that the greater extent of tropical grassland north of the Amazon lowland and in the northern fringes of it as compared with areas south of the Amazon is in part a reflection of greater aboriginal population density in the northern areas and of a greater use of fire for making clearings and driving game. The comparable areas south of the Amazon have a cover of semideciduous woodland that has been much less affected by fire.

The largest more or less uniform vegetation region is that of the tropical lowland rain forest in the Amazon Basin and the Guianas, with an outlier along the east-central Brazilian coast. It is the largest area of its kind in the world and has several distinct subtypes. The number of plant species growing together in this forest is greater than in any other plant association in South America. Additional characteristics are dense growth, generally slender trees, storied vegetation, lianas, parasites, and epiphytes, and, at times, buttressed trunks.

A semideciduous phase of the tropical forest is found in the Venezuelan part of the Guiana Highlands and in some inland sections of central and southern Brazil. It is somewhat less luxuriant than the true lowland rain forest, and some of the plants that compose it drop their leaves seasonally because of dryness or low temperatures. Tropical highland forest, in part cloud forest, is found on the eastern flanks of the Andes from latitude 20° S northward. Mangrove forests fringe most of the warm-water coasts of the continent from the Gulf of Guayaquil clockwise almost to Rio de Janeiro. A palm forest, also tropical, is characteristic of a portion of the state of Maranhão.

Temperate or mid-latitude forests are of two main types: the araucaria evergreen forest of southern Brazil and the more extensive deciduous forests with small araucaria enclaves in southern Chile. In southern Brazil the araucaria, a coniferous softwood, is the basis of a significant lumber and paper industry.

One step below the forests in magnitude are the woodlands of the continent, composed chiefly of lower trees, more widely spaced, with an admixture of shrubs and grass. As compared with the forests, they are found in areas of less total and less uniform rainfall. Semideciduous woodlands known as *campos cerrados* are characteristic of central Brazil. Especially thorny woods are found in northeastern Brazil, in much of the Chaco, and in northernmost Colombia and Venezuela. Park landscapes occur along the Paraguay River and in the hill lands west of the Chaco. They are dotted with well-spaced trees, the intervening areas being covered with grass and shrubs, and represent a most attractive plant association.

Shrub lands display a further diminution in scale, and the typical plants are bushy rather than treelike. They are characteristic of generally dry areas and of those with pronounced seasonal drought. Various forms are found in the Andean foothills behind the tropical coastal desert of Peru, in central Chile, in the lands west of the Argentine pampas, and in eastern Patagonia.

Tropical grasslands are characteristic of the llanos, of the coastal fringe of the Guianas behind the mangrove forest, of parts of the Guiana Highlands in northernmost Brazil, and of certain sections of the Amazon, including the northern and interior parts of Marajó Island. South of the Amazon forest, tropical grasslands are rare, but two large areas of palm-savanna swampland are found.

Temperate grasslands are characteristic of the pampas of Argentina and of adjacent areas in Uruguay and southernmost Brazil.

At intermediate elevations they occur near the araucaria forests of southern Brazil, as well as in the Peruvian Andes. At high Andean levels moist alpine grassland is found in equatorial and mid-latitude areas, whereas a dry phase (*puna*) covers most of the intervening section in the Peruvian and Bolivian altiplanos and in parts of northern Chile and Argentina. Extreme aridity in the Atacama Desert of coastal Peru and northern Chile causes an almost complete absence of vegetation.

E. Animal Life

The native animals of South America belong largely to the so-called neotropic realm of animal geography. For obvious reasons there is a fair correspondence between the occurrence of specific animals and the distribution of native vegetation.

The tropical forests contain a fauna characterized by several kinds of monkeys, the tapir, rodents (including the capybara, the largest rodent in the world), birds (parrots, egrets, hummingbirds, etc.), and reptiles (especially crocodiles). Large members of the cat family are relatively rare, as are the large herbivores. The insect world is represented in great numbers. A most characteristic, though not common, member of the Amazonian fauna is the manatee, a vegetarian, aquatic mammal. Just as characteristic, and unfortunately all too common, is the piranha, a small, fierce freshwater fish that attacks other animals that venture into its waters. A school of piranhas has been known to kill a cow and to consume it completely, except for the skeleton, in less than five minutes.

The greatest contrast to this tropical lowland development of animal life is found in the high Andes and in the cold steppes and deserts of Patagonia. Among the important animals native to the central Andes are the four members of the American branch of the camel family: the guanaco, llama, alpaca, and vicuña. The guanaco and vicuña have not been domesticated, whereas the llama and alpaca are found only in the care of man. The llama is the common beast of burden of the Andean Highland Indian, the alpaca is a prolific bearer of fine wool, and the vicuña has a fleece of unbelievable softness and thickness. A rodent found in parts of the high Andes and famous for its magnificent fur is the chinchilla.

The guanaco ranges beyond the Andes into Patagonia and at one time was fairly common on the Argentine pampas. Even wider in range is the rhea, the three-toed South American ostrich, which extends from Tierra del Fuego into central Brazil.

The shrub lands and woodlands south of the Amazon have a fauna that well expresses their transitional character between the Amazon on the one hand and the Chaco and pampas on the other. Truly tropical forms occur there, such as monkeys, anteaters, sloths, tapirs, and parrots, but animals from the colder south are also found, including the rhea, the armadillo, and the raccoon. Large members of the cat family, including jaguars, leopards, and pumas, occur in the Brazilian state of Mato Grosso and other inland areas. Some of these animals are called *tigres* by the South Americans, but true tigers in the wild state are confined to the Eastern Hemisphere.

II. NATURAL RESOURCES

1. Water.—Because of diverse topography and rainfall conditions, the water resources of the continent vary enormously in different areas. In the Andes, navigation possibilities are slight except on the Magdalena River, Lake Titicaca, and the Chilean-Argentine lakes in the south. Irrigation for agriculture is of importance from northwestern Peru to Patagonia. Less than 10% of the Andes' estimated hydroelectric power potential had been harnessed by the mid-1960s.

The Brazilian Shield has better hydroelectric possibilities than the Andes because of the large size of many rivers, the harder rocks, and the occurrence of cliff margins. The harnessing of the Paulo Afonso Falls provides electricity for a large part of northeastern Brazil. Waters impounded near São Paulo are made to fall over the Serra do Mar to provide cheap power and a reliable municipal water supply.

The Amazon River system has more than 8,000 mi. of navigable waterways but presents few possibilities for power development because of the gentle slopes throughout the region.

2. Soils.—Thousands of square miles of dark prairie soil of aeolian and alluvial origin occur in the pampas of Argentina and Uruguay. They are among the best farming soils in the world. Smaller, disconnected areas of good soils are located in various Andean and west-coast valleys, notably in the Valle Longitudinal of Chile, the Guayas lowland of Ecuador, and the Cauca Valley of Colombia.

The Paraná Basin is deeply covered with a rich, purplish-red soil called *terra roxa*, derived by weathering from basaltic outflows. This soil is especially suited to coffee growing, as is the volcanic soil of the Colombian Andes.

In the Amazon Basin the soils are generally poor. Small areas of bottomlands have fertile soils, but these are subject to flooding. Infertility and high acidity make most tropical lowland soils submarginal for agriculture. To improve them with lime and fertilizer is rarely feasible.

3. Minerals.—South America is well endowed with petroleum. Roughly one-fourth of the area contains sedimentary basins that might, other conditions being favourable, permit the occurrence of oil-bearing strata. As of 1965 only minor strikes had been made at distances greater than 300 mi. from the Andes, most of the operating oil fields being located in structural basins flanked by Andean ranges, especially in the vicinity of the Caribbean.

Venezuela, with 17,000,000,000 bbl. of proved reserves as of the end of 1963, had by far the greatest wealth of petroleum in South America. It ranked second in the Western Hemisphere and seventh in the world, being exceeded in addition to the United States by four countries in the Middle East and by the Soviet Union. Other South American countries with large amounts of proved petroleum reserves include Argentina, Colombia, Peru, and Bolivia.

Natural gas is commonly associated with petroleum, but its commercial utilization has been limited by the remoteness of fields from major population centres. Gas is beginning to be used commercially in Venezuela, but its greatest use occurs in Argentina, where a 1,000-mi. pipeline extends northward from the Comodoro Rivadavia field in Patagonia to supply greater Buenos Aires.

The dearth of good coal in South America represents one of the greatest resource contrasts between the two American continents. Although scattered deposits are found in many parts of the Andes, as well as in southern Brazil, only three countries, Colombia, Chile, and Brazil, have a modest production. Brazil's coal is excessively high in ash content and must be mixed with higher-grade imported coal before it can be utilized industrially. The paucity of coal is a serious disadvantage to the growth of heavy industry on the continent, notwithstanding the plentiful supply of iron ore.

South America is extraordinarily well endowed with deposits of iron ore. The deposits are associated mainly with the ancient core areas of the Guiana Highlands and the Brazilian Shield. Very extensive, high-grade ore bodies have been discovered at or immediately below the surface, and they can be mined directly and cheaply in open pits through the use of power shovels of large size. As a result of the investment in Venezuela of hundreds of millions of dollars by United States corporations, the production of iron ore in that country increased from 2,500,000 tons in 1953 to 12,500,000 tons in 1960, followed by a sizable decline in the early 1960s because of political unrest in the producing area. The main iron districts lie near the Caroní River, a southern tributary of the Orinoco, and their economic development has been greatly facilitated by the availability of cheap water transport.

In Brazil the situation is less favourable, the large ore bodies lying mostly in the interior of the state of Minas Gerais, far from navigable waterways. Although the deposits are rich and more extensive than those in Venezuela, large-scale utilization awaits the improvement of land transport and coastal docking facilities.

Copper occurs mainly in the geologically young lands of the central Andes, in Chile and Peru. Only the great copper mines of North America and central Africa have greater world significance than those of Chile. Commercial tin ores in the Western Hemisphere are confined mainly to Bolivia. The mines are located in

the high Andes, some of them being earlier silver mines that were converted to tin mining when the richer silver ores were exhausted. Total tin resources in the Bolivian Andes are estimated to rank third in the world, following southeast Asia and equatorial Africa.

Brazil ranks high in manganese reserves, one of the world's great deposits being located in southwestern Mato Grosso. A smaller but more easily accessible deposit in Amapá lies immediately north of the mouth of the Amazon. Brazil ranks with the Soviet Union and India as a leading producer of manganese.

Other significant minerals in South America are bauxite (aluminum ore), which is especially plentiful in the Guianas and adjacent parts of Brazil; platinum, in Colombia; silver, in Peru and Bolivia; and nitrate, in Chile.

4. Land Use.—Most South Americans live within a few hundred miles of the coasts, and most regions of intensive land use are confined to the same peripheral locations. For the continent as a whole, less than 5% of the land is cultivated, whereas about 17% is permanent grazing land and fully 47% is forest. In the early 1960s the proportion of cultivated land by countries varied from a high of 14% for Uruguay to a low of 1% for Paraguay and one-thirtieth of 1% for French Guiana.

Regions of commercial agriculture are widely scattered. The irrigated coastal valleys of Peru are noted for their long-staple cotton, and the Guayas lowland of Ecuador has become the world's leading area of banana production for export. In Brazil more than 2,500,000 ac. are devoted to sugarcane, mainly along the south-central coast and in nearby interior valleys, and tobacco, bananas, and cacao occupy other low-lying districts not far from the sea. The world's greatest coffee-growing district is located on the upland *terra roxa* soils of the Brazilian states of São Paulo and Paraná. The other major coffee area of South America lies in the Colombian Andes, where higher-quality mild coffees are grown.

Mid-latitude agriculture, in part for export, is dominant in the southern part of the continent. The greatest agricultural region of South America is the pampas of Argentina and Uruguay. Wheat is the chief crop, but corn and flax also enter significantly into export trade.

Notwithstanding the importance of South America's trade in agricultural commodities or its significance as a source of foreign exchange for the South American countries, it must be emphasized that the vast bulk of the continent's agriculture is devoted to supplying local needs. For subsistence and local commercial markets corn (maize) is grown over much of the continent and leads in total acreage planted. In the tropical lowlands cassava (manioc) and sweet potatoes are widely planted for local consumption, and in the Andes potatoes and other cold-country root and tuber crops are cultivated. Rice and wheat have long been staple food crops and are widely grown wherever conditions are favourable.

Pastoral industries with emphasis on cattle are dominant on the Argentine pampas and adjacent grasslands of Uruguay and southern Brazil, as well as on the llanos of Colombia and Venezuela. Sheep raising for wool is commercially significant in Uruguay, Patagonia, and Tierra del Fuego.

Forest industries based on planted eucalyptus and native araucaria have been organized in southern Brazil; south-central Chile appears to have possibilities of similar development in the future.

Extensive manufacturing districts are not yet common on the continent. Their greatest development has taken place in the environs of São Paulo and Buenos Aires. Examples of individual industrial centres are numerous, however, and include such settlements as Volta Redonda (Brazil), Concepción (Chile), Córdoba (Argentina), Medellín (Colombia), and San Félix (Venezuela).

The great Amazon Basin, the largest undifferentiated region on the continent, is still mainly restricted to a collecting economy, with emphasis on Brazil nuts. (H. J. Bx.)

III. ANTHROPOLOGY

Recent archaeological discoveries indicate that man first arrived in South America more than 20,000 years ago. The predominance of large stone projectile points in some Paleo-Indian sites, and of chopping tools in others, suggests that some of the early migrants were primarily hunters, while others specialized in the gathering of wild plant foods.

The introduction of agriculture, which dates from at least 2600 B.C., initiated a relatively rapid cultural evolution that reached different levels in different parts of South America. Some areas remained at the hunting-fishing-gathering level of subsistence. Others adopted agriculture and made certain other cultural advances but continued to live in relatively small and independent communities. Still others developed further, both technologically and socially, and attained the level of chiefdoms. The greatest cultural development of all, which occurred only in the central Andes, was the rise of a full-scale civilization. In the *Handbook of South American Indians*, Julian Steward called these four culture types Marginal, Tropical Forest, Circum-Caribbean, and Andean, respectively, and this usage will be employed in the following discussion.

1. Racial Type.—The Indians of South America all belong to the Mongoloid racial stock and came originally from Asia by way of North America. The distribution of various morphological characteristics among them has been studied, but as yet no proposed classification of physical types has been generally accepted. The study of blood groups has revealed a high predominance of O, a small amount of A, and practically no B or AB.

2. Linguistic Groups.—The number of language families recognized for South America has varied. A. F. Chamberlain distinguished 83 families in 1913, Paul Rivet 77 in 1924, Čestmír Loukotka 94 in 1935, and J. Alden Mason 91 in 1950. It was generally conceded, however, that more detailed study would decrease the number of independent linguistic groups recognized. In 1956 Joseph Greenberg proposed a tentative classification that reduced the languages of South America to three phyla or superfamilies: Macro-Chibchan (including Chibchan); Andean-Equatorial (including Araucanian, Arawak, Aymara, Chon, Quechua, and Tupí-Guaraní); and Ge-Pano-Carib (including Carib, Ge, Guaycuruan, and Panoan).

3. Marginal Culture.—Societies subsisting entirely or largely on wild food sources were found over large areas of South America in aboriginal times. These groups occupied habitats unsuitable for agriculture. From the Chaco south, it was generally too dry or too cold for cultivated plants; in eastern Brazil the savanna could not be readily farmed with the tools available. The differences in culture level among Marginal peoples resulted largely from differences in the richness of the wild foods in their respective environments.

The relative importance of hunting, fishing, and gathering varied with the habitat. Among the Tehuelche of Patagonia the hunting of guanacos and rheas provided almost the only source of food. Other tribes, especially those of the Chaco, were more generalized in their subsistence activities and consumed a wide variety of vegetable products along with meat and fish. North of the pampas the principal hunting weapon was the bow and arrow; from the pampas south to Tierra del Fuego it was the bolas (*q.v.*).

Community size was generally small among the Marginal societies. The Yahgan, living in a difficult environment, had local groups of about 15 or 20 persons. On the other hand, after the acquisition of the horse some of the tribes of the Chaco, pampas, and Patagonia had bands of 500 or more. The average size for a Marginal local group was probably about 50. The exploitative nature of their subsistence economy meant that Marginal peoples had to move their campsites periodically to stay abreast of their food supply. With almost no exception, each local group was economically and politically independent. Intertribal gatherings occurred for ceremonial occasions, but they were never permanent. Chieftainship might or might not be hereditary. The chief was usually weak except in societies where achievement in warfare served to increase his power.

Dwellings varied according to the materials available. The guanaco hunters of the south used tents made of animal skins, while the Amazonian Marginals had thatched huts.

Pottery was usually absent, and containers were leather bags, baskets, or gourds. Loom weaving occurred only in the Chaco, where it had diffused from the Andes. Land travel was by foot

except among tribes that adopted horse riding from the Spaniards. Watercraft were absent among most Marginal peoples, but a few had bark canoes, rafts, and, in Patagonia, coracles. The band constituted a kinship as well as a local group. Sibs and moieties occurred only among the Ge groups of Brazil, and social classes only among a few horse-riding Indians of the Chaco. Marriage tended to be exogamous, and postmarital residence patrilocal. Matrilocality was found where women's share in subsistence was important, as in the Chaco. There were no specialists except the shaman and the chief, who were nevertheless not exempt from subsistence.

Available evidence indicates that warfare was sporadic and was waged for revenge, witchcraft, and trespass. It rarely involved the capture of prisoners, the taking of land, or the exacting of tribute. The basis of religion was a belief in a number of gods and spirits. Rituals and ceremonies were concerned principally with matters of subsistence and with life crises. The shaman was a diviner and curer.

4. Tropical Forest Culture.—The occurrence of Tropical Forest culture coincided closely with the distribution of the tropical rain forest in South America. In the absence of all but the crudest farming tools, the slash-and-burn cultivation characteristic of this culture could be carried on only in areas of loose forest soil. A plot of land was cleared during the dry season, burned just before the rains, and planted with a digging stick. Cultivated plants included corn, sweet potatoes, beans, and peanuts, but the staple crop was cassava. The hydrocyanic acid found in the poisonous variety of this tuber was removed with the help of a number of devices, the best known of which was a sleevelike press called a *tipití*. Cakes, gruel, and a fermented drink were made from cassava meal.

Fishing was an important supplement to the diet and was carried on with bow and arrow, fish traps, and by drugging with poisonous vines. The hunting of large game was done chiefly with the bow and arrow and with traps; the blowgun, when present, was used mainly against birds and small mammals.

Tropical Forest villages were semipermanent and were located in or near forests and close to watercourses. Houses had a pole framework and were covered with thatch, generally of palm. Usually several families lived in a house, and in some areas the entire village might consist of a single large house. The hammock was used for sleeping by almost all tribes. Community size varied within fairly wide limits. In the Peruvian Montaña, Campa and Amahuaca settlements averaged about 15 persons; along the coast of Brazil, Tupinambá villages often had 600 or 800. The average for the entire area was perhaps 150 to 200. Each village had a chief and occasionally subchiefs or house chiefs. Political authority was not strong except in warlike societies. Chieftainship was sometimes hereditary but was generally only loosely so. Villages were autonomous, and confederacies between them were rare and never permanent.

Baskets were of many kinds and were made by a variety of techniques. Wood carving was common, the most representative item being the one-piece wooden stool. Featherwork was universal and was often beautiful and spectacular. Most tribes had some form of water transport. The dugout canoe was the most typical watercraft, but some tribes had only bark canoes or rafts.

Marriage was generally exogamous, although a number of groups permitted local endogamy. Polygyny occurred, especially for such persons as the chief or shaman. Clans were found among some tribes, such as the Mundurucú and Tucuna, but were not characteristic of the area as a whole. The practice of couvade was very prevalent. According to this custom, a man took to his hammock or underwent other restrictions at the time his wife gave birth.

Warfare was common to most tribes of the region. Some, like the Mundurucú and the Island Carib, devoted a great deal of time to it. The securing of head trophies and the practice of cannibalism were frequent adjuncts of warfare. Offensive weapons consisted primarily of bows and arrows and clubs. For defense, villages were often palisaded and trails sometimes set with poisoned stakes. Characteristically, warfare was for revenge rather than for economic gain.

Ceremonialism and ritual were well developed. Puberty rites often included tests of courage for the initiates, such as the ant ordeal, whipping, and being bitten by an anaconda. Narcotics of many kinds were used, including tobacco, coca, and *paricá* snuff. Tobacco was particularly important because it was frequently smoked by the shaman in connection with divination and curing.

5. Circum-Caribbean Culture.—Around the Caribbean Sea from Cuba to Nicaragua and in the northern Andes there were societies that resembled Tropical Forest tribes in certain details, *e.g.*, use of the hammock, arm and leg ligatures, and drinking bouts, but were at a distinctly higher cultural level. Agriculture was carried out intensively. The digging stick was still the basic farming tool, but such techniques as planting in mounds and the use of urine fertilizer increased productivity. Irrigation canals and terracing were employed in some places. Corn was the most important crop in the mountain valleys, while cassava was the principal crop in the lowlands. Fruit orchards were also common. Fishing was of importance locally, but hunting was of little significance.

Communities of 2,000 or 3,000 were not uncommon. Towns were generally permanent and well laid out, usually around a large central plaza. For defensive purposes they were often palisaded. Flagstone-paved roads sometimes connected larger towns. Houses were of pole-and-thatch construction rather than of stone. Settlement was often dense: one valley occupied by the Tairona was reported to contain 250 towns. The Anserma had a total population of more than 40,000.

Confederacies and petty states were common in the area. The chief of such political units had considerable authority and prestige. He lived in a special house and was carried about in a litter or on the shoulders of retainers. Some of his many wives and concubines were buried with him after being stupefied with *chicha*, an alcoholic drink made from fermented corn. Society was stratified, with nobles, commoners, and slaves ranking below the chief. Slaves were prisoners of war or persons caught stealing. Commoners could rise to the status of nobles by valorous exploits in war.

There were craft specialists in pottery making, weaving, goldworking, and stone carving. Metallurgy was particularly well developed. The Quimbaya of Colombia and the inhabitants of Coclé in Panama produced outstanding goldwork. *Tumbaga*, an alloy of gold and copper, was made into utilitarian and luxury goods. Trade was very important, and markets existed. The principal items of exchange were salt, fish, gold, and cotton. Water travel was in dugout canoes along the coasts and on the larger rivers. To cross narrow mountain valleys suspension bridges of vines were constructed. Attire varied considerably. Some tribes went naked or used a simple breechclout. Others wore elaborate cotton mantles or tunics. Class distinctions were reflected in differences in dress.

Marriage was customarily by purchase, and wealthier individuals had several wives. Ceremonies surrounded life crises, especially death. The dead were disposed of in various ways, depending upon locale and social status. Commoners were often cremated and buried in urns; chiefs were interred in stone-lined shaft graves. Funerals were often accompanied by drinking bouts in which huge quantities of *chicha* were consumed.

Warfare prevailed except among a few tribes, such as the Taino. It was waged for conquest of land, to exact tribute, and to capture slaves. Organized military tactics were employed, in some cases directed by special war chiefs. When confederacies fought each other, thousands of warriors might be involved. The favorite weapon was the *macana*, a wooden sword-club, but poisoned arrows and fire arrows were also used. Captured trophy heads were displayed on posts, and the bodies of notable enemies were dried over the fire and stuffed with ashes. Cannibalism was widespread.

The sun and moon were commonly regarded as important deities, and shrines dedicated to them were erected on mountain tops. In trying to influence the supernatural, priests made offerings and sacrifices. Idols, like the *zemis* of the Taino, were used to com-

municate with the spirit world.

6. Central Andean Culture.—The Inca Empire of the central Andes was the largest and best organized state in aboriginal America. It encompassed about 1,000,000 sq.mi. and had a population conservatively estimated at 6,000,000.

At the base of Inca civilization was a system of intensive agriculture. Cultivated land was extended to the fullest by terracing mountain sides and irrigating deserts. The *taclla,* or foot plow, was used to turn the soil. Llama dung and guano were employed as fertilizer. Able-bodied men farmed plots for the state and the church in addition to their own. Surpluses went into granaries to support priests and other specialists, to maintain the army, and to serve as emergency rations in times of famine. More than 40 different species of crop plants were grown, corn being the most important. In high altitudes, quinoa and potatoes were raised. The principal domestic animal was the llama, which served as a beast of burden and as a source of meat, wool, hides, and fertilizer.

The household head, called a *puric,* was allotted a certain portion of land at marriage, which was added to at the birth of each child. The basic unit of settlement was the *ayllu,* an endogamous agricultural community of about 100 families. *Ayllus* were grouped into districts, districts into provinces, and provinces into quarters. The Inca Empire as a whole was called Tawantinsuyu, "four parts of the world." An official governed each administrative unit of the empire. The smaller administrative units were governed by *curacas,* or lower nobility, while the larger units were ruled by members of the higher nobility, called *orejones* ("big ears") by the Spaniards because of the earplugs they wore as symbols of their office. At the top of the hierarchy was the Sapa Inca, or "sole lord." The Sapa Inca was considered the descendant of the sun-god and, thus, divine himself. No one could look directly at the Sapa Inca, and he never drank from the same vessel or wore the same garment twice. He had hundreds of concubines, but his rightful heir was his son by his eldest sister.

There existed a system of forced labour known as the *mita,* by means of which men were drafted for public works. As many as 30,000 men were mobilized at one time to work on roads, fortresses, etc. The *mitimae* system involved forced resettlement and was employed to ease population pressure and to disperse rebellious peoples.

The army, the instrument for expanding the empire, was also recruited by draft. Draftees served either as bearers or as fighting men. On the march the army used the excellent highways and were quartered and provisioned in *tambos,* or fortified storehouses, at points along the road. Military formations and tactics were used during warfare. Soldiers were divided into slingsmen, bowmen, spearmen, etc.

Arts and crafts were practised by full-time specialists and showed great technical excellence. Pottery making, metallurgy, and weaving were especially well developed. Inca architects built solid and imposing structures of dry stone masonry. No system of writing was known, but the *quipu* permitted the keeping of detailed records.

The system of roads consisted of two arterial highways, one along the coast and the other through the highlands. Shorter roads connected the two at various points. Couriers running in relays carried messages or goods rapidly. The 1,400 mi. between Cuzco and Quito could be traveled in five days.

The Inca believed in a creator god, Viracocha, and in a number of other deities, including the sun. The assistance of the gods was invoked at public ceremonies connected with such events as the agricultural cycle and warfare. Many shrines and temples existed throughout the empire. *See* ANDEAN CIVILIZATION; INDIAN, LATIN-AMERICAN. (R. L. C.)

IV. POPULATION

The population of South America was estimated (1963–64) to be 160,000,000. Together with the population of Middle America this gave a total population for Latin America of approximately 235,000,000, which was about 23,000,000 more than that of Anglo-America. Of the population of South America, nearly 50% is in Portuguese-speaking Brazil; and 0.6% is in the three Guiana colonies.

The population of South America is arranged in a pattern of more or less separate clusters. In the centre of each cluster there is a central city, of which eight in South America (and ten in Latin America as a whole) had passed 1,000,000 in 1963. Between these clusters there are only scanty populations. A large part of the continent has less than one person per square mile. Most of the clusters are located near a coast, and the vast interior, mostly in Brazil, is one of the world's larger areas of very thin population. Table I shows how this population is divided among the ten countries and three colonies.

TABLE I.—*Area and Population of South America*

Country	Area (io sq.mi.)	Population (1963–64 est.)	Densities per sq.mi.
Total	6,879,122	159,942,730	23.4
Argentina	1,072,067	22,023,431	20.5
Bolivia	424,162	3,653,000	8.6
Brazil	3,286,470	79,937,000	24.3
British Guiana	83,000	621,390	7.5
Chile	292,256	8,369,000	28.6
Colombia	439,512	15,434,090	35.1
Ecuador	105,685	4,877,000	42.2
French Guiana	35,135	35,000	1.0
Paraguay	157,047	1,936,000	12.3
Peru	496,222	11,854,000	23.9
Surinam	63,251	320,000	5.1
Uruguay	72,172	2,556,020	35.1
Venezuela	352,143	8,426,799	23.9

The population of Latin America is growing faster than that of any other part of the world. Between 1920 and 1960 the population increased 126.3%, as compared with 92.4% for second-place Australia and New Zealand. Table II shows the medium

TABLE II.—*Medium Estimates of Future Population by Subregions*
(in 000,000)

Region	1960	1975	2000
Low-latitude South America	107.0	163.0	339.0
Middle-latitude South America	32.9	41.7	55.8
Latin America as a whole	205.8	304.1	592.8

Source: Population Branch, United Nations.

estimates made by the United Nations for the future population of Latin America by subregions. It is important to note that in all of Latin America the most rapid jump in the rate of increase took place between the pre-World War II period and the period after the war. The decline in the death rate that began generally in the late 1930s is the major cause of this accelerated gain in population. During the 1920s the death rate for 11 Latin American countries for which adequate and comparable data exist was lowered by 6%, but during the period from 1945 to 1954 the death rate was lowered by 17.1%. With the continued high birth rates in these countries the result was a "population explosion." The decrease in the death rate that has taken place in all the Latin American countries is essentially the result of the improvement of medical services and of public hygiene and has accompanied the decrease of illiteracy. The virtual elimination of malaria, the great decrease in smallpox cases, the effective attack on yellow fever and hookworm—all these medical gains have had the effect of saving the lives of children.

The population of South America is quite diverse in its racial composition. There are certain countries in which the proportion of Indians remains high. In Ecuador, Peru, Bolivia, Paraguay, and the southern part of Colombia, people of unmixed or almost unmixed Indian ancestry make up more than half of the population. In Argentina, Uruguay, and southern Brazil (from São Paulo state southward) the population is more than 95% pure European in origin. It is to this part of Brazil and to Argentina that the great stream of European immigration into South America has gone. Since 1800 an estimated 12,000,000 immigrants have arrived in Latin America. Of these about 4,000,000 came from Spain, 4,000,000 from Italy, and 2,000,000 from Portugal. The remainder include a variety of peoples: Germans, Poles, and other Europeans; Syrians; Chinese; and Japanese. There are also parts of South America in which there is a high proportion of Negroes. The Negroes came first to the Portuguese colonies of Salvador

and Recife in northeastern Brazil to work the sugarcane plantations established there during the 16th century. This area still has a large proportion of Negroes and mulattoes. After the beginning of the 18th century, when gold and diamonds were discovered in Minas Gerais, many Negroes were brought as slaves to that part of Brazil; this region thus contains a considerable proportion of Negroes in its present-day population. Similarly, deep in the interior, near Cuiabá, there is a concentration of Negroes around old colonial goldworkings. Negroes were also brought as slaves to sugarcane areas in Venezuela and in the Cauca Valley and Caribbean coastal lowlands of Colombia. Elsewhere, the population of South America is a mixture chiefly of Indians and Spaniards known in Spanish America as mestizo. Table III gives the best available estimates of racial proportions.

TABLE III.—*Estimated Racial Proportions*
(in %)

Country	European	Indian	Negro	Mestizo	Oriental
Argentina	97	3			
Bolivia	15	53		32	
Chile	30	5		65	
Colombia	20	7	5	68	
Ecuador	15	60	3	22	
Paraguay		(97% Indian and mestizo)			
Peru	15	46		38	1
Uruguay	90		2	8	
Venezuela	20	7	8	65	

Throughout Latin America there is a tendency for the densely populated areas to increase in population more rapidly than the average and for the thinly peopled areas to become even more sparsely populated. There are, to be sure, a few zones of pioneer settlement where new lands are being cleared and settled, but by far the greater part of the population movement is away from these frontiers rather than toward them. The movement generally is from the rural areas into the cities. Urban growth, especially after World War II, increased rapidly. The first city in South America to reach a population of 1,000,000 was Buenos Aires in 1906. Since then Rio de Janeiro, São Paulo, Santiago, Lima, Caracas, Bogotá, and Montevideo have all passed this mark. Not only have the cities increased in size, but also the proportion of the total population of each country living in the central cities has increased. This urban growth is a measure of the rapidity with which the Industrial Revolution is sweeping over Latin America. (P. E. J.)

V. HISTORY

When the mainland of South America was sighted by Columbus on his third voyage, in 1498, it was inhabited by numerous Indian tribes of widely varying cultures. By 1560 Spain had overrun much of the continent, conquered a large majority of the native inhabitants, discovered mines that gave the area a firm economic base, established the Roman Catholic Church, and evolved an administrative system that did not undergo any major modification until the 18th century, when threats from Great Britain forced a reorganization. Meanwhile, the Portuguese in Brazil developed a tropical-agriculture colony based upon Negro slavery. The British, French, and Dutch established their claims to certain of the "leftover" areas of the continent in the 17th and 18th centuries. The Spanish colonies won their independence as a result of a costly 15-year struggle (1810–25), while Brazil separated from Portugal with a minimum of effort in 1822–23 and formed an empire, which lasted until 1889.

For approximately one century after political freedom had been won, the Spanish republics were ruled—often by force—in the interest of a narrow elite. After 1875, European immigrants and foreign capital and technicians contributed significantly to the material development of the nations.

Since World War I, several of the republics have undergone major transformations. New middle classes and working classes have emerged and have been incorporated into the political life. Economically, the republics have sought, with some success, to widen the range of their production and have made the modern factory the symbol of progress. For a full treatment of the history of South America, *see* LATIN AMERICA. (J. J. J.)

SOUTH AMERICAN LANGUAGES. South America and the West Indies apparently had more than twice as many aboriginal languages as were found in Central and North America. It seems also that the South American languages were more diverse than those to the north. By the late 1950s linguists had enumerated no more than 40 distinct stocks from Central and North America, but the estimated number of stocks for South America ranged from 70 to more than 100. The aboriginal languages of the West Indies belonged to two stocks, Arawakan and Cariban, both of which became widespread in South America. (For a general discussion of the languages of the Americas and a brief description of the methods of classifying languages into stocks or families, *see* AMERICAN ABORIGINAL LANGUAGES; *see* also CENTRAL AND NORTH AMERICAN LANGUAGES.)

It is probable that the extreme linguistic diversity noted for South America rested in large part on two factors: (1) a lack of detailed knowledge of most South American languages; and (2) a method of classification based mainly on the comparison of brief vocabularies which was wholly inadequate for the establishment of more distant relationships. It appeared that when more was known of the languages of South America and proper methods could be applied to their classification, it would be quite certain that the number of stocks would be considerably reduced.

The larger and the more widely distributed linguistic stocks of South America and the West Indies would appear to fall into the following groups:

1. Chibchan. These languages, in the aboriginal period (*i.e.*, roughly A.D. 1500 and before), occupied most of Colombia, and extended southward to the latitude of Guayaquil in Ecuador and northward to include Panamá and most of Costa Rica and Nicaragua. One of the languages served a powerful Indian empire, with its centre near Bogotá. Later, but before the Spanish conquest in 1533, this empire was defeated by the Incas of Peru and the southernmost Chibchan languages were driven back or replaced by Quechuan languages (*see* no. 6, below). A number of Chibchan languages were still spoken at mid-20th century, mainly by small Indian groups in Colombia and Central America, but these were gradually giving way to Spanish.

Some investigators (notably J. Alden Mason, 1950) proposed a Macro-Chibchan stock, composed of Chibchan plus the Misumalpan languages of Central America (*see* CENTRAL AND NORTH AMERICAN LANGUAGES). The evidence for this grouping was weak in the late 1950s.

2. Cariban. Spread, in preconquest times, from the Greater Antilles to central Mato Grosso in Brazil, and from eastern Peru to central Pará, Braz. Most of the languages, however, were found north of the Amazon river, in the territory now occupied by northern Brazil, the inland areas of the Guianas and Venezuela and lowland Colombia.

West Indian Cariban was extinct but a considerable number of Cariban languages, perhaps as many as 200, survived in South America at mid-20th century.

3. Arawakan. This is the largest and most widespread of all South American linguistic stocks. During preconquest times, Arawakan languages were found, in a number of disconnected areas, from Cuba and the Bahamas southward to the Gran Chaco and the sources of the Xingu river in southern Brazil, and from the mouth of the Amazon river to the eastern foothills of the Andes. Some of the languages of this group were extinct or nearly so (this applied particularly to the West Indies) at mid-20th century, but more than 200 were still spoken in South America.

4. Ge. The languages of this group, in modern and aboriginal times, occupied a solid large area in eastern Brazil. About 50 languages survived at mid-20th century although most of these were slowly being replaced by Portuguese. Mason (1950) combined Ge and eight smaller stocks spoken near it into a Macro-Ge grouping, which, in his terms, was still "far from proven." The eight added stocks were Caingangan, Camacanian, Machacalian, Purian, Patacho, Malali, Caropa and Botocudo.

5. Tupi-Guarani is the second largest linguistic stock of South America. It has two major divisions: Tupian, which was spoken in eastern Brazil, mainly along the coast from Uruguay to the mouth

of the Amazon, and inland along the Amazon to the foothills of the Andes; and Guaranian, which was spoken in Paraguay and the neighbouring region of Argentina.

Languages of both divisions were used by the first European traders and missionaries as contact languages in their dealings with the Indians. About 120 Tupí-Guaraní languages were still spoken at mid-20th century, but these, as well as the two more widespread contact languages, were slowly giving way to Portuguese and Spanish.

Macro-Tupí-Guaraní, proposed very tentatively by Mason (1950), combines Tupí-Guaraní with two smaller stocks, Huitoto-Miranyan and Zaparoan in eastern Ecuador and Peru.

6. Quechuan. Archaeological and historical studies suggest that the Quechuan languages probably occupied a small area in the southern Peruvian highlands until, about 1450, they were spread over a far larger region by the rapid conquests of the Inca empire. When Pizarro conquered the Incas in 1533, Quechuan languages were spoken over most of the west coast of South America, from southern Colombia to central Chile, and eastward from the coast to the borders of the Amazonian basin. In the 1950s Quechuan languages were spoken by several million people in roughly the same area, many of whom spoke no other language. Nevertheless, the Quechuan languages were slowly losing ground to Spanish which, by reason of its position as the language of government and education, plays an increasingly important role throughout the area. The modern Quechuan languages, about 28 in number, are said to be very much alike.

7. Aymaran. The languages of this group were spoken over a fairly large region in the southern Peruvian highlands and adjacent areas of Bolivia. Inca conquests probably reduced the Aymaran area somewhat, but there remain almost 500,000 speakers of the 14 Aymaran languages.

J. Alden Mason (1950) suggested that Quechuan and Aymaran might be combined into a single stock, which he called Quechumaran. The evidence for this relationship was not impressive in the late 1950s.

Many of the remaining languages of South America are grouped into 45 stocks, which are more restricted in size and distribution than the ones listed above. These are:

In Venezuela: Chirianan, Cuica-Timote, Guahiban, Guararan, Jirajaran and Salivan.

In Colombia: Mocoa, Pamiguan, Puinavean and Tucanoan.

In Brazil: Arauan, Borotuque, Carajan, Cariri an, Catuquinan, Guaitacan, Huanhaman, Muran, Nambicuran, Panoan, Peba-Yaguan and Zamucoan.

In Ecuador: Canelan, Jivaroan and Omurano.

In Peru: Aguanoan, Cahupanan, Munichean, Sabelan and Yunca-Puruhan.

In Bolivia: Chapucuran, Chiquitoan, Mosetenan, Tucanan and Yuracarean.

In Paraguay (and adjacent parts of Brazil and Argentina): Guaicuruan, Mascoian and Mataco-Maca.

In Chile: Alacalufan, Araucanian and Ataguitan (also in adjacent parts of Argentina).

In Argentina: Chon (or Tehuelchean), Comechingonan, Huarpean and Lule-Vilelan.

This listing by no means accounts for all the languages of South America. P. Rivet and C. Loukotka (1952) listed 108 stocks—the list above contains only 52. But the additional "stocks" were actually very little known: McQuown estimated (1955) that there were over 600 languages in South America on which there were too little data to permit even a tentative classification. (*See* also INDIAN, LATIN-AMERICAN.)

South America, as late as the mid 1950s, was unmatched in its linguistic diversity. The classification given above must therefore be regarded as highly tentative, and subject to drastic revision when better knowledge of the area becomes available.

BIBLIOGRAPHY.—J. Alden Mason, "The Languages of South American Indians," vol. 6, pp. 157–317 of *Handbook of South American Indians*, U.S. Bureau of American Ethnology *Bull. 143* (Washington, D.C., 1950); N. A. McQuown, "The Indigenous Languages of Latin America," *Amer. Anthrop.*, 57:501–570 (1955); P. Rivet, G. Stresser-Pean and C. Loukotka, "Langues de l'Amerique" in A. Meillet and M. Cohen, *Les Langues du monde* (Paris, 1952). (H. HR.)

SOUTHAMPTON, HENRY WRIOTHESLEY, 3RD EARL OF (1573–1624), Shakespeare's patron, was born at Cowdray, Sussex, on Oct. 6, 1573. He succeeded his father in 1581 and became a royal ward under the care of Lord Burghley. Educated at St. John's college, Cambridge, and Gray's Inn, London, at 17 he was presented at court, where he was favoured by the queen and befriended by the earl of Essex. Southampton became a munificent patron of poets: Barnabe Barnes appended a sonnet in his honour to *Parthenophil and Parthenophe* (1593); Thomas Nashe dedicated *The Unfortunate Traveller, or, the Life of Jacke Wilton* (1594) to him; and Gervase Markham his poem on Sir Richard Grenville's last fight (1595).

Southampton is best known, however, as the patron of Shakespeare. When he dedicated *Venus and Adonis* (1593) to the earl he wrote with respect but no special intimacy, but the dedication of *Lucrece* (1594) is very different: "The love I dedicate to your lordship is without end . . . What I have done is yours; what I have to do is yours; being part in all I have, devoted yours." Sir William Davenant stated that Southampton gave Shakespeare £1,000 to complete a purchase. If the sonnets were addressed to Southampton, the earlier ones, urging marriage, must have been written before the beginning in 1595 of his intrigue with Elizabeth Vernon, one of the queen's waiting women, which culminated with their hasty marriage in 1598, incurring the queen's wrath and leading to their brief imprisonment.

Meanwhile in 1596 and 1597 Southampton had accompanied Essex on his expeditions to Cádiz and to the Azores. In 1599 he went to Ireland with Essex, who made him general of his horse, but the queen insisted that the appointment be canceled and Southampton returned to London. He was deeply involved in the Essex rebellion, on the eve of which he induced players at the Globe theatre to revive *Richard II*, a play dealing with the deposition of a king, in order to stir up the populace. He was tried for treason on Feb. 19, 1601, and condemned to death, but his sentence was commuted to life imprisonment through the intervention of Sir Robert Cecil.

On the accession of James I, Southampton resumed his place at court. He was made a knight of the Garter and captain of the Isle of Wight in 1603, and was restored in blood by act of parliament in 1604. In 1603 he entertained Queen Anne with a performance of *Love's Labour's Lost* by Richard Burbage and his company, to which Shakespeare belonged.

Southampton was an active member of the Virginia and East India companies. A born fighter, he was engaged in more than one serious quarrel at court and was imprisoned for a short time in 1603. He was a volunteer in support of German Protestants in 1614, and in 1617 proposed fitting out an expedition against the Barbary pirates. He became a privy councilor in 1619 but fell into disgrace through his determined opposition to the favourite, the duke of Buckingham. In 1624 he and his elder son volunteered to fight for the United Provinces against Spain, but on landing they were attacked with fever and Southampton died at Bergen op Zoom on Nov. 10, 1624, a few days after his son.

See "Memoirs of Henry Wriothesley, the third Earl of Southampton," in James Boswell (ed.), *Shakespeare* xx, pp. 427 ff. (1821).

(RD. C. G.)

SOUTHAMPTON, THOMAS WRIOTHESLEY, 1ST EARL OF (1505–1550), was lord high chancellor of England under Henry VIII and Edward VI. He was born on Dec. 21, 1505. In 1537–38 he obtained extensive lands between Southampton and Winchester at the dissolution of the monasteries. He was made one of the king's principal secretaries in 1540, was knighted in the same year, and was created Baron Wriothesley in 1544. He became chancellor in 1544 in which capacity he became notorious by his proceedings against Anne Askew. He was one of the executors of Henry VIII's will, and was created earl of Southampton in Feb. 1547. However, he had committed a technical offense in appointing four persons to relieve him of his judicial duties as lord chancellor and he was deprived of his office in March, when he left the privy council. Readmitted to the council by 1548, Southampton helped Warwick in Oct. 1549 to bring about the fall

of the protector Somerset. He was dismissed from the privy council again in Feb. 1550 and died on July 30, 1550.

SOUTHAMPTON, a port, city (1964), county borough and parliamentary borough of Hampshire, Eng., lies 76 mi. S.W. of London by road. Pop. (1961) 204,822. Area 15.1 sq.mi. The town stands near the head of Southampton water, an inlet of the English channel, on a peninsula formed by the rivers Test and Itchen. It is the third largest in importance among English ports and the first in regard to passenger traffic. In A.D. 43 there was a Roman settlement, Clausentum, on the east bank of the Itchen, and inscribed stones, coins, pottery, etc., have been found. Southampton (Hamtune, Suhampton) superseded the Saxon Hamtune and was a royal borough before 1086. The earliest town charter was given by Henry II, probably in 1154–55, but the borough was not incorporated until 1445. Henry VI, in 1447, created Southampton a county in itself. A charter which remained in force until 1835 was granted by Charles I in 1640. It was from Southampton that the Mayflower sailed with the Pilgrim Fathers to America. The town has returned two members to parliament since 1295. It is a quarter sessions borough. Southampton gives name to a bishopric suffragan to Winchester. Natives of the town include Isaac Watts (1674–1748), Sir John Millais (1829–96) and Earl Jellicoe (1859–1935).

There are remains of the Norman walls which were strengthened by towers and had six main gates. An old house known as King John's palace is one of the earliest 12th-century domestic buildings in England. The south (God's House gate) and west town gates date from the early 14th century; Bar gate (the north gate), as it stands, is mostly later and retains good Decorated work. St. Michael's church has an 11th-century tower and arches and a font of Tournai marble (*c.* 1150). The spire (1745) was raised a further 9 ft. in 1877 as a mark for mariners. The priory of St. Denys, an Augustinian foundation of 1124, gives its name to a suburb by the Itchen. Netley (*q.v.*) abbey is 3 mi. southeast.

The old guildhall, long used as a courthouse, is over the archway of the Bar gate. Tudor house, a mediaeval merchant's residence, is the museum. King Edward VI Grammar school was founded in 1550 but now occupies modern buildings near the common. Taunton's school was founded in 1760 and is now in large modern buildings at Highfield. The University College of Southampton, evolved from the former Hartley institute, was removed to Highfield in 1914. It received its charter as a university in 1952. The town suffered severely from air raids in World War II, many houses and three-quarters of the old district in the centre having been devastated and important port installations destroyed. The civic centre, completed in 1939, comprises the new guildhall; municipal offices; police headquarters and law courts; and the public library and art gallery. The tower is a landmark for shipping and contains a clock and bells. There are a large sports ground (268 ac.), a common and several parks.

The Port and Its Trade.—The modern importance of the port begins with the creation of a pier and harbour commission in 1803 and the erection of the Royal pier in 1831. The prosperity of the town was enhanced by the opening of railway communication with London in 1840. The fine natural harbour has the advantage of a double tide, the tide of the English channel giving it high water first by way of the Solent and two hours later by way of Spithead. In 1950 the docks were taken over by the Docks and Inland Waterways executive. The older docks lie at the confluence of the Test and Itchen rivers; large modern extensions lie to the northwest of these alongside the Test river. The old docks cover more than 200 ac. and in addition to the river quays there are three large tidal basins, the Ocean dock, Empress dock and the Outer dock, leading to the Inner dock. The new docks, with 1½ mi. of deep-water quay facing the Test river, include the King George V dry dock, one of the largest in the world.

Southampton is England's chief port for express transatlantic services and is also used by ships bound for South America, East Africa, Australia, New Zealand and the far east. There are cross-channel services by British Railways ships and a ferry to the Isle of Wight. Southampton airport is at Eastleigh, 5 mi. northeast. Flying boat services operate from the marine air terminal.

Principal industries are shipbuilding and ship repairing, building and contracting, dock and harbour services, distribution services and marine engineering. There is also a wide range of light industries. A large new oil refinery was opened at Fawley on Southampton water in 1951.

SOUTH ARABIA, FEDERATION OF, comprises a number of states belonging to the Protectorate of South Arabia (*see* SOUTH ARABIA, PROTECTORATE OF), and Aden (*q.v.*). Originally known as the Federation of Arab Amirates of the South, it was formed in February 1959 when the rulers of six of the states in the western area of the protectorate—the amirates of Beihan (Baihan) and ad Dali', the Fadhli, 'Audhali, and Lower Yafa'i (Yafi') sultanates, and the Upper 'Aulaqi sheikhdom—signed a federal constitution and also a treaty of friendship and protection with the United Kingdom, which agreed to provide military and financial assistance. Lahej (Lahij, Lahj, or 'Abdali sultanate) joined the federation in October 1959, and in February 1960 the state of Dathinah, the Lower 'Aulaqi sultanate, and the 'Aqrabi sheikhdom acceded. In April 1962 the Wahidi sultanates of Balhaf and Bir'Ali (in the eastern area of the protectorate) joined, and in May 1962 the federation changed its name to the Federation of South Arabia. Aden acceded to the federation, not without some opposition, in January 1963, followed in March of that year by the Haushabi sultanate and the Saqladi sheikhdom of Sha'ib. In the mid-1960s the federation had an area of about 60,000 sq.mi. (155,000 sq.km.) and a population of 760,000. Its administrative capital is at al Ittihad.

Most member states are equally represented in the federation's legislative body, the Federal Council, but Aden has more than one-quarter of the total membership. Each state selects its representatives by its own means; in Aden they are elected on a franchise confined to Adeni citizens—a category difficult to define because of the high rate of immigration and a largely floating population. The executive federal Supreme Council consists of 14 ministers elected from the Federal Council. The federation has exclusive legislative and executive authority in such matters as external affairs, defense, internal security, posts, currency, and health, subject to an order in council providing that the authority of the high commissioner for Aden shall prevail in regard to Aden's defense, internal security, and external affairs. The federation's first constitutional conference, held in London in June–July 1964, envisaged independence in 1968 with some modification of the constitution.

The armed forces consist of the Federal Regular Army (formerly the Aden levies) and the Federal National Guard, both recruited from the tribes. In these and the public services Arabization is being effected. Annual revenue in the mid-1960s amounted to nearly £7,000,000, of which about £4,000,000 consisted of grants by the U. K. government. Excise is levied, mainly at Aden, on alcoholic liquor, tobacco, gasoline, kat, and carbonated beverages. A major step toward integration was the abolition of customs duties between member states in April 1963. The federation's development schemes include an ambitious road construction program.

Outside Aden occupations are mainly subsistence farming, stock raising, and fishing. Cotton, notably from the Abyan Scheme, and the provision of fruit and vegetables for the Aden market, mainly through cooperative marketing societies, are successful examples of development. These are supplemented by the remittances from the federal forces and workers in Aden and abroad. Over most areas the standard of living and public health have greatly improved since World War II.

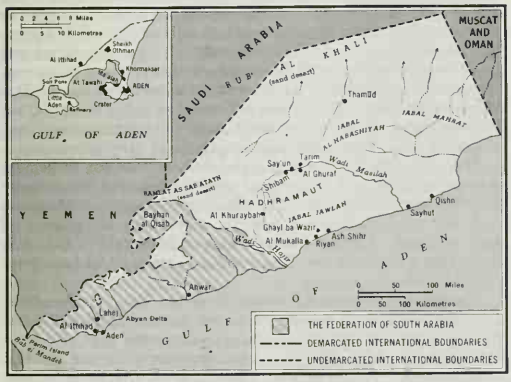
The first proposals for federation, put forward in January 1954, were alleged by Yemen to be counter to the Anglo-Yemeni treaty of 1934 and the 1950 *status quo* agreement. This argument had little justification, but Yemen had long laid claim to the South Arabian territories, and the ruler of Yemen, Imam Ahmad, embarked on a policy of infiltrating dissidents into the protectorate to undermine security. In 1958 his accession to the United Arab Republic determined the rulers of the states of the protectorate to reconsider the moribund federation proposals. For the time being, however, acts of aggression by Yemen ceased. Following

the military coup in San'a' in 1962, the republican part of Yemen, where Egyptian influence was strong, resumed hostile activities. Egypto-Yemeni propaganda intensified its vigorous radio campaign of vilification of the federation as "feudalist." An organization was formed under Egyptian aegis to infiltrate trained terrorist bands into the federation but with little sustained success.

BIBLIOGRAPHY.—*Treaty of Friendship and Protection Between the United Kingdom and Federation of the Arab Emirates of the South,* Cmd. 665, *Accession of Aden to the Federation of South Arabia,* Cmd. 1814 (H.M.S.O., 1962); *Handbook of the Federation and Protectorate of South Arabia Prepared in Accordance With Article 73(e) of the United Nations Charter* (1964); W. H. Ingrams, *The Yemen* (1963); G. King, *Imperial Outpost—Aden* (1964).

SOUTH ARABIA, PROTECTORATE OF, formerly known as the Aden Protectorate, comprises a group of Arab states or tribal districts under the protection of the United Kingdom. The protectorate extends along the southern coast of Arabia east of the Red Sea, bounded on the west and north by Yemen, on the north by the desert plateau and Saudi Arabia, and on the south by the Gulf of Aden. It is divided into two areas: the Western Protectorate, which consists of a generally rather narrow coastal plain, behind which rise high, broken mountain ranges dissected by fertile valleys, some with perennial streams; and the Eastern Protectorate, which is largely barren mountain plateau, through which runs Hadhramaut (Hadramawt) Valley and its side branches, where some cultivation is possible, irrigation being provided by floodwater and wells. The mountains in the north give way to steppe, and the steppe to the great sand desert of the Rub' al Khali (ar Rab' al Khali), or Empty Quarter. The population, which is mostly Muslim, was (mid-1960s est.): the Western Protectorate approximately 600,000; the Eastern Protectorate 400,000. The area is about 112,000 sq.mi. (290,000 sq.km.), including the island of Socotra.

The climate is tropical, with hot, humid summers on the coastal plain, which also experiences sandstorms, high winds, and extremes of heat and cold, with scanty rainfall on the mountains; the air is less humid than that of Aden, and the temperature range is greater. There are virtually no made-up roads, but the rough tracks are used by motor traffic though much transport is still by camel or donkey. There is an airport at Riyan, in the Eastern Protectorate, about 300 mi. (480 km.) NE of Aden, as well as several landing grounds, and there are ports at Al Mukalla, Qishn, and Ash Shihr, also in the Eastern Protectorate, for small craft and coastal steamers. The principal towns are Lahej (Lahij), in the Western, and Al Mukalla and Say'un (Saiwun), in the Eastern Protectorate. Most of the inhabitants are employed in agriculture, cereals (millet, wheat, and maize [corn]) and cotton being the chief crops. Cotton is ginned, there is fish curing, and the fishing industry is being developed. The Abyan Scheme, in the Western Protectorate, is an irrigation project utilizing floodwaters from the mountains; much of the acreage is planted in cotton.



THE PROTECTORATE OF SOUTH ARABIA, SHOWING THE LIMITS OF THE FEDERATION OF SOUTH ARABIA AND (IN THE INSET) ADEN

There are three farmers' associations and a date farmers' association.

In the mid-1960s a U.S. oil company held concessions for the prospecting and production of oil in the Qu'aiti, Kathiri, and Mahri states and had conducted a core-drilling survey. The main imports include foodstuffs, kerosene, cotton piece goods, cotton-ginning and agricultural machinery, and motor vehicles. The chief exports include cotton, coffee, and skins and hides. Most of the trade passes through the port of Aden. East African shillings and riyals (Maria Theresa dollars) are used as currency. Health services are provided in several hospitals and many health units. Each area has a health training centre and mobile sanitation unit. There are many primary schools, mostly for boys, and several intermediate schools.

History and Administration.—After Aden was occupied by the British in 1839, the tribal chiefs from neighbouring territories entered into protective treaty relations with the British government, all the states being covered by treaty in 1914. From 1915 until the World War I Armistice in 1918, three districts of the Western Protectorate were occupied by Turkey, which had conquered Yemen in 1872. In 1934 an Anglo-Yemeni treaty was signed to protect the frontier (then undemarcated) between Aden and Yemen, and in 1951 an agreement to set up a frontier commission followed. Yemen, however, has continued to claim the whole of the protectorate. Between 1937 and 1955, chiefs of both areas of the protectorate (including the sultan of Socotra) signed new treaties bringing them into closer relationship with the British government.

In the Eastern Protectorate the Qu'aiti sultan, who had gradually come to rival the Kathiri sultan in power and wealth, first entered into treaty relations with the British government in 1882. These ties were strengthened in 1888 by the signing of a protectorate treaty, which the Kathiri sultan acknowledged in 1918 as binding on his people also. By the present treaties the chiefs agreed to accept the advice of a British resident officer in all matters except those concerning Islam and Islamic customs.

The high commissioner for Aden (*q.v.*) is also high commissioner for the Protectorate of South Arabia, but the protectorate is administered by the local rulers with the help of the British advisory staff and not by the British government directly. The various sultanates, sheikhdoms, and amirates of the Western Protectorate nominate their chiefs, who have to be recognized by the British government. Several of the tribal units possess either state or legislative councils. The two principal states in the Eastern Protectorate, the Qu'aiti and the Kathiri, are administered by their sultans, who are constitutional rulers with state councils. Both states have well-developed systems of local government based on town and village councils. These and the other states are advised by a British agent stationed at Al Mukalla. In the mid-1960s the British government's annual expenditure on the protectorate exceeded £6,000,000. In 1959 certain states of the protectorate formed the Federation of Arab Amirates of the South (*see* SOUTH ARABIA, FEDERATION OF).

BIBLIOGRAPHY.—H. O. Forbes, *The Natural History of Socotra and Abd-el-kuri* (1903); D. A. Ingrams, *Survey of Social and Economic Conditions in the Aden Protectorate* (1949); Lord Belhaven and Stenton, *The Kingdom of Melchior* (1949), *The Uneven Road* (1955); D. van der Meulen, *Aden to the Hadhramaut* (1947); H. St. John B. Philby, *Sheba's Daughters* (1939); W. H. Ingrams, *Arabia and the Isles,* 2nd ed. rev. (1952); Wendell Phillips, *Qataban and Sheba* (1955); R. B. Serjeant, *The Portuguese off the South Arabian Coast* (1963); Admiralty Geographical Handbooks, *Western Arabia and the Red Sea* (1964). Current history and statistics are summarized annually in the *Britannica Book of the Year.* (W. C. BU.; X.)

SOUTH AUSTRALIA, a state of the commonwealth of Australia, occupying 380,070 sq.mi. (12.8%) in the central southern portion of the continent.

PHYSICAL FEATURES

Natural Regions.—Situated athwart the junction line of the great western plateau and the great central plains of Australia, South Australia can be divided into a number of different physical areas. Its relief is nowhere so marked as that of the eastern states, and there is relatively little land above 2,000 ft. The northwest

quarter is occupied by the southwestern extension of the western plateau. The northeastern quarter is occupied by the Lake Eyre depression, with its horseshoe of mud and the marsh-filled depressions—"Lakes" Frome, Callabonna, Blanche, Gregory, Eyre (3,700 sq.mi.) and Torrens—surrounding the northern Flinders ranges. Southward from these lies an area where recent (Pliocene) cross-patterned warping and faulting, elevation and submergence have provided the most decisive features of the state: the Mt. Lofty-Flinders ranges, the northern part of Yorke peninsula, the Spencer gulf-Lake Torrens (semidrowned) valley; the Flinders Barrier spur or ridge, the extremities of the peninsulas, Kangaroo Island and perhaps also—though it is older—the Gawler range. South and east of this lie the Murray river and the southeastern lowland plain, the western fringe of the great Murray (Tertiary sea basin) lowland and the counterpart of the Lake Eyre basin in the north.

The northeastern corner of the state (118,000 sq.mi.) forms part of the Great Artesian basin, the water-bearing beds lying at average depths of 4,000–5,000 ft. in the northeast but shallowing until, along the edge of the west Australian platform, the waters well up in a line of remarkable "mound springs." The soft strata of the lower Murray basin also yield plentiful supplies, often of good quality, from depths of 50–350 ft. and, toward the extreme southeast (Penola, Millicent) the underground water table intersects the surface and has resulted in extensive swamps and coastal lagoons.

Climate.—The state may be roughly divided into two climatic regions, a northern warm, dry region with high summer temperatures and less than 10 in. of rainfall per annum, much of which is erratic and unreliable, and a southern region which is typically Mediterranean with moist winters and hot dry summers. Very few areas in this latter region receive more than 35 in. of rain each year. The settled part of the state averages between 15 and 25 in. Adelaide, for example, has a mean annual rainfall of just over 21 in. and is the driest state capital in the commonwealth. North of the 32nd parallel and in the eastern part of the state north of the Murray river, the mean annual rainfall is less than 10 in. but, as in most dry areas, distribution is very erratic. The driest known part of the whole continent occurs around Lake Eyre. There the annual average rainfall is only 5 in., and it is rare for more than 10 in. to be recorded in any 12-month period. Generally the lake itself is dry and covered with salt.

South Australia is warm and sunny. The range of average temperature is greater in the summer than in the winter. In June, July and August most of the state experiences mean monthly temperatures between 50° F. and 60° F., whereas in the summer the southern part of the state has mean monthly temperatures around 65° F.,while in the northern part of the state the monthly average is between 80° F. and 85° F. It is not uncommon at Oodnadatta for the thermometer to register 120° F. in the shade. Adelaide, on the average, has 13 days per annum when the temperature is more than 100° F.

Vegetation.—The natural vegetation consists mainly of five major types. Eucalypt woodland and forest is on the higher and wetter parts of the Mt. Lofty and Flinders ranges and in the southeastern part of the state. Mallee (dwarf eucalypt), she-oak and tea-tree used to cover the remainder of the southern portion of the state but large areas of mallee have been cleared for agriculture. In the drier north, the mallee gives way to mulga scrub (acacia) and on the fringe of the Nullarbor plain in the west, saltbush is dominant. In the northwest and northeast spinifex grass is the chief plant of the very dry areas, but in places there is no vegetation at all, but only sandy ridges or rough gibber (stony) plains.

One of the greatest changes in the vegetation pattern of the state has been the establishment of pine forests in the southeast. *Pinus radiata* was expected to cover 90,000 ac. by 1961. By the 1950s the forests already produced raw material for a cellulose factory, timber for packing cases as well as for buildings, and the Victorian match industry depended upon a supply of veneers produced at the government peeling plant at Mt. Burr and at a private plant at Nangwarry, 20 mi. N. of Mt. Gambier.

Soils.—The distribution of the different types of soils is naturally closely related to the climate and vegetation patterns. The podsols, or ash gray soil, are found in the wetter parts of the state where iron oxide and other soluble materials have been leached out leaving the surface soil ash gray in colour and poor in plant nutrients. Where the rainfall is slightly less and the evaporation higher the podsols give way to the red brown earths which extend north of Adelaide beyond Jamestown and in the southeast around Naracoorte and Bordertown and on Eyre peninsula. All these areas are important for agriculture and grazing. The mallee soils, rich in lime but poor in humus and phosphate, are closely associated with the mallee type of vegetation. With the application of superphosphate and some of the important trace elements (copper, zinc and molybdenum) many of these mallee soils, once thought to be of little economic value, are gradually being sown down with subterranean clover and it is expected that they will add considerably to the pasture acreage of the state. Desert-steppe soils and desert sand hills extend over the northern half of the state, and are used for extensive grazing where artesian water is available.

Water Supply.—Climatic conditions made the provision of an adequate water supply an early problem for settlers in South Australia. Throughout the arid areas of the state, settlers were forced to rely on surface catchments and later on artesian supplies. The Murray river water was used for irrigation purposes in isolated settlements along the banks. It was not until 1944 that a double pipe line of 230 mi. was completed from Morgan to Whyalla, being used on the way for boosting several rural reservoirs. An additional pipe line from the Murray was begun in 1951 to supplement the inadequate water supply of Adelaide.

HISTORY

The coast line of South Australia was explored by Matthew Flinders and Nicolas Baudin in 1802. Apart from occasional visits by whalers little was done about the settlement of the area until 1830 when preliminary plans were drafted by the National Colonisation society, founded by Anthony Bacon and Edward Gibbon Wakefield. These proposals were submitted to the colonial office in the form of the South Australian Land company, the members of which were philanthropic bankers and Whig politicians. After a year's negotiations, the company's proposal was rejected by Lord Goderich because it was "too republican." With the passing of the Reform bill however, a new society called the South Australian association was formed, the committee being mostly "advanced-liberal" members of parliament. In spite of antagonism from those opposed to the theories of systematic colonization, an act establishing South Australia as a British province was passed in Aug. 1834.

Among other things the Foundation act protected South Australia from becoming a convict colony and fixed a minimum price for land, the revenue from which was to be used entirely for sending emigrant labourers to the new province. It also divided the home administration of the colony between the colonial office and a board of commissioners. A sum of £20,000 was to be raised as a guarantee fund and £35,000 worth of land sold before the province could be gazetted. The fulfilment of these provisions was made possible by the South Australian company, a joint stock venture unconnected with either government or commissioners. This company not only made the first settlement in South Australia, but also, through its banking, mercantile, land and shipping operations, played an active part in forming the policies of early governors.

The surveying party sent out to select the site for settlement was understaffed and ill-equipped and was followed too soon by the main party of settlers. The official proclamation of the province was read at Holdfast bay, by Gov. John Hindmarsh on Dec. 28, 1836. Through the resolution and foresight of Col. William Light a town site was chosen and the survey of Adelaide completed by March 1837.

The division of powers between Governor Hindmarsh and the commissioner's agent rapidly developed into factional strife, but this was greatly diminished by the appointment in 1838 of George Gawler to the dual office of queen's representative and resident commissioner. As a result of administrative difficulties, almost three years passed before the surveys were sufficiently advanced

to permit country settlement. The colonists meanwhile spent their time and substance in speculating on town lands, a process accentuated by the public works program of Governor Gawler. Immigrants continued to flow into Adelaide, and because provisions had to be imported, prices soared. The commissioners' resources were exhausted by the end of 1839 in spite of the fact that they had borrowed from the sacrosanct Land and Emigration fund. An appeal to the colonial office for assistance led to an inquiry into South Australian affairs by a select committee of parliament, the report of which was not completed till June 1841. In the meantime, the commissioners were obliged to defer payment on drafts from Adelaide, immigration ceased, and the province suffered a financial collapse. By an act of 1842 South Australia was placed wholly under the control of the colonial office.

Capt. George Grey, appointed as Gawler's successor, embarked on a program of strict economy. Before he arrived the exodus of settlers to the country had begun, and several large flocks of cattle and sheep had been overlanded from New South Wales. By 1842, the settlers had surplus grain for export. With the development of copper mining in 1845, a second, though more limited, speculation mania broke out. The flow of migrants recommenced, hundreds of Cornish miners and their families receiving free passages to the colony. The prospects of copper production were somewhat dimmed by the discovery of gold in Victoria in 1851—more than 18,000 South Australians proceeded to the diggings. Most of them returned however to the more dependable income from the wheat fields and copper mines of their home colony.

Throughout the foundation and early years of settlement, religious and civil liberty had been given great prominence. Religious freedom found expression in "a fair field for all churches" and, except for a brief period ending in 1850, there was no state aid to religion, the denominations in South Australia being supported entirely by the voluntary principle. Secular education subsidized by the state was also introduced in 1850. Religious equality was used as an additional inducement to prospective migrants, with the result that the proportion of nonconformist sects, especially Methodists, has always been higher in South Australia than in the other Australian states and helps to account for Adelaide's reputation as "a city of churches."

The first goal of civil liberty was self-government, which the Foundation act promised when the population reached 50,000. In 1843, with the population at about 17,000, four nonofficial colonists were admitted to the legislative council, and in 1850, the population being 63,700, the province was granted a chamber of 8 crown nominees and 16 elected representatives. This blended council prepared a constitution in 1853 providing for a legislature composed of a nominated council and an elective assembly. The impatience of the colonists with the nominee system, however, was expressed in a petition to the queen bearing more than 5,000 signatures. The authorities yielded to pressure and the Constitution act of 1856 gave South Australia two elective chambers. Although a small property qualification was retained for electors to the council, adult manhood suffrage was adopted for the assembly. Other principles of civil liberty which placed South Australia in the van of empire in 1856 were vote by ballot, one man one vote and triennial parliaments. A few years later, Robert Torrens introduced a simplified system for the transfer of land titles which was later copied in many countries. In 1887 payment of members was introduced, and votes for women followed in 1894. Though the colonists maintained a scrupulous loyalty to the crown, their resentment of interference in local affairs found expression in prolonged legal disputes such as those which led to the Colonial Laws Validity act of 1865. This act gave validity to the laws passed by all colonial legislatures provided that they were not repugnant to British statutes. It remained as the legal keystone of the empire until the Statute of Westminster, 1931.

The control of crown land revenue, secured in 1856, enabled the colonial legislature to begin a more systematic exploitation of natural resources. By a vigorous immigration policy, an average of 3,000 persons a year were brought to the colony between 1860 and 1880. A number of Germans were sent out by George Fife Angas and others in the early years of settlement but the great majority of South Australians are of British stock.

By 1860 most of the colony had been explored and pioneer pastoralists had spread as far as water supplies would permit. Behind them agriculturists speedily extended the areas of crop. A severe drought in 1865 checked the expansion temporarily, and some attention was diverted to the settlement of the Northern Territory. The development of this project proved both costly and cumbersome to the South Australian government, although a few individuals profited through land speculation and mining, and others by cattle raising round Alice Springs. Few objections were raised when the responsibility for the Territory was taken over by the commonwealth government in 1911.

New copper mines were discovered at Wallaroo and Moonta in 1860 and the Murray river shipping trade was developed to reach a peak of over £1,250,000 in 1882, after which it declined through the competition of Victorian and New South Wales railways. Most of the state's capital and labour, however, was directed toward agriculture and wool growing. A network of some 2,000 mi. of railway connected the centres of rural production with easily accessible ports, from which a large proportion of the grain was taken to Europe by sailing vessels, the annual race of these ships being followed with keen interest. In the year 1880 South Australia through its abundant wheat production earned the title of the "farinaceous state."

Among the many new adaptations made by the early settlers for wheat growing in South Australia, John Ridley's reaping machine deserves special mention as a labour-saving device which came into general use after 1850. Dry-farming techniques and the application of superphosphates were also important factors in bringing new land under cultivation and increasing agricultural production. As the supply of wasteland diminished, however, more attention was given to scientific research, aided both by state grants and generous gifts from private individuals. The instruction given at Roseworthy Agricultural college (established in 1885) and experiments at the Waite institute (1924) have benefited the whole continent. Improved methods of tillage and a ceaseless war against pests and plant diseases have greatly reduced the risks of rural production. The pastoralists, too, have benefited. The development of a large-framed Merino sheep suited for semiarid conditions, experiments with the top-dressing of pastures and the introduction of English breeds have made possible a flourishing export trade in fat lambs.

In the first 40 years of the colony the progress of education was impeded by the wide spread of population and the emphasis on rural production. State subsidies were first granted in 1846, and a Central Board of Education established a few years later but no spectacular advance was made till 1874. In that year the University of Adelaide was founded and a few months later compulsory and secular education was introduced. This was followed by the establishment of the Teachers' Training college in 1876 and the School of Mines ten years later. In the '90s primary education was made free, and attention was given to the creation of secondary schools. Until the introduction of motor transport, however, secondary education was largely confined to the vicinity of Adelaide. The city remained the centre, not only for mercantile interests but also for culture, entertainment and sport.

Until the federation, South Australia was noted for its frequent changes of government and the independence of its elected representatives. This was encouraged in part by the attitude of colonists outside parliament, expressed through a profusion of small and short-lived societies for social and political reform, and in part by the extreme difficulty of reshaping electoral divisions to conform with the continual shift of population. From the beginniug, care was taken to assure a high proportion of representation to the rural districts to counterbalance the indirect influence of urban voters through their control of the press and the situation of the parliament in Adelaide. The system of multi-member electorates also encouraged voters to support individual candidates rather than a party program. The emergence of a Labour party after the industrial troubles of the early 1890s, therefore, made but little impression on the tradition of independence, especially after the outstanding leaders transferred their attention to the federal sphere. South Australian politicians played a prominent part in the drafting of the federal constitution, some of them making a strong bid for the retention of state rights. The activity of leading millers and bankers in establishing commonwealth-wide combines after the federal government took over the control of tariffs was perhaps the most important single cause of the development of the party system in South Australia. To defeat the schemes of the "monopolist conservatives" in 1905, a party of Independent Liberals

was formed, and with this group the Labour party made a compact. The Conservatives were defeated but the Labour-Liberal combination proved unsatisfactory. It was soon realized that the control of fiscal policy by the commonwealth greatly reduced the power and influence of state governments. The Labour party's insistence on adult suffrage for the legislative council and the ultimate abolition of that chamber was too radical for the Liberals but it nevertheless remained the central plank of Labour's platform.

In 1910 the Labour party was returned with a majority of two in the assembly and under the leadership of John Verran formed its first South Australian ministry. This government held office for two years, but its reforms made little headway against the Conservatives in the legislative council. It had, however, become closely identified with the Australian Labour party and had taken a stand on the closer unification of states within the commonwealth. Though strongly supported in the urban electorates it came to grief in 1912 on the issue of state rights and did not regain office until the outbreak of World War I when its ranks were again split on the conscription issue, more than half the members going over to the new Nationalist party which then came to terms with the Liberals.

In 1921 four parties contested the state election. There were the remnants of Labour, the Nationalists who had parted company with the Liberals, and those who supported the Farmers and Settlers' party. After the Australian Labour party adopted the "socialization of industry, production, distribution and exchange" as its central objective, there was increasing combination between the anti-Labour groups. For the next 12 years parliamentary power alternated between Labour and Liberal. In 1933 there was a strong reaction to the old tradition of independent members, chiefly because of dissatisfactions caused by the depression. It was soon realized that it was impossible to isolate the state from the parties of federal politics, especially after the creation of the commonwealth loan council.

In 1941, therefore, fewer Independents were returned and, with the backing of the Liberals and Country party, Thomas Playford became premier. His masterly grasp of the complexities of state-federal finance earned him the leadership of the less influential states in the premiers' conferences at Canberra, while at home his vigorous and imaginative developmental policies broke, to a large extent, the almost complete economic dependence of the state on rural production. In 1953 his party was elected to office for the sixth time in succession.

POPULATION

The growth of population in South Australia is shown in the following table:

Population of South Australia, 1850–1950

Year	Population (in thousands)		
	Metropolitan	Other	Total
1850	32	31	63
1870	61	123	184
1890	132	186	318
1910	199	207	406
1930	310	264	574
1950	415	285	700

Adelaide, the capital of the state, had a population of 646,073 in 1947; it was estimated to have 739,563 in 1952.

AGRICULTURE AND INDUSTRY

Agriculture.—South Australia, although third in size of the Australian states, has a relatively small area that can be utilized for agriculture. More than four-fifths of the total area has an average rainfall of less than 10 in. a year and supports fewer than 2,000,000 sheep and 100,000 cattle. The following table showing the percentage of the total sheep population carried outside the 10 in. rainfall country indicates the trend that has been taking place over the last 50 years.

Distribution of Sheep, 1900–1950

Year	Number of sheep	Percentage outside 10 in. line
1900	5,667,000	28.2
1910	6,432,000	24.9
1930	6,186,000	18.7
1940	9,941,000	20.1
1950	9,477,000	14.8

Goyder's line of rainfall marking out the southern extent of the 1865 drought was for many years regarded as a line of safety beyond which agriculture could not be carried out. The use of dry-farming methods and artificial fertilizers has, however, brought large areas outside the line into cereal production. In the higher rainfall areas sheep rearing and cereal production predominate. Many wheat farmers turned to fat lamb production after the depression of the 1930s and by 1940 the sheep population of the state had increased to 10,000,000. More land is being put under pasture for livestock as the result of a wider use of subterranean clover, superphosphate and modern mechanical methods of clearing. Land previously considered useless, or able to carry as little as a quarter of a sheep to the acre, is being converted to two sheep to the acre country or better. The area sown to pasture doubled between 1940 and 1950, and if the trend in the early 1950s of 100,000 ac. a year were to continue, it would have doubled again by 1960.

Wheat for many years has been the dominating cereal in South Australia. A peak of 4,650,000 ac. was sown to cereals in 1930–31. Of this 4,180,000 ac. were sown to wheat, 252,000 ac. to barley and 218,000 ac. to oats. Livestock then began to be more important in areas formerly devoted exclusively to cereals, with the result that by 1951 the wheat acreage had declined to less than half that of the peak year. The yield, however, increased slightly to an average of 12 bu. per acre for the 1940–50 period, resulting in an average production of about 26,000,000 bu. There has been a considerable increase in the acreage under barley in recent years.

Dairying is carried on mainly on the plains and in the hills within reach of Adelaide where markets and export facilities exist. Vines and olives thrive on the lower western slopes (below 500 ft.) of the Adelaide hills, while further northeast and north the Angaston, Tanunda and Clare districts are noted for their excellent wines. Oranges, lemons, peaches, apricots and almonds abound on the plains and in the valleys where water, shelter and warmth are available. Around Renmark, Loxton, Berri and Barmera on the Murray river, citrus fruits, vine crops and vegetables are grown under irrigation.

The barrage at the mouth of the Murray has maintained satisfactory levels in Lakes Albert and Alexandrina. Around these lakes and along the lower Murray there are areas that could be used for the production of fodder crops for dairy cows or for market gardens to meet the increasing demands of the growing urban population for fresh milk and vegetables.

Mining.—Although South Australia has never been rich in precious minerals there are a number of mineral resources that are becoming increasingly important not only for the state, but for the commonwealth as well. The discovery of silver lead at Broken hill in the neighbouring colony of New South Wales was exploited by the South Australians. The mines were connected by rail with Port Pirie where treatment plants were established. Among the metallic minerals iron will be significant for a long time to come. In the dry Middleback ranges west of the northern part of Spencer gulf the ore occurs in Iron Knob, Iron Monarch and several other hills and provides the raw material for the heavy industries at the near-by port of Whyalla as well as for the steel industries at Newcastle and Port Kembla in New South Wales. Copper declined in importance after the Moonta mines were closed in 1923, but in the 1950s investigations were being made to discover new workable deposits.

A later nonmetallic substance to be developed is sulphur, an important raw material for the manufacture of sulphuric acid needed in ever increasing quantities for the production of superphosphate. More than 30,000,000 tons of pyrites—enough to produce 3,000,000 tons of sulphur—had been discovered at Nairne in the Adelaide hills by 1953. Gypsum, one of the state's mineral products since the 19th century, had become the basis for the manufacture of plaster of paris for the building industry. The state is the main producer of salt in the commonwealth, and besides the deposits on Yorke and Eyre peninsulas, large areas of flat land on the coast north of Adelaide are being used for the production of salt by flooding and evaporation on a large scale. The coal deposits at Leigh Creek, 350 mi. N. of Adelaide, have been developed as a state undertaking, and in 1953 it was hoped that the annual output would soon reach 1,000,000 tons, and so supply the state with a fair percentage of its needs. The field has been proved to contain not less than 380,000,000 tons of subbituminous coal with a moisture content between 30% and 40%. Limestone in all grades and of the purest quality is worked in various parts of the state. Important recent discoveries include high grade cosmetic talc in the north Flinders range and uranium ore at Radium hill and Mt. Painter.

Industry.—Until its centenary in 1936, South Australia was essentially a primary producing state. However, after 1936 there was a remarkable development of secondary industry, and employment in manufacturing in 1953 far exceeded that in agriculture. This development received active encouragement from the South Australian government and was stimulated during World War II when munition factories of many kinds were established.

Manufactures of every description are undertaken, ranging from motor bodies to textiles, from consumers' durables to steel tubes. At Whyalla, a modern blast furnace and shipyards have been established, and at Salisbury there is the headquarters of the Long-Range Weapons establishment. The latest developments include industries associated with the processing of uranium ores. All this growth in manufacturing production has been assisted by improved power supplies, made possible by the use of Leigh Creek coal deposits.

The rapid increase in factory employment between 1938–39 and 1949–50 illustrates the industrialization of the state. Over this period factory employment increased by 84%, as compared with an increase of 17% in the state's population. By the 1947 census 37% of male breadwinners were engaged in secondary production, whereas only 22% were in primary production, the remainder being involved in the service industries. This emphasizes the remarkable shift in the balance

between primary and secondary production.

BIBLIOGRAPHY.—The most complete work on South Australia is *The Centenary History of South Australia,* prepared by the Royal Geographical Society of Australasia, South Australian branch (1936). *The Proceedings of the South Australian Royal Geographical Society* (1885–) contain many valuable studies. *See* also E. Hodder, *The History of South Australia from Its Foundation to the Year of Its Jubilee,* 2 vol. (1894); A. Grenfell Price, *The Foundation and Settlement of South Australia, 1829–1845* (1924); C. E. Fenner, *South Australia, A Geographical Study* (1931). (P. H. K.; G. H. LN.; D. H. PE.)

SOUTH BEND, a city of northern Indiana, U.S., the seat of St. Joseph county, is located 90 mi. E. of Chicago on the south bend of the St. Joseph river where it turns northward to Lake Michigan. The locale was visited by Robert de la Salle in 1679; the American Fur company in 1820 and Alexis Coquillard in 1823 established fur trading posts at the site. South Bend was platted in 1831, incorporated in 1835 and chartered in 1865.

The population of the city in 1960 was 132,445, an increase of 14.3% in the decade; that of the standard metropolitan statistical area (St. Joseph county) was 238,614, an increase of 16.4%. (For comparative population figures *see* table in INDIANA: *Population.*)

South Bend's industry is diversified, including automobile and truck manufacture, aviation equipment, guided missiles, farm machinery, tools and dies, clothing, toys, paints, varnishes, lathes, baits and fishing rods, steel ranges, folding cartons, bearings, wallpaper, elastic and surgical goods, electrical equipment, brass, iron and semisteel castings and dress patterns. In the early 1900s it was known as the home of the largest carriage and wagon works in the world, those of the Studebaker Bros. Manufacturing company and of the Oliver Chilled Plow Works and the woodworking department of the Singer Sewing Machine company. About 45% of the working force is employed in manufacturing industries and 25% in transportation equipment fields. Industrial advantages of the city include excellent transportation facilities, highly skilled workers and reasonably priced electric power and natural gas. The city is surrounded by rich peppermint-growing muck which also produces potatoes, onions and truck produce.

To the immediate north of South Bend lie two nationally known Roman Catholic educational institutions, St. Mary's college and the University of Notre Dame. Notre Dame, established in 1842 by the Very Rev. Edward Sorin, C.S.C., has a substantial academic reputation in the natural sciences. Its world-renowned Lobund Germ Free laboratory, its unsurpassed football teams and its famous "gold dome" on the administration building are hallmarks of the university. St. Mary's, a small liberal arts college for women, chartered in 1851, is the mother house of the Sisters of the Holy Cross. South Bend also has an extension centre of Indiana university. Other community cultural programs are provided by the Civic Music association, the South Bend Symphony orchestra, the Presbyterian Players, the South Bend Art association, the Northern Indiana Historical society and the public library. South Bend lies near many small lakes and the Lake Michigan resort area. (J. J. D.)

SOUTH CAROLINA, popularly called the "Palmetto state," is an Atlantic coast state of the United States, one of the original 13. The eighth state to ratify the constitution, it entered the union on May 23, 1788. It is bordered on the north by North Carolina, on the south and west by Georgia and on the east by the Atlantic ocean. Its area is 31,055 sq.mi., of which 783 sq.mi. are water; in size it ranks 40th among the states. The South Carolina state flag has a blue field in the centre of which is represented a palmetto tree and in the upper corner of which, next to the staff, is a crescent. The state flower is the yellow or Carolina jessamine (*Gelsemium sempervirens*), the state tree the cabbage palmetto, the state bird the Carolina wren (*Thryothorus ludovicianus*), the state song "Carolina." The capital is at Columbia (*q.v.*).

PHYSICAL GEOGRAPHY

Physical Features.—South Carolina (lying between the extremes of approximately 32° 4′ and 35° 12′ N. lat. and 78° 31′ and 83° 23′ W. long.) is mainly in the coastal plain and Piedmont plateau regions, but in the northwest it extends slightly into the Appalachian mountain region. Locally the coastal plain region is known as the low country, and the Piedmont plateau and Appalachian mountain regions are called the upcountry. The coast is low, and islands increase in size and number toward the Georgia border. For about 10 mi. inland the coastal plain is occupied largely by salt marshes. Although continuing flat, the surface rises at the rate of approximately 2 ft. per mile for 50 mi. or more. Parts of the Blue ridge (*q.v.*) rise abruptly from the foothills to 3,425 ft. in Mt. Pinnacle, 3,208 ft. in Caesar's Head and 3,124 ft. in Table Rock. The highest point in the state is Sassafras mountain (3,560 ft.) in the Blue ridge and on the North Carolina state line. The mean elevation of the state is about 350 ft.

The principal rivers rise in the Appalachian mountains and flow southeast into the Atlantic ocean. Santee river is formed by the confluence of the Wateree (known above Wateree creek north of Camden as the Catawba; *see* SANTEE-WATEREE-CATAWBA RIVER SYSTEM) and the Congaree, which is in turn formed by the Broad and the Saluda. The basin of this system embraces about one-half the area of the state. In the northeast the Waccamaw and the Pee Dee and its tributaries—the Little Pee Dee and Lynches—are wholly within the coastal plain, but the main stream of the Pee Dee is a continuation below the fall line of the Yadkin river which rises in the mountains of North Carolina (*q.v.*). The Edisto is the principal stream in the southeast.

In the Piedmont plateau there are rapids in the rivers, but in the coastal plain the current becomes sluggish, and in time of high water the rivers spread over wide areas.

Climate.—Along the coast the climate is mild and equable. At Charleston, the mean winter temperature is 51° F., the mean summer temperature 80°, the mean annual temperature 66° and the range of extremes from 104° to 7°. For the state the mean temperature is about 63°. In nearly all sections January is the coldest month and July the warmest. The mean annual rainfall for the state is about 47.75 in. and its distribution is excellent, 17 in. falling during the summer. Snow is uncommon in the southeast, but elsewhere there may be several inches, occasionally more than a foot.

The frost-free season ranges from 245 days on the coast to 204 days on the uplands. Tornadoes sometimes occur in the west, and the coast suffers from hurricanes.

Soil.—In general the soils of the Piedmont plateau are such as have been formed by the disintegration of the underlying rocks. These consist mostly of granite and gneiss, but in the north-central section there is trap and in the southeast section some slate. On the Piedmont plateau the subsoil is a reddish or yellowish clay. In the upper section of the coastal plain the soil is for the most part a loose sand, but near the coast it becomes more fertile, much being underlain by marl.

Vegetation.—Palmettos grow on the coast but only artificially inland. For some distance from the coast there are magnolias, live oaks draped with long gray moss and reed-covered marshes. In the swamps are cypress, gum and bay trees.

In most of the uplands of the coastal plain the longleaf pine is predominant, but large water oaks and undergrowths of several other oaks and of hickories are not uncommon. On the Piedmont plateau and in some of the more hilly sections below the fall line there is some shortleaf pine, but most of the trees in these areas are hardwoods. Deciduous oaks are commonest, but beech, birch, ash, maple, black walnut, sycamore and yellow poplar also abound. On the mountains are oak, chestnut, laurel, white pine and hemlock. Among indigenous trees, shrubs and vines are the blackberry, grape, persimmon, plum, crab apple, hickory, chestnut and hazelnut.

Animal Life.—Because of temperate climate and good rainfall, South Carolina has a wide variety of wildlife. Hunting has long been a favourite sport of the residents of the state, and many out-of-state persons are attracted by it. A number of large hunting preserves and lodges are maintained, especially in the low country. A state wildlife resources commission maintains a conservation program and supervises hunting and fishing laws. The laws specify the following as game animals, and, except during the open season for hunting, generally protected: deer, fox, mink, muskrat, opossum, otter, rabbit, raccoon, skunk and squirrel. Hunters also find

duck, geese, quail and wild turkey in the state.

Though a relatively small state, South Carolina has a variety of fish, ranging from game fish in the offshore waters to trout in the northwestern mountains. The commonest game fish are largemouth bass, white bass, bluegill, crappie, jackfish or pickerel, yellow perch and redbreast sunfish. Salt-water species found close to shore and in the inlets include channel bass, croaker, flounder, spot, weakfish and whiting. Farther offshore are amberjack, barracuda, blackfish, bluefish, grouper, mackerel, snapper and tarpon.

Parks, Historic Sites and Recreation.—The state maintains 27 state parks, many of which include facilities for boating and swimming as well as camping. Resorts and vacation places attract both South Carolinians and others in all seasons, especially to the mountains and seashore. Much of the coast of the state is marked by a strand and beaches that draw thousands every summer. An abundance of woods and waters is accompanied by a climate that permits year-round activities in most recreational pursuits.

Historic sites dating from days of Spanish and French exploration are to be found throughout the state. Many battles of the American Revolution and the American Civil War took place in the state. Of greatest interest to both historian and tourist is Charleston (*q.v.*), site of the first permanent settlement and a city that still maintains much of its ante-bellum charm and character along with scores of historic landmarks and structures.

HISTORY

The history of South Carolina may be divided into the periods of discovery and exploration (1521–1670), proprietary rule (1670–1719), royal rule (1720–75) and statehood (from 1776).

The first Europeans to visit the coast were a party of Spaniards from Santo Domingo (Hispaniola) in 1521. The first settlement, most probably at Winyah Bay, was made by Spaniards under Lucas Vásquez de Ayllón in 1526, but after a few months it was abandoned. The Spaniards, again settling in 1566, maintained a fort on Parris Island for about 20 years. In the meantime (1562) French Protestants under Jean Ribault made an unsuccessful attempt to establish a colony on Parris Island (*see* PORT ROYAL). In 1629 Charles I granted to his attorney general, Sir Robert Heath, all the territory lying between the 31st and 36th parallels and extending through from sea to sea, but no settlement was made, and in 1663 the same territory was granted to the earl of Clarendon and seven other favourites of Charles II. A second charter in 1665 extended the limits to 29° and 36° 30′. The proprietors were to legislate for the province "by and with the advice, assent and approbation of the freemen." They were empowered, though not required, to grant religious freedom to Dissenters. Circulars were issued in 1663 and 1665 offering most liberal terms to prospective colonists. In the fundamental constitutions adopted by the proprietary board in 1669, John Locke and Lord Ashley prepared for the province an elaborate feudal system of government that would have been obsolete even in Europe. The colonial assembly refused (as the charter gave it the right to do) to adopt them. They were, nevertheless, an element in arousing the feeling of discontent among the colonists which culminated in the overthrow of proprietary rule, and they encouraged the large plantation system which constituted the foundation of the slaveholding aristocracy.

Proprietary Period.—The first permanent English settlement was made in April 1670 at Albemarle point on the west bank of the Ashley river, but as the situation proved unfavourable the government and most of the people moved over in 1680 to the point between the Ashley and Cooper rivers, the site of the present city of Charleston. The area of settlement was gradually extended along the coast in both directions, but did not penetrate far into the interior until after 1730. There were many English from Barbados and French Protestants, both of whom strongly influenced the history of the province.

South Carolina's political history during the colonial era is the story of a struggle between popular and prerogative interests, between the people and the lords proprietor, then, later, between the people and the crown. From 1670 to 1700 the principal questions at issue were the refusal of the settlers to subscribe to the numerous editions of the fundamental constitutions and disputes over the collection of quitrents. Concessions were finally made which brought the government more directly under popular control. In 1693 the commons house elected by the people, secured the privilege of initiating legislation. The truce was followed by a controversy between Anglicans and Dissenters. A test act requiring members of the assembly to conform to the Church of England and to take the Sacrament of the Eucharist according to the rites and usages of that church (1704) was defeated only through the intervention of the Whig house of lords in England. By an act of Nov. 30, 1706, which remained in force until 1778, the Church of England was made the established religion. After a few years of peace and prosperity, the proprietors, acting on the advice of Chief Justice Nicholas Trott, adopted a reactionary policy, vetoed several popular laws and could not give aid in the desperate 1715 Indian war. In 1719 the people rebelled, overthrew the existing government and elected their leader, James Moore, governor.

Royal Rule.—The result of the revolution was accepted in England, and the province at once came under royal control, although the rights of the proprietors were not extinguished by purchase until 1729.

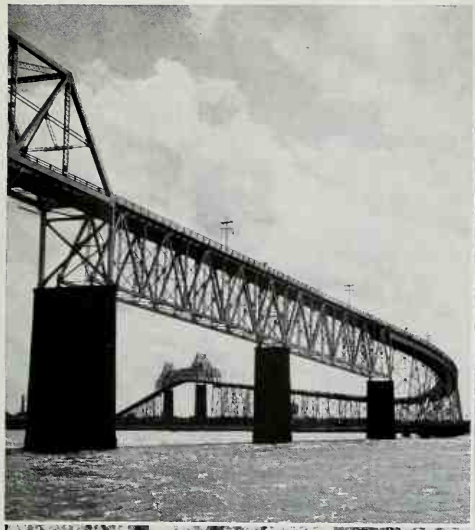
Theoretically South Carolina and North Carolina constituted a single province, but, as the settlements were far apart, their governments were always separate. Until 1691 each had its own governor. From 1691 to 1712 there was usually a governor at Charleston and a deputy for the northern settlements, and after 1712 there were again separate governors. An order of 1730 for a survey of part of the boundary was carried out in 1735, but the definition of the whole boundary was not completed until 1815.

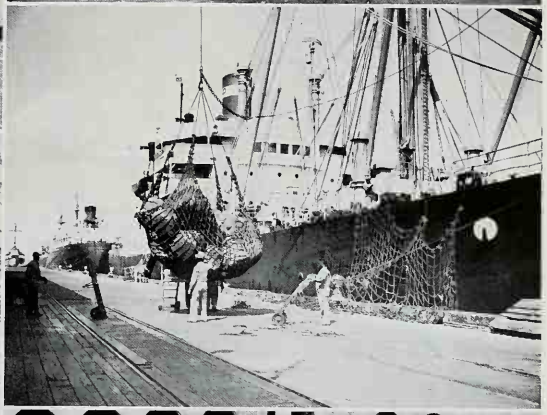
The period 1725–75 was a period of great prosperity, based upon the trade in pelts (mainly deerskins) and more permanently on the rapidly expanding production of rice and indigo. The southern colonies' Indian trade centred mainly in Charleston. This, and a rich agriculture and the large commercial business to serve them, supplied the economic basis of a coast country society of notable intelligence and culture. One of the earliest theatres in America, musical culture, libraries and the education of a large number of youths in England gave the society around Charleston a tone and finish at that time rare in the new world.

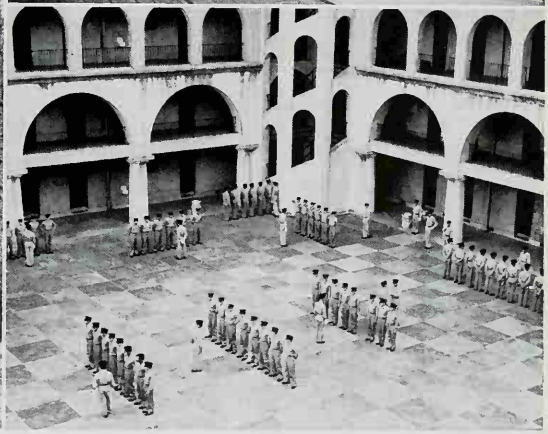
Passing under the royal government did not check the encroachments upon the governor and council by the commons house of assembly, which defiantly announced the powers of the British commons as its model. By 1760 the council had almost ceased to exercise any real control over legislation. It rarely initiated or amended a bill and never attempted to change a money bill without incurring violent denunciation from the popular house and paralyzing legislative action, once for years.

The province was unconsciously preparing for independence. Though measures of the British government after 1763 were not especially oppressive to the province, the people were too long accustomed to having their own way. They had, especially in the bitter dispute over the Wilkes fund (1769–75), involving the right of the commons alone to control finance, developed too fully a sense of general American and even British freedom to submit to measures which they regarded as subversive of the principles of liberty. Delegates were sent to the Stamp Act congress (1765) and to the Continental congresses (1774 and 1775). A council of safety appointed by an extralegal provincial congress virtually took over the government in June 1775. Royal administration ended when, on Sept. 15, the governor dissolved the assembly and fled.

Revolution.—Wars with the Spanish in 1686, 1702–04 and 1740, with the Spanish and French in 1706, with pirates in 1718, with the Yamasee Indians in 1715 and the Cherokees in 1760–61, with aid from British troops only in 1760–61, and a slave rising in 1739 had accustomed the people to arms. The state suffered severely during the American Revolution both from British troops and from the presence of numerous loyalists. A British fleet attempting to capture Charleston was repulsed by Ft. Moultrie, June 28, 1776. Calm prevailed until Gen. Sir Henry Clinton returned in 1780 with an overwhelming force and obtained the sur-







BY COURTESY OF (CENTRE RIGHT) SOUTH CAROLINA STATE DEVELOPMENT BOARD, (BOTTOM RIGHT) THE CITADEL, CHARLESTON S.C., PHOTOGRAPHS, (TOP LEFT) HAYS FROM BLACK STAR (TOP RIGHT) EWING GALLOWAY, (BOTTOM LEFT) A. DEVANEY

SCENES IN SOUTH CAROLINA

Top left: The Cooper river bridge gives access from Charleston to Mt. Pleasant, Sullivan's Island and the Isle of Palms. Completed in 1929, it is 10,000 ft. long and rises 150 ft. above the water

Top right: The state capitol at Columbia, built of local granite, is in a hilltop park containing trees of more than 50 species. There are many historic monuments on the grounds

Centre right: Stevedores unloading cargo at Charleston, a major southern Atlantic port

Bottom left: Lumberjacks selecting trees to be cut for pulpwood. Lumber products rank second among the state's manufactures

Bottom right: Squad leaders instructing freshmen in the rudiments of infantry drill at the Citadel, The Military College of South Carolina, Charleston

SOUTH CAROLINA

PHOTOGRAPHS, (TOP LEFT, BOTTOM LEFT, BOTTOM CENTRE AND BOTTOM RIGHT) SAMUEL CHAMBERLAIN, (TOP RIGHT) ARTHUR GRIFFIN FROM FPG, (CENTRE RIGHT) A. DEVANEY

BUILDINGS IN HISTORIC CHARLESTON

Top left: Dock Street theatre and St. Philip's church. The original Dock Street theatre (1736) is said to be the first in America used exclusively for stage productions

Top right: The Miles Brewton, or Pringle, house, built about 1765, was used as headquarters by British troops, 1781–82, and Federal troops, 1865

Centre right: The Josiah Smith house, built about 1800 for a great-grandson of one of the state's original landgraves

Bottom left: The Huguenot church, remodeled in 1845 on the walls of a church burned in 1797. The congregation was founded in 1680 by Huguenot refugees

Bottom centre: The William Blacklock house, built about 1800

Bottom right: Meeting street, with St. Michael's church in the background

render of the city from Gen. Benjamin Lincoln on May 12.

Completely overrun, the state experienced two years of fighting involving more battles, though most of them small, than occurred in any other state. A continental army under Gen. Nathanael Greene, assisted by state troops under Thomas Sumter, Francis Marion, Andrew Pickens and others, slowly drove the British back into Charleston and wrecked the plan of British troops from the south to join those from the north to crush Washington. The chief battles fought in the state were at Ft. Moultrie (June 28, 1776), the siege of Charleston (March 12–May 12, 1780), Camden (Aug. 16, 1780), King's mountain (Oct. 7, 1780), Cowpens (Jan. 17, 1781), Hobkirk's hill (April 25, 1781) and Eutaw Springs (Sept. 8, 1781).

Early 19th Century.—The period following the Revolutionary War witnessed an ambitious program of transportation development. The Santee canal, connecting Charleston with the whole Santee-Wateree-Broad-Saluda River system, was opened in 1800. Highway building followed. The latter was dropped, however, and the canal was ruined by railroad development. The South Carolina railroad, 136 mi. from Charleston to Hamburg opposite Augusta, Ga., when completed in 1833, was the longest railroad in the world.

The early state period was characterized by a bitter struggle between the older low country and the newer upcountry, the latter settled largely by Scotch-Irish coming down the Piedmont belt from Pennsylvania, Virginia and North Carolina. In 1786 it was necessary, to allay discontent, to consent to the removal of the capital to a newly located site to be called Columbia. Although removal took place in 1790, some state officers until 1865 kept offices both in Columbia and Charleston and the supreme court met in each city to hear appeals from the two sections respectively. The upcountry (then comprising a larger area than now thus designated), containing four-fifths of the white population and one-fifth the wealth, in 1808, with the help of a liberal low country minority led by Joseph Alston, secured a constitutional amendment apportioning one senator to each county and representatives to each county in proportion equally to white population and wealth. This left the control of the upper house to the low country and that of the lower house to the upcountry. Manhood suffrage followed in 1810. The low country's fear for its slave interests was allayed as slavery, fostered by the increasing importance of cotton, spread up the state.

The South Carolina college was chartered in 1801 largely to allay sectional antagonisms. The necessity of standing unitedly against antislavery and later radical Reconstruction agitation from the North prevented the masses from forcing democratic changes until late in the 19th century.

Northern antislavery agitation and the adoption of a tariff harmful to southern agriculture united both sections in the passage of the ordinance of nullification, Nov. 24, 1832, forbidding the execution of the tariff in South Carolina (*see* NULLIFICATION). The readiness of a powerful minority to assist Pres. Andrew Jackson, combined with a reduction of the tariff, operated to deter the majority from forcing the issue to armed conflict, but left a bitterness within the state that disappeared only after 1840 in the face of a graver danger from outside.

After a serious slave conspiracy in 1822 and northern abolitionism in the 1830s, the slave code was severely revised (1840–44), even to forbidding the sending of a slave to freedom anywhere in the world. The years 1847–52 saw most of the people of South Carolina deterred from secession only because of lack of cooperation from other southern states. After 1828 Virginia's southern leadership was superseded by the more uncompromising leadership of South Carolina under John C. Calhoun. South Carolina was the first state to secede, its ordinance of Dec. 20, 1860, being almost unanimously approved by its citizens.

Civil War and Reconstruction.—With a white population of 291,300, the state put into the field an estimated 62,838 effectives, of whom 22% were killed or died in prison. Gen. William T. Sherman's march across the state, Feb.–March 1865, was accompanied by enormous destruction, including the burning of Columbia.

The misfortunes of war were more easily borne than the humiliation of Reconstruction. Under Pres. Andrew Johnson's guidance, the white population elected officers who were soon ejected under the congressional plan of Reconstruction, disfranchising many whites and transferring power to the Negroes, northern adventurers (carpetbaggers) and native whites (scalawags) joining them. In the spring of 1868 the state was readmitted to the union under the control of these elements, and entered upon a period of eight years of crime and corruption, accompanied by arrogance on the part of the Negroes that the whites found galling. Much legislation was by bribery. Stealing extended from large blocks of state property to the price of a politician's whisky. Two hundred trial justices were said to be unable to read. Daniel H. Chamberlain, able carpetbagger and would-be reformer of his own party (governor, 1874–76), declared in 1901 that if he had been re-elected in 1876 his party, even with white assistance, could not have given government "fit to be endured." The most urgent pleas on a nonpartisan basis failing to draw the least sympathy in Washington, the whites in 1876, by combined fraud, intimidation and persuasion of Negroes, elected Gen. Wade Hampton (1818–1902) governor by a narrow majority. Pres. Rutherford B. Hayes's withdrawal of troops in March 1877 marked the collapse of "radical" rule.

Political Developments.—In 1878–80, under the leadership of M. W. Gary (1831–81), the old conflict between upcountry and low country became a fight mainly between the poorer masses and the propertied classes.

The triumph of the former under Benjamin R. Tillman (governor, 1890–94; U.S. senator, 1895–1918), prominent farmers' movement leader, was facilitated by acute agricultural distress. Tillman's victory had come in the state Democratic convention, which had been captured by his agrarian rebels. In the subsequent general election, Col. A. C. Haskell, a conservative representing the Bourbon class who had ruled the state since 1876, bolted the party and campaigned against Tillman, making his appeals to both conservatives and Negroes. His efforts failed and served only to embitter feelings among the whites. The farmers', or more broadly "reform," movement was marked by the establishment of Clemson Agricultural college (1889), the chartering of Winthrop college for women (1891), the dispensary system of state liquor monopoly (later abandoned because of corruption and failure to restrain intemperance) and the work of the constitutional convention of 1895 disfranchising so far as possible the Negro. Direct primary elections, long in use in most of the counties, beginning in some in 1878, were adopted in 1896 for naming the U.S. senators, governor and state officers.

State politics after 1890 presented a strange combination of rancorous personal politics, mass conservatism and dislike for "aristocratic" influence. Class feeling was strongly manifested in the campaigns of Coleman L. Blease (governor, 1911–15). R. I. Manning's administrations (1915–19) were notable for progressive legislation. Factionalism subsiding, Blease was in 1924 elected senator, after having been badly defeated in 1918 on a platform condemning U.S. participation in World War I.

Agriculture and banking suffered severely in the deflation of 1920–21, and all business still more from 1929 to 1937. North and South Carolina were the only states voting against the repeal of national prohibition in 1933. In 1935 Gov. O. D. Johnston climaxed his long hostility to the highway department and its chief commissioner, Ben M. Sawyer, by military seizure of the department until condemned by both legislature and supreme court. Public esteem for the construction by the department seemed justified by the failure to discover any wrong. Extensive violent strikes in 1935–36 caused a conservative reaction. Pres. Franklin D. Roosevelt's attempt to have Sen. E. D. Smith defeated resulted in Smith's overwhelming re-election in 1938, although the state supported Roosevelt's general policies.

From 1876 almost the entire white population was Democratic, partly from historical reasons and partly because of the conviction that union under the only national party that had defended them against Negro rule in the past was necessary for the maintenance of white supremacy. In 1948 the state gave its presidential electoral votes to Gov. J. Strom Thurmond, States' Rights Democrat. The Republican presidential candidates gained strong sup-

port in 1952, 1956 and 1960, but Democratic electors won in each year. However, in 1964 South Carolina voted Republican for the first time since Reconstruction (1876), supporting Barry Goldwater.

In 1950 new election laws were passed controlling general elections and providing the secret (Australian) ballot in general elections for the first time in the history of the state. Conduct of party primaries was again put under state control (after a 1944 ruling by the U.S. supreme court that had opened primaries to Negroes if conducted under state authority the legislature had repealed all primary election laws to make primaries voluntary and not subject to federal interference). In 1951 the state abolished the poll tax as a voting requirement and also undertook to curb activity of the Ku Klux Klan.

Economic Changes.—As other parts of the south, the state began an industrial revolution in the 1940s, to a large extent attributable to expansion and building related to World War II. This trend in South Carolina was accelerated following the war and was officially encouraged by the State Development board (established 1945). Changing the character of the state has been the growing diversification of its industry. Considerable growth occurred after 1939 in clothing, furniture, meat packing and the paper and pulp industry. The U.S. Atomic Energy commission also established its huge Savannah River plant in the state. (*See* also *The Economy* below.) Accompanying the 20th-century industrial revolution was a relative decline of the agriculture which once dominated the state's economy (*see* also *Agriculture* below). Despite the rapid expansion of industry the large number of marginal agricultural workers kept the per capita income of the state low ($1,332 in 1959 as compared with the U.S. average of $2,166).

GOVERNMENT

South Carolina was governed from 1670 to 1719 under the provincial charter of 1665, from 1719 to 1775 under commissions and instructions from the crown and from 1776 under constitutions of 1776, 1778, 1790, 1865, 1868 and 1895. An amendment to the constitution may be proposed in either house of the legislature. If approved by two-thirds of the members elected to each it must be submitted to the people at the next election for members of the house of representatives. If it is approved by a majority of those voting upon it and subsequently by a majority vote in each house of the general assembly chosen at that election, it becomes a part of the constitution. A convention to revise the constitution may be called by a two-thirds vote in each house ratified by a majority of those voting for representatives at the next election.

A voter must be 21 years old, a resident of the state two years, of the county one year and of the voting precinct four months; and must obtain a certificate of registration conditioned on showing that he or she is able to read and write the constitution, or the past year paid taxes on property assessed at $300 or more. (Assessments are usually at a fraction of actual value.) Paupers, idiots, the insane, prisoners and persons ever convicted of certain crimes, unless pardoned, are disfranchised.

The Democratic nomination is equivalent to election. Voting in the Democratic primary was confined to whites without other qualification than age and residence until federal courts in 1948 ordered Negroes admitted to party membership. An increasing number of Negroes voted after that date, with voting qualifications the same after 1950 in both primaries and general election.

Executive.—Constitutional amendments changed the terms of the governor and other state officers from two to four years, beginning with those elected in 1926. The governor may not succeed himself. He may veto "any one or more of the items or sections contained in any bill," but may be overruled by a two-thirds majority of those voting in each house. In practice the item veto has been used only for appropriation bills. The governor cannot grant pardons except on the recommendation of the pardon board, but he may commute a death sentence to life imprisonment.

Legislative.—The general assembly, meeting annually, is composed of the senate (one member from each of the 46 counties elected for four years) and the house of representatives (124 members elected for two years from the counties in proportion to population). The state elects six congressmen.

Judiciary.—The 5 supreme court judges and the 14 circuit court judges are elected by the legislature for ten and four years respectively. Magistrates (justices of the peace) are appointed by the governor with senate confirmation, although the recommendation of the senator from the county or of the primary election practically determines the selection.

In a number of counties, a county court (with a judge usually elected) disposes of a large part of the lesser civil and criminal cases elsewhere burdening the state circuit courts. Special juvenile courts have been established in some counties; elsewhere the probate judge has this duty. The state is divided into two federal court districts.

Local Government.—The unit of local government is the county. Counties are divided into townships, which have no governments and exist largely for the assessment of taxes. Although the forms of municipal government are provided under general statutes and enjoy certain constitutional rights, they are considered subject to state legislative authority. New industries are encouraged by exemptions from certain municipal and county taxes under conditions set by the state constitution and statutes.

The county is administered by a commission, almost always popularly elected. Road or hospital commissions, etc., are frequently appointed by the governor, usually upon recommendation of the county legislative delegation. Most of the county officers are elected, with the exception of the auditor and treasurer, who are appointed by the governor with senate confirmation; however, the governor almost always names the successful primary election candidates without question.

Since there is no legislative body on the county level, the state legislature is oppressed with a mass of local legislation. In actual practice it enacts without question whatever the senator and representatives of the county concerned agree upon. The delegation thus forms in effect a county legislature. The senator of each county, by controlling action in the senate, exercises an enormous power and in effect has a veto over any legislation concerning his county.

Finances.—The state tax commission, created in 1915, supervises all tax administration and directly administers those taxes from which most of the state revenue is derived, such as taxes on incomes, gasoline, retail sales, soft drinks, beer and liquor, electric power, corporations and business. There exists in practice a considerable separation of state and local tax sources, the local governments depending mainly on the general property tax; however, the counties receive a share of state gasoline and income taxes, and both county and municipality share the state tax on alcoholic beverages.

POPULATION

The population of South Carolina in 1790 was 249,073; in 1840 it was 594,398; in 1880, 995,577; in 1910, 1,515,400; in 1940, 1,899,804; in 1950, 2,117,027; and in 1960, 2,382,594. The 1960 figure represents an increase of 12.5% over the population in 1950. The population per square mile in 1960 was 76.7, as compared with 68.2 in 1950 and with 49.6 for the United States in 1960.

Of the 1960 population, 739,300 or 31.0% lived in incorporated places of 2,500 or more, as compared with 28.8% in 1950. The entire urban population amounted to 981,386, or 41.2% of the state total. The state has 3 standard metropolitan statistical areas, which are Charleston, Columbia and Greenville and part of a fourth (Atlanta, Ga.). These areas had a total population of 852,050, or 35.8% of the total population of the state in 1960.

The number of households in 1960 was 603,551, as compared with 514,660 in 1950. The average population per household had declined from 4.1 in 1950 to 3.8 in 1960.

The population of the state was distributed by colour and nativity in 1950 as follows: 60.7% native white; 0.4% foreign-born white; and 38.9% nonwhite, practically all Negro. There were 99.1 males per 100 females in the native white population and 93.0 in the Negro population; 5.4% of the population was 65 years old or over; and 56.2% of the population 14 years old and over was in the labour force. Of the total number of em-

*South Carolina: Places of 5,000 or More Population (1960 census)**

Place	Population				
	1960	1950	1940	1920	1900
Total state	2,382,594	2,117,027	1,899,804	1,683,724	1,340,316
Abbeville	5,436	5,395	4,930	4,570	3,766
Aiken	11,243	7,083	6,168	4,103	3,414
Anderson	41,316	19,770	19,424	10,570	5,498
Beaufort	6,298	5,081	3,185	2,831	4,110
Belton	5,106	3,371	2,119	1,780	826
Bennettsville	6,963	5,140	4,895	3,197	1,929
Camden	6,842	6,986	5,747	3,930	2,441
Cayce	8,517	3,294	1,476	746	—
Charleston	65,925	70,174	71,275	67,957	55,807
Cheraw	5,171	4,836	4,497	3,150	1,151
Chester	6,906	6,893	6,392	5,557	4,075
Clinton	7,937	7,168	5,704	3,767	1,869
Columbia	97,433	86,914	62,396	37,524	21,108
Conway	8,563	6,073	5,066	1,969	703
Darlington	6,710	6,619	6,236	4,669	3,028
Dillon	6,173	5,171	3,867	2,205	1,015
Easley	8,283	6,316	5,183	3,568	903
Florence	24,722	22,313	16,054	10,968	4,647
Gaffney	10,435	8,123	7,636	5,068	3,937
Georgetown	12,261	6,004	5,559	4,579	4,138
Greenville	66,188	58,161	34,734	23,127	11,860
Greenwood	16,644	13,806	13,020	8,703	4,824
Greer	8,967	5,050	2,940	2,292	648
Hartsville	6,392	5,658	5,399	3,624	704
Lake City	6,059	5,112	2,522	1,606	375
Lancaster	7,999	7,159	4,430	3,032	1,477
Laurens	9,598	8,658	6,894	4,629	4,029
Marion	7,174	6,834	5,746	3,892	1,831
Mount Pleasant	5,116	1,857	1,698	1,575	2,252
Mullins	6,229	4,916	4,392	2,379	828
Myrtle Beach	7,834	3,345	1,597	—	—
Newberry	8,208	7,546	7,510	5,894	4,607
North Augusta	10,348	3,659	2,629	1,742	—
Orangeburg	13,852	15,322	10,521	7,290	4,455
Rock Hill	29,404	24,502	15,009	8,809	5,485
Seneca	5,227	3,649	2,155	1,460	920
Shannontown	7,064	5,828	—	—	—
Spartanburg	44,352	36,795	32,249	22,638	11,395
Sumter	23,062	20,185	15,874	9,508	5,673
Union	10,191	9,730	8,478	6,141	5,400
Walterboro	5,117	4,616	3,373	1,853	1,491
West Columbia†	6,410	1,543	1,744	1,793	1,089

*Populations are reported as constituted at date of each census. †Name changed from Brookland in 1938.
Note: Dash indicates place did not exist during reported census, or data not available.

ployed males, 31.1% were engaged in agriculture, 7.9% in construction, 27.7% in manufacturing and 17.3% in transportation and trade.

EDUCATION

Public Schools.—As early as 1710 the state made meagre provision for the education of indigent children, and in 1811 an effort was made to inaugurate a genuine public school system. Because of small appropriations and a law that paupers be given first preference, this system proved very inadequate and most education continued to be by private schools and academies in the antebellum period.

The modern free school system was established in 1868. The educational system is under the supervision of a state superintendent of education, with the assistance of a board composed of the governor, the superintendent of education and seven other persons appointed by the governor. State support and regulation of high schools were greatly increased in 1907.

In 1951 Gov. James F. Byrnes successfully sponsored a long-range expansion program designed to improve and equalize physical facilities for white and Negro schools. State bond issues up to $75,000,000 (later increased) were authorized and a 3% retail sales tax earmarked for education was adopted. Prior to this, white children, who in 1950 constituted 55% of the school population, had occupied school property constituting 71.3% of the total value; in 1950 about 67.5% of new construction had been for white schools. Both this ratio and the total expenditure were altered by the 1951 reform efforts. In the first year of the program 78% of the funds for school construction was spent on Negro schools. The value of school property increased from $90,720,547 in 1950 to $333,754,446 in 1958. Per pupil expenditure increased 274.4% between 1929 and 1956.

The 1954 decision of the U.S. supreme court holding compulsory racial segregation in public schools unconstitutional centred political activity and public interest on the racial issue. The state constitution was immediately amended to remove the requirement of state support for schools. In 1955 the legislature repealed the compulsory attendance law and the law regulating school terms. Local school boards of trustees were given increased authority and freedom, including the exclusive authority to operate or not operate schools. It was also decreed that appropriations should cease for any school when any pupil was transferred to or from that school pursuant to a court order.

Enrollment in the public elementary schools during the late 1950s was approximately 400,000, of whom 220,000 were white and 180,000 Negro; in high schools there were about 190,000, including 120,000 white and 70,000 Negro. The school construction program launched in 1951 was reflected in the Negro high school enrollment, which doubled between 1950 and 1958.

Higher Education.—*State Institutions.*—The state government supports several institutions of higher learning, chief among which is the University of South Carolina at Columbia. The university, chartered as South Carolina college in 1801 and opened for instruction in 1805, comprises schools of arts and sciences, business administration, education, engineering, journalism, law, nursing and pharmacy. It also conducts junior college programs at Florence, Lancaster and Beaufort, S.C. By the early 1960s the university was enrolling more than 6,000 students annually; the teaching staff comprised more than 270. The university library included 379,000 volumes.

Clemson Agricultural college at Clemson (incorporated 1889), with 3,500 students, is the state agricultural and mechanical college, with schools of agriculture, architecture, engineering and textiles, as well as arts and sciences. Winthrop college at Rock Hill (established 1886) is a state college of liberal arts for women, specializing in teacher training. At Charleston is located the Citadel, the Military College of South Carolina (1842), one of the few state-supported military colleges. The South Carolina State college at Orangeburg (1896), for Negroes, has schools of arts and sciences, education, agriculture, home economics and engineering.

In 1956 the legislature provided for the closing of state-supported white colleges if served with a court order to admit Negroes; if this occurred, the state college for Negroes must also cease operation.

Private Institutions.—Major Protestant denominations operate several accredited colleges for white persons, among them Columbia college, for women, at Columbia (Methodist, 1854); Erskine college, at Due West (Presbyterian, 1839); Furman university, at Greenville (Baptist, 1825); Newberry college, at Newberry (Lutheran, 1856); Presbyterian college, at Clinton (Presbyterian, 1880); Wesleyan Central college, at Central (Wesleyan Methodist, 1906); Wofford college, for men, at Spartanburg (Methodist, 1854). There are also several nondenominational colleges: Bob Jones university, at Greenville (1927); Coker college, at Hartsville, for women (1908); College of Charleston (1785); Converse college, at Spartanburg, for women (1889); Lander college, at Greenwood (1872); and Limestone college, at Gaffney, for women (1845).

In addition to the state college, there are a number of private colleges for Negroes, among them Allen university, at Columbia (African Methodist, 1870); Benedict college, at Columbia (Baptist, 1870); Claflin college, at Orangeburg (Methodist, 1869); and Morris college, at Sumter (Baptist, 1908).

HEALTH, WELFARE AND CORRECTIONS

The state has maintained a State Board of Health since 1879, and beginning in the 1880s similar agencies were established on the municipal level and later by counties. The state board conducts programs for the control of preventable diseases (notably venereal and heart diseases, poliomyelitis, cancer and tuberculosis) and for maternal and child care. Its program was expanded considerably following passage of the Social Security act in 1935. In the early 1960s the number of existing acceptable hospital beds in the state exceeded 10,000. The state also maintains a mental health commission. Most mental patients are treated at the state hospital, Columbia, although there are four other smaller institutions for such care as well as five mental health clinics.

The state department of public welfare (established 1937) is concerned mainly with dependent children, the blind and the aged.

The state and counties share in maintaining penal institutions

for convicted criminals. The state maintains a penitentiary, reformatories and industrial training schools for delinquent minors.

THE ECONOMY

Agriculture.—South Carolina's farm products are normally valued at much less than its factory output. Thus, in the late 1950s, the value of manufactures was over six times greater than the gross income from agriculture. In the period 1950–58 gross income from agriculture increased only 1% while value of manufactured products increased about 64%.

During the 1940s and 1950s the average size of farms increased, mechanization was widely adopted and farmers generally shared in prosperity. From 1920 through 1950, the number of farms declined by 27.7% and the farm population by 34.5%. South Carolina agriculture, like industry, became more diversified in the 20th century. This was particularly noticeable in the case of cotton, which declined 68.6% in value during the period, and in that of peaches, production of which increased by 737.8%.

Total farm acreage harvested in 1958 was 3,070,000, in comparison with 4,097,000 in 1950. By the mid-1950s about 40% of farm operators were tenants, as compared with 45% in 1950 and 65% in 1930. Of the total number employed in agriculture, almost 65% were Negroes.

Cotton, long dominant in agriculture in South Carolina, took second place to tobacco in crop value for the first time in 1955. Third crop in value is corn. Among the other major crops are wheat, oats, rye and barley; four legumes that yield both hay and grain (cowpeas, soybeans, velvet beans and peanuts); and two sirup crops (sorghum and sugar cane). Potatoes, both white and sweet, fruit and vegetables also add to the wealth of the state.

The coastal region, with a mild climate and a soil responding quickly to intensive cultivation, is especially suited to commercial truck farming. The chief products, exclusive of potatoes and cane grown for sirup, are watermelons, cabbages, cucumbers, string beans, beets, asparagus, tomatoes and lettuce. Strawberries, dewberries, blackberries, figs and grapes add to the fruit wealth of the state.

Dairying and the livestock industry were receiving much greater attention at mid-century. Between 1920 and 1950 the value of cattle and hogs increased 62.4% and 26.9%, respectively. There also was a large increase in poultry.

Forests.—Forests make up around 60% of the area of South Carolina, and timber resources are estimated at about 15,000,000,000 bd.ft. in softwoods and 10,700,000,000 bd.ft. in hardwoods. Emphasis has been placed on pulpwood production for use in the paper industry.

Minerals.—South Carolina's mining output is very small, amounting to 0.12% of the total mineral value of the United States. Chief minerals are kaolin (china clay), vermiculite (produced in only one other state), stone, sand and gravel. The stone quarried is chiefly granite and granite-gneiss. High-grade granite, the Winnsboro Blue, is quarried in Fairfield county.

Manufactures.—The Industrial Revolution that began in the 1940s continued through the 1950s, and in the late 1950s there were more than 2,000 manufacturing establishments in South Carolina employing nearly 200,000 wage earners and annually producing goods worth over $3,350,000,000 (1960). Besides new local industries, many manufacturing operations were moved from elsewhere to South Carolina, and many industries added new plants in the state. Older industries also undertook expansion. With more than 35% of the spindle hours in the nation, textiles dominate industry in the state and make it the largest textile producer in the United States. In terms of the total value of product, the six leading industries of the state in the late 1950s were textiles (more than ten times as much as the next most valuable product), clothing, paper and pulp, electricity, meat packing and barrels, boxes, baskets and veneers.

Textile manufacturing plants consume more than 2,300,000 bales of raw cotton annually. Almost every grade of cotton cloth is produced, and there is also widespread use of synthetic materials, many of which are made in the state. Factors contributing to the rapid development of the cotton textile industry in South Carolina were a suitable climate, dependable native labour and an abundance of cheap hydroelectric power. The Piedmont section has led in the textile industry, chief producing counties being Spartanburg, Greenville, Anderson, York, Kershaw and Greenwood.

Charleston and Richland counties are important for diversified industry. Industries showing the greatest growth between 1936 and 1958 were clothing, furniture, meat packing and paper and pulp.

Hydroelectric development and new steam generating plants greatly aided industry and improved living conditions. Of the numerous hydroelectric installations the greatest is the publicly owned Santee-Cooper project, producing approximately 133,000 kw. Its two reservoirs (Lakes Marion and Moultrie) have a combined area of 160,000 ac. One-half of the 280,000 kw. installed at Clark Hill (publicly owned) on the Savannah river is also available to South Carolina. Lake Murray dam, about 10 mi. above Columbia on the Saluda river, 208 ft. high and 7,000 ft. long (one of the largest earth dams), forms a lake about 50,000 ac. in area and develops 130,000 kw. The state's total installed generating capacity (both privately and publicly owned) is more than 1,500,000 kw.

Transportation, Commerce and Communication.—The chief railroad systems of South Carolina are the Southern, the Atlantic Coast Line and the Seaboard Air Line. Mainline railway mileage was reduced from 3,780 mi. at the end of 1930 to 3,257 in 1957. During the same period there was a steady increase in the mileage of paved roads built by the state government. By the late 1950s mileage of state highways reached almost 27,000; the post-World War II period brought 15,000 mi. of paved secondary roads also. Automobile registrations were 230,000 in 1929 and more than 600,000 in the late 1950s.

Ocean traffic through Charleston (*q.v.*) and Georgetown expanded greatly after World War II, promoted by the State Ports authority, which began a large expansion of port facilities in 1958. Charleston ranked 15th among the nation's ports in the late 1950s, with more than $200,000,000 in export-import trade annually. Between 1946 and 1957 the dollar value of exports and imports through South Carolina harbours increased 671%.

See also Index references under "South Carolina" in the Index volume.

BIBLIOGRAPHY.—J. H. Easterby, *South Carolina Bibliographies* (No. 1, 1949, Topical Lists; No. 2, 1950, General Classified Bibliography); D. D. Wallace, *History of South Carolina,* 4 vol. (1934) and *South Carolina* (1951); Edward McCrady, *History of South Carolina, 1670–1783,* 4 vol. (1897, 1899, 1901–02); W. R. Smith, *South Carolina as a Royal Province, 1719–1776* (1903); R. L. Meriwether, *The Expansion of South Carolina, 1729–1765* (1940); C. G. Singer, *South Carolina in the Confederation* (1941); J. H. Wolfe, *Jeffersonian Democracy in South Carolina* (1940); C. S. Boucher, *Nullification Controversy in South Carolina* (1916); A. G. Smith, *Economic Readjustment of an Old Cotton State: South Carolina, 1820–1860* (1958); P. M. Hamer, *Secession Movement in South Carolina, 1847–1852* (1918); H. S. Schultz, *Nationalism and Sectionalism in South Carolina, 1852–1860* (1950); C. E. Cauthen, *South Carolina Goes to War, 1860–1865* (1950); F. B. Simkins and R. H. Woody, *South Carolina During Reconstruction* (1931); F. B. Simkins, *Tillman Movement in South Carolina* (1926) and *Pitchfork Ben Tillman, South Carolinian* (1944); S. M. Derrick, *Centennial History of the South Carolina Railroad* (1930); C. J. Milling, *Red Carolinians* (1940); W. A. Schaper, "Sectionalism and Representation in South Carolina," in Amer. Historical Assocn. *Annual Report,* I, 1900; R. H. Taylor, *Ante-Bellum South Carolina* (1942); G. B. Tindall, *South Carolina Negroes, 1877–1900* (1952); J. G. Van Deusen, *Economic Bases of Disunion in South Carolina* (1928); D. D. Wallace, *South Carolina Constitution of 1895* (1927); J. K. Williams, *Vogues in Villainy* (1959). Selected biographies: A. K. Gregorie, *Thomas Sumter* (1931); T. D. Jervey, *Robert Y. Hayne and His Times* (1909); L. A. Kibler, *Benjamin F. Perry, South Carolina Unionist* (1946); Broadus Mitchell, *William Gregg, Factory Master of the Old South* (1928); D. D. Wallace, *Life of Henry Laurens* (1915); Laura A. White, *Robert Barnwell Rhett: Father of Secession* (1931); C. M. Wiltse, *John C. Calhoun,* 3 vol. (1944–51); Margaret L. Coit, *John C. Calhoun, American Portrait* (1950); H. T. Cook, *Life and Legacy of David Rogerson Williams* (1916). *Elementary books:* H. K. Hennig, *Great South Carolinians,* 2 vol. (1940–1949); Mary C. S. Oliphant, *History of South Carolina* (numerous late editions); Federal Writers' Project, WPA, *South Carolina* (American Guide, 1941).

Current statistics on production, employment, industry, etc., may be obtained from the pertinent state departments; the principal figures,

together with the current history, are summarized annually in the *Britannica Book of the Year*, American edition.
(D. D. W.; C. E. Cn.; L. P. J.)

SOUTH CHARLESTON, a city of West Virginia, U.S., across the Kanawha river from Charleston (*q.v.*), is an important chemical manufacturing centre. It was founded in 1916 and incorporated in 1917. Two chemical companies were established there in 1916, and a plant built in 1920 to manufacture antifreeze has become the city's largest single industry, employing over 10,000 workers and participating in the manufacture of about 450 articles. A U.S. naval ordnance plant located in South Charleston in 1917 has also been a stimulus to the economy. In the heart of the city is an Indian mound 33 ft. in height and 550 ft. in circumference. In 1883 it was explored to a depth of 31 ft. by archaeologists under the direction of the Smithsonian institution at Washington, D.C. For comparative population figures *see* table in West Virginia: *Population.* (K. K. McC.)

SOUTHCOTT, JOANNA (1750–1814), English religious fanatic, was born at Gittisham in Devonshire. Her father was a farmer and she herself was for a considerable time a domestic servant. She was originally a Methodist, but about 1792, becoming persuaded that she possessed supernatural gifts, she wrote and dictated prophecies in rhyme, and then announced herself as the woman spoken of in Revelation xii. Coming to London at the request of William Sharp (1749–1824), the engraver, she began to "seal" the 144,000 elect at a charge varying from 12*s.* to a guinea. When over 60 she affirmed that she would be delivered of Shiloh on Oct. 19, 1814, but Shiloh failed to appear, and it was given out that she was in a trance. She died of brain disease on Oct. 29, 1814. Her followers are said to have numbered more than 100,000, and are not yet extinct.

Among her 60 publications, all equally incoherent in thought and grammar, may be mentioned: *Strange Effects of Faith* (1801–02); *True Explanation of the Bible* (1804); *The Book of Wonders* (1813–14); and *Prophecies Announcing the Birth of the Prince of Peace* (1814). A lady named Essam left large sums of money for printing the *Sacred Writings of Joanna Southcott.*

See D. Roberts. *Observations on the Divine Mission of Joanna Southcott* (1807), R. Reece, *Correct Statement of the Circumstances attending the Death of Joanna Southcott* (1815).

SOUTH DAKOTA, popularly called the "Coyote" or "Sunshine state," is one of the north-central plains states of the United States, bordered on the north by North Dakota, east by Minnesota and Iowa, south by Nebraska and west by Wyoming and Montana. It is 372 mi. (east-west) by 204 mi., with an area of 77,047 sq.mi., of which 669 sq.mi. are covered by rivers and lakes; in size it ranks 16th among the states. Its name is taken from the Dakota or Sioux Indians. The geographical centre of the United States is located in the northwestern corner of the state. South Dakota was admitted to the union in 1889 as the 40th state. Its capital is at Pierre (*q.v.*). The state motto is "Under God the People Rule," the flower is the American pasqueflower, the bird, the ring-necked pheasant, the song (unofficial) "South Dakota." The state flag has a field of blue on which is represented a blazing sun in gold surrounded by the words "South Dakota the Sunshine State."

PHYSICAL GEOGRAPHY

Physical Features.—South Dakota (extreme points 42° 29′ and 45° 57′ N. lat. and 96° 26′ and 104° 3′ W. long.) has some of the features of regions that bound it: the mountains of Wyoming to the west, the lakes of Minnesota to the east and the rich corn lands of Iowa to the southeast. With its other border states, Nebraska and North Dakota, it is part of the continental Great Plains (*q.v.*), a land that rises gradually to the foothills of the Rockies, accompanied by a climatic change from humid to semiarid. The half of the state east of the Missouri, which originally flowed in the valley of the modern James river, was subjected to glacial action and is largely flat prairie. In the northeastern part of the state there are numerous glacial lakes. Oddly, the lowest point in the state, at the northeastern edge, is a continental divide. From this point waters flow northward to Hudson bay and southward into the Gulf of Mexico. Apart from this extreme corner, the entire state is drained by the Missouri river and its tributaries.

The Missouri, which bisects South Dakota north-south, flows through a veritable gorge except for a short stretch above Sioux City, Ia., at the southeastern corner of the state. The channel is 300 to 500 ft. below the bordering uplands, with the river plain seldom more than two miles and often less than one mile in width. This feature has made possible the location of four federal dams, primarily for hydroelectric power and flood control, along the South Dakota stretches of the Missouri. Total installed capacity is set at 1,232,000 kw. The dams form a series of lakes roughly equal in area to the state's natural water area.

The valley of the James river, on the other hand, averages 60 mi. in width, and the river drops only 100 ft. in its course from north to south across the state (east of the Missouri). The *coteaux* on either side of the James are largely undrained, reflecting their glacial history. The Big Sioux river is at the eastern edge of the state, where, for the last part of its course, it forms the state boundary.

The land west of the Missouri is largely rolling grassland, cut by a number of streams of considerable size. On either side of the White river are the Bad Lands (*q.v.*), the eroded sections of which, banded in soft colours and carved into unusual shapes, are interspersed with numerous well-grassed flat-topped tablelands.

From the James river to the Black hills at the western edge of the state the land rises gently from 2,000 ft. above sea level to 3,700 ft. This area is subject to recurrent periods of drought, the worst occurring in the early 1890s and in the 1930s.

The Black hills (*q.v.*) are actually well-eroded mountains, Harney peak (7,242 ft.) being the highest point in the United States east of the Rocky mountains. The region is virtually the only wooded section of the state and the source of a wide variety of minerals. There are three national forests in the western part of the state, covering nearly 2,000,000 ac.

Climate.—South Dakota straddles the 100th meridian, the usual dividing line between the arid and more humid sections of the United States. Rainfall is heaviest in the southeast, averaging 24 in., but tapers off to the west and north to a low of 14 in. Variability in the weather and a wide range of temperature help to make the climate invigorating. Highest temperature recorded is 120° F.; the lowest −58°. Droughts of varying intensity marked the periods 1886–94, 1910–11, 1930–36 and the mid-1950s. It is estimated that much of the state receives less than 75% of its normal rainfall 20% of the time. The average snowfall is 36 in., ranging from 22 in. to about 59 in. Three-fourths of all precipitation occurs during the growing season. Contrary to popular belief, blizzards are not common. The severest storms on record are those of 1881 and 1949.

Soil.—Most of the eastern half of the state lies in the chernozem (black earth) belt of the Great Plains. The soil varies in depth from 18 to 28 in., is rich in lime and highly fertile. The chestnut soils of western and northwestern South Dakota range from 6 to 14 in. in depth and are generally fertile, but climatic factors limit their use. Soils in the Black hills are thin and vary widely in type. The parent materials of most soils in the state are soft rocks, loess, glacial till and alluvium. Clay and clay shale are important parent materials in the west-central section of the state where gumbo soil or Pierre clay is commonly found. Fringe areas on the northern and southern borders of the state have soils derived from sand.

Vegetation and Animal Life.—A wide diversity of trees is native to the eastern section, including elm, oak, box elder, cottonwood and hackberry. Small cedars are found in much of the rough country west of the Missouri. In the Black hills are ponderosa pine, spruce, aspen, white birch and cedar. Animals widely found include deer and antelope. Buffalo and prairie dogs can still be seen in Black hills preserves. Pheasants can also be found in the state.

Parks, Historic Sites and Recreation.—The state has 170 state and federal parks, recreational areas, camps and picnic grounds. Included in this number are the Mt. Rushmore National

memorial, Wind Cave National park, Custer State park in the Black hills area, the Badlands National monument and the Gavins Point dam and Fort Randall dam recreation areas.

The gigantic sculptured faces of Washington, Jefferson, Lincoln and Theodore Roosevelt at Mt. Rushmore are the greatest single tourist attraction in the state, followed by the weird landscape of the Bad Lands of the White river. Camping is especially popular in the Black hills and at federal dams along the Missouri. There are extensive facilities for boating and fishing in the Black hills, in the northeastern region of the glacial lakes, at Lewis and Clark lake above Gavins Point dam and at the Fort Randall dam. The state boasts some of the best pheasant hunting in the nation and is also on the flyway of ducks and geese. Deer are hunted in the Black hills and in some eastern counties. Antelope, once the most popular game animal at frontier army posts, are hunted in western sections of the state.

Sites of great historic interest include a number connected with the periods of exploration and Indian warfare. At Ft. Pierre a small monument marks the place where Louis and François Verendrye buried a lead plate in 1743, claiming the region for the king of France. Ft. Wadsworth, near the Sisseton Indian reservation, has been restored and is maintained by the state. Points visited by Lewis and Clark include Spirit mound, near the town of Vermillion. The site of the Wounded Knee massacre of 1890 is on the Pine Ridge reservation. Yankton, the capital of Dakota Territory from 1861 to 1883 and onetime river port, has many fine old homes as well as a Territorial museum.

HISTORY

Mound Builders resided in the area that is now South Dakota prior to A.D. 1200. Evidence of their occupation has been found in 22 counties east of the Missouri. The Arikara, an agricultural and village dwelling Indian people, inhabited the region after 1600. The first white men to visit the area were the Verendrye brothers, who came into northwestern South Dakota via the Mandan villages of North Dakota in 1742 and may have visited the Black hills. On March 30, 1743, they planted a lead plate on a hilltop on the site of what was later Ft. Pierre, claiming the land for France. The plate was recovered by three school children in 1913.

In 1762 the area was ceded to Spain, and it was under Spanish rule that fur traders began visiting this part of the St. Louis trade empire. The arrival of fur traders coincided with the arrival of the Sioux or Dakota Indians, pushing westward from the woodlands of Minnesota and driving the Arikara upriver before them. Under Napoleon France reclaimed Louisiana, but held it only three years before selling it to the United States in 1803. Lewis and Clark crossed the state in the course of their explorations.

The fur trade on the upper Missouri reached its peak in the 1830s and 1840s. Ft. Pierre, started as a fur trading post in 1817, became the capital of a fur territory of 250,000 sq.mi. Among the principal traders were Manuel Lisa, Pierre Chouteau and Registre Loisel. In support of fur traders Col. Henry Leavenworth in 1823 led a small party of U.S. soldiers and a motley array of trappers, traders and Indian allies against the Arikara at the mouth of the Grand river. In 1831 the "Yellowstone," the first steamboat on the upper Missouri, traversed the state. For more than 50 years thereafter steamboats provided the readiest means of access to the Dakota frontier.

Settlement.—No attempt at settlement by persons other than traders was made until 1856, when land speculators from Iowa and Minnesota, drawn by the prospects of water power at the falls of the Big Sioux river, located there. Indian title had been surrendered by the treaty of Traverse des Sioux of 1851. These settlements were abandoned at the time of the War of the Indian Outbreak in 1862. A permanent settlement was effected at Yankton in 1859, following a treaty with the Yankton Sioux. Three years later, with the creation of Dakota Territory, Yankton became its capital. In the next decade fewer than 12,000 people migrated to the territory, establishing settlements largely along the Missouri and its lower tributaries, the Big Sioux, Vermillion and James rivers.

Rate of settlement increased with the completion of a railroad from Sioux City to Yankton in 1872 and to Sioux Falls and Watertown in 1878. The Black hills expedition of Gen. George A. Custer in 1874, which confirmed rumours that gold was to be found there, also hastened the development of the territory; this invasion of unceded lands contributed directly to the warfare in which Custer lost his life in 1876. The richest gold deposits were found in the area of Deadwood (*q.v.*) and Lead. With favouring rains, settlement reached boom proportions in the 1880s, tripling the territorial population and filling in the lands east of the Missouri, largely with north Europeans.

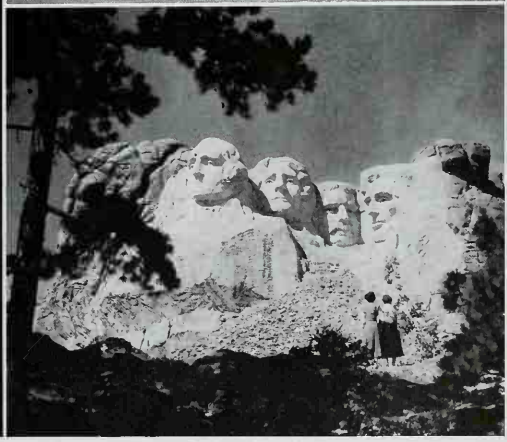
Statehood.—Efforts to organize as a state were hampered by Democratic opposition in congress, it being known that South Dakota as a state probably would vote Republican. The Republican party had emerged victorious in 1876 by such a narrow margin that its opponents were reluctant to add to Republican strength in congress or in the electoral college. In the election of 1888, however, Republicans captured control of both houses and the presidency, making statehood a certainty and influencing Pres. Grover Cleveland to sign the enabling act under which North and South Dakota were subsequently admitted to the union, simultaneously, on Nov. 2, 1889.

The next decade was marked by drought and a national economic depression. The distress of Dakota farmers is well described in the works of Hamlin Garland, whose parents had settled near Aberdeen. Farm discontent found expression in wide support of the Populist movement, including the election of a Populist governor and a sweep of the state for William Jennings Bryan in the "free silver" presidential campaign of 1896 (*see* POPULIST PARTY). With the return of prosperity, however, the state resumed its normal Republican course in politics, to be disturbed again only by a flurry for Theodore Roosevelt when he ran as a Progressive in 1912 and by Franklin D. Roosevelt in 1932 and 1936. Democratic governors were elected in periods of farm protest, W. I. Bulow in 1926, re-elected in 1928, Tom Berry in 1932, re-elected in 1934 and 1936, and Ralph Herseth in 1958.

The Republican party frequently has been divided between standpatters and progressives. The latter faction was responsible for constitutional amendments that permitted the state (1915–25) to enter into a variety of business affairs, including rural credits, state bonding, state coal mining, state hail insurance, state cement plant and state guarantee of bank deposits. The state cement plant proved to be highly remunerative, but the other ventures, particularly rural credits, left the state in debt for the next 30 years.

Settlement of the lands west of the Missouri awaited the extension of rail lines and the concentration of Indians upon smaller reservations in the decade before World War I. Further development of the state also was advanced by the discovery of dry farming techniques and the development of plant types adapted to the region. Highway bridges across the Missouri were not constructed until the 1920s. South Dakota suffered greatly during the period of drought, grasshoppers and low farm prices of the 1930s. The dust storms, whipped up from areas where several years of above-normal rainfall had encouraged farmers to plow under the native buffalo grass, were worst in 1934 and 1935. Cash farm income fell to one-third of its former level. In a single year (1933) one farm out of every ten in the state was taken from its owner through foreclosure or tax sale. After World War II, on the other hand, only 1 farm in 1,000 was ordinarily foreclosed or sold for taxes in any one year. At one time nearly two-fifths of the people of the state were on government relief.

By mid-century life on South Dakota farms had improved in many ways. Tenancy and mortgage indebtedness were at a record low. In 1930 seven-eighths of all farms were located on dirt or unimproved roads. With more adequate revenues, the state legislature after 1945 appropriated large sums for interstate and farm-to-market roads. Under Gov. Joe Foss the state greatly increased its support to higher education. Also during the post-World War II period, the Black hills became a major tourist attraction. The gigantic sculptured monument of Washington, Jefferson, Lincoln and Roosevelt, carved on the face of Mt. Rushmore by Gutzon Borglum, was opened to the public in 1942.



BY COURTESY OF (TOP RIGHT) RAPID CITY CHAMBER OF COMMERCE, (CENTRE RIGHT) MITCHELL CHAMBER OF COMMERCE, (BOTTOM RIGHT) SOUTH DAKOTA DEPARTMENT OF HIGHWAYS; PHOTOGRAPH, (LEFT) GRANT M. HAIST

SCENES IN SOUTH DAKOTA

Left: Tunnel on Needles highway cut through the solid granite upthrust of the Harney range in the Black Hills

Top right: Two of the five life-sized models of Mesozoic reptiles in Dinosaur park, Rapid City

Centre right: Corn palace, Mitchell, scene of an annual harvest festival. The exterior of the building is decorated with corn and other grain products of the state

Bottom right: Mt. Rushmore National memorial, Black Hills. The heads of U.S. presidents Washington, Jefferson, Theodore Roosevelt and Lincoln were carved out of the mountain's granite face by Gutzon and Lincoln Borglum from 1927 through 1941

SOUTH DAKOTA

PHOTOGRAPHS, (TOP) HINE FROM FREDERIC LEWIS, (BOTTOM LEFT) HERBERT LANKS FROM BLACK STAR, (BOTTOM RIGHT) DAVID W. CORSON FROM A. DEVANEY

THE BAD LANDS AND OTHER SCENES IN SOUTH DAKOTA

Top: Typical formation in the Bad Lands, an area of eroded rock in the White river region of southwestern South Dakota

Bottom left: The state capitol at Pierre. Built chiefly of sandstone and limestone, it was erected in 1907–10, with an annex added in 1931

Bottom right: View along a street in Rockerville, a late 19th-century mining boom town, deserted then restored as a tourist attraction. Between 1876 and 1882 more than $1,000,000 in gold ore was extracted from nearby deposits

GOVERNMENT

The 1889 constitution provides for a bicameral legislature (35 senators and 75 representatives) which is elected and meets biannually for a 60-day session. The governor is nominated by direct primary while most other state officers are nominated by party convention. The constitution is notable for its detailed provisions, a reflection of a populistic distrust of legislatures. In addition, the constitution can be amended by a simple majority vote of legislature and electors. Amendments have been added at the rate of approximately one per year. Most notable is the 1898 amendment which provides for initiative and referendum. South Dakota having been the first state to adopt such measures. The state's 5 supreme court judges, serving for 6 years, and its 20 circuit judges, serving for 4 years, are elected on a nonpolitical ballot.

The state has no personal income tax and ordinarily levies no general property tax. Among the states, South Dakota ranks fourth to seventh from the bottom in per capita payment of state taxes. Sales and use taxes and motor fuel taxes account for more than 60% of all revenue from taxes. After World War II expenditures for highways amounted to roughly one-third of all state expenditures. The other leading items of expenditure were education and social welfare. With the retirement of rural credit bonds in 1954, the state became free of debt.

POPULATION

The area which now forms South Dakota had 11,776 inhabitants in 1870 and 348,600 in 1890. The population of the state in 1910 was 583,888; in 1940, 642,961; in 1950, 652,740; and in 1960 680,514. This last figure represented an increase of 4.3% over the population in 1950. The population per square mile in 1960 was 8.8, as compared with 8.5 in 1950 and with 49.6 for the United States in 1960.

Of the 1960 population 267,180, or 39.3%, lived in incorporated places of 2,500 or more, as compared with 33.1% in 1950 and

*South Dakota: Places of 5,000 or More Population (1960 census)**

Place	Population				
	1960	1950	1940	1920	1900
Total state	680,514	652,740	642,961	636,547	401,570
Aberdeen	23,073	21,051	17,015	14,537	4,087
Brookings	10,558	7,764	5,346	3,924	2,346
Huron	14,180	12,788	10,843	8,302	2,793
Lead	6,211	6,422	7,520	5,013	6,210
Madison	5,420	5,153	5,018	4,144	2,550
Mitchell	12,555	12,123	10,633	8,478	4,955
Pierre	10,088	5,715	4,322	3,209	2,306
Rapid City	42,399	25,310	13,844	5,777	1,342
Sioux Falls	65,466	52,696	40,832	25,202	10,266
Vermillion	6,102	5,337	3,324	2,590	2,183
Watertown	14,077	12,699	10,617	9,400	3,352
Yankton	9,279	7,709	6,798	5,024	4,125

*Populations are reported as constituted at date of each census.

24.6% in 1940. The state has one standard metropolitan statistical area, which is Sioux Falls. This area had a total population of 86,575 or 12.7% of the total population of the state in 1960.

The number of occupied dwelling units (or households) in 1960 was approximately 217,000, as compared with 190,000 in 1950 and 165,000 in 1940. The average population per household had declined from 3.9 in 1940 to 3.6 in 1950 and to 3.1 in 1960.

The population of the state was distributed by colour and nativity in 1950 as follows: 89.4% native white; 6.9% foreign-born white; and 3.7% nonwhite, almost all Indians. Most of the Indians (nearly all Sioux) live on the state's eight reservations. Of the foreign-born white population a large proportion were born in Scandinavian countries. Germany and Russia also provided significant numbers of immigrants.

In 1950 there were 105.4 males per 100 females in the native white population, and 132.0 in the foreign-born white. Of the total population, 6.8% were 65 years old or over, and 50.3% of the population 14 years old and over were in the labour force. Of the total number of employed males, 57.5% were engaged in agriculture, 3.7% in construction, 4.8% in manufacturing and 17.7% in transportation and trade.

EDUCATION

Public Schools.—Approximately nine-tenths of the state's primary and secondary school pupils attend public schools. State aid amounts to about 10% of the cost of public education. A high proportion of local property taxes goes to its support, and in many localities school authorities have found it difficult to operate within prescribed mill levy limitations. School reorganization in rural districts is not compulsory, and many common school districts do not operate schools, preferring to pay tuition at other schools for the limited number of children in their own districts. This is particularly true for high school students.

Higher Education.—The State University of South Dakota, located at Vermillion, is chief among the state institutions of higher education. Established in 1882, it is governed, along with the other state institutions, by the Regents of Education, State of South Dakota, a seven-member board appointed by the governor with the consent of the state senate. The university comprises schools and colleges of arts and sciences, business, education, fine arts, law, medicine and nursing. By the second half of the 20th century it enrolled about 2,500 students annually; the faculty numbered more than 200.

The state also supports the South Dakota State College of Agriculture and Mechanic Arts at Brookings (1883); the South Dakota School of Mines and Technology at Rapid City (1885); and state teachers' colleges at Aberdeen (1901), Spearfish (1883), Springfield (1881) and Madison (1881).

Private institutions of higher education include Augustana college at Sioux Falls (Lutheran, 1860), Dakota Wesleyan university at Mitchell (Methodist, 1885), Huron college at Huron (Presbyterian, 1883), Sioux Falls college (Baptist, 1883) and Yankton college at Yankton (Congregational, 1881).

HEALTH, WELFARE AND CORRECTIONS

The South Dakota department of public welfare provides old-age assistance, aid to dependent children, the disabled and the blind, and general assistance. The state maintains one prison and one training school.

THE ECONOMY

Agriculture.—South Dakota leads all the states in the proportion of personal income derived from agriculture (more than 28%). The economy of the state is thus subject to wide fluctuations, varying with both the market and the weather. Approximately two-thirds of cash farm income is derived from livestock and products; alfalfa, corn and oats are commonly sold "on the hoof" rather than as cash crops. Only in the northeastern and north-central sections do cash crops amount to as much as one-half of cash farm income. The average size of farms in the southeastern section is more than 225 ac.; in the northern wheat areas it is 800 ac.; and in the western range sections, 3,000 ac. In all sections the trend is toward larger farm units and fewer operators. Rural farm population decreased one-third between 1930 and 1950, while the number of farms was reduced by one-fourth. Farm tenancy during the same period declined from a high of 53% in 1940 to 30% in 1950.

The state's leading crop is corn, nearly all hybrid strains that have helped to increase yields and to extend the corn area north and west within the state. Sorghum grain is planted as a substitute for corn during extended periods of drought, and in small amounts at other times. Production of corn averages a little over 100,000,000 bu. on approximately 4,000,000 ac. of land. The state usually ranks ninth in the production of corn.

Wheat production, both in acres harvested and in total yield, varies more widely than does corn production. During years of good rain large areas of range land have been plowed up and planted to wheat. This was particularly true during World Wars I and II, which were also periods of high wheat prices. As much as 4,000,000 ac. of land have been planted to wheat, but government quotas reduced this to between 2,000,000 and 3,000,000 ac. In spite of these limitations, total production increased. The state usually ranks tenth in the production of all wheat and second in the production of durum wheat. South Dakota is usually among

the first five states in numbers of beef cattle and calves and among the first ten in numbers of hogs and sheep. Milk production is concentrated in the eastern quarter of the state, near to urban centres, while sheep are raised primarily in the northwestern counties. The state ranks tenth in wool production.

Irrigation is carried on extensively in only a few localities. The largest federal reclamation project is Oahe dam on the Missouri, designed to provide water for lands in the James river valley. The Belle Fourche project dates from 1908, the Shadehill dam from 1950 and the Angostura dam from 1952. Other irrigation is carried on along the arid fringes of the Black hills and, by means of wells, in scattered localities in the humid eastern parts of the state.

Manufacturing.—Meat packing and the processing of dairy and poultry products are the leading industries, accounting for nearly two-thirds of all workers engaged in manufacturing (total: 12,000–13,000). Printing and publishing rank next in importance, followed by lumber and wood products.

Mining.—South Dakota ranks first among the states in the production of gold and beryl, second in feldspar, columbium-tantalum and lithium minerals and third in lignite. Between $16,000,000 and $20,000,000 worth of gold is mined annually. The Homestake mine at Lead is the largest gold mine in the western hemisphere. Sioux quartzite and a mahogany-coloured granite are quarried in the eastern part of the state. Uranium ores were discovered in the Black hills in 1951. There are a small number of oil wells, the first dating from 1953. There are undeveloped deposits of low-grade manganese ore near Chamberlain and low-grade uraniferous lignite in the northwestern section of the state.

Transportation and Communication.—The first railroad reached the southeastern corner of South Dakota in 1872, but it was 1881 before the Chicago and North Western and Milwaukee lines reached the central part of the state; it was 1907 before the Missouri river was bridged and east-west rail connections established to the Black hills. There are nearly 4,000 mi. of rail lines in the state. Yankton, Dakota Territory, was a thriving port for steamboats prior to 1881, but the new railroad facilities of that year reduced river traffic to short hauls. The construction of a network of highways, begun in the 1920s, was interrupted by depression and war, but after World War II highway construction claimed the largest share of the state budget. The completion of five Missouri river bridges (1924–26) was a major milestone in establishing the state trunk system which, by the early 1960s, consisted of more than 7,000 mi. of highway. Eleven cities are served by commercial airlines; the state has more than 30 public airports.

See also Index references under "South Dakota" in the Index volume.

BIBLIOGRAPHY.—Robert J. Casey, *The Black Hills and Their Incredible Characters* (1949); George E. Hyde, *Red Cloud's Folk* (1937) and *A Sioux Chronicle* (1956); G. W. Kingsbury, *History of Dakota Territory* (1915); Bruce Nelson, *Land of the Dacotahs* (1946); M. L. Reese, *South Dakota, a Guide to the State*, American Guide Series, rev. ed. (1952); Herbert S. Schell, *History of South Dakota* (1961); S. S. Visher, *Geography of South Dakota* (1918).

Current statistics on production, employment, industry, etc., may be obtained from the pertinent state departments; the principal figures, together with the current history, are summarized annually in the *Britannica Book of the Year*, American edition. (E. W. St.)

SOUTHEAST ASIA. This term usually denotes the area comprising the East Indies (Greater and Lesser Sunda islands, Moluccas and Philippines) and the Indo-Chinese peninsula of continental Asia. Indonesia, the Philippines, British territories in Borneo (Sarawak, Brunei, North Borneo) and Portuguese Timor are political units in the former section; Burma, Thailand, Cambodia, Laos, Vietnam (North and South), Malaya and Singapore in the latter.

Mainland Southeast Asia consists principally of a series of roughly parallel, north-south-trending ranges and valleys, associated with some of Asia's larger rivers; *e.g.*, Mekong and Irrawaddy. Insular Southeast Asia includes several geologically unstable island-arcs, possessing an abundance of active volcanoes, as well as a stable zone (Borneo, eastern Sumatra and nearby areas) resting on the Asiatic continental platform. The climate is predominantly tropical, with annual rainfall generally above 40 in. The natural vegetation is forest.

The area is highly diversified culturally, although Malaysia (Malaya and the East Indies), with strong affinities among its languages, is a partial exception. Rice and maize are the principal dietary staples. Southeast Asia is an important source of such agricultural products as rubber, rice, coconut oil, sugar, teak, Manila hemp, tea, coffee, tobacco and pepper, and of tin, petroleum, bauxite and other minerals. Industry is poorly developed. Prior to the 16th century, Southeast Asia had important commercial and other contacts with India and China, and in several places, *e.g.*, Cambodia, Sumatra, Java, had evolved high civilizations. From the 16th to the 20th centuries the area was influenced by the west, all portions except Thailand becoming colonies of western powers. By 1961, however, all parts of Southeast Asia except British Borneo and Portuguese Timor had achieved independence.

BIBLIOGRAPHY.—E. H. G. Dobby, *Southeast Asia* (1950); Karl J. Pelzer, *Pioneer Settlement in the Asiatic Tropics* (1945); Charles Robequain, *Malaya, Indonesia, Borneo, and the Philippines* (1954); Rupert Emerson, *Representative Government in Southeast Asia* (1955); Russell H. Fifield, *The Diplomacy of Southeast Asia: 1945–1958* (1958); D. G. E. Hall, *A History of South-east Asia* (1955). (J. M. Br.)

SOUTHEAST ASIAN LANGUAGES. Southeast Asia is one of the great linguistic treasure houses of the world, both in number of languages found and in the complexity of their interrelationships. Burma alone is variously reported to have between 125 and 140 languages spoken within its borders. Although national languages like Vietnamese, Cambodian, Thai, Malay and Burmese are literary languages with a centuries-old tradition, others, like the Karen languages and Shan, have acquired a writing system only within the past hundred years. Little is known about most of the rest of the languages. Missionaries have compiled dictionaries of some, travelers and officials have written vocabularies of others, but many have never been recorded at all.

Three important linguistic families are represented: (1) Austronesian or Malayo-Polynesian; (2) Mon-Khmer; and (3) Sino-Tibetan. Actual linguistic reconstructions based on well-attested phonetic correspondences have been worked out only for the Austronesian family. Malay (Malaya) is the principal representative of this family on the southeast Asian mainland. Mon-Khmer, in its narrowest sense, includes Mon (Burma) and Cambodian or Khmer (Cambodia). Sino-Tibetan may be conservatively described as having three branches: (1) Tibeto-Burman; (2) Chinese; and (3) Tai. Tibeto-Burman, named after the two great literary languages, Tibetan (Tibet) and Burmese (Burma), also includes many nonliterary languages, *e.g.*, Lushai, Garo and Bodo (Assam); Jingphaw or Kachin (Burma) and possibly the languages of the Chins and Nagas (Burma) and the Lolo (southwestern China). The Karen languages (Burma and Thailand) may also be related. The best-known Tai languages are Thai or Siamese (Thailand), Lao (Laos) and Shan (Burma). Many lesser-known languages also unquestionably belong here, including the extinct Ahom (Assam), and Dioi, Tho, Nung, Lung-chow and Wu-ming (northern Vietnam and southern China), as well as others. The consonants, consonant clusters and tone classes of Proto-Tai are now fairly well known, and the tonal system bears a surprisingly detailed resemblance to that of Ancient Chinese.

The above description of family affiliations is conservative and leaves many questions unanswered. Vietnamese, for example, has never been proved related to any other languages. Other theories of affiliation have been proposed, but, as a glance at any two handbooks on southeast Asia soon reveals, many of them are contradictory. One theory places Chinese and Tai together in a single subfamily on a par with Tibeto-Burman in Sino-Tibetan. According to another theory, however, Tai is not related to Chinese at all (or to Tibeto-Burman) but is affiliated with Austronesian. Much the same kind of confusion is found with other southeast Asian languages. Thus Vietnamese is classified now as a Mon-Khmer language, now as a Tai affiliate. The much-debated Austroasiatic hypothesis attempts to encompass the Munda languages (India), Khasi (Assam), Nicobarese (Nicobar Islands), Semang and Sakai (Malaya and southern Thailand), as well as Mon-Khmer, Vietnamese and Muong. Miao and Yao (China) are sometimes added, but other students favor a Tai or Tibeto-Burman affiliation.

The Austroasiatic hypothesis, tenuous as it is, has appealed to many as the starting point for other connections; *e.g.*, Austroasiatic plus Austronesian (the "Austric" theory), Austroasiatic plus Japanese and Austric plus Sino-Tibetan. On an even wider scale, transpacific relationships with American Indian languages have also been suggested. Possibly deserving of further investigation is Edward Sapir's proposal, in 1925, that Sino-Tibetan (including Tai) may be related to Nadene (the Athapascan languages plus Tlingit and Haida).

Only when linguistic studies of southeast Asia have been greatly intensified can we hope to reach a stage at which the interrelationships of these languages can be cogently described. This will require, among other things, the linguistic reconstruction of such families as Tibeto-Burman and Mon-Khmer according to the same rigorous methods that have been used so successfully to arrive at Proto-Austronesian and Proto-Indo-European.

See also the articles on individual languages and language groups.

BIBLIOGRAPHY.—Paul K. Benedict, "Thai, Kadai and Indonesian: a New Alignment in Southeastern Asia," *Amer. Anthrop.*, vol. 44, pp. 576–601 (1942); John F. Embree and Lillian Ota Dotson, *Bibliography of the Peoples and Cultures of Mainland Southeast Asia* (1950) (includes languages); Thomas A. Sebeok, "An Examination of the Austroasiatic Language Family," *Language*, vol. 18, pp. 206–217 (1942).

(M. R. H.)

SOUTHEAST ASIA TREATY ORGANIZATION, usually known as SEATO, is the regional defense organization created by the Southeast Asia Collective Defense treaty signed at Manila, Sept. 8, 1954, by the representatives of Australia, France, New Zealand, Pakistan, the Philippines, Thailand, the United Kingdom and the United States. The treaty entered into force Feb. 19, 1955.

The formation of SEATO responded to a demand that the area be protected against Communist expansionism especially as manifested through military aggression in Korea and Indochina and through subversion backed by organized armed forces in Malaya and the Philippines. The Republic of Vietnam and the kingdoms of Cambodia and Laos (the successor states of Indochina) were not considered for membership in SEATO for reasons largely relating to the 1954 Geneva agreements ending the fighting in Vietnam. These states were, however, accorded military protection by a protocol. Other nations of south and southeast Asia preferred to retain their foreign policies of nonalignment.

The treaty area is the general area of southeast Asia, including the entire territories of the Asian members and the general area of the Southwest Pacific south of latitude 21° 30′ N. The treaty defines its purposes as defensive only and includes provision for self-help and mutual aid in preventing and countering subversive activities directed from without, and co-operation in promoting economic progress and social well-being.

SEATO has no standing forces but is based upon the mobile striking forces of its member states, which regularly engage in combined military exercises. A secretary-general, acting under the direction of the council or its representatives, heads an international staff in Bangkok which supplements the work of the organization's expert committees on public relations, area security, cultural affairs and certain economic matters. (R. W. Z.)

SOUTHEND-ON-SEA, a county and parliamentary borough and seaside town of Essex, Eng., 39 mi. E. of London bridge by road, on the Thames estuary facing south. Pop. (1961) 164,976. Area 16.1 sq.mi. The borough includes Leigh-on-Sea, Westcliff-on-Sea, Thorpe Bay and Shoeburyness. With its 7-mi. sea front it is one of England's largest seaside towns and the nearest to London. It has gardens, beaches, an open-air theatre, yachting, a carnival week, magnificent illuminations from August to late October, and a pier (1¼ mi.) which accommodates pleasure steamers. The oldest part of the borough is Prittlewell where the priory (1110) has housed the museum since 1917. Southchurch hall (13th century) is a public library and Porters (Tudor) is the civic house. Southend became fashionable as a seaside resort when visited by Princess Charlotte of Wales in 1801 and her mother Princess Caroline in 1803.

It was incorporated in 1892, became a county borough in 1914 and a parliamentary borough in 1918. It returns two members. Besides providing for more than 5,000,000 annual holiday makers, Southend's many industries, mostly on the new industrial estate, include radios and televisions, electrical apparatus, glass blowing and engineering.

SOUTHERN AND ANTARCTIC TERRITORIES (TERRES AUSTRALES ET ANTARCTIQUES FRANÇAISES). These barren lands, which were linked for administrative purposes with Madagascar (now the Malagasy Republic) from 1924, have been effectively occupied (mainly by scientific personnel) only since 1949. In Aug. 1955 they became a territory of the French Union, later French Community, governed by a senior administrator who is partially resident in Paris. The territories comprise: in the south Indian ocean, and in a temperate cold climate, the islands of Saint Paul and Nouvelle Amsterdam; in a subantarctic climate the Kerguelen and Crozet groups of islands; and Adélie land on the antarctic continent. The scientific research station of the territories played a significant part in the International Geophysical Year program (1957–58).

Nouvelle (New) Amsterdam (area 19.3 sq.mi. in latitude 37° 52′ S., longitude 77° 32′ E.) and **Saint Paul** (2.7 sq.mi. in 38° 43′ S., 77° 31′ E.) are two volcanic islands. Nouvelle Amsterdam is an extinct volcano (2,953 ft.); the crater of Saint Paul has been infilled by the sea. The temperature varies from 41° to 68° F. with an average of 53.6°. Fog is frequent and the westerly winds are strong. Mosses, bracken, grasses and a scattering of trees (*Phylica arborea*) grow there. Animal life consists mainly of sea birds, fish and crayfish. Access to the islands is difficult. Magellan's companions discovered Nouvelle Amsterdam in 1522. Fishermen from Réunion used to visit the islands and France annexed them in 1843. In 1949 a permanent radio-meteorological station, Camp Heurtin, was established at Nouvelle Amsterdam. Research stations, an administrative office and a hospital were also established there.

Kerguelen, lying latitude 49°–50° S. and longitude 69°–70° E., covers about 2,700 sq.mi. It is volcanic and has a certain similarity to Iceland. It comprises one main island with an area of about 2,700 sq.mi. and many small islets. The main island is broken into ravines by glacial erosion, and has many peninsulas and fjords, chiefly in the east. The Cook glacier occupies the west and Mt. Ross in the south reaches 6,430 ft. Temperatures are not excessively low (averages: winter 32° F., summer 42.8°). Rain and snow are frequent, the west wind strong and the sea frequently rough. The vegetation consists solely of two flowering bushes (*Acaena* and *Azorella*), some enormous sea weeds and the edible "Kerguelen cabbage" (*Pringlea antiscorbutica*). There are numerous sea birds: penguins, petrels, cormorants, sheathbills, skuas and albatrosses. The chief of the species of seal which go to Kerguelen to breed is the huge sea elephant (*Mirounga leonina*). The rabbit, introduced by man, has become a pest. Kerguelen was discovered in 1772 by the French navigator Yves Joseph de Kerguélen-Trémarec who took possession of it. The island was explored in 1776 by Capt. James Cook who called it the Island of Desolation, and since then it has been frequented by seal hunters. In 1950 a permanent base and scientific centre, Port aux Français, was established on the west coast. It has a hospital, military camp and seal-oil plant. Reindeer, sheep, poultry and vegetables have been introduced. There are lignite beds and peat bogs, and some semiprecious stones have also been found. About 100 persons (relieved at regular intervals) live on the island.

The Crozet archipelago, discovered in 1772, comprises two groups of volcanic islands between latitude 45°–47° S. and longitude 50°–52° E.: (1) the islands of Pigs, Penguins and Apostles, together with several small islands; (2) Possession island, which rises to 4,921 ft., and Eastern Island which is glacier covered. The total area is 193 sq.mi. The coasts are steep. The climate, vegetation and animal life are similar to those of Kerguelen. The islands were discovered by Capt. Marion Dufresne and his mate, Crozet, in 1772. They are uninhabited, but have now been created a national park.

Adélie Land (Terre Adélie), discovered by Dumont d'Urville in 1840, is situated south of Australia on the edge of Antarctica between longitude 136° and 142° E. It is a plateau of ancient rocks

covered by ice and has an area of about 155,000 sq.mi. It can only be reached in summer (average summer temperature 32° F.). In winter there are more than 54° of frost and blizzards reach 124 m.p.h. Whales, seals, petrels and penguins were the sole inhabitants until 1949, when a scientific mission was established at Pointe Géologie on the Île de Petrels (the Dumont d'Urville base). Exploratory teams have penetrated 342 mi. into the interior.

See also ANTARCTICA.

See Aubert de la Rue, *Les Terres Australes* (1953); *Terres Australes et Antarctiques Françaises* (quarterly). (HU. DE.)

SOUTHERN DESERT TERRITORY (AL BADIYAH AL JANUBIYAH), a region in the extreme southwest of Iraq, bordering Saudi Arabia. The region was in 1947 formed into one of three *liwa's* or provinces (along with the Northern Desert territory and Al Jazirah) from what had previously been termed tribal areas and left undemarcated. Area 29,399 sq.mi. The population is of mixed tribal allegiance, mainly Anaiza ('Unayzah) and Amarat ('Imarat), with many Dhafir and Muntafiq and some others. The region lies astride a main pilgrimage route southwest from Karbala but otherwise has few international connections and trade links. The population is almost entirely nomadic, and there are no important settled centres. By 1960 the territory had ceased to be a *liwa'*, and its area and population had been administratively attached to the settled *liwa's*. (W. B. FR.)

SOUTHERNE, THOMAS (1660–1746), English dramatist, was born at Oxmantown, near Dublin, in 1660. His first play, *The Persian Prince, or the Loyal Brother* (1682), contained a flattering portrait of James II. In 1692 he revised and completed *Cleomenes* for John Dryden; and in 1694 he scored a success in *The Fatal Marriage, or the Innocent Adultery*, based on Aphra Behn's *The Nun*, with a comic underplot. The general spirit of his comedies is well exemplified by a line from *Sir Anthony Love* (1691): "Every day a new mistress and a new quarrel." Southerne scored another success in *Oroonoko, or the Royal Slave* (1696). He died on May 22, 1746.

SOUTHERN OVERLAND MAIL. Early United States mail service to the Pacific coast was slow and unreliable. Beginning in 1849, the year of the gold rush, mail was carried by steamboat to the Isthmus of Panama, over the isthmus by mule or canoe and then up to California by steamboat. Inadequate overland mail service was provided along a central route through Salt Lake City, Utah, in 1851. Thereafter, demands mounted for greater government support for overland service. The problem was complicated by the bitter quarrels of the period between southerners and northerners in congress. In March 1857 congress passed an overland mail bill that left the choice of route to the postmaster general, Aaron V. Brown of Tennessee. He chose a southern route, defending it as the best all-weather route. In Sept. 1857 John Butterfield of Utica, N.Y., and associates signed a contract to transport letter mail from St. Louis, Mo., and Memphis, Tenn., by way of Fort Smith, Ark., El Paso, Tex., Tucson and Yuma, Ariz., and Los Angeles to San Francisco, Calif., and back "twice a week in good four-horse post-coaches or spring wagons, suitable for the conveyance of passengers as well as the safety and security of the mails." Butterfield and his associates were to receive a government subsidy of $600,000 a year for six years, 1858–64. The first eastbound mail on the new route left San Francisco on Sept. 15, 1858, and the first westbound mail left St. Louis on the following day.

Although northern critics objected to the circuitous oxbow route, 2,795 mi. from St. Louis to San Francisco, the service proved satisfactory. The contractors were able to provide the 25-day service they promised. Stations were built at intervals of 15 mi. or less, and Concord coaches carried passengers as well as mail. Unfortunately for the contractors the American Civil War interrupted service and made it necessary to move the line to a central route on July 1, 1861. The contract was transferred to Ben Holladay in 1862, and he in turn transferred his interests to Wells, Fargo and company in 1866. The completion of a transcontinental railroad in 1869 brought cancellation of the last contract for carrying overland mail by stagecoach.

BIBLIOGRAPHY.—R. P. and M. B. Conkling, *The Butterfield Overland Mail, 1857–1869*, 3 vol. (1947); L. R. Hafen, *The Overland Mail 1849–1869* (1926); L. H. Wright and Josephine M. Bynum (eds.), *The Butterfield Overland Mail* (1942). (T. A. LN.)

SOUTHERN PROVINCE, CEYLON, covers an area of 2,146 sq.mi. of southernmost Ceylon and had a population (1953) of 1,129,308. It consists for the most part of coastal lowland crossed by rivers, the more important of which are the Nilwala Ganga and Walawe Ganga, but in the western part the relief is more prominent and more confused, and includes the southern part of the Rakwana hill mass, which rises to more than 4,000 ft. Apart from coastal sands and alluvium, Southern province is made up entirely of ancient crystalline rocks.

The western part of the province (Galle and Matara districts and Giruwa Pattu west of Hambantota district) lies in the wet zone of Ceylon and has a rainfall of 70 in. or more, well distributed throughout the year; the hillier areas of the interior are particularly wet. The rest of the province is in the dry zone and is subject to a searing seasonal drought from May to September. The extreme east is particularly dry. Natural vegetation has largely disappeared from the wet zone, though some forests and fernlands remain; much of the dry zone is still under tropophilous forest, which passes into poor scrub as the driest areas are approached. In this region is Ruhuna (Yala) National park (50 sq.mi.), where elephant, bear, leopard, deer and other wild animals may be studied. (*See* also CEYLON: *Physical Geography.*)

Ruhuna is the traditional name for an area wider than the present Southern province and it functioned as one of the two main core areas of ancient Sinhalese civilization. The part of Southern province lying in the dry zone, together with the belt transitional to the wet zone, was of great importance, as is witnessed by the many surviving ruins, irrigation works and inscriptions. The ruins of the ancient city of Magama are at Tissamaharama. Wet zone places like Matara later became more prominent. It was the area in the wet zone that was particularly valuable to the Portuguese and Dutch, who used Galle and Matara, in particular, as fortified trading posts.

Low-country Sinhalese form the majority of the population. The wet zone is densely peopled, but apart from occasional areas of denser settlement the dry zone has a sparse population, the extreme east being virtually empty. The main towns are Galle (*q.v.*), capital of the province (55,848), Matara (27,641) and Hambantota (4,299), all district capitals; others are Weligama (12,834), Ambalangoda (10,554) and Tangalla (6,823).

The province has a varied economy. The wet zone produces rice, coconut, rubber, tea, cinnamon and citronella. In the dry zone rice-growing under new or restored irrigation works is of increasing importance. Fishing is also carried on. A coastal railway runs from Colombo to Galle and Matara, the latter being the railhead. The road network is adequate in the wet zone and rather rudimentary in the dry zone. (B. H. F.)

SOUTHEY, CAROLINE ANNE (1786–1854), second wife of Robert Southey (*q.v.*) and the daughter of an East Indian captain, Charles Bowles, was born at Lymington, Hants, on Oct. 7, 1786. After her mother died in 1816 she was left in dire financial straits through the fraudulence of a guardian. In order to support herself, she determined to turn her talents to account in literature. She sent anonymously to Southey a narrative poem called "Ellen Fitzarthur," which may be taken as typical, in its prosy simplicity, of the rest of the author's work. Southey became interested in her, and their acquaintanceship and long friendship developed through a correspondence that lasted 20 years.

A combination of prose and verse in eight volumes, *Solitary Hours*, was published in 1826. This was succeeded by her work *Chapters on Churchyards*, a group of stories first published in *Blackwood's Magazine* and later in complete form in 1829. Her writing was simple yet had power and a strong element of pathos. She was at her best in prose or blank verse rather than in lyric.

Other prose writings which were published included *Tales of the Moors* (1828) and *Selwyn in Search of a Daughter* (1835). Besides the works already mentioned, Mrs. Southey wrote *The Widow's Tale, and Other Poems* (1822) and *Robin Hood* in conjunction with Southey, at whose death the poem was incomplete.

In 1839 she was married to Southey. Soon after the marriage, however, her husband's mental state became hopeless. From that time till his death in 1843, and indeed till her own, her life was a sad one. She was not on good terms with her stepchildren, and her share in Southey's life was hardly noticed in Charles Cuthbert Southey's *Life and Correspondence* of his father. But with Edith Southey, who was the wife of J. W. Warter, she was always in friendly relations, and she supplied the valuable additions to Southey's correspondence published by Warter.

Her correspondence with Southey, neglected in the official biography, was edited by Edward Dowden in 1881. Mrs. Southey died at Buckland cottage, Lymington, on July 20, 1854, two years after Queen Victoria had granted her an annual pension of £200.

SOUTHEY, ROBERT (1774–1843), English poet and man of letters, was born at Bristol on Aug. 12, 1774. His father, Robert Southey, an unsuccessful linendraper, married Margaret Hill in 1772.

When he was three, Southey passed into the care of Elizabeth Tyler, his mother's half-sister, at Bath, where most of his childhood was spent. Miss Tyler was a whimsical and despotic person, of whose household he left an amusing account in the fragment of autobiography written in a series of letters to his friend John May. Before Southey was eight years old he had read Shakespeare and Beaumont and Fletcher, while his love of romance was fostered by the reading of Hoole's translations of Tasso and Ariosto, and of the *Faerie Queene*.

In 1788 he was entered at Westminster school. After four years there he was privately expelled by William Vincent (1739–1815) for an essay against flogging called "The Flagellant," written for a school magazine. At Westminster he made two friends who proved faithful and helpful to him through life—Charles Watkyn Williams Wynn and Grosvenor Bedford. His uncle, Herbert Hill, chaplain of the British factory at Lisbon, who had paid for his education at Westminster, determined to send him to Oxford with a view to his taking holy orders, but the news of his escapade at Westminster had preceded him, and he was refused at Christ Church. Finally he was admitted at Balliol, where he matriculated on Nov. 3, 1792, and took up his residence in the following January. His father died soon after his matriculation. At Oxford he lived a life apart and gained little or nothing from the university except a liking for swimming and a knowledge of Epictetus.

Revolutionary Youth.—In the vacation of 1793 Southey's enthusiasm for the French Revolution found vent in writing an epic poem, *Joan of Arc*, published in 1796 by Joseph Cottle, the Bristol bookseller. In 1794 Coleridge, then on a visit to Oxford, was introduced to Southey, and filled his head with dreams of an American Utopia on the banks of the Susquehanna. The members of the "pantisocracy" were to earn their living by tilling the soil, while their wives cared for the house and children. Coleridge and Southey soon met again at Bristol, and with Robert Lovell developed the emigration scheme. Lovell had married Mary Fricker, whose sister Sara married Coleridge, and Southey became engaged to a third sister, Edith. Miss Tyler, however, would have none of "pantisocracy" and "aspheterism," and drove Southey from her house. To raise the necessary funds for the enterprise Coleridge and he turned to lecturing and journalism. Cottle generously gave Southey £50 for *Joan of Arc;* and, with Coleridge and Lovell, Southey dashed off the drama, printed as the work of Coleridge, *The Fall of Robespierre*. A volume of *Poems* by R. Southey and R. Lovell was also published by Cottle in 1795.

Southey's uncle, Hill, desired him to go with him to Portugal. Before he started for Corunna he was married secretly (Nov. 14, 1795) to Edith Fricker. On his return to England his marriage was acknowledged, and he and his wife had lodgings for some time at Bristol. He was urged to undertake a profession, but the church was closed to him by the Unitarian views he then held, and medicine was distasteful to him. He was entered at Gray's Inn in Feb. 1797 and made a serious attempt at legal study, but with small results. At the end of 1797 his friend Wynn began an allowance of £160 a year, which was continued until 1806, when Southey relinquished it on Wynn's marriage. His *Letters Written During a Short Residence in Spain and Portugal* was printed by Cottle in 1797, and in 1797–99 appeared two volumes of *Minor Poems* from the same press.

In 1798 he paid a visit to Norwich, where he met Frank Sayers and William Taylor, with whose translations from the German he was already acquainted. He then took a cottage for himself and his wife at Westbury near Bristol, and afterward at Burton in Hampshire. At Burton he was seized with an illness which had been threatening for some time. He moved to Bristol, and after preparing for the press his edition of the works of Thomas Chatterton, undertaken for the relief of the poet's sister and her child, he sailed in 1800 for Portugal, where he began to accumulate materials for his history of Portugal. He also had taken with him the first six books of *Thalaba the Destroyer* (1801), and the remaining six were completed at Cintra. The unrhymed, irregular metre of the poem was borrowed from Sayers.

Life at Greta Hall.—In 1801 the Southeys returned to England, and at the invitation of Coleridge, who held out as an inducement the society of Wordsworth, they visited Keswick. After a short experience as private secretary to Isaac Corry, chancellor of the exchequer for Ireland, Southey in 1803 settled at Greta hall, Keswick, which he and his family shared thenceforward with the Coleridges and Mrs. Lovell. There he accumulated a library of more than 14,000 volumes, including valuable manuscripts and a collection of Portuguese authorities probably unique in England. After 1809, when Coleridge left his family, the whole household was dependent on Southey's exertions. His nervous temperament suffered under the strain, and he found relief in keeping different kinds of work on hand at the same time, in turning from the *History of Portugal* to poetry. *Madoc* and *Metrical Tales and Other Poems* appeared in 1805, *The Curse of Kehama* in 1810, *Roderick, the Last of the Goths*, in 1814.

This constant application was lightened by a happy family life. Southey was devoted to his children and was hospitable to the many friends and even strangers who found their way to Keswick. His friendship for Coleridge was qualified by a natural appreciation of his failings, the results of which fell heavily on his own shoulders, and he had a great admiration for Wordsworth. He met Walter Savage Landor in 1808, and their friendship lasted until Southey's death.

From the establishment of the Tory *Quarterly Review* Southey, whose revolutionary opinions had changed, was one of its most regular and useful writers. He supported church and state, opposed parliamentary reform, Roman Catholic emancipation and free trade. He did not cease, however, to advocate measures for the immediate amelioration of the condition of the poor. With William Gifford, his editor, he was never on very good terms, and would have nothing to do with his harsh criticisms on living authors. His relations with Gifford's successors, Sir J. T. Coleridge and Lockhart, were not much better. In 1813 the laureateship became vacant on the death of Pye. The post was offered to Scott, who refused it and secured it for Southey. A government pension of some £160 had been secured for him, through Wynn, in 1807, increased to £300 in 1835. In 1817 the unauthorized publication of an early poem on *Wat Tyler*, full of his youthful republican enthusiasm, brought many attacks on Southey. He was also engaged in a bitter controversy with Byron, whose first attack on the "ballad-monger" Southey in *English Bards and Scotch Reviewers* nevertheless did not prevent them from meeting on friendly terms. Southey makes little reference to Byron in his letters, but Byron asserts (*Letters and Journals*, ed. Prothero, iv. 271) that he was responsible for scandal spread about himself and Shelley. In this frame of mind, due as much to personal anger as to natural antipathy to Southey's principles, Byron dedicated *Don Juan* to the laureate, in what he himself called "good, simple, savage verse." In the introduction to his *Vision of Judgment* (1821) Southey inserted a homily on the "Satanic School" of poetry, unmistakably directed at Byron, who replied in the satire of the same name. The unfortunate controversy was renewed even after Byron's death, in consequence of a passage in Medwin's *Conversations of Lord Byron*.

Meanwhile the household at Greta Hall was growing smaller.

Southey's eldest son, Herbert, died in 1816, and a favourite daughter in 1826; Sara Coleridge married in 1829; in 1834 his eldest daughter, Edith, also married; and in the same year Mrs. Southey, whose health had long given cause for anxiety, became insane. She died in 1837, and Southey went abroad the next year with Henry Crabb Robinson and others. In 1839 he married his friend Caroline Bowles. (*See* SOUTHEY, CAROLINE.) But his health was failing, and he died on March 21, 1843.

A collected edition of his *Poetical Works* (10 vols., 1837–38) was followed by a one volume edition in 1847. Southey's letters were edited by his son Charles Cuthbert Southey as *The Life and Correspondence of the late Robert Southey* (6 vols., 1849–50); further selections were published in *Selections from the Letters of Robert Southey* (4 vols., 1856), edited by J. W. Warter; and *The Correspondence of Robert Southey with Caroline Bowles. To which are added: Correspondence with Shelley, and Southey's Dreams* (1881), was edited, with an introduction, by Professor E. Dowden. An excellent selection from his whole correspondence, edited by Mr. John Dennis, as *Robert Southey, the story of his life written in his letters* (Boston, Massachusetts, 1887), was reprinted in Bohn's Standard Library (1894). *See* also *Southey* (1879) in the English Men of Letters Series, by Professor E. Dowden, who also made the selection of *Poems by Robert Southey* (1895) in the Golden Treasury Series. W. Haller, *Early Life of Robert Southey (1774–1803)* (Columbia, 1917). A full account of his relations with Byron is given in *The Letters and Journals of Lord Byron* (vol. vi., 1901, edited by R. E. Prothero), in an appendix entitled "Quarrel between Byron and Southey," pp. 377–399.

SOUTHGATE, a municipal borough (1933) of Middlesex, Eng., which under the London Government Act 1963 was to be amalgamated on April 1, 1965, with Edmonton and Enfield to form the new London borough of Enfield (*see* LONDON). Southgate municipal borough, on rising ground 9 mi. N. of Charing Cross, includes parts of Winchmore Hill, Palmers Green, Bowes Park, Cockfosters and New Southgate (formerly Colney Hatch). Pop. (1961) 72,051. Area 5.9 sq.mi. Southgate is a parliamentary constituency.

Southgate grew up as a settlement at the south gate of Enfield chase in the great Middlesex forest; Oakwood park preserves 64 ac. of Enfield chase. The other three big public parks are Broomfield park (54 ac.) with the Jacobean Broomfield house, now a local museum; Grovelands park (91 ac.) with a lake covering 7 ac.; and Arnos park (44 ac.). Formerly part of the parish of Edmonton, Southgate achieved a separate existence in 1881. Leigh Hunt, the poet, was born in Southgate and Thomas Hood, the humorist and poet, lived there for a few years. It is now almost wholly a residential area, the chief industries being the making of metal boxes and die casting machine tools. (Go. H. T.)

SOUTH GATE, an industrial and residential city of central Los Angeles county, Calif., U.S., is located 12 mi. S. of the Los Angeles civic centre and a part of the Los Angeles metropolitan area (*see* LOS ANGELES). Prior to 1915 the area of South Gate was devoted to truck farming. Enterprising real estate men saw the possibilities of developing the area and encouraged heavy industry to come in; South Gate's growth has been continuous since. It was incorporated as a city Jan. 20, 1923, and has a council-manager form of government in effect since 1958. Manufactured products include automobiles, fibreboard products, auto tires, metal drums, chemical products and airplane parts. Pop. (1960) 53,831. For comparative population figures *see* table in CALIFORNIA: *Population.* (J. M. Wo.)

SOUTH GEORGIA, an island in the South Atlantic ocean, 800 mi. E. of the Falkland Islands, of which it is a dependency. Length is 105 mi., maximum breadth 20 mi. and area 1,450 sq.mi.; est. pop. (1961) 521. Chief settlement is Grytviken, in Cumberland bay, on the north coast; it has a wireless station and is served by a vessel of the Falkland Islands company. Whaling is the sole industry. Captain Cook claimed the island for Great Britain in 1775. "Discovery II" made coastal surveys during 1926–30.

See L. H. Matthews, *South Georgia* (1931); D. Ommaney, *South Latitude* (1938), in the U.S. titled *Below the Roaring Forties* (1938).

SOUTH HOLLAND, the most crowded province of the Netherlands; area 1,234 sq.mi.; pop. (1962 est.) 2,755,724. The province has strips of dunes, geest lands, low fen and clay, but the southwest area consists of ramifications of the Lek, Maas and Waal producing a mosaic of islands such as Dordrecht, Voorne-Putten, etc.; Goeree-Over Flakkee is wholly detached. Traffic from East Britain converges on the coastal Hook of Holland harbour (built 1866–1872) en route by water or rail for the Rotterdam (*q.v.*), Schiedam (80,756 [mun.]) industrial area with Dordrecht (84,807 [mun.]) farther southeast on the Rotterdam-Antwerp route. Many old sites along the "dune-foot" line have decayed—the famed Teilengen castle (near Sassenheim) where Jacqueline of Bavaria died (1433) has almost disappeared—but The Hague (*q.v.*) remains the capital of both province and kingdom. Eastward is the triangle of towns: Leiden (97,409 [mun.]), university, and Delft (74,742 [mun.]), pottery, with Gouda (43,696 [mun.]), cheese and earthenware. East and south on the borders are the historic river sites of Vianen (Lek), the reputed *Fanum Dianae* of Ptolemy and Gorichem (Waal)—taken by the "Sea Beggars" from the Spaniards (1572).

South Holland suffered perhaps more than any other province in Holland during World War II. Rotterdam was subjected to widespread destruction by aerial bombing on May 14, 1940; the whole business centre was wiped out and at least 25,000–30,000 persons were killed. During the winter of 1942–43 the Germans tore down at least 20,000 homes and the famous 17th-century wood, Haagsche Bosch, in The Hague, in order to build defense works. The sea resort of Scheveningen was also razed to a large extent.

SOUTH MOLTON, a municipal borough in Devonshire, Eng., 12 mi. S.E. of Barnstaple by road, on the river Mole. Pop. (1961) 2,994. Area 9.2 sq.mi. There are flour mills and factories for shirts and collars. It is a tourist centre for Exmoor with a Perpendicular church. The weekly market and one fair date from 1246, and the fair held on St. John the Baptist's day dates from 1490.

SOUTHPORT, a municipal, county and parliamentary borough, Lancashire, Eng. Pop. (1961) 82,004. Area 15.1 sq.mi. Extending 8 mi. along the coast between Liverpool and Preston, it is a popular holiday resort with a large residential population. The present town was built on wind-blown sand and reclaimed land during the 19th century. Southward the coast is bordered by high sand dunes while inland the ground is marshy. Southport was known as Meols (Scandinavian "sand dunes") until the early 19th century.

Liberal planning, with emphasis on gardens and flowers, produced handsome thoroughfares such as Lord street and several parks. The annual flower show is famous. Engineering is the most important industry. Southport returns one member to parliament.

SOUTH PORTLAND, a city of Cumberland county in southwestern Maine, U.S., separated from Portland (*q.v.*) by Portland harbour and tidal Fore river. South Portland was originally part of Cape Elizabeth, which divided over the offer of the Portland Water district to supply the town with Sebago lake water; some of the Cape Elizabeth people were opposed but residents of the salt-water-lapped southern section, who were unable to dig enough fresh-water wells, welcomed the offer. The region was set off and incorporated as the town of South Portland on March 15, 1895, and was granted a city charter a week later. A city-manager system was instituted Jan. 1, 1934. During World War II Liberty ships were built in South Portland; the shipyards were later taken over by three steel-fabrication firms and by a Canadian company for an oil pipeline terminal. The city also became the site of one of the largest Atlantic coast oil-tank farms. For comparative population figures *see* table in MAINE: *Population.* (A. K. GL.)

SOUTH SAN FRANCISCO, an industrial city of California, U.S., 9 mi. S. of San Francisco (*q.v.*) and just north of San Francisco International airport, on San Francisco bay. The city is built upon Spanish land bought in 1869 for the "cattle empire" of Henry Miller and Charles Lux; in the 1890s financiers and industrialists led by G. F. Swift of the meat-packing firm planned a workers' community there. The tract was subsequently subdivided and incorporated in 1908. The location offered excellent rail and water transportation and inexpensive factory sites, and industrialization continued. By the city's 50th anniversary important industries included steel (Bethlehem-Pacific Coast; U.S. Steel), paint

(W. P. Fuller) and chemicals, and meat packing (Swift; Armour). Public schools, churches, libraries, the weekly *Enterprise-Journal* (est. 1895), and recreation centres as well as new privately financed housing expanded rapidly during and following World War II. For comparative population figures *see* table in CALIFORNIA: *Population.* (J. H. St.)

SOUTHSEA: *see* PORTSMOUTH.

SOUTH SEA BUBBLE. This name was given to the mania of speculation which ruined many English investors in 1720, closely following the French financial crisis of the previous year. This so-called "bubble" (or hoax) centred on the fortunes of the South Sea company which was founded by Robert Harley in 1711. This company was formed with the object of trading to Spanish America on the assumption that the War of the Spanish Succession would end with a treaty under which such a trade (mainly in slaves) might be opened. Holders of navy and ordnance bonds were offered stock, with a guaranteed interest of 6%, in this government-sponsored venture, a company given a monopoly of trade to South America and the Pacific. The stock was accepted willingly but the peace treaty, when made, conveyed only a very limited trading concession and that restricted to a single annual ship which could not enter the Pacific. The success of the first voyage in 1717 was only moderate but the king became governor of the company in the following year, creating fresh confidence in an enterprise which soon yielded up to 100%. There had been a first wave of speculation when the Bank of England was founded in 1694 but the postwar boom of 1713–19, paralleled by the Mississippi company mania in France, now gave rise to extravagant hopes. Those who were unable to buy South Sea stock (which was genuine, if over-capitalized) were consoled by optimistic inventors and swindlers with the offer of shares in a hundred dubious ventures.

South Sea stock itself stood at 128½ in Jan. 1720, rose to 330 in March, 550 in May and 890 in June. It stood at 1,000 in July and August, and then dropped to 175 in September and 124 in December. Other shares followed a similar trend, the Bank of England stock falling from 263 to 145.

The occasion of the boom in South Sea shares was the scheme by which the company took over the whole of the national debt except that owed to the Bank of England and the East India company in 1720 and persuaded the investors to exchange their state annuities for South Sea stock, sold at a tremendous premium. A few investors profited greatly, far more being ruined. Some fled from their creditors and not a few committed suicide. This financial disaster was followed by an inquiry undertaken by a committee of the house of commons. Its report, rendered in Feb. 1721, showed that at least three ministers had both accepted bribes and speculated in the stock. John Aislabie (1670–1742), chancellor of the exchequer, was expelled from the house and imprisoned. The Earl of Sunderland and Charles Stanhope were acquitted by small majorities, and the estates of the company's directors were mostly confiscated. The company itself continued to exist, making fair profit until 1732, selling most of its rights to the Spanish government in 1750 but retaining some privileges until 1807 and its existence until 1853. It was latterly concerned mainly with the Greenland whale fishery and for a short time numbered Charles Lamb among its clerks. It was the scandal surrounding this company in 1720 which weakened the whole system of chartered companies and provided arguments for free trade. Another result of this affair was the destruction, indirectly, of the London slave trade, to the immediate profit of Bristol and the eventual profit (on a far greater scale) of Liverpool.

BIBLIOGRAPHY.—Jean O. McLachlan, *Trade and Peace With Old Spain, 1667–1750* (1940); Lord Erleigh, *The South Sea Bubble* (1933); L. Melville, *The South Sea Bubble* (1921); D. Hannay, *The Great Chartered Companies* (1926). (C. N. P.)

SOUTH SHIELDS, seaport, municipal, parliamentary and county borough, Durham, England. Pop. (1961) 109,521. Area 7.6 sq.mi. It is on the south side of the mouth of the river Tyne, there protected by the massive South pier, nearly 1 mi. long. Ferries connect the town with North Shields. Shipbuilding and repairing and marine engineering are important, and there is considerable export of coal and iron. The Tyne dock has a water area of 50 ac. and the tidal basin is 5¾ ac. Roman Remains park, near the northern end of the town, may have been an island in the Tyne; the remains comprise storehouses, etc. The town was founded by the convent of Durham in the 13th century. In the 17th and 18th centuries it had large glassworks and salt pans. The first self-righting lifeboat was launched there in 1790. The town has returned one member to parliament since 1832 and in 1850 it was incorporated, being raised to county borough status in 1889.

SOUTHWARK, a metropolitan borough of London, Eng., which under the London Government Act 1963 was to be amalgamated on April 1, 1965, with Bermondsey and Camberwell to form the new London borough of Southwark (*see* LONDON). Southwark metropolitan borough is on the south bank of the Thames and is connected with the City by Blackfriars, London and Southwark bridges. Pop. (1961) 86,249. Area 1.8 sq.mi. It is a parliamentary constituency. The convergence of roads to cross London bridge gave Southwark its medieval importance and its inns became famous; these included the Tabard of the *Canterbury Tales* and the George, the last surviving galleried inn in London (now National trust). In 1327 Edward III granted Southwark to the citizens of London forever and it is still known as "The Borough"; the charter was renewed by Edward IV and Edward VI, when the ward of Bridge Without was constituted (1550). The authority of the City is now nominal.

The name Southwark is taken from the "southward works," or fortifications, of London. Numerous Roman remains have been found. The hospital of St. Thomas, founded in 1213, was moved to Lambeth in 1868; its chapel became the chapterhouse of the cathedral. Opposite is Guy's hospital, founded by Thomas Guy and opened in 1725. Various bishops had town houses in Southwark; the grandest was Winchester house, used for five centuries from 1107. Paris garden was a centre of amusement in the 16th and 17th centuries, and in it were bear gardens and the Swan theatre (1595, burned down 1633). The Globe theatre, where many of Shakespeare's plays were acted, was built in Bankside in 1599 and was burned down in 1613.

The Augustinian priory of St. Mary Overie (*over rie*, "over the water"), founded in 1106, became the parish church of St. Saviour in 1540. In 1905, when the diocese of Southwark was created, it became the cathedral church and in 1937 was renamed the Cathedral and Collegiate Church of St. Saviour and St. Mary Overie. It lies well below the level of the streets at the southern end of London bridge. Built in the 13th and 14th centuries (the original having been burned down), it was restored in the 19th century. The choir and Lady chapel are Early English, the transepts are Decorated and Perpendicular and the nave is 19th century. The Harvard memorial chapel was restored in 1907 in memory of John Harvard, the founder of Harvard university who was born in the parish and baptized in the church on Nov. 29, 1607. There is also the Roman Catholic St. George's cathedral. The area sustained heavy air raid damage during World War II and among the demolished buildings was the Pilgrims Memorial church on the site of an old Congregational church (founded 1616).

SOUTHWELL, ROBERT (*c.* 1561–1595), English Jesuit and poet, son of Richard Southwell of Horsham St. Faith's, Norfolk, was born in 1560/61. He was educated at Douai and at Paris, joining the Society of Jesus. In 1584 an act was passed forbidding any English-born subject of the queen who had taken priest's orders in the Roman Catholic Church since her accession to remain in England longer than 40 days on pain of death. But Southwell at his own request was sent to England in 1586 as a Jesuit missionary with Henry Garnett. He went from one Catholic family to another, administering the rites of his church, and in 1589 became domestic chaplain to Ann Howard, whose husband, the first earl of Arundel, was in prison convicted of treason. It was to him that Southwell addressed his *Epistle of Comfort*. This and other of his religious tracts, *A Short Rule of Good Life, Triumphs over Death, Mary Magdalen's Tears* and a *Humble Supplication to Queen Elizabeth*, were widely circulated in manuscript. That they found favour outside Catholic circles is proved by Thomas Nashe's imitation of *Mary Magdalen's Tears* in

Christ's Tears over Jerusalem.

After six years of successful labour Southwell was arrested (1592). He was imprisoned at first in Richard Topcliffe's house, where he was repeatedly put to the torture in the vain hope of extracting evidence about other priests. Transferred to the gatehouse at Westminster, he was so abominably treated that his father petitioned Elizabeth that he might either be brought to trial and put to death, if found guilty, or removed in any case from "that filthy hole." Southwell was then lodged in the Tower, but he was not brought to trial until Feb. 1595. Much of his poetry, none of which was published during his lifetime, was probably written in prison.

On Feb. 20, 1595, he was tried before the court of King's Bench on the charge of treason, and was hanged at Tyburn on the 21st. In 1929 he was beatified.

St. Peter's Complaint with other Poems, published anonymously in 1595, was reprinted 13 times during the next 40 years. A supplementary volume entitled *Maeoniae* appeared later in 1595, and *A Foure fould Meditation of the foure last things* in 1606. This, which is not included in Grosart's reprint (1872) in the Fuller Worthies Library, was published by Charles Edmonds in his *Isham Reprints* (1895). *A Hundred Meditations of the Love of God,* in prose, was first printed from a manuscript at Stonyhurst college in 1873.

See Sidney Lee's account in the *Dict. Nat. Biog.;* Alexis Possoz, *Vie du Père R. Southwell* (1866); and a life in Henry Foley's *Records of the English Province of the Society of Jesus,* i, 301–387 (1877). Foley's narrative includes copies of the most important documents connected with his trial, and gives full information of the sources.

SOUTHWELL, a town and rural district of Nottinghamshire, England, 14 mi. N.E. of Nottingham by road. The rural district of 55 parishes had a population (1961) of 45,818; area 185.2 sq.mi. It has hosiery works, a flour mill and an interest in the neighbouring market-garden area (Bramley seedling apples are said to come from Southwell), but is chiefly known for its minster, a successor of a 7th-century church generally attributed to Paulinus; certainly there was a church there in the 8th century.

The present building was begun in the 12th century and has a Norman nave, transepts and towers, an Early English choir and a Decorated octagonal chapter house. It is cruciform and is 306 ft. long. There are no cloisters. Particularly noteworthy are the magnificent 14th-century stone rood screen, intricately carved, and the carved animals and foliage on the doors of (and within) the chapter house.

In 958 land at Southwell was granted to the archbishop of York by Edwy and a detailed description of the manor appears in Domesday Book. Southwell was a collegiate establishment and remained under the lordship of the see of York until it was taken over by the ecclesiastical commissioners. Remains of the archbishop's palace (15th-century; "great chamber" restored 1880) are incorporated in the bishop's palace.

The episcopal see founded in 1884 included the counties of Nottingham and Derby, but the latter was detached in 1926 when the see of Derby was established.

SOUTHWEST, THE, has long been an elastic term for a slowly shifting geographic area of the United States. After the War of 1812, the southwest generally meant Missouri, Arkansas and Louisiana. Following Texas' annexation it, too, was included. In the wake of the war with Mexico, the southwest embraced most, but not all, of the territory acquired under the treaty of Guadalupe-Hidalgo (1848). California's littoral and great interior valleys were so different in physiography and development that the term never settled easily upon them. Rather, the southwest came to mean, by and large, the 900-mi. stretch of arid land between the 100th meridian and the deserts athwart the lower Colorado river; *i.e.,* New Mexico, Arizona and such contiguous parts of Texas, Oklahoma, Colorado, Utah and Nevada as suited the convenience of the person using the term.

Early explorers found a diverse and difficult topography. The high, dry plains of Texas, sliced in the north by such unexpected abysses as Palo Duro canyon, rose imperceptibly to the Pecos valley of New Mexico, a dividing line that the myriads of buffalo farther east seldom crossed. Beyond the Pecos and embracing its headwaters were the southern spurs of the Rocky mountains, cool with evergreens. West of the Rockies came vast horizontal strata of highly coloured sandstone; in places these beds formed broad peneplains whose dry monotony was broken by occasional cliff-girt mesas or buttes. Where the beds were higher they were trenched by spectacular gorges, three of which—Zion and Bryce canyons in Utah and the Grand canyon of the Colorado river in Arizona—became national parks. Still westward, beyond a forested uplift that wrinkled Arizona from northwest to southeast, lay the true deserts marked by prickly plant growths and parallel chains of north-south mountains, gaunt and almost devoid of vegetation yet achieving in the amethyst glow of sunset a strange and haunting beauty.

The common denominator of the land was aridity. Although the highest peaks of New Mexico and Arizona caught 30 in. of rain or more a year, the annual average was less than half that amount; indeed, the western deserts might receive only 2 in. or less. Except for the snow on the mountains, the bulk of this scanty precipitation came during violent thunderstorms whose flash floods did much to give the region its scarred and angular topography.

High temperatures and rapid evaporation intensified the dryness. Precipitation that would support dry farming in the Dakotas or Nebraska proved, generally speaking, inadequate in the southwest: crops could be grown there only with irrigation. Yet the region's two principal rivers, the Colorado and the Rio Grande (both of which rose outside the area), as well as the smaller streams of the local mountains could be tapped at only a handful of widely scattered spots. This overwhelming geographic fact turned some of the prehistoric Indians into sedentary agriculturalists, determined the shape of the Spanish occupation and ultimately led the American inhabitants to an alliterative folk description of their economy as one of "copper, cattle, cotton, and climate."

Despite oversimplification, there was much to justify this summation. Geologic evidence even suggested that aridity might have had some bearing on the southwest's huge deposits of low-grade copper ore, the so-called porphyry coppers. Nevertheless, Mexicans were working the richer streaks of the long-productive ore body at Santa Rita, N.M., before 1800; other veins were opened in southeastern Arizona following the Civil War. However, the great development came after 1910 when the handling of enormous tonnages by power shovels in open-pit mines lifted Arizona to first place among the copper-producing states. Compared with this outpouring of red metal, the southwest's brief stampedes for gold and silver and later for spotty showings of oil and uranium achieved little more than regional significance.

Agriculturally, the dryness of the land enforced a pastoral economy that began in 1598 when Juan de Oñate established in the Rio Grande valley of New Mexico the first permanent colonies of the western United States. (San Antonio, Texas and Xavier del Bac near Tucson, Ariz., were not founded for another century; San Diego, Calif., not until 1769.) Because of supervisional difficulties, cattle were not popular with the first settlers. Horses, however, multiplied rapidly and were soon spread among the Indians by barter and theft. Sheep ranches grew to enormous size—in the early 1800s New Mexico's governor, Bartolomé Baca, owned 2,000,000 head—and social cleavage was largely between those who owned sheep and those who herded them. The Pueblo Indians took readily to the animals and learned to weave wool just as they were already weaving a native cotton. The adaptable Navajos learned the craft from the Pueblos, developed their famous blankets and as late as World War II accounted for half the sheep raised in Arizona.

During the period of Spanish ascendancy the sheep were driven to market among the silver mines of Mexico. In the 1850s the California gold fields offered an outlet for those owners willing to brave the long trek across mountains and desert to the coast. After the Civil War the demand of the southwest's new Indian reservations and army posts for beef lured herds of longhorns west from Texas. When the Apaches of the mountains were brought under control, permanent establishments took root in New Mexico and Arizona, sometimes in violent conflict with rival sheep ranches.

High profits made overgrazing of the public domain a more serious threat to the economy of the southwest than the Indians ever were. Only federal control, exerted through an increasing number of agencies, was able to check the trend and start restoring the ranges.

Other federal intervention came in the management of water resources. Irrigation had been practised before A.D. 1000 by the Pueblo Indians in New Mexico and central Arizona. The Spanish added a few *acequias* or "ditches" of their own, often as part of the huge land grants belatedly made by the Mexican government in the hope of establishing agricultural communities as buffers against the encroaching Americans. In their turn these Americans, individually and through cooperatives, struggled to increase the paltry stream diversions; but measures commensurate with the problems were not possible until the Reclamation act of 1902 made available the resources of the national government.

The reclamation service's first successful large-scale project in the nation was Theodore Roosevelt dam near Phoenix, completed in 1911; the largest project in the southwest was Hoover dam on the Colorado river, completed in 1936. But the water proved expensive and led to intensive cultivation of high-income crops: alfalfa, citrus and long-staple cotton. It also led, in the case of streams crossing state boundaries, to complex legal disputes, the bitterest of which concerned the diversion of the Colorado river water far from its natural channels. Various state compacts were entered into by the claimants, but sectional jealousies militated against a satisfactory solution. After World War II the water shortage was made even more critical when developments associated with aviation and nuclear energy brought an unprecedented influx of industry and population into all of the southwest's urban areas.

Though the crisp climate and scenic landscapes were a curse to agriculture, they were a boon to businesses catering to tourists and health-seekers. These visitors, it was shortly discovered, had a lively interest in the once-despised cultures of the Indians and Spanish-Americans. The heavily beamed, flat-roofed Indian pueblo, as modified by the Spanish, became the inspiration for the region's new architecture. Indian dances, Spanish fiestas and cow town rodeos, catching hold first as tourist attractions, became the year's social and economic high points for scores of communities. Although the showier forms of this evolving culture were not truly typical of any of its founding strains, they fixed themselves in the popular mind at least as an even more distinctive feature of the American southwest than were the ancient angular topography and dazzling sunlight. (D. S. L.)

SOUTH WEST AFRICA (SUIDWES-AFRIKA), a territory of southern Africa which came into being as a German protectorate in 1890 and developed under German control until it was invaded by forces from the Union of South Africa in 1915. After the treaty of Versailles the territory was mandated to the Union. It is still administered as a mandate of the Republic of South Africa. Extending from the Orange river in the south to the Kunene and Okovango (Okavango) rivers and the parallel of 17° 23′ S. in the north, the territory extends to the Chobe and Zambezi rivers in the Caprivi Zipfel (Caprivi strip) between northern Bechuanaland and southeastern Angola. Its eastern boundary to the south of this strip lies on the meridian of 21° E. as far south as the 22nd parallel and then on the meridian of 20° E. to the Orange river. The area of the territory is 317,887 sq.mi., not including the Walvis Bay enclave of 374 sq.mi. which is an integral part of Cape province in the Republic of South Africa, although administered as a part of South West Africa.

Physical Geography.—The territory includes four of the major geological formations of southern Africa. The Basement complex, of old granite and gneiss, with metamorphosed sediments, forms much of the plateau surface between Windhoek and Outjo, extending from this main area northwestward into the Kaokoveld mountains and southward along the plateau edge to the Orange river valley. The Pre-Cambrian Damara system, of dolomite, limestone, sandstone and metamorphosed rocks, occupies most of the plateau surface from the Khomas highland in the Windhoek area to the Kunene on the northern border. Much of this system is strongly folded, the axes trending generally north by east. In the south of the territory the Nama system, of quartzite, shale and black limestone, dips gently eastward. The Karroo system, comprising almost horizontal beds of tillite sandstone, shale and overlying lava, forms the surface in the area extending from the Orange river in the south to near Gobabis in the middle of the territory. In the north it forms local plateaus such as the Waterberg and great areas of lava in the Kaokoveld. Young granites have been intruded into these beds, forming such prominent features as the Brandberg, in the Namib, and the Erongo mountains.

The youngest formations are in the east and west. In the east the Tertiary Kalahari system, of thin limestone and sandstone at the base and a varying thickness of overlying sand, covers about half the territory to the north of 22° S. and a considerable area to the west of the southeastern boundary. In the west of the territory the sands and clays of the Namib, together with the diamondiferous shingle of the southern coastal strip, are probably the youngest formations.

Two major divisions can be recognized: the plateau surface, and the marginal zone between the plateau and the coast. Separating these divisions is the plateau edge or Great Escarpment, the highest parts of which are known in the south as the Huib (highest point 5,860 ft.), the Tsaris (6,486 ft.) and the Naukluft (6,473 ft.) escarpments and, in the north, as the Joubertsberge (5,852 ft.) and the Baynes mountains (6,496 ft.).

BY COURTESY OF THE SOUTH AFRICAN INFORMATION SERVICE

VIEW OF WINDHOEK, CAPITAL OF SOUTH WEST AFRICA

On the plateau the highest area is the Khomas highland whose ridges trend east-northeast through the Windhoek area. The main watershed of the territory is formed in this highland by the Auas mountains (Moltkeblick 8,153 ft.). In the north the Otaviberge (Nageib 7,001 ft.) trend east-west in the Grootfontein area, bending northward in the Joubertsberge. In the south of the plateau area the main orographic features are the Klein and Gross Karas mountains (Schroffenstein Berg 7,224 ft.) and Mount Brukkaros (5,207 ft.), an ancient volcanic crater, about 9 mi. N. of Berseba, which rises 1,500 ft. above the Namaqualand plain.

Most of the marginal zone is known as the Namib which is generally flat and true desert near the coast, passing, as it ascends to the plateau edge, into a semidesert zone generally known as the Inner Namib. (*See* SOUTH AFRICA, REPUBLIC OF: *Geographical Regions.*) The highest point in the territory is the Brandberg in the Namib, a young granite mass which rises precipitously to an altitude of 8,481 ft.

Five geographical regions can be recognized. (1) The Etosha plain in Ovamboland, to the north of the Otaviberge, extending eastward to the south of the Okovango river. (2) Damaraland (or Hereroland) between the Otaviberge and the Auas mountains, to the south of Windhoek. This is a plain of fairly even surface which, however, is broken by many island-mountains, of which the highest is the Great Omatako (7,475 ft.). (3) Namaqualand

(or Namaland), lying between the Auas mountains and the Orange river, rather flat in the north but deeply dissected in the south by the Fish river and other smaller tributaries of the Orange. (4) The Kalahari surface along the eastern border, which there forms the western part of the southern Kalahari sandveld. (5) The Namib between the plateau edge and the coast, predominantly sandy in the south but behind the coastal belt becoming rocky in the north where it forms the western part of the Kaokoveld.

Drainage.—Except for the Orange in the south and the Kunene and Okovango at the northern boundary all the rivers are of intermittent flow, "coming down" generally during two or three months of the year after the seasonal rains. The most spectacular scenery in the territory is the canyon of the Fish river, the southward-flowing tributary of the Orange. Cut in Nama beds and Archean rocks the canyon is between 1,500 and 2,000 ft. deep and lies between about 27° and 27° 50′ S. In the east the Auob river has artesian conditions in its upper course. Pans (shallow temporary ponds or lakes) occur in many areas, especially on the Kalahari sand. The largest is the Etosha Pan, about 70 by 30 mi. which receives the summer drainage of Ovamboland.

Climate.—Because of the northward sweep of the cold Benguela current, from the south Atlantic, along the coast, temperatures are consistently low in the Namib. Walvis Bay has a mean temperature for the coolest month (August) of 14° C. (57° F.) and for the warmest month (February) of 19° C. (67° F.). On the plateau the temperature range is greater. Windhoek, at 5,700 ft., has a mean for the coldest month (July) of 13° C. (55° F.) and for the warmest month (January) 23° C. (74° F.). Rainfall is least in the western Namib and in the lower Orange river valley where the annual precipitation is less than 2 in. On the plateau there is an increase of rainfall from southwest to northeast. Warmbad, in southern Namaqualand, has a mean annual rainfall of 3.4 in., Windhoek 15.1 in., Tsumeb 20.7 in. and Katima Mulilo in the Caprivi strip 26.6 in. Except in the extreme southwest the rainfall of the territory occurs mainly in the summer months, the percentage of summer precipitation increasing steadily from southwest to northeast.

Vegetation.—In the Namib coastal zone the only vegetal types are those adapted to extreme drought. Among these are the indigenous tumboa (*Welwitschia bainesii*), the tortoise bush (*Zygophyllum stapfii*) and the edible narras (*Acanthosicyos horrida*) which grows between the dunes near Walvis bay. Behind the coastal zone various succulents occur which give place near the plateau edge to desert grasses and aloes. In the southern part of the plateau the vegetation is predominantly of the Karroo type with desert shrubs and succulents passing northward into grass steppe. In Damaraland and on the eastern sandveld of the Kalahari, thorn savanna (Acacia) with grass is the predominant type, merging to the north of the Etosha Pan into dry forest in which the mopane (*Colophospermum mopane*) is associated with large fig trees (*Ficus hereroensis*), *Hyphaene* palms and baobabs. In the *oshanas*, or shallow floodwater channels, of Ovamboland the floors are covered with *Aristida* grasses.

Animal Life.—Primates are represented by the gray vervet monkey, the chacma baboon and, in the north, by the bush baby (*Galago senegalensis*). Carnivores include the lion in the northwest and the northeast. The leopard and the cheetah are widespread, but the most numerous of all the larger carnivores is the brown hyena which roams even the coastal belt of the Namib where it lives on marine refuse and shellfish. Jackals (black-backed and side-striped) are widespread. The elephant roams the Kaokoveld, the Caprivi and, occasionally, the Okovango area. The black rhinoceros (*Diceros bicornis*) is found in small numbers in the Kaokoveld. The giraffe and the hippopotamus also are found only in the north. Of the antelopes some, such as the koodoo (kudu), gemsbok and springbok, are found over most of the territory; others are limited to the northern areas, including the sassaby (tsessebe), lechwe, puku, impala, roan antelope and the sitatunga, or marshbuck (*Tragelaphus spekei*). The hartebeest is found in the northeast and the buffalo (*Syncerus caffer*) only in the northern areas. On many of the islands off the coast to the south of Walvis Bay, and at Cape Cross farther north, the Cape sea lion or fur seal (*Arctocephalus pusillus*) is abundant.

The People.—The nonwhite people of the territory may be divided into two main racial types, the Southwestern Bantu and the Khoisan types (the yellow-skinned Hottentots and the primitive Bushmen). The Bastards of the Rehoboth Gebiet are the descendants of mixed European and Hottentot unions. The strongest Bantu groups in pre-European times were the Ambo of the north and the Herero, called Damara by the Hottentot. Occupying Hereroland (or Damaraland) to the north of the Khomas highland, the Herero were at constant strife with the Hottentots who occupied areas to the south of the Khomas. In the early period of German occupation the Herero, with their northern offshoot, the Ovashimba, numbered probably about 100,000; they were seriously reduced under German occupation and now number about 30,000. A more primitive Bantu group, the Bergdama, or Berg (Mountain) Damaras, formerly vassals of the Herero and Hottentots, had no territory of their own, but have been allocated reserves. They number about 38,000 and are probably the one native group that openly rejoices at the European occupation of the territory. They speak the language of the Nama Hottentots who subjugated them. Hottentots, in reserves mainly in the southern part of the territory, number about 29,000. The total Bushmen population in the territory is estimated at 20,000. Under German rule the area allocated as native reserves was about 4,030 sq.mi. Under the administration of South Africa the reserves were enlarged to about 79,500 sq.mi., in addition to the Rehoboth Gebiet, 4,800 sq.mi., containing about 8,000 Bastards. These people speak Afrikaans. (*See* also AFRICA: *Ethnography (Anthropology): Southern Africa;* AMBO; BERGDAMA; BUSHMAN; HERERO; HOTTENTOT.)

Population.—At the census of 1921 the total population was 227,988, made up of 19,372 whites and 208,616 Africans and Coloured. In 1951 the total population was 430,575. In 1960 the total population of 512,496 comprised 68,347 whites, 444,147 Africans and Coloured and 2 Asians. In 1960 Windhoek (*q.v.*), the capital, had a population of 35,916, including 19,021 whites. Most of the other towns have considerably increased populations since 1951. The Bantu population may be divided into two groups: those in the police zone and those in the northern native territories consisting of Ovamboland and the Okovango native territory, with small groups in the Kaokoveld and the Caprivi Strip. These territories are generally referred to as "outside the police zone." In Ovamboland (Amboland) the Ambo number about 198,000; in the Okovango territory the tribal Bantu number about 18,000. Within the police zone the native reserves in 1950 had about 26,000 Bantu, Hottentots and Coloureds. (J. H. WN.)

History.—The first European to set foot in the territory was the Portuguese navigator Diogo (Cão) Cam who erected a pillar at Cape Cross in his voyage of 1485–86. At the end of 1487 his compatriot Bartolomeu Dias de Novaes touched the coast at Walvis Bay and at Dias point, where he erected a cross near Angra Pequena (Lüderitz). The Dutch East India Company's ship "Grundel" sailed up the coast in 1670 and was followed by the same company's "Bode" in 1677. Though the interior still remained unknown, the coastal waters were not neglected; sealing and whaling attracted ships of various nationalities. In 1793 Capt. F. R. Duminy, of the Dutch ship "Meermin," proclaimed certain areas to be the property of the Dutch government. In 1795 Capt. Thomas Alexander of H.M.S. "Star" hoisted the British flag at various points on the coast but his action was later disowned by his government. Meanwhile, in 1760, Jacobus Coetzee found a way into Great Namaqualand by the overland route from the Cape. He was followed in 1761–62 by Hendrik Hop and thereafter by a fairly steady stream of explorers of whom the most famous was C. J. Andersson, a Swede, who in 1853 crossed from Walvis Bay to Lake Ngami in Bechuanaland. The London Missionary society was represented on the Orange river as early as 1802 and in 1807 established a station at Warmbad. This society withdrew and was succeeded by the Wesleyans. In the 1840s the German connection with the territory began with the arrival of the Rhenish Missionary society.

In 1868 tribal wars prompted the missionaries to appeal for

British protection. In this they were actually supported by the German government, but the British government turned the proposal down. In 1876 the Cape colony sent a special commissioner, W. C. Palgrave, who made treaties with chiefs anxious to secure British protection. This placed the whole territory as far as the Portuguese colony of Angola under British control. The British high commissioner in South Africa approved Palgrave's action but the British government was unwilling to assume these new responsibilities and was only with difficulty persuaded to consent to the annexation of Walvis Bay and some adjacent territory (1878). The guano islands off the coast had been annexed in 1867. In 1880 tribal war again broke out and the German government itself asked Great Britain to protect the life and property of its subjects but the British reply was negative.

German Possession.—In 1883 Heinrich Vogelsang, agent of F. A. E. Lüderitz, a merchant of Bremen, obtained from the Hottentot chief, Joseph Frederick of Bethany, a cession of land at Angra Pequena, later known as Lüderitz. Bismarck, still unconverted to a colonial policy, informed the British government beforehand of the project in words that were almost an invitation to Great Britain to assume sovereignty over the territory and to act as a protecting power. This opportunity, like the previous one, was allowed to slip away and in April 1884 Bismarck took the initiative and assured Lüderitz and his establishments of German protection which later developed into full-fledged German annexation of the whole territory.

The first German representative, sent to the territory in 1885, was H. E. Göring. He was succeeded by Kurt von François, who transferred the capital from Otjimbingue to Windhoek. It was there that the first German farmers settled in 1892. In 1893 there was trouble with Hendrik Witbooi, the Hottentot chief. The Germans attacked Witbooi's village and 150 of his subjects, including women and children, were killed. Von François was succeeded in 1896 by Theodor Leutwein who tried to treat the natives with consideration. However, the latter found much cause for dissatisfaction, and it required only a spark to set the country ablaze. The Bondelswaartz Hottentots rose in 1903. This rising was suppressed, but early in 1904 the Herero revolted and killed a number of German (but not British or Boer) settlers. The rebellion was speedily quelled but in the "cleaning up" operations the Herero were ferociously harried by Gen. Lothar von Trotha, Leutwein's successor from 1904, and were in the end reduced from a tribe of 80,000 people to 15,000 starving refugees; various groups of Hottentots who had joined in the rebellion still held out until early in 1907. In the years that followed these troubles the depopulation resulting from the methods used in suppressing the rebellion caused a labour shortage which hampered the development of the territory. The discovery of diamonds in 1908 near Lüderitz bay led to a considerable increase in the European population, which rose to nearly 15,000 by 1913.

On Aug. 6, 1914, the government of the Union of South Africa undertook to assume all obligations resting upon the British regular garrison in South Africa, and on Aug. 10 to send a military expedition of its own to German South West Africa. The surrender of the Germans was forced on July 9, 1915. At the end of the campaign the German troops were interned and later (in 1919) repatriated with about 600 undesirables. German civilians, on the other hand, were allowed to return to their homes in the territory and to continue their ordinary pursuits.

The South African Mandate.—Under the treaty of Versailles, Germany ceded its colonies to the principal Allied and associated powers. South West Africa was designated as a C mandate and was conferred on "his Britannic majesty for and on behalf of the government of the Union of South Africa." The newly created League of Nations was charged, *inter alia,* with the supervision of mandates. Subject to the mandate, South Africa was given full powers of legislation and administration over the territory as an integral part of the Union. The exercise of these powers was vested by the house of assembly in the governor general (Union act no. 49 of 1919), who later delegated them to a resident administrator. The S.W.A. Naturalization of Aliens act of 1924 provided for the automatic naturalization of all German adult males domiciled in the territory on Jan. 1, 1924, unless they specifically objected.

During the first few years of the mandate the territory was administered after the fashion of a crown colony, with an administrator assisted by an advisory council. In 1926 a legislative assembly of 12 elected and 6 nominated members, with an executive committee, was set up to legislate for needs such as roads and bridges, agriculture and taxation, while the rest of the administrative functions remained in the hands of the administrator.

Until the rise of Hitler the Germans in the territory had been reasonably co-operative; but by 1939 the grip of the Nazi party on all Germans in the territory was practically complete. At the outbreak of World War II the Union government acted with dispatch. By Dec. 1939 about 150 Nazi leaders were arrested and the organizations themselves proscribed. By Oct. 1940 the number of persons interned was 1,200 or one-third of the German adult male population. Under the Naturalization and Status of Aliens Amendment act of 1942 most of the automatically naturalized Germans reverted to their former German status.

Trusteeship or Incorporation?—With the demise of the League of Nations the Union government held that there was nothing in the covenant that empowered it to transfer or delegate its powers, as regarded mandates, to any organization. In this contention South Africa was supported in 1950 by an opinion of the International Court of Justice, which held that the Union was under no obligation to place the territory under a United Nations trusteeship, but that on the other hand an obligation rested on South Africa to submit reports on the administration of the territory to the UN. In spite of pressure South Africa firmly refused to agree to the inclusion of the territory in the international trusteeship system, but was willing to submit reports to Great Britain, the United States and France as a matter of courtesy.

The proposal that South West Africa should be incorporated in or should become a fifth province of the Union arose in 1946 when the subject was referred to a series of tribal meetings and it was reported that the majority of natives was in favour of the proposal, the Herero being the only substantial body against it. Later in the same year Gen. J. C. Smuts asked the United Nations that the territory be incorporated in accordance with the wish of its people but the request was refused. The South African government did not press the demand but declared its intention, and this applied both to trusteeship and to incorporation, to continue to administer South West Africa in the spirit of the mandate. However, the association of the two countries, already close, was made much closer by the South West African Affairs (Amendment) act of 1949 (see *Government* below). The South African Citizenship act of 1949 made it possible for those who had been "denaturalized" in 1942 to apply once more for naturalization. In 1955 a commission was appointed to consider the question of German language rights. Its report, produced in 1957, stated that the 20,000 German-speaking people, *i.e.,* 35% of the white population, wanted equal status for German with English and Afrikaans. A German-speaking people may address the South African state departments in German but cannot demand a reply in the same language. German is, however, fully recognized in the local legislative assembly.

The deadlock regarding trusteeship continued in discussions at the UN in the following years. A UN committee (appointed to consult with the South African government on a possible basis for agreement) suggested in its report in Sept. 1958 that the territory might be partitioned. The northern portion of South West Africa, containing the great majority of the Bantu population, was to be administered by the South African government as an integral part of the Union under a trusteeship agreement with the UN, while the rest of the territory was to be annexed to the Union. The Union government announced that this did not mean that it had agreed to partitioning the territory, but that it would be prepared to investigate the possibilities at the request of the UN. However, the UN subsequently rejected partition as a possible basis for agreement.

A deputy minister for South West Africa was appointed for the first time in Aug. 1961. Debates on the future of the mandate

continued in the UN and a UN special committee was refused entry to South West Africa. The Union was charged before the International court of having failed in the terms of its original mandate. Eric Louw, the Union's foreign minister, defended his government's record and announced a proposal for "three persons of international standing" to visit the area and report on their conclusions. Two members of the UN committee on South West Africa were admitted to the territory in 1962 and at the same time the South African government (a republic after 1961) proposed a five-year development plan for the area. (Ay. Sy.)

Administration and Social Conditions.—In 1920 the Council of the League of Nations conferred on the king of Great Britain a mandate over the former German colony of South West Africa to be exercised by the governor general of the Union of South Africa who, by proclamation in 1921, delegated his powers to an administrator appointed by the Union government. After May 31, 1961, the state president of the Republic of South Africa exercised the mandate and delegated his powers to an administrator appointed by the government of the republic. An executive committee of 4 members (presided over by the administrator) and a legislative assembly of 18 members are elected as in the republic's provinces and with similar powers. (*See* SOUTH AFRICA, REPUBLIC OF.) In addition, the South West African Affairs Amendment act, 1949, provides for six electoral divisions of the territory, each returning one representative to the South African house of assembly, and for representation in the senate by four senators, two of whom are nominated (one for the Coloured population).

Urban local government is undertaken by elected town councils and appointed village management boards, while rural administration is a matter for local magistrates. Windhoek is the seat of the administration. Administration of Bantu affairs (subject to the Native Urban Areas Proclamation, 1951, which followed the Union Native [Urban Areas] act, 1945) was taken over by the South African department in 1955. Extensive Native reserves have been established in the rural areas.

Generally speaking, taxation accords with that of the Republic of South Africa, a state income tax having been introduced in 1942 and the old municipal income tax replaced by rates on property in 1920. The main sources of revenue are income tax, customs and excise duties and mining taxation.

Living Conditions.—Machinery for regulating wages, registering trade unions and settling disputes was established in 1953. In 1959 there were 3 unions with about 600 members. There was full employment and no unemployment benefit plans were found necessary. African labour is generally adequate but in some cases there is a shortage. Housing for the white population is financed by building societies. In the African townships municipalities provide subsidized housing with state aid. (*See* SOUTH AFRICA, REPUBLIC OF.)

Health.—Health and hospital services parallel those of South Africa. Hospitals for whites are state-aided and run by boards responsible to a director of health services. Hospitals for Africans are directly controlled and maintained by the administration.

Justice.—The judiciary consists of a high court and circuit courts (from which appeals lie to the appellate division of the South African court) and magistrates' courts. The police and defense forces are incorporated in the South African forces.

Education.—Primary and secondary education is compulsory for all white children between 7 and 16 years of age, and German is a medium in some schools. The Cape syllabus and system are followed. Education of Coloureds and Africans is undertaken mainly by mission schools subsidized by the administration.

(L. H.)

The Economy.—*Agriculture.*—There is little crop cultivation in the territory except in the northern areas where the rainfall is just sufficient in most years for small crops of maize (corn) and kaffir corn varying from about 30,000 to a few hundred bags (of 200 lb.). Under irrigation a few thousand bags of wheat are produced, especially in the artesian valleys of the Nossob-Auob river system. The territory is mainly a pastoral area with about 1,500,000 cattle, 4,000,000 karakul sheep, 100,000 other sheep and about 1,500,000 goats. The pelts of the newborn karakul lambs are among the most valuable of the territory's exports.

Fisheries.—Sea fisheries are based mainly on Walvis bay, with crawfish or rock lobster fishing mainly at Lüderitz. The pilchard is the all-important shoal fish. The fishing fleet consists of about 100 well-equipped boats with diesel engines, echo-sounding devices, radio telephones and radio direction-finders. The annual catch is about 400,000 to 500,000 tons. Whitefish and snoek are also taken in small quantities. At Walvis Bay there are six factories for processing pilchards and three for whitefish and snoek, the fish meal and body oil being of more value than the canning product.

At Lüderitz six factories can rock lobster and produce meal and frozen tails. Products of sealing on the offshore islands and at Cape Cross include annually about 10,000 pups (for furs) and 3,000 wigs (skins of bulls) and between 10,000 and 20,000 gal. of oil.

Mining.—The most valuable mineral is the diamond which is found along the coastal strip and offshore between Lüderitz and the Orange river mouth. The principal fields are now the Oranjemund diggings, along the coast for about 60 mi. N. of the Orange river. The annual production is about 1,000,000 carats, of which about 98% are gem stones and 2% industrials, the value being about £15,000,000 ($42,000,000). Copper, lead and zinc are found in considerable quantities in the Otavi limestone of the Tsumeb mine, with small quantities of silver and cadmium. The annual production is about 250,000 tons of "lead complex concentrates" valued at about £8,000,000 ($22,400,000). The production of manganese ore (about 48%) from the Primitive limestones of Hereroland amounts to about 100,000 tons a year, valued at rather more than £1,000,000 ($2,800,000). Tin and wolfram concentrates, from the vicinity of the Erongo mountains and the Brandberg, vary in quantity from about 450 tons to about 1,500 tons, the value being about £300 ($840) a ton. Other minerals have much lower values: lithium about £60,000 ($168,000), beryllium £35,000 ($98,000), sillimanite £30,000 ($84,000). Salt is an important mineral from the coastal pans, the annual production of coarse and refined salt being about 70,000 tons, the value between £150,000 and £200,000 ($420,000 and $560,000).

Trade.—The bulkier items of export include, except for diamonds, the minerals mentioned above, the export being mainly from Walvis Bay, where 300,000 to 500,000 tons of ore and other minerals are shipped annually, mainly for European and American destinations. Karakul pelts are usually sent to the European auctions by air, their value usually exceeding £5,000,000 ($14,000,000). Fish products (canned, salted, fish meal and oil) realize about £7,000,000 ($19,600,000).

To the Republic of South Africa the territory sends supplies of dairy products (butter and cheese) and cattle on the hoof. Imports include machinery, metal goods, building materials, manufactured goods and oil. Of the cargoes landed at Walvis Bay the main commodities are gasoline, oil, timber, cement and coal.

Finance.—In the late 1950s and early 1960s the territorial revenue fluctuated between about £12,000,000 ($33,600,000) and £16,500,000 ($46,200,000), the main items of revenue being income tax (most of which was paid by the mining companies and only about 10% by the non-Bantu inhabitants), customs and excise, and diamond export duty and profit tax.

Expenditure fluctuated during the same period between about £10,500,000 ($29,400,000) and £21,000,000 ($58,800,000), but these amounts included the annual appropriations to the Territorial Development and Reserve fund which varied between £5,000,000 ($14,000,000) and £12,500,000 ($35,000,000). Apart from these appropriations the main items of expenditure were: administration, education, public works, health, posts and telegraphs. The net accumulated surplus in the territory revenue fund fluctuated during the same period between £7,500,000 ($21,000,000) and £11,000,000 ($30,800,000).

Communications.—Of the 1,887 mi. of railway line in South West Africa, the widening of the Usakos-Tsumeb narrow-gauge system brought the total length of 3 ft. 6 in. gauge to 1,535 mi., the remaining 2 ft. gauge being 352 mi. There is also a transition from coal-burning locomotives to diesel traction. More than 34,000 mi. of trunk, main and district roads have been reported. Road motor services include a bus route from the railhead at Grootfontein to the Angola border, and from Gobabis to the Bechuanaland protectorate. There are frequent air services between Windhoek and Keetmanshoop in the territory and Johannesburg and Cape Town, linking up with South Africa's international air services. The port of Walvis Bay, with an entrance channel 33 ft. deep, accommodates about 250 ocean-going vessels a year, with a gross tonnage of about 1,500,000. South West Africa is a member of the African Postal and Telecommunications union. Telex services were introduced in 1956. There are radiotelephone links between point-to-point radio stations where the use of wire communication is uneconomical.

BIBLIOGRAPHY.—*Yearbook of the Republic of South Africa;* G. N. G. Hamilton and H. B. S. Cooke, *Geology for South African Students,* 4th ed. (1960); P. A. Wagner, *The Geology and Mineral Resources of South-West Africa,* Geological Survey Memoir no. 7 (1916); F. Jaeger and L. Waibel, *Beiträge zur Ländeskunde von Südwestafrika* Mitt. Deutsch. Schutzgebieten, Ergänzungsheft, nr. 14 (1920); Weather Bureau, Pretoria, *Climate of South Africa,* parts i–iv (1957); Drought Commission Report, *South West Africa* (1924); H. Wedder, *Native Tribes of South-West Africa* (1928); H. Walter and O. H. Volk, *Grundlagen der Weidewirtschaft in Südwestafrika* (1954); H. Walter, *Die*

Farmwirtschaft in Deutsch-Südwestafrika, Heft 1–4 (1941); South-West Afrika, Long Term Agricultural Policy Commission, *Report* (1949); Sir Charles Dundas, *South-West Africa: the Factual Background* (1946). Since 1955 all official South West African statistics have been merged in government reports of the Union or Republic of South Africa. (J. H. Wn.)

SOUTINE, CHAIM (Haim) (1894–1943), Lithuanian-Jewish Expressionist painter, was born in Smilovitchi, near Minsk in White Russia, the 10th of 11 children of a poor tailor. As a child he caricatured his family and neighbours. At the age of 16 he went to Vilnius, where a friendly doctor helped him to attend the school of fine arts for three years. In 1911 he emigrated to Paris, and there met the sculptor Jacques Lipchitz and the painter Amedeo Modigliani, whose portraits he painted. In Céret and elsewhere in the south of France he developed his characteristic style of painting with thick pigment and convulsive rhythms, twisting the human face and figure and tilting the forms of nature as in an earthquake. Though a leader of the Expressionist movement, he only admired earlier masters, especially Tintoretto, Rembrandt and Gustave Courbet, whose compositions he adapted in certain famous canvases. He exhibited little during his lifetime, but his jewel-bright landscapes, studies of valets, cooks and choirboys, and still-life paintings of poultry and carcasses of beef entered into French and American private collections and U.S. museums. While hiding in Touraine during World War II, he fell ill and was brought to Paris for an operation, where he died on Aug. 9, 1943. The Museum of Modern Art, New York city, exhibited his lifework in 1950, and 100 of his paintings were shown in the Galerie Charpentier in Paris in the summer of 1959.

See Monroe Wheeler, *Soutine* (1950). (M. Wh.)

SOUTSOS, ALEXANDROS (1803–1863), Greek poet, born in Constantinople, was the founder of the Greek romantic school of poetry. He studied in Chios and later in Paris, where he was influenced by the French romantics and by liberal political opinion. His verse satires, modeled on those of Béranger, are his liveliest writings and inspired the early development of modern political liberalism in Greece. His dramas and one long prose work, the *Exoristos* ("The Exile"), were considered frigid and artificial, but his numerous lyrics were admired by his contemporaries in spite of their lack of originality and their rhetorical exuberance. Byron's *Childe Harold* was the model for his longest poem, *Periplanomenos* ("The Wanderer").

Soutsos' collected works were published in 1916.

See C. A. Trypanis, *Medieval and Modern Greek Poetry* (1951). (Ce. A. T.)

SOUZA BOTELHO, ADÈLE MARIE ÉMILIE FILLEUL, Marquise de (1761–1836), French writer of sentimental novels of a high moral tone, was born in Paris on May 14, 1761. Her mother was the wife of one of Louis XV's secretaries and may have been one of the king's mistresses. After the marriage of her sister Julie in 1767 to the marquis de Marigny (Abel Poisson, Mme. de Pompadour's brother), Adèle in 1779 was married to the old comte de Flahaut (Alexandre de Flahaut de La Billarderie, 1726–1794). One of the many visitors to her salon, however, was Talleyrand, who was generally recognized to be the father of her son Auguste, later celebrated as comte de Flahaut (*q.v.*). She went to England in 1792 (her husband stayed in France and died imprisoned by the revolutionaries), and published a novel, *Adèle de Senange* (1794), which had some success. Passing then to Switzerland, she attached herself to the duc d'Orléans (Louis-Philippe, the future king of the French) and went with him in 1795 to Hamburg, where she had to work as a milliner. She returned to Paris in 1798. Married in 1802 to the Portuguese envoy, José-Maria, marques de Souza-Botelho (1758–1825), she enjoyed some influence in society under Napoleon I. Later she helped in the education of the future duc de Morny (*q.v.*), the child of her son Flahaut and Queen Hortense. Her collected works, 6 vol. (1821–22), contained *Adèle* and six other stories. Her last work, a play, *La Duchesse de Guise,* was published in 1832. She died in Paris on April 19, 1836.

See A. de Maricourt, *Madame de Souza et sa famille* (1907).

SOVEREIGNTY. The concept of sovereignty is one of the most controversial ideas in political science and international law. It is also closely related to the difficult concepts of state and government, of independence and democracy. Originally, as derived from the Latin term *superanus* through the French term *souveraineté,* sovereignty was meant to be the equivalent of supreme power. It has departed, however, quite often from this traditional meaning.

Historical Development.—In 16th-century France, Jean Bodin (*q.v.*) used the new concept of sovereignty to bolster the power of the French king over the rebellious feudal lords; the transition from feudalism to nationalism was thus facilitated. The theories of John Locke at the end of the 17th century, and of Jean Jacques Rousseau in the 18th century, that the state is based upon a compact of its citizens, through which they entrust such powers to a government as may be necessary for common protection, led to the development of the doctrine of popular sovereignty that found expression in the American Declaration of Independence in 1776. (*See* Social Contract.) Another twist was given to this concept by the statement in the French constitution of 1791 that "Sovereignty is one, indivisible, unalienable and imprescriptible; it belongs to the Nation; no group can attribute sovereignty to itself nor can an individual arrogate it to himself." The idea of popular sovereignty exercised primarily by the people became thus combined with the idea of national sovereignty exercised not by an unorganized people in the state of nature, but by a nation embodied in an organized state. Going one step further and investigating who in the name of the people or of the state exercises sovereignty, John Austin (*q.v.*) came to the conclusion that sovereignty was vested in a nation's parliament. This was the supreme organ that enacted laws binding upon everybody else, but that was not bound by the laws and could change these laws at will. This particular description again fitted only a particular system of government, such as prevailed in Great Britain in the 19th century.

When this idea of legislative sovereignty crossed the Atlantic ocean, it did not really fit the American situation. The constitution of the United States, being the fundamental law of a federal union, did not endow the national legislature with supreme power but imposed important restrictions upon it. A further complication was added when the supreme court of the United States asserted successfully its right to declare laws unconstitutional. While this development did not lead to a concept of judicial sovereignty, it seemed to vest the sovereign power in the fundamental document itself, the constitution. This system of constitutional sovereignty was made more complex by the fact that the authority to propose changes in the constitution and to approve them was vested not only in congress but also in the several states and special conventions called for that purpose. It could be argued, therefore, that sovereignty continued to reside in the states or in the people, who under the terms of the 10th amendment to the constitution retained all powers not delegated to the United States by the constitution or expressly prohibited by it. Consequently, the claims of the advocates of states' rights (*q.v.*) that states continued to be sovereign were bolstered by the difficulty of finding a sole repository of sovereignty in a complex federal structure; and the concept of dual sovereignty of both the union and the component units found a theoretical basis. Even if the competing theory of popular sovereignty were accepted, resting sovereignty in the people of the United States, it still might be argued that this sovereignty need not be exercised on behalf of the people solely by the national government, but could be divided on a functional basis between the federal and state authorities.

Another assault from within on the doctrine of state sovereignty was made in the 20th century by those political scientists (*e.g.,* Léon Duguit, Hugo Krabbe and Harold J. Laski) who developed the theory of pluralistic sovereignty exercised by various political, economic, social and religious groups that dominate the government of each state. According to this doctrine, sovereignty in each society does not reside in any particular place but shifts constantly from one group (or alliance of groups) to another. This pluralistic theory went even so far as to contend that the state is but one of many examples of social solidarity and pos-

sesses no special authority in comparison with other components of society.

Sovereignty and International Law.—While the doctrine of sovereignty has had an important impact on developments within states, its greatest influence has been in the relationships between states. The difficulties here can be traced back to Bodin's statement in 1576 that the sovereign who makes the laws cannot be bound by the laws he makes (*majestas est summa in cives ac subditos legibusque soluta potestas*). This statement has been often interpreted as meaning that a sovereign is not responsible to anybody and is not bound by any laws whatever. But closer reading of Bodin's writings does not support this interpretation. He emphasizes that even with respect to his own citizens a sovereign is bound to observe certain basic rules derived from the divine law, the law of nature or reason and the law that is common to all nations (*ius gentium*); as well as the fundamental laws of the state that determine who is the sovereign, succession to sovereignty and limitations on the sovereign power. Thus Bodin's sovereign was in fact restricted by the constitutional law of the state and by the higher law that was considered as binding upon every human being. In fact, Bodin discussed as binding upon states many of those rules that were later woven into the fabric of international law. Nevertheless, his theories have been used as justifying absolutism in the internal political order and irresponsible anarchy in the international sphere.

This interpretation was developed to its logical conclusion by Thomas Hobbes in the *Leviathan* (1651), in which the sovereign was identified with might rather than law. Law is what the sovereign commands, and it cannot limit his power; sovereign power is as absolute as men can make it. In the international sphere this condition led to a perpetual state of war, one sovereign trying to impose his will by force on all other sovereigns. This situation changed but little over the next two centuries; and sovereign states have continued to claim the right to be judges in their own controversies, to enforce by war their own conception of their rights, to treat their own citizens in any way that suited them and to regulate their economic life with complete disregard for possible repercussions in other states.

During the 20th century important restrictions on the freedom of action of states started to appear. The Hague conferences (*q.v.*) of 1899 and 1907 established detailed rules governing the conduct of wars on land and at sea. The covenant of the League of Nations (*q.v.*) restricted the right to wage war, and the Briand-Kellogg pact of 1928 condemned recourse to war for the solution of international controversies and its use as an instrument of national policy. They were followed by the charter of the United Nations (art. 2) which imposed the duty on member states to "settle their international disputes by peaceful means in such a manner that international peace and security, and justice, are not endangered," and supplemented it with the injunction that all members "shall refrain in their international relations from the threat or use of force. . . ."

At the same time, however, the charter listed as one of the basic principles of the UN "the principle of sovereign equality of all its Members."

In consequence of these developments, sovereignty by mid-20th century ceased to be considered as synonymous with unrestricted power. In the first place, all states being equal, no state was allowed to impose its will upon another by the use of force. Sovereignty of each state became entitled to protection by the international community, as organized in the UN, against the designs of greedy neighbours; and the organization was authorized "to take effective collective measures for the prevention and removal of threats to the peace, and for the suppression of acts of aggression . . ." (*United Nations Charter*, art. 1).

Though the means given to the UN to protect the sovereignty of its members were not sufficient to prevent all violations of national sovereignty and to settle all disputes by peaceful means, it cannot be doubted that important limitations have been imposed on the right of each state to do what it pleases in relation to other states.

These limitations on the sovereignty of each state were balanced, however, by the increase in the freedom of each state from interference by other states. As far as matters essentially within the domestic jurisdiction of a state were concerned, each state regained that measure of internal sovereignty that had suffered so many actual restrictions in the period of constant use of force by one nation against another.

Only to the extent that use of force had been effectively restricted, was it possible for a state to expect that its personality, territory and political independence would be properly respected, and that it would be able to enjoy unmolested the rights inherent in its external sovereignty.

It is, therefore, in the interest of every state that all the other states should comply faithfully with all those international obligations that safeguard its sovereignty. A state cannot, however, expect such compliance unless it reciprocally fulfills in a scrupulous manner all its obligations to other states. Those obligations are derived from principles of customary law developed painstakingly over several centuries, and from hundreds of treaties that are concluded every year.

Consequently, by mid-20th century, states had accepted a considerable body of law limiting their sovereign right of acting as they please. Those restrictions on sovereignty are usually explained as deriving from consent or autolimitation, but it can be easily demonstrated that in some cases states have been considered as bound by certain rules of international law despite the lack of satisfactory proof that these rules were expressly or impliedly accepted by them. On the other hand, new rules cannot ordinarily be imposed upon a state, without its consent, by the will of other states.

In this way a balance has been achieved between the needs of the international society and the desire of states to protect their sovereignty to the maximum possible extent.

Nonsovereign States.—The 19th-century distinction between fully sovereign states and several categories of less sovereign units lost its importance under the law of the United Nations. In the 1950s emphasis was placed not on legal differences among colonies, protected states, protectorates and states under the suzerainty of another state, but on the practical distinction between self-governing and nonself-governing territories. Nonself-governing territories became under the charter of the UN "a sacred trust," and the states administering them promised to develop them towards self-government. Some of these territories were placed under a UN trusteeship that resulted in a closer supervision of their administration by the UN and in their speedier progress towards self-government or independence. Once a territory achieved self-government, as defined in resolutions of the general assembly, supervision by the UN ceased, even though independent status was not reached. This happened, for instance, when Puerto Rico achieved the status of a self-governing commonwealth in 1953.

Conclusion.—The concept of absolute, unlimited sovereignty did not last long after its adoption, either domestically or internationally. The growth of the democratic form of government imposed important limitations upon the power of the sovereign and of the ruling classes. The increase in the interdependence of states restricted the principle that might is right in international affairs. The peoples of the world have recognized that there can be no peace without law, and that there can be no law without some limitations on sovereignty. They have started, therefore, to pool their sovereignties to the extent needed to maintain peace; and in increasing degree sovereignty is being exercised on behalf of the peoples of the world not only by national governments but also by organs of the world community. Thus the theory of divided sovereignty, first developed in federal states, has become applicable also in the international sphere.

See INTERNATIONAL LAW, PUBLIC; SPHERES OF INFLUENCE. *See* also Index references under Sovereignty in vol. 24.

BIBLIOGRAPHY.—The literature on sovereignty is abundant; almost every book on political science, government, international law and legal theory deals with this problem. A few of the modern special studies are: C. E. Merriam, *History of the Theory of Sovereignty Since Rousseau* (1900); H. Kelsen, *Das Problem der Souveränität und die Theorie des Völkerrechts* (1920); L. Duguit, *Souveraineté et Liberté*

(1922); H. J. Laski, *Authority in the Modern State,* ch. i (1919); H. Krabbe, *The Modern Idea of the State,* Eng. trans. by G. H. Sabine and W. J. Shepard (1922); J. Dickinson, "A Working Theory of Sovereignty," *Political Science Quarterly,* vol. xlii–xliii (1927–28); W. Sukiennicki, *La Souveraineté des États en Droit International Moderne* (1927); J. Mattern, *Concepts of State, Sovereignty and International Law* (1928); G. W. Keeton, *National Sovereignty and International Order* (1939); J. Marshall, *Swords and Symbols: the Technique of Sovereignty* (1939); M. S. Korowicz, *La Souveraineté des États et l'Avenir du Droit International* (1945); J. Maritain, "The Concept of Sovereignty," *American Political Science Review,* vol. xliv (1950); K. W. B. Middleton, "Sovereignty in Theory and Practice," *Juridical Review,* vol. lxiv (1952); J. Calvez, *Droit International et Souveraineté en U.R.S.S.* (1953); E. N. van Kleffens, "Sovereignty in International Law," *Académie de Droit International, Recueil des Cours,* vol. lxxxii (1953); J. L. Brierly, *The Law of Nations,* 5th ed., ch. i–ii (1955).

(L. B. Sx.)

SOVIET (Russian, "council"), a policy-making and administrative agency established at various levels of government in the Union of Soviet Socialist Republics under the direction of the Communist party.

Soviets were organized in 1905 by Marxist-stimulated intellectuals in Russian cities as strike committees to co-ordinate worker opposition to tsarist policies. Suppressed after the 1905 revolution, soviets of workers and soldiers were re-established by left-wing political leaders to influence governmental policy on the tsar's abdication in March 1917. Lenin demanded that they assume power when he returned from exile in April 1917. A national federation of soviets was created in June 1917, but moderate leftists held control until November. After seizure of power Lenin turned to a congress of deputies of local soviets to provide a semblance of popular support for Bolshevik leadership. By spring 1918, he had control, and thereafter the various soviets became unchallenged instruments of Bolshevik policy.

Structurally, soviets were large assemblies elected at village, city, county, province, republic and all-union levels. From 1917 to 1936 the franchise was denied employers of labour for profit, private merchants, *rentiers,* priests and monks.

The 1936 constitution opened the franchise to all on secret ballot, but in practice the one-candidate election was continued, permitting citizens only to reject candidates nominated by professional groups under Communist party direction. Soviets chose executive committees, called councils of ministers, at republic and all-union levels, to supervise administrative departments. Local soviets operated public housing, local transportation and industry, schools and health facilities. Republic and all-union soviets approved economic plans and budgets. No critical debate or adverse votes occurred because deputies were largely Communist party members bound to resolve disputes privately.

BIBLIOGRAPHY.—Sir E. D. Simon, Lady Simon, W. A. Robson and J. Jewkes, *Moscow in the Making* (1937); George Barr Carson, *Electoral Practices in the U.S.S.R.* (1955); John N. Hazard, *The Soviet System of Government* (1957).

(J. N. H.)

SOVIET LAW is a system of public order devised by communist politicians after the Russian Revolution of 1917 and emulated subsequently in Asian and eastern European states where Communist parties became dominant. Although it was classified by some legal scholars as a subdivision of the Romanist or civil-law system, its adherents claimed independent status for it. This claim was recognized in practice after World War II by election to the International Court of Justice of judges representative of the Soviet legal system in accordance with requirements of the International court's statute (art. ix) that it include judges representative of the main forms of civilization and of the principal legal systems of the world.

HISTORY

Origin.—Since Karl Marx and Friedrich Engels prescribed no pattern of law for a state organized to achieve the aims set by their *Communist Manifesto* of 1848, it was left for Lenin and his colleagues in the Russian Communist party after the Bolshevik Revolution of 1917 to improvise a legal system. General guides provided by Marx and Engels were that law be regarded as an instrument of the state, not as a limitation upon those who make policy, and that it enunciate rules of public order facilitating the transition to socialism and ultimately communism. Its major tasks were to be two: (1) elimination of the political power of the *bourgeoisie* by depriving them of their ownership of productive resources; and (2) education of citizens in the disciplined pattern of life claimed by communists to be requisite to achievement of the social order they desired. Completion of these two tasks would, in the communist view, assure abundant production suitable to realization of the communist aim of distribution according to need and also the self-discipline of citizens necessary to make coercion unnecessary in preserving order. The police, the army and courts would become unnecessary, and law would in Marxist terminology "wither away." Citizens would perform social obligations, expressed in morals and unsanctioned administrative regulations, out of conviction of their desirability rather than out of fear of punishment for their violation.

The new principles were embodied in the first constitution of the Russian Soviet Federated Socialist Republic, promulgated July 10, 1918 (art. ix), as follows:

> The basic task placed during the present transitional moment on the constitution of the R.S.F.S.R. is the establishment of the dictatorship of the city and village proletariat and of the poorest peasantry in the form of a powerful all-Russian Soviet authority with the objective of complete suppression of the *bourgeoisie,* the exploitation of man by man and the installation of socialism, under which there will be neither division into classes nor a State authority.

Lacking any precise pattern for a legal system designed to achieve the purposes stated in the first constitution, the new government issued only a few decrees designed to establish a framework for a new society and a primitive institutional structure to enforce it. The decrees deprived private individuals of the ownership of land, banks, insurance companies, merchant fleets and large-scale industry in implementation of the policy of expropriation of the *bourgeoisie;* created restrictions on the employers of labour; and secularized marriage and divorce. The enforcement instruments, called "people's courts," operated like tribunals in a tribal society, without benefit of professional prosecutor or bar.

To permit regulation of the multitudinous social relationships for which no new law was prescribed, judges were directed at the start to apply Russian imperial laws, but only to the extent that they had not been revoked by the revolution and were not contrary to the revolutionary conscience of the judges. Perhaps because Lenin and his colleagues were too busy to do otherwise, or perhaps because they preferred to let develop in practice a new legal order, no all-inclusive systematized body of law was prescribed until 1922. Local judges were guided during the first years, apart from the basic decrees indicated, only by suggestions issuing from the people's commissar of justice in the form of instructions, procedural manuals and journal articles praising some court decisions as the proper application of revolutionary conscience and denouncing others as improper.

Political enemies of the communists were not brought before the courts but were condemned by political bodies created at the same time as the people's courts and called revolutionary tribunals, or they were arrested and imprisoned without public hearings by the political police (Cheka). Lenin and his jurists professed only temporary resort to treatment of opponents outside the regular court system. Nevertheless, the conduct of revolutionary tribunals and of the Cheka greatly influenced the course of Soviet law long after their formal abolition in 1922, for they stimulated lack of respect for formality and strict adherence to law even among those charged with enforcing and administering the law.

New instrumentalities emerged within the ministry of internal affairs (M.V.D.), under the personal dictatorship of Lenin's successor, Joseph Stalin, to rid him of opponents. Following the latter's death in March 1953 the Communist party's first secretary, N. S. Khrushchev, disclosed that many innocent persons had been convicted and sentenced to long prison terms and even to death by special boards of the M.V.D. The Soviet method of preserving order was revealed officially to have been dual, with courts on one side dealing publicly with nonpolitical offenses and social disputes, and administrative boards and the security police punishing secretly on the other side those whom Stalin chose to call his enemies.

Codification.—Reintroduction of private enterprise in limited areas and subject to strict controls followed the declaration, in 1921, of the New Economic Policy (*see* RUSSIA) aimed to speed reconstruction to overcome the devastating effects of world and civil war. This reintroduction necessitated, in communist view, stabilization of law if capitalists were to be induced to invest. Lenin's commissar of justice, D. I. Kurskii, considered these measures retrogressive but necessary, while N. V. Krylenko, later to become federal commissar of justice, called them the natural evolution of five years of Soviet law. Whatever their origin, stabilization efforts took form in the first systematic codification, and a complex institutional framework was constructed in evolution of the primitive people's courts established by the first decree of 1917.

A judiciary act of Oct. 31, 1922, established within the Russian republic a three-stepped system of general courts with civil and criminal jurisdiction.

1. The name people's court was retained for the lowest level, but the institution was less primitive than that of 1917. Its bench comprised a full-time judge appointed annually and two lay judges selected for service of a few days in rotation from a panel of intelligent and politically trustworthy citizens. Lay judges shared with the professional judge in deciding issues of law and fact.

2. Appeals from this court lay to a new provincial court evolved from the congress of people's judges which under the prior rules had gathered periodically in each province to survey the work of the local courts. The provincial courts also received jurisdiction to try offenders against the security of the regime and otherwise serious civil and criminal cases. As an appellate court the provincial court presented a bench of three professional judges. As a trial court it was composed like the people's court, except that the lay judges were selected from a panel of more experienced and politically sound citizens.

3. To co-ordinate policy in all provinces there was created the supreme court of the republic, evolving from a control department established earlier in the commissariat of justice. This court was authorized not only to hear appeals from cases tried in provincial courts but also to discipline lower courts, issue rulings interpreting the codes and even to try cases of an unusually important character.

To provide specialized treatment of crimes relating to military matters and also to the disruption of transport, then deemed critical to the success of the regime, specialized courts which had evolved earlier were continued by the judiciary act and subordinated to the supreme court. An office of public prosecutor and a college of defenders to perform the functions of a bar were established in aid of the judges, following experimentation with institutions designed to conduct prosecution and defense on a nonprofessional basis.

Substantive law and procedure to be applied by the new courts were established by codes enacted throughout 1922 and early in 1923, these being criminal, civil, family, land and labour codes and also codes of criminal and civil procedure. Drafters of these codes followed generally patterns essentially similar to those of the Romanist states of the European continent.

Patterns of legal institutions and substantive and procedural law enacted by the R.S.F.S.R. were copied with little variation in the other Soviet-type republics that had emerged in the years following the revolution in peripheral regions of what had been the Russian empire. Federation of the Soviet republics into the Union of Soviet Socialist Republics on Dec. 29, 1922, caused no change in the judicial system except to establish at the top in the federal government, as a co-ordinator of all practice, a supreme court of the U.S.S.R. The first federal constitution adopted provisionally on July 6, 1923, and permanently on Jan. 31, 1924, extended to the federal government authority to enact general principles of law to be followed by the republics in maintaining their codes. While this development might have caused a change in the Soviet legal system, it did not; the first federal judiciary act of Oct. 29, 1924, confirmed the system of courts then existing in each of the republics, and no change was made in substantive or procedural law. The sole innovation was that the military and transport courts were placed under sole control of the supreme court of the U.S.S.R.; thus was created a self-contained system of federal courts, with lower instances functioning throughout the republics but beyond the reach of republic officials. A sharp line thus was drawn between courts having to do with matters vital to the security of the regime and those concerned with social disputes.

The pattern established for Soviet courts and codes of law in 1922 and 1923 remained with minor amendment until Stalin's death, although the second federal constitution of Dec. 5, 1936, transferred to the federal level, authority to enact codes of law to replace those of the republics. Although a second federal judiciary act was issued on Aug. 16, 1938, new codes were not enacted. On Feb. 11, 1957, the constitution was amended to restore the original form of relationship between republics and federation, namely that they enact their own codes but conform to general principles established federally. The first of the new general principles were enacted for criminal law and procedure on Dec. 25, 1958, while others were enacted at intervals thereafter, and republic codes were revised subsequently. The major innovation of the 1958 enactments was the requirement that punishment be ordered only by a court in accordance with the rules of the procedural code. This provision was heralded by Soviet jurists as preventing return to extralegal procedures such as those exercised by the M.V.D. during Stalin's dictatorship.

CHARACTER OF SOVIET LAW

Source.—Montesquieu's concept of separation of powers as embodied in the constitution of the United States of America was rejected by the founders of the Soviet legal system. The legislature was made supreme over the executive and the judiciary. In principle only the legislature's enactments were a source of law.

Practice made the executive a source of law, however, for the council of ministers often acted without convening the legislature. The second federal constitution of 1936 attempted to stop this practice in violation of the constitution of 1923, but the executive soon again usurped legislative authority. The post-Stalin reforms again sought to restore monopoly lawmaking power to the legislature. In practice it was not the legislature in full assembly but its derivative body, the presidium, elected from its membership to legislate during intervals between full sessions, that created the day-to-day changes in law. Although subsequent ratification from the full body was sought, revocation of presidium action was virtually impossible because of passage of time. The presidium thus became the most active source of law, except for the state economic plan, the budget and enactment of general principles for republic codes.

Communist party orders were in the technical sense not a source of law, for to be binding on courts these had to be enacted by state agencies. Still, the party provided the initiative for legislative action. Some vital laws, such as those reforming agriculture or industry, were given pre-eminent status by signature by the party secretary as well as by the chairman of the presidium when they were published as law.

Having rejected judicial lawmaking with assertion of the legislature's role as the sole lawgiver, Soviet jurists accepted no concept of judicial precedent like that of the Anglo-U.S. common law. Nevertheless, the supreme court of the U.S.S.R., in exercise of its authority to enunciate principles of interpretation based upon examination by its full body of a course of judicial practice, enunciated such far-reaching principles as to make of these a source of law. Even individual decisions of the supreme court, although lacking any authoritative force such as that of rulings generalizing practice, were given careful attention by lower courts, since they were published as guides. The Soviet legal system in practice adopted much the same approach to judicial decisions as have Romanist systems generally.

The supreme court of the U.S.S.R., being subordinate to the legislature, had no right to declare a law unconstitutional, nor to create any limitation upon the legislature or the executive on the ground that some higher or general principles of law had been violated.

Control over the judiciary was preserved for the legislature at all levels above the people's court by giving it the appointive and recall power, through the soviet operating at the court's level. Only people's judges were elected by the people under provisions introduced by the 1936 constitution, although elections were not begun until after World War II. Elections followed the pattern established for all representative bodies in the U.S.S.R. by Stalin, namely the one-candidate ballot, the candidate being chosen by local professional groups with the guidance of the Communist party members in the groups. In practice, judicial decisions had to accord generally with Communist party policy or the judge risked recall.

In light of the foregoing, the constitutional provision that "judges are independent and subject only to the law" (art. cxii) had to be understood as denying to local officials the right of intervention in the formation of an individual decision on personal grounds, but as not prohibiting communist party intervention when such intervention was properly formulated in criticism of a line of decisions out of keeping with party wishes.

Judges were not deprived entirely of initiative by the rule of legislative supremacy or by the leading role played by the Communist party. In both the civil and criminal codes adopted in 1922 there was authorization to depart from the rigid provisions of the code under special circumstances. The civil code opened with art. i, reading, "Civil rights are protected by law except in cases where they are exercised contrary to their social and economic purpose." This rule permitted judges to apply their concept of state policy, when it was not specifically stated, to achieve an appropriate result in a given situation, but the decision created no rule of law for future cases. The article was used frequently to justify judicial refusal to protect the rights of capitalists during the period of reconstruction (1922–28), but after 1930 it fell into disuse, and Soviet jurists argued for its deletion from the code as unrelated to a stable society from which all capitalist elements had been eradicated.

The criminal code of 1922, and as revised in 1926, permitted the judge to exercise discretion not only in withholding conviction where a crime had clearly been committed, provided that he found no social danger, but also in punishing a citizen who had committed no crime defined as such by the code. The formula for such judicial initiative was that the act be found socially dangerous in a manner analogous to another act specifically defined as a crime. This provision, which led to unpredictability, was eliminated from the criminal law by the general principles adopted in Dec. 1958.

Substantive Provisions.—Soviet jurists claimed uniqueness for their system because of its constitutionally defined function of promoting socialism. The end in their view transformed the means.

Civil Rights.—Civil rights as defined in the U.S.S.R. constitution of 1936 (ch. x) provided the primary example of the Soviet approach. Guarantees of freedom of speech, press and assembly were limited to such exercise as might strengthen socialism (art. cxxv). Judicial practice indicated that speech evoking hostility to Soviet authority on grounds of disagreement with the elements of social order declared by the constitution as fundamental to socialism—such as state ownership of the productive resources, collectivization of agriculture, the one-party political system and equality of races—was punished as a crime against the state if intent to disrupt the system was found. In the practice of the security police prior to the death of Stalin, disruptive intent was presumed from social origin (as a prerevolutionary landowner, industrialist, merchant, army officer or priest, and sometimes in children of such persons).

State Enterprise.—Among the various fundamentals of socialism, primary place was given to state ownership of productive property, and law codes reflected this fundamental in various provisions. The criminal code established more severe penalties for theft or destruction of state property than for that of personal property. The labour code was concerned with state employment. The provisions of the civil code of 1922 governing incorporation of private businesses became obsolete with the passing of the New Economic Policy, and even those provisions relating to sales, contracts, property damage and inheritance governed only a small segment of social activity because the volume of personal property relationships was sharply reduced. Nothing monopolized by the state could be sold or leased by private persons or even by state enterprises and co-operatives, with limited exception for land allocated to peasant households disrupted temporarily by war or industrial employment.

A major new area of law that accompanied socialism was the law governing state enterprise. Its task was to provide the order and accountability required for performance of the national economic plan. After experimentation for a period with direct operation by ministerial departments, state enterprise passed in 1923 to newly created public corporations, having juridical status with power to sue and be sued. State property was assigned to each enterprise, and business-type accounting procedures were introduced to assure strict accountability. The function of law in the accountability process was performed by an adaptation of the principle of contract. Interrelations of public corporations took the form of written agreements to buy, sell and perform services, these agreements being executed in performance of the planned objective of each corporation.

For scarce commodities, such as aluminum, copper and steel, the plan was detailed, establishing the precise corporations to produce and consume the materials. Contracts in such cases became documentation of detail in implementation of planned relationships; management had no choice but to execute the contract. With regard to less critical commodities, the plan provided no such detail of corporate interrelationship; management was permitted to seek out opposite parties producing needed raw materials, and to enter into relations with distributing units prepared to sell its products to ultimate consumers. With the administrative reform of 1957, which abolished many centralized ministries and substituted regional economic councils to guide the economy, the area of relatively free choice in the establishment of productive relationships among public corporations was enlarged, and contracts took on greater meaning as the documentation of relationships established with some exercise of free will by the parties. Nevertheless, the aim of plan performance was dominant, and a party evidencing reluctance to make a contract with another could be taken before a state agency for determination of the necessity of execution.

The state agency designated to hear precontract disputes as well as those arising over failure of performance was a specialized court designated as "state arbitration." Created on various levels to facilitate quick response to demand for a decision, its arbitrators utilized the law of the civil code in reaching a decision, but more than money damages was required. Since the primary purpose for the contract relationship was execution of the plan, arbitrators were directed to assure specific performance, if possible, and only if that were out of the question, to order the payment of damages; such payment set in balance the profit and loss statement of the injured party and attracted the attention of state inspectors to the offending party by upsetting its profit and loss statement.

Personal Property and Personal Injury.—Provisions of the civil code relating to personal property matters of private citizens, as well as to personal injury actions, provided no surprises for those familiar with the Romanist provisions of the civil codes of the European continent. There was nothing socialist about these provisions, although Soviet codifiers sought to incorporate some hardship provisions to avoid exploitation of the poorer party to a bargain. Such a provision, already in use even in western Europe, was the one prohibiting enforcement of a contract if one party in negotiating it had taken advantage of the "extreme want" of the other. Likewise, uneconomic maintenance of a dwelling could result in its confiscation by municipal authorities if the owner had resources to make repairs and had been notified of the necessity of repair at an appropriate time.

Labour.—The provisions of labour law reflected state ownership of productive resources, especially after 1936, when the second federal constitution prohibited private employment from which a private citizen anticipated a profit as entrepreneur. Labour law then came to resemble civil-service law in many non-Soviet-type

states, its distinguishing feature being that it was the law of all employment rather than of a small segment. Statutes established wages and hours in accordance with a system of standard classification of all jobs. Employers were not permitted to depart from the standards after 1938.

The labour code defined circumstances under which dismissal was permitted (art. xlvii), such as incompetence, malfeasance, termination of a position or unauthorized absence for a given period of time. In regard to unauthorized absence, the period was changed as discipline became stricter with the approach of World War II, until finally dismissal was eliminated as a sanction and by law of June 26, 1940, workmen were subjected to criminal prosecution if absent without authorization for more than 20 minutes. A labour grievance procedure comprising tribunals on which representatives of management and the labour union shared equally in the decision provided a preliminary opportunity for an aggrieved employee to contest management action. A dissatisfied employee might proceed to court in appeal of the tribunal's decision unless he was in the supervisory category or in a position requiring security check.

Collective agreements between management and labour unions fell into disuse in the mid-1930s, when wages and hours became fixed by statute, but they were reintroduced in new form in 1947. They then became instruments to stimulate production by restating workers' obligations and assuring distribution, in accordance with the workers' desire, of the bonuses granted by statute to employees for exceeding planned quotas or reducing costs of production. Labour unions were not deprived specifically of the right to lead strikes, but Communist party leaders in 1928 declared strikes unthinkable as being against the workers' own state.

Land.—Land use was governed by the land code of 1922, by supplementary decrees, such as the Model Charter for the Collective Farm as enacted initially in 1930 and subsequently amended, and by specialized decrees on the use of land by public corporations and individual houseowners. These reflected socialism in that title to land remained in the state and only the use was allocated. Use passed in perpetuity to the individual houseowner, peasant household, collective farm or public corporation, subject to withdrawal in the event of misuse or requirement for other state activity such as railroad or highway construction. Safeguards against arbitrary withdrawal of farm lands were instituted, it being the desire of policy makers to preserve the psychological advantage provided by capitalism in encouraging maintenance and improvement of land held in perpetuity. No sale or leasing of perpetual right to use was permitted, except as already indicated.

Family.—Family law provided for secularization of marriage and divorce. Early law permitted registration of divorce at the request of one spouse and without notification to the other, but this procedure was replaced in 1944 by a stricter system thought to be in keeping with Engels' dictum that socialism would see the introduction of the truly monogamous family.

Thereafter, the people's court was obligated to attempt reconciliation, and only in the event of failure could a party proceed to the provincial court, which was required to satisfy itself that restoration of the marriage relationship was impossible before granting the divorce.

Procedural Law.—Romanist influence was strong on procedural law, the codes of criminal and civil procedure copying the main features of the French codes, although retaining more possibilities of control by the state (*see* also FRENCH LAW).

Emphasis in criminal procedure was placed on investigation before trial, as in French practice. The examiner had to establish more than a prima-facie case required of the grand jury in the traditional common-law system. The suspect had to be given an opportunity to testify and to produce witnesses and evidence in his defense. The examiner became a preliminary judge authorized to forward the case to trial only if he became convinced of guilt, so that the public trial constituted primarily verification of the examiner's work rather than a hearing for the first time of the defendant's case. Contrast with the French model was provided by the fact that the examiner was not a magistrate but subordinate to the office of prosecutor, and until 1958 no attorney was permitted the defendant during the preliminary investigation. Even thereafter only juveniles were given an absolute right to an attorney.

Trial procedure followed the French model closely in that the judges were instructed to establish their intimate conviction of guilt. To do so they might go beyond evidence presented by the prosecution and defense. The court might call witnesses, search out its own experts, examine material evidence and take a view of places. The judges did most of the interrogating in the courtroom, relying on the prosecutor and the defense only to bring out points overlooked. No rule of evidence of any kind bound the court, except the rule of relevance.

Civil procedure required observance of rules of evidence only when necessary to sanction a rule of formality established by the civil code, such as those for documents like contracts or wills. Otherwise, the judges were as free in civil as in criminal matters to seek or hear what they desired to establish their intimate conviction. Soviet jurists claimed an innovation in the prosecutor's authority to intervene in a civil case at any stage thought necessary to the interests of society. This sometimes brought in the prosecutor to protect a party whose attorney appeared incompetent.

Appellate courts finding violation of procedural requirements, substantive law or unconvincing evidence were required to remand for new trial. Revision of a sentence or decision without retrial in a court of original jurisdiction was permitted only when the original court had no jurisdiction, where an amnesty had been improperly applied or where there had been no basis for trial at all. A penalty could not be increased on defendant's appeal.

A singular feature of Soviet procedure was the authority given to supreme-court presidents and to the prosecutors of the republics and of the U.S.S.R. to protest a decision, whether civil or criminal, after it had become final on appeal or in absence of appeal. This procedure permitted reopening of convictions or even of acquittals if the highest legal authorities found in their periodical audit of inferior-court activities grave error at the trial or in the review. In practice the prosecutors requested frequently such reopenings, even of convictions obtained by inferior prosecutors. This created the opportunity of petition to superior prosecutors or court presidents after all remedies as of right had been exhausted. At the same time it gave state authorities a second chance, not given the convict or the party to a civil suit, to change an otherwise final decision.

PEOPLE'S DEMOCRACIES

Application of the pattern of law evolved within the U.S.S.R. to other areas began when Outer Mongolia declared establishment of its "people's government" on July 11, 1921. Mongolia was then a nomadic society without industry. Nationalization of productive resources had little effect on nomadic life, an exception to nationalization being made in the area of primary concern, namely the private ownership of cattle herds. Nevertheless, the base was laid for emulation of Soviet legal patterns in the first constitution of the Mongolian People's Republic promulgated on Nov. 26, 1924. Thereafter, state-owned industry was developed in emulation of Soviet patterns, and following adoption in the U.S.S.R. of the second federal constitution, a second Mongolian constitution was promulgated on June 30, 1940. Its provisions on courts and prosecutors were like those of the U.S.S.R., and the subsequently adopted law on land use of Feb. 6, 1942, provided the same formula of state ownership of land allocated for use gratis and in perpetuity to the nomadic herdsmen for their privately owned cattle. A labour law of Feb. 14, 1941, followed the Soviet pattern, as did a social insurance law of June 22, 1942.

The major opportunity to install the Soviet legal system outside the U.S.S.R. came only after World War II, when Soviet-type governments were created for European states occupied by the Soviet army in pursuit of retreating German forces. The system was also established in China, North Korea and North Vietnam as military victories placed the communists in power. In each country there was some variation from the Soviet pattern evolved in the U.S.S.R., this variation being related to the degree of economic development.

Eastern European states permitted retention of limited forms of

private enterprise in industry and agriculture in forms reminiscent of the economic pattern of the U.S.S.R. during the 1920s. In China, capitalists, except for landlords, were permitted for a time to function widely, though subject to strict state controls and heavy taxation. In North Vietnam even private landowning was permitted except for large estates, which were divided and distributed among tenants and poor peasants. Law in these communities reflected the presence of capitalist elements in the economies. Yet in all the people's democracies the banks, large-scale industry, insurance, mining and transport were nationalized and monopolized by the state. Law relating to this sector of the economy, termed the socialized sector, followed the U.S.S.R. model.

Throughout all the states coming under communist domination law was given the functions previously established in the U.S.S.R. Thus, the constitution of the Hungarian People's Republic of Aug. 20, 1949, declared (art. xli):

> Courts of the Hungarian People's Republic punish enemies of the working people, defend and secure the state, economic and social structure of the people's democracy, its offices and the rights of the toilers, educate the toilers in the spirit of observance of the rules of socialist intercourse.

Regulations governing the organization of the people's courts in the People's Republic of China, promulgated Sept. 3, 1951, specified that courts were "to consolidate the people's democratic dictatorship, uphold the new democratic social order and safeguard the fruit of the people's revolution." The People's Republic of Bulgaria was even more outspoken than the U.S.S.R. in clarifying the function of law; in its constitution of Dec 4, 1947, it granted the right to citizens to create associations, but (art. lxxxvii) only "if they are not directed against state and public order established by the present constitution," and then stated:

> It is forbidden and will be punished by law to form organizations having as their purpose taking from the Bulgarian people the rights and freedoms won by the people's uprising of Sept. 9, 1944, and guaranteed by the present constitution, or to limit these rights and freedoms, to place under threat the national independence and state sovereignty of the country or to propagate open or concealed fascist or antidemocratic ideology, or to facilitate imperialist aggression, and also to participate in these organizations.

The Chinese Communist party leader, Mao Tse-tung, in his 1949 speech "On People's Democratic Dictatorship" indicated that he had adopted in effect what had been Stalin's interpretation, enunciated in 1930, of the Marxist concept of the withering away of the state. Mao declared that there could be no withering in China as long as there was opposition to the state. The constitution of the People's Republic of China, promulgated on Sept. 20, 1954, nearly five years after proclamation of the people's republic on Oct. 1, 1949, fixed as final the system of courts evolved in emulation of those of the U.S.S.R. There were to be lay assessors, full-time professional judges, Soviet-type prosecutors, three levels of courts and "judges were independent and subject only to the law" as they had been declared to be in the U.S.S.R. constitution.

Judicial patterns of the same type were established in the constitutions of the other people's democracies in addition to those already mentioned, namely the Democratic People's Republic of Vietnam (Nov. 8, 1946); the People's Republic of Albania (March 14, 1946); the Rumanian People's Republic (April 13, 1948, and Sept. 24, 1952); the Republic of Czechoslovakia (May 9, 1948); the Democratic People's Republic of Korea (Sept. 8, 1948); the German Democratic Republic (Oct. 7, 1949); and the Polish People's Republic (July 22, 1952).

The applicable law in the new Soviet-type republics was stated in a small number of specific decrees in each republic, but otherwise determination of law was left to the judges. The Chinese government among its first actions abrogated all legal codes of the prior government. In Oct. 1950 the Chinese premier, Chou En-lai, listing the documents to serve as guides to the judges, included not only specialized decrees of the government but also the Common Program of the Chinese Communist party. The provisional regulations on people's courts of 1951 (art. iv) read, "Where no regulation is applicable, the policy of the Central People's Government will be adhered to." A law codification commission was established in the spring of 1950 but no basic codes emerged, only specialized statutes such as the land reform law of June 28, 1950; the law on marriage, April 13, 1950; the law on questions of employment, July 25, 1952; and the organic law for the court system, Sept. 21, 1954.

The Soviet-type states in eastern Europe moved more rapidly toward codification than had China, and in contrast to China confirmed the existence of prewar civil codes, although with modifications. Poland adopted a General Statue on Civil Law with 119 articles on July 18, 1950. As in the Soviet civil code of 1922, the codifiers sought to avoid application of the law to possible injury of working-class elements, for art. iii provided: "It is forbidden to use one's rights to violate the fundamentals of socialist intercourse of the people's government." A Polish family code was adopted on June 27, 1950, and a decree on planned contracts on April 26, 1950.

Prior to the expulsion of the Yugoslav Communist party from the Cominform on June 28, 1948, Yugoslavia, under the constitution of Jan. 31, 1946, had followed the Soviet legal pattern. Under Tito, Yugoslavia accepted the concept of monopoly of power in a Communist party and the necessity of state ownership of productive resources, but refused to accept dictation from the Soviet Union on every detail, especially on centralization of industrial control and the speed of collectivization. The Yugoslav government experimented with bringing workmen into the managerial process of public corporations and in broadened autonomy for local government councils. Yugoslav legal philosophers rejected the view that the state could not begin to wither away until full achievement of communism; they wanted to start relaxation of coercion immediately. Stalin's attempts to unseat Tito failed, and on Stalin's death his successors attempted reconciliation for a time, indicating acceptance of the idea that there could be many roads to socialism.

Polish influence upon development of Soviet legal theory was noticeable after the risings of 1956. Of primary importance to Poles, including communists, was preservation of legal procedures evolved over the centuries to assure that only the guilty were punished. Polish legal philosophers claimed no incompatibility between communist goals and humanist attitudes toward trial. Polish courts were reorganized to require strict adherence to procedural rules, and the position of the bar was popularized as necessary to the proper functioning of justice. Soviet legal philosophers endorsed the new humanism in 1958 and declared that it had always been an implicit fundamental of the Soviet system. Nonetheless, the fundamentals of Soviet law remained unchanged. Lenin had declared after the Russian Revolution that there was no basis under Soviet conditions for the Roman law division of law into public and private spheres. To Lenin the Soviet state's concern with every detail of social intercourse required maintenance of the state's right and opportunity of intervention in any matter at any time. Consequently he held that in the Soviet system all law must be public, that is, reflect the state's vital concern in the social relationships governed by law. This attitude of state direction of all matters, sometimes called *dirigisme* by French scholars, remained characteristic of Soviet law.

The industrialization characteristic of the 20th century required legal systems in all states to accept increasing state intervention in social relationships. Employers were no longer left free to execute such labour contracts as they wished; property owners were subjected to increasing restraints on use, and in some countries were subjected even to the loss of farms or dwellings if these were left idle or not used productively; nationalization in banking, industry and transport was widespread; compulsory insurance against injury to third parties became commonplace, as did social insurance against industrial accident; and secularization of marriage and divorce nearly everywhere prevented citizens from conducting their marital relationships solely within religious rules.

Soviet law, in its reflection of the state's pervading presence in human relationships, appeared to be but an extension in extreme degree of trends common to legal systems everywhere. Yet many legal philosophers concluded that Soviet law differed in more than degree from other systems. While its techniques were those of the Romanist system, its change in degree was interpreted as in-

troducing a change in quality, justifying the conclusion that Soviet law was a unique system requiring separate categorization.

BIBLIOGRAPHY.—John N. Hazard, *Settling Disputes in Soviet Society: the Formative Years of Legal Institutions* (1960) and *Law and Social Change in the U.S.S.R.* (1953); Peter S. H. Tang, *Communist China Today* (1957); Vladimir Gsovski and Kazimierz Grzybowski, *Government, Laws and Courts in the Soviet Union and Eastern Europe*, 2 vol. (1959); Charles P. McVicker, *Titoism: Pattern for International Communism* (1957); George C. Guins, *Soviet Law and Soviet Society* (1954); Harold J. Berman and Miroslav Kerner (eds.), *Soviet Military Law and Administration* (1955); Boris A. Konstantinovsky, *Soviet Law in Action: the Recollected Cases of a Soviet Lawyer* (1953); R. R. Betts (ed.), *Central and South East Europe 1945–1948* (1950); Harold J. Berman, *Justice in Russia: an Interpretation of Soviet Law* (1950); Vladimir Gsovski, *Soviet Civil Law: Private Rights and Their Background Under the Soviet Regime*, 2 vol. (1948–49); Rudolf Schlesinger, *Soviet Legal Theory: Its Social Background and Development* (1945); V. I. Lenin *et al.*, *Soviet Legal Philosophy*, trans. by Hugh W. Babb (1951); Andrei Y. Vyshinskii, *The Law of the Soviet State* (1948). (J. N. H.)

SOWER, CHRISTOPHER (1693–1758), printer and publisher, was born at Laasphe, Ger., and emigrated to Pennsylvania in 1724 with his wife and small son Christopher (1721–84). Sower worked first as a tailor, then took up farming and later bought six acres of land in Germantown, where he settled and built a large house. There, in 1738, he set up a print shop to become the first German printer and publisher in the American colonies. It is not known where he secured his training or the equipment he used—the press, the type, etc. He made his own ink and it is possible that he cast his own type, but the better fonts he used were imported from the Frankfurt foundry of Heinrich Luther. In 1744, it is believed, he also built a paper mill. Most of the Sower publications dealt with religious or educational subjects. The first to bear the Sower imprint was *Eine Ernstliche Ermahnung, an Junge und Alte* (1738), followed by the first issue of his almanac (1739). In 1739 he also issued his first complete book, a collection of hymns, and the first number of the newspaper *Der Hoch-Deutsch Pensylvanische Geschicht-Schreiber*. Like the almanac, the newspaper continued to appear until 1777. The two publications were available in all the colonies and made Sower and his son persons of note in the German communities of Pennsylvania and Maryland. Best known of Sower's publishing efforts was the *Biblia; Das ist, Die Heilige Schrift Altes und Neues Testaments, nach der Deutschen Übersetzung D. Martin Luthers* (1743), the second edition of the Bible to appear in America. Two later editions were published by his son Christopher, in 1763 and 1776. The first English publication carrying Sower's imprint was *Extract From the Laws of William Penn* (1740). As a popular journalist Sower was outspoken in his opposition to slavery and war. His practical knowledge of contemporary handicrafts and his social zeal brought him a considerable audience and influence in his time. His son Christopher had been trained very thoroughly, and at Sower's death in 1758 was well qualified to inherit the Sower press.

SOWERBY, LEO (1895–), U.S. composer, teacher and organist, was born at Grand Rapids, Mich., May 1, 1895. He studied in Chicago, where in 1917 he presented a program of his orchestral works and songs and in 1921 was named the first recipient of the American Prix de Rome. He taught composition and theory at the American Conservatory of Music in Chicago from 1925 to 1962 and was organist at St. James's church (now the Cathedral of St. James) there from 1927 to 1962. In 1962 he was appointed director of the College of Church Music in Washington, D.C.

Sowerby's early works were influenced by Max Reger and Grieg. His piano pieces, *The Irish Washerwoman* and *Money Musk*, later orchestrated, are in the manner of Percy Grainger. In *Synconata* (1924) and *Monotony* (1925) he experimented with jazz. His numerous vocal works include cantatas, services, anthems and *Canticle of the Sun* (performed 1945), which won a Pulitzer prize. His orchestral works include four symphonies. Besides concertos for piano, cello and organ Sowerby wrote many chamber works including sonatas for wind instruments and piano.

See B. C. Tuthill, "Leo Sowerby," *The Musical Quarterly* (April, 1938). (N. Sy.)

SOWING: *see* PLANTING MACHINERY.

SOYBEAN (*Glycine soja;* syn. *Glycine max*), also known in some countries as the soja bean and the soya bean, is an annual, summer leguminous plant native to eastern Asia and extensively cultivated in the United States. Cultivated in China and Japan long before written history, it is from the standpoint of uses and value the most important legume grown in those countries. Many records of soybean culture in China, some going back as far as 2207 B.C., give advice on soil preference, time of planting, varieties for different purposes and numerous uses indicating that the soybean was among the early crops grown by man. The crop was considered the most important cultivated legume and one of the five sacred grains (rice, soybeans, wheat, barley and millet) essential to the existence of Chinese civilization. In a Chinese dictionary dating from about the beginning of the Christian era the soybean is called *sou*, which very probably was the source of the names soi, soy, soya and soja.

Europeans first knew of the soybean through Engelbert Kaempfer, a German botanist who spent three years (1690 to 1693) in Japan. Kaempfer gave the Japanese name *daidsu mame*, and described it as an erect bean with the pod of a lupine and the seeds of a large pea. Although he discussed the many food products prepared from the soybean by the Japanese, little interest was taken in the crop for almost another century.

In 1875 a great impetus was given to soybean culture in Europe through the experiments of Friedrich Haberlandt of Austria who grew the seed of several varieties obtained at the Vienna exposition in 1873. Seeds of the four varieties that matured were distributed widely in Austria-Hungary and other countries of Europe and gave promising results. Haberlandt published the results of his investigations in 1878, stimulating further research on the crop. While records show packets of seed being received in France as early as 1739 from missionaries in China, no record of successful culture of the soybean was established in that country until 1879 when a Chinese variety matured satisfactorily at Marseilles. In 1880 Vilmorin-Andrieux and company introduced into France one of the yellow-seeded varieties tested by Haberlandt, which proved to be well adapted to French conditions. Soybeans were grown as early as 1790 in the Royal Botanic gardens, Kew, Eng., but the crop is not well adapted to the English climate and there is no commercial production at present. Cultivation of the soybean in Italy dates from about 1840 but it has not become an important crop in any part of the country.

J. HORACE MCFARLAND COMPANY

PODS AND LEAVES OF THE SOYBEAN

The first mention of the soybean in the United States was by James Mease in 1804, who noted that it did well in the climate of Pennsylvania and recommended that it be cultivated. The Perry expedition to Japan in 1854 brought back two varieties of "soja bean"; for several years after that, the soybean was referred to as the Japan pea. In 1882, the North Carolina agricultural experiment station grew a yellow-seeded soybean, probably the Mammoth Yellow, which remained a leading variety in that area for many years. In 1889 W. P. Brooks of the Massachusetts agricultural experiment station brought with him from Japan a number of varieties. The next year C. C. Georgeson secured three lots of soybeans from Japan and grew them at the Kansas agricultural experiment station. Since then most of the agricultural experiment stations in the United States have conducted extensive tests with the soybean crop.

In 1898 the U.S. department of agriculture began introducing large numbers of soybean varieties, mostly from Asian countries with a few from Europe and other parts of the world, and since then has grown and evaluated about 10,000 varieties of foreign

origin. Probably among the first varieties secured from Europe were some used by Haberlandt in his early trials in Austria. Beginning in 1936 the U.S. department of agriculture co-operated with the agricultural experiment stations of states and Canadian provinces in developing improved soybean varieties through hybridization and selection (*see* CHEMURGY). These improved varieties have higher yield and oil content and growth habits more suitable to modern farming methods than the Asian varieties. Increase of acreage and production has been closely correlated with the introduction of these varieties and their improvement through breeding and selection. The improved varieties developed through co-operative research have also contributed very substantially to the increased industrial value of the crop.

Areas of Production.—*Asia and Australia.*—The soybean is grown intensively in China including Manchuria. It is the cash crop of the Manchurian farmer and occupies about 25% of the total cultivated area of that region. In the second half of the 20th century China (with Manchuria) produced nearly 340,000,000 bu. a year or about 40% of the world total. Other important soybean producing countries of Asia are Indonesia, Japan, Korea, Formosa (Taiwan), Thailand and India. The soybean is cultivated in the Philippines and grown successfully in parts of Australia and New Zealand.

Europe and Africa: The climate of Europe is not generally suited to soybean culture although there is limited production in Italy, Yugoslavia, Rumania and southern U.S.S.R. There is some soybean culture in Africa, mainly in Tanganyika, the Congo, Nigeria and the Republic of South Africa.

North and South America: In the western hemisphere, the United States was the leading soybean-producing country, accounting for nearly 55% of world production. Approximately 83% of the U.S. soybean production is in the north central states. Other areas of heavy production are in the Arkansas and Mississippi delta area of the Mississippi river and along the middle Atlantic coast. Soybeans became a major crop in the United States, with production increasing from 5,000,000 bu. in the mid-1920s to about 490,000,000 bu. in the mid-1950s. Another important soybean-producing area in North America was in the province of Ontario, Can. In South America, Brazil and Argentina were becoming important soybean-producing countries.

Soybean Production in Specified Countries and the World, Average 1935–39 and Annual 1957
(in 000 bu.)

Continent and country	Average 1935–39	1957
North America		
Canada	207	6,524
United States	56,167	479,841
Europe		
Italy	1	22
Yugoslavia	71	147
Others	1,065	965
Asia		
China	207,666	335,000
Manchuria	151,294	
Formosa (Taiwan)	151	1,152
Indonesia	9,731	12,860
Japan	12,499	17,978
Korea*	17,054	5,879
Thailand	232	—
Turkey	37	92
South America		
Argentina	—	380
Brazil	—	2,939
Africa		
The Congo	—	147
Nigeria	—	597
Rep. of South Africa	—	—
World total	464,300	877,185

*South Korea only.

Plant Characteristics.—The soybean is an erect, branching plant, resembling in its early growth ordinary field and navy beans. Varieties range in maturity from early to late, and in height from less than one foot to five or six feet. The root system constitutes about 10% of the weight of the plant but is well branched and often extends to a depth of four or five feet. The characteristic deep roots are responsible in part for the great resistance to drought offered by soybeans. Except for the pair of simple leaves on the first node above the cotyledons, the leaves are trifoliate and borne one to a node. They vary widely in shape, size, colour and degree of persistence (leaf remaining attached to the stem) though they usually turn yellow as the pods ripen and are shed by the time the plant is ready to harvest. Nearly all varieties are pubescent; *i.e.*, the stems, leaves and pods are covered with fine tawny (brown) or gray hairs.

Several yellow-seeded varieties in the orient are entirely glabrous, that is, lack pubescence; these usually are smaller, shorter, and yield less seed than most pubescent varieties, and when grown in North America are subject to considerable injury from leafhoppers. In the orient, the glabrous varieties have been found highly resistant to attacks by the pod borer, a serious insect pest, which as yet has not been found in North America.

In the axil of each leaf is a bud that may develop into a branch or flower. The flowers are small and inconspicuous and are either white or different shades of purple. They contain both male and female parts and are almost completely self-fertilized although in the field there is usually around 1% cross-fertilization.

The pods, usually containing two or three seeds, range in colour from light straw or tan through shades of gray and brown to nearly black. Those of the wild soybean and a few hay-type varieties dehisce (open and scatter their seed) readily as the seed matures, although most commercial varieties have been selected for the ability to hold their seed for two or more weeks after maturity.

The seeds are usually yellow, green, brown or black but may be bicoloured. No truly white or red seeds are known in soybeans. The most common bicoloured patterns are green or yellow with a saddlelike patch of black or brown extending down on each side of the hilum or seed scar. Some seeds are brown and black with the patterns concentrically arranged. The cotyledons of soybeans may be yellow or green. Varieties with yellow seed coats and yellow cotyledons are preferred for industrial processing. Seed size varies greatly from the 30,000 seeds per pound of the wild bean to as many as 1,200 seeds per pound of some vegetable varieties. Most commercial varieties range in seed size from 2,500 to 3,500 seeds per pound.

Cultivation.—The soybean will succeed on nearly all types of soils but does best on fertile or sandy loams. In general, soil requirements are about the same as those for corn or cotton. Like other legumes, soybeans obtain part of their nitrogen requirement from the air through the action of certain nodule-forming bacteria living on the roots of the plant. Where soybeans have not been grown previously it is advisable to inoculate the seed with the soybean bacteria just before sowing. (*See* SOIL: *Soil Microbiology.*) In the United States soybeans are usually planted during May or early June in rows 30 to 40 in. apart. Early cultivations to kill small weeds or to break a soil crust are usually done with a rotary hoe or harrow starting soon after the plants emerge. Later cultivations are done with regular row equipment used for corn or cotton. Chemical weed control holds promise but is dependent to some extent on weather conditions.

Approximately 50 diseases are known to attack soybeans, but only a few have caused serious economic losses. Varieties resistant to leaf diseases have been developed, and diseases can be controlled to some extent by planting resistant varieties and following proper cultural practices. Several insects attack soybeans and cause severe losses, but good control may be obtained by timely dusting or spraying with a number of different insecticides. *See* SPRAYS AND DUSTS IN AGRICULTURE.

When soybean varieties are grown in their area of best adaptation they mature during September or October, utilizing the full growing season. Flowering and ripening in soybeans are controlled by length of day and night (photoperiod) in which the plants are growing. Soybeans are called short-day plants because under normal conditions they will not start blooming or maturing until the day length is shorter than a certain minimum value which is specific for each variety. This is the reason that a soybean variety is adapted to a rather narrow belt of latitude; and why, for example, in the northern hemisphere, if a variety is planted north of its proper latitude, blooming and ripening will be delayed in the longer summer days and the plant may be caught by frost. Conversely if a variety adapted in the north is moved to a lower

latitude it will bloom and mature too early and will not take advantage of the full growing season.

Most soybeans are harvested by a combine, which threshes as it reaps, when the plants are mature, the leaves have fallen and the seed has dried to a moisture content of about 13%, at which time the crop can be stored safely.

Uses.—*Food and Food Products.*—In Asiatic countries the soybean is grown primarily for the seed, which is used largely in the preparation of hundreds of fresh, fermented and dried food products. These various food preparations not only give flavour to the oriental diet but supply to a considerable extent the necessary protein in the diet. This has been especially true inland, where sea food is not readily obtainable. Nutritionally the protein of the soybean is similar to that of animal protein—even the amino acid analysis of soybean protein and casein are remarkably similar.

Sprouts grown from soybean seed are eaten by the Chinese as a green vegetable throughout the year, either in salads or as a cooked vegetable. They are a good source of vitamin C. Soybean "milk" is said to have been made long before the Christian era. It is used in the fresh state and also in the manufacture of bean curd or vegetable cheese *teou fu* or *tōfu* used so extensively in the orient. When soybean milk is boiled a film called *yuba* forms on the surface. This film is rich in protein and has been a popular food in China and Japan since ancient times. In cooking, *yuba* is either fried or used in soups. *Nattō* and *miso*, prepared by steaming and fermenting soybeans, are two staple foods in Japan. Red *miso* contains a large amount of salt and will keep for several years. *Shōyū* or soy sauce, a fermented product made from soybeans and parched wheat, is a very salty dark-brown liquid used extensively both in the orient and the west as a seasoning. In China and Japan soybeans are boiled in syrup and eaten as candied beans and also soaked in salt water and roasted as a nut.

In the U.S. soybeans are used as a green vegetable and are preserved by canning or freezing. The mature or dry soybeans are baked or boiled in a manner similar to navy beans. For this purpose special vegetable varieties have been developed that are bland or nutty in flavour.

The soybean was first used in the U.S. as a forage crop and for green manure. When properly harvested and cured the soybean plant makes excellent hay. Soybeans are especially valuable for plowing under before planting certain crops such as sugar cane. The first commercial processing of soybean seed to obtain oil and cake was in 1911 when a small hydraulic-press mill in Seattle, Wash., processed a shipment of beans from Manchuria. In 1915 a cotton-oil mill in Elizabeth City, N.C., using screw-press equipment, made a successful start in processing locally grown beans. In 1917 a plant in Chicago Heights, Ill., began crushing experiments using both hydraulic and screw-press equipment, although the supply of soybeans was not adequate at that time to keep the mill in operation. A corn-processing company in Decatur, Ill., began crushing soybeans in 1922. Several other processing companies built soybean mills in Decatur in the late 1920s and 1930s, and the city soon became recognized as the leading soybean-processing centre in the U.S. It was not until 1935 that the acreage of beans for processing equaled that for forage but thereafter the volume of processing increased very rapidly.

Early processing of the beans for oil and meal was by hydraulic pressing or by expeller or continuous pressing (screw presses), although by the mid-1950s most processing plants in the U.S. and Europe had changed to hydrocarbon solvent extraction, which leaves around 1% oil in the meal. The older methods left from 4 to 5% oil in the press cake. With modern extraction equipment a bushel (60 lb.) of soybeans produces about 11 lb. of crude oil and 47 lb. of meal.

Soybean oil contains from 80% to 85% unsaturated glycerides and is classed as a semidrying oil. With an iodine number (an index of drying rate) of 130 to 135 it lies between rapid-drying oils such as linseed and tung, and the nondrying oils such as peanut oil. Methods of refining have been developed that make soybean oil very good for edible products such as shortening, margarine and salad oils. The generally low free fatty-acid content of soybean oil results in a low refining loss and high yield of refined oil. Cooking oils and salad oils may be prepared by refining, deodorizing and bleaching soybean oil. By controlled hydrogenation that partly solidifies the oil, an excellent solid fat can be prepared for use in the manufacture of shortening and margarine. The amount of soybean oil used in edible products in the United States increased from 1% in the mid-1930s to, in the mid-1950s, 69% of the total oil consumed in margarine, 47% of the total in shortening and 35% of the total in other edible products including cooking and salad oils. (*See* OILS, FATS AND WAXES.)

Crude soybean oil contains from 1% to 3% lecithin—an antioxidant, emulsifier and softener used in food manufacturing and many other industries.

Soybean meal remaining after the oil has been extracted is a valuable product with many uses, particularly as a high-protein feed for livestock and poultry. In the second half of the 20th century about 95% of the soybean meal consumed in the U.S. was used for feed, supplying over 6,000,000 tons or 66% of the total vegetable protein meal consumed by livestock and poultry (*see* ENSILAGE; FEEDS, ANIMAL). Soybean flour, grits and flakes are used in bread, doughnuts, cakes and cookies; as emulsifiers and binders in sausages and related meat products; in breakfast foods, salad dressings, low-starch health foods, macaroni and noodles, confectionery products, ice creams and as whip toppings. Soybean flour proved very successful in meeting food shortages during emergencies. Cereal proteins are low in lysine content, and soybean protein supplies the deficiency. The addition of 5% soybean protein to wheat flour makes a remarkable improvement in nutritive value. It has been stated that soybean flour is one of the cheapest foods available to man when judged by the amount of protein, minerals, vitamins and energy obtainable per unit of cost. Soybean meal also is used in the brewing industry to provide nutrients for the growing yeast and to improve the flavour and body of beer. While whole soybeans are generally used in the preparation of soy sauce by fermentation, soybean meal is used by many manufacturers in its preparation by the more rapid acid hydrolysis method. Monosodium glutamate, an important condiment in the orient and in the U.S., may be prepared by acid hydrolysis of soybean meal. Glutamic acid is an essential constituent of soy sauce.

Uses in Industry.—Although food manufacture absorbs about 85% of the soybean oil consumed in the U.S., the oil has many applications in industry, one being in the production of high-grade industrial enamels, particularly white enamels for kitchen appliances and coloured enamels for automobile finishes. It is used extensively in the production of varnishes and alkyd resin paints and also is used in inks and stains, sealing and caulking compounds, linoleum and oilcloth, pharmaceuticals, cosmetics, synthetic rubber and core oils. The protein in soybean meal is readily soluble in water and when properly processed is similar to casein in viscosity, adhesive strength and many other characteristics. Soybean protein is used in many industrial applications such as adhesives, paper coatings, water-thinned paints, plastics, textile fibres, fire foam stabilizers, printing inks, fillers, core binders, stickers for agricultural sprays, linoleum, emulsifying agents, paper and textile sizing and leather finishes. In the plywood industry soybean glues represent the largest tonnage of any type of adhesive used. Industrial research is rapidly widening the industrial uses of the soybean and its products.

See also BEAN.

BIBLIOGRAPHY.—C. V. Piper and W. J. Morse, *The Soybean* (1923); K. S. Markley and W. G. Goss, *Soybean Chemistry and Technology* (1944); K. S. Markley (ed.), *Soybeans and Soybean Products*, 2 vol. (1950–51). *See* also *Bulletins* of the U.S. Department of Agriculture and State Agricultural Experiment Stations. (J. L. CR.)

SOZOMEN (*c.* 400–443), the name of a famous 5th-century church historian, HERMIAS SALAMANES (SALAMINIUS) SOZOMENUS, who came of a wealthy family of Palestine. He tells us that he was brought up under monkish influences. After studying law in Beirut he settled down as an advocate in Constantinople, where he wrote his *Ecclesiastice Historia* about the year 440. The nine books begin with Constantine (323) and come down to the death of Honorius (423). The work, dedicated to Theodosius II, has reached us only in a mutilated condition, at least half a book

being wanting. It is a plagiarism of the ecclesiastical historian Socrates (*q.v.*), but Sozomen has referred to the principal sources used by Socrates (Rufinus, Eusebius, Athanasius, Sabinus, the collections of epistles, Palladius), and has not unfrequently supplemented Socrates from them. The ninth book is drawn from Olympiodorus. Sozomen wished to present a picture in which monasticism should be brought into still stronger prominence.

Sozomen also wrote an *Epitome of History* from the Ascension of Christ to the defeat of Licinius (323) which is not now extant. (*See* his *History*, i, 1.)

For bibliography *see* the article on the church historian, SOCRATES.

SPA, a town in the province of Liège, Belgium, less than 20 mi. S.E. of Liège, famous for mineral springs first discovered in 1326. The springs have given the common name of "spa" to such resorts. The town is 850 ft. above sea level. In the 18th century it was the most fashionable resort in Europe for the medicinal use of such waters. In 1807 much of the town was burned down, and the principal buildings, the Casino and the Pouhon, are quite modern. Spa in ordinary times attracts about 20,000 visitors annually. Pop. (1961) 9,055.

SPA, CONFERENCE OF (July 5–16, 1920), the first occasion after World War I when the heads of the German and Allied governments negotiated on a formal footing of equality. The crucial questions at issue were the execution of the disarmament clauses and the coal delivery clauses of the treaty of Versailles, and in both cases the Germans were compelled to sign protocols dictated by the Allies under threat of military sanctions which were set out in the documents, and included an eventual Allied occupation of the Ruhr basin.

A German proposal for the general settlement of the reparations problem was rejected, but no general Allied counterproposal was put forward. An inter-Allied agreement was made for the allocation of prospective German payments. It was provided that France should receive 52% of sums obtained from Germany, the British empire 22%, Italy 10% and Belgium 8%, while the remaining 8% was to be divided between the other Allied powers. By a subsequent agreement among the several members of the British commonwealth 86.85% of the British empire's share was allocated to Great Britain. Another article in the Spa agreement provided for the allocation of reparation payments from Austria, Bulgaria and Hungary, and "cost of liberation" payments from Italy, Czechoslovakia, Rumania and Yugoslavia.

At this conference the Allies also came to an agreement with Germany on the trial of war criminals.

SPAAK, PAUL HENRI (1899–), Belgian statesman who played a major role in the early history of the United Nations and in promoting European unity after World War II. He was born in Schaerbeek, near Brussels, on Jan. 25, 1899, of parents who were prominent in political and civic affairs. He studied law after World War I and practised in Brussels during the 1920s. His political career began in 1932 when he was elected to the Chamber of Representatives as a Socialist member from Brussels. He served as Belgium's foreign minister from 1936 to 1938 and as prime minister from 1938 to 1939. When Belgium was overrun by German troops in 1940, Spaak escaped to France and then to England where he became foreign minister of the Belgian government in exile.

Returning to Belgium after the war he held many important government posts, including the premiership from 1947 to 1950. He also took an active part in world affairs, representing Belgium at the United Nations Conference on International Organization in San Francisco (1945); the next year in London he was elected the first president of the UN General Assembly. He later gave more attention to promoting European unity, playing a major role in negotiating trade agreements and in fostering the development of the European common market. From 1950 to 1955 he was chairman of the International Council of the European Movement. From 1957 to 1961 he was secretary-general of the North Atlantic Treaty Organization (NATO). He then returned to Belgium and served as assistant prime minister and as minister of foreign and African affairs. *See* also BELGIUM: *History*.

SPAATZ, CARL (1891–), U.S. air force officer, the leading U.S. combat air commander of World War II and first chief of staff of the independent United States air force, was born at Boyertown, Pa., on June 28, 1891. After a distinguished career as a combat pilot during World War I, Spaatz acquired rich staff and command experience between 1919 and 1942. As commander of the 8th air force in England during World War II, he initiated the American daylight bombing offensive against Germany in 1942. In Jan. 1944 he became commander of the U.S. strategic air forces in Europe and directed the daylight strategic bombing of Germany from both England and Italy until the end of the war in Europe. He moved to the Pacific in July 1945 in time to direct the final strategic bombing of Japan, which culminated in the dropping of atomic bombs on Hiroshima and Nagasaki. Spaatz became chief of staff of the newly independent air force in Sept. 1947, but finding the administrative burden not to his liking, he retired in 1948. (A. Gg.)

SPACE EXPLORATION. Space is not empty—not even those regions that lie between the stars and the planets. Throughout the vast reaches of space, matter (largely hydrogen) is scattered at the extremely low density of perhaps one atom per cubic centimetre in interstellar space and about ten particles per cubic centimetre in interplanetary space. Space is permeated by gravitational fields, traversed by an extremely broad spectrum of electromagnetic radiation and by cosmic rays, and is occupied by magnetic fields of unknown intensities and distributions. Although the scientist has deduced much about space from astronomical and other observations made at the surface of the Earth, only with the advent of the artificial satellite and the deep-space probe has he been able to discover at first hand the tremendous complexity of space phenomena.

This article is concerned with the exploration of space by those methods made possible by the large modern rocket. For the purposes of this discussion, the term space should be regarded generally as meaning those reaches of the universe lying beyond the Earth's sensible atmosphere. This definition of space is imprecise, but efforts to sharpen it become extremely complicated. For the biologist outer space means one thing, for the legal expert another, and for the astronomer something still different. The lawyer, who is concerned with the question of where existing international agreements concerning aviation should be considered to apply and where new agreements concerning astronautics should be developed, has an extremely difficult task when he attempts to develop a precise definition of space. Many of the applications of space technology apply primarily to activities on the surface of the Earth—for example, in the making of weather observations, in communications and in navigation. Even in the area of scientific exploration it becomes impossible for the scientist to separate the study of near space from the science of the Earth's atmosphere and ionosphere, and satellite investigations of the Earth's gravitational field by means of artificial satellites lead to important results in geodesy and in the study of the internal structure of the Earth. Tables I and II list some of the properties and characteristics of various regions of space.

The rocket in one form or another furnishes the only practicable means for transportation through space. Late in the 19th century the Russian mathematician Konstantin Tsiolkovski called attention to the usefulness of rockets for this purpose. Robert Goddard, a U.S. physicist, was motivated in his pioneering work on rockets, during the 1920s and 1930s, by the desire to send research instruments into the Earth's upper atmosphere. A German group, led by Hermann Oberth in the 1920s, looked to interplanetary travel as one of its principal objectives.

As these early pioneers clearly foresaw, space activities did not, indeed could not, come into being until the rocket had been further developed. Although the beginning of the space age is associated with the launching of the first Soviet Sputnik on Oct. 4, 1957, the groundwork for this event was laid by the high-altitude rocket exploration of the preceding decade. Beginning in 1945 with the launching of a WAC Corporal rocket, the United States undertook a series of scientific experiments in the Earth's upper atmosphere. For some years captured German V-2 rockets furnished the principal means of carrying on this upper-air research.

TABLE I.—*Regions of Space: Earth's Atmosphere and Near Space*

Distance from Earth's centre (Earth radii)	Altitude (km.)	Environmental characteristics						Operational aspects	
		Dynamic	Chemical	Electrical	Thermal	Meteors	Gravitational	Technical	Medical
1–1.0024	0–15	Winds Turbulence	20.9% O_2, 78% N_2 0.9% Ar, 0.03% CO_2, variable amount H_2O (volume %)	Thunderstorms	Average temperature gradient about −6.5° C./km.	Very rare	Terrestrial	Air-breathing engines	Ceiling for use of an oxygen mask: 3 km.: O_2 required 7.5 km.: ceiling for operational crew (U.S. air force) 11 km.: ceiling for all personnel (U.S. navy) 12 km.: ceiling for passengers at rest (U.S. air force) 13 km.: ceiling for pressure breathing (U.S. navy)
1.0024–1.0048	15–30	Lower region: jet streams	Ozone appears; maximum concentration at about 25 km.	None	Approximately isothermal atmosphere about −50° C.	Very rare	Terrestrial	30 km. is limit for most air-breathing engines; hypersonic ram jets may extend beyond Slow-down region of ICBM nose-cones	10–25 km.: pressurized cabin 25 km.: sealed cabin; beginning of space equivalence
1.0048–1.0126	30–80	Turbulence in region of negative temperature gradient	Above 50 km., atomic oxygen appears	D layer (daytime) NO^+ predominant ion	30 to 50 km.: +2 to 3° C./km. 50 to 80 km.: −1.5 to −4° C./km. Temp. at 75 km.: −73° C.	Most meteors disappear Accumulation of meteoritic dust	Terrestrial	Slip-flow region Slow-down region of entry gliders from space Radicals for propulsion?	Space equivalence Aerodynamic heating Cosmic radiation (maximum atmospheric secondaries)
1.0126–1.063	80–400	Hydromagnetic effects on dynamics Pressure about 10^{-8} atmosphere at 100 km.	O_2 dissociated Slight N_2 dissociation Airglow processes Aurorae	E, F_1, F_2 layers O^+, NO^+, O_2^+ major ions present	Temperature increases rapidly with altitude 100 km.: −53° C. 150 km.: +672° C. 300 km.: +1,170° C.	Meteors visible below about 100 km.	Terrestrial Perturbation of satellite orbits due to Earth's oblateness	Free molecule flow region Lower satellite orbit region; departure orbit for lunar and interplanetary probes	Space equivalence
1.063–1.157	400–1,000	Pressure about 10^{-12} atmosphere at 500 km. Lower edge of magnetosphere over equator	Helium becomes significant	O^+ ion predominates He^+ present Lowest portion of Van Allen radiation belts	Temperature becomes isothermal with altitude, but is subject to large diurnal variations	Micrometeorites and meteorites	Same as above	Manned space stations Instrumented satellites	Space conditions Corpuscular radiation
1.157–1.5	1,000–3,200	Some particles escape from Earth's atmosphere Magnetosphere	Hydrogen is predominant	H^+ present Charged particles travel along magnetic force lines Van Allen radiation belts	Diurnal effects predominate	Micrometeorites and meteorites	Same as above	Instrumented satellites	Space conditions Increasingly intense corpuscular radiation

With the development of the Aerobee, Viking and other rockets more suited to upper-atmosphere exploration, the work with sounding rockets accelerated, reaching a peak of activity during the International Geophysical Year (1957–58). Upper-air research in the Soviet Union followed a similar chronology.

In both the U.S. and the Soviet Union the development of missile technology was essential to the larger satellite and space probe applications. As preparations for the International Geophysical Year (IGY) progressed, the possibility of launching artificial Earth satellites for scientific investigations became more and more apparent. As a result, the planning committee for the IGY in 1954 passed a formal resolution calling attention to the desirability of using artificial satellites in the IGY program. Both the United States and the Soviet Union responded with announcements that they would prepare scientific satellites for launching during the IGY. The first Soviet Sputniks and the early Explorer and Vanguard satellites of the U.S. were the direct outcome.

Although the possibilities offered by artificial satellites had been apparent to scientists and engineers working in these fields, general recognition of their potentialities did not take place until the first satellite was in orbit. After the launching of the first satellites, the U.S. created the National Aeronautics and Space administration (NASA) by the Space act of 1958. Under this act NASA took over the personnel, facilities and aeronautical responsibilities of the former National Advisory Committee for Aeronautics and also assumed responsibility for the civilian space activities. NASA began functioning on Oct. 1, 1958.

Thereafter, U.S. space activities expanded tremendously. The space budget grew rapidly from its initial total of $331,000,000 in 1959 to more than $5,000,000,000 for 1964. The program encompassed research efforts in aeronautics, space science, space applications, manned space flight and research on advanced technology. In addition, the U.S. department of defense developed a program for the application of space techniques to defense uses, and by 1964 the budget for these purposes was about $1,600,000,000. Also, the Atomic Energy commission was co-operating with NASA in the development of nuclear techniques for space, while the department of commerce was playing an ever-increasing role in the use of satellites in meteorology and geodesy.

This article is divided into the following sections:

I. Astrodynamics
  A. Two-Body Motion
    1. Effect of Asymmetrical Bodies
    2. Additional Forces
  B. Three-Body Problem
  C. Spacecraft Rotations
  D. Applications
    1. Angular Motion and Controls
    2. In-Space Maneuvering
II. Launch Vehicles
    1. History
    2. Configurations
    3. Guidance
    4. Launch Facilities
    5. Operation
    6. Vehicle and Spacecraft Communication
    7. Launch Vehicle Characteristics

III. Spacecraft
- A. Function
- B. Typical Spacecraft
  1. Atmospheric Structure Satellite
  2. Interplanetary Explorer
  3. Orbiting Solar Observatory
  4. Transit
  5. Mariner
  6. Orbiting Geophysical Observatory
  7. Ranger
- C. Spacecraft Design
- D. Spacecraft Testing
  1. Acceleration and Vibration Tests
  2. Pressure and Temperature Tests
  3. Radiation Testing
  4. Guidance and Control Testing

IV. Space Science: Physical Sciences
- A. Solar Phenomena
- B. Interplanetary Phenomena
- C. The Earth
  1. Geodesy
  2. Upper Atmosphere
  3. Ionosphere
  4. Magnetosphere
- D. The Moon and Planets
- E. Astronomical Phenomena

V. Space Science: Life Sciences
- A. Origin of Bioastronautics
- B. Space Biology
  1. Environmental Biology
  2. Exobiology
- C. Problems of Man in Space
  1. Biological Space
  2. Hazards and Stresses of Space Flight
  3. Biologistics

VI. Manned Space Flight
1. Initial Programs
2. Gemini
3. Voskhod
4. Lunar Landing Program
5. Manned Spacecraft Design

VII. Applications
1. Meteorological Satellites
2. Communications Satellites

VIII. International Space Activities
1. U.S. Programs of Co-operation
2. U.S.-U.S.S.R. Relationships
3. European Regional Co-operation
4. UN Activities
5. Prospects

IX. Space Law

In reading these sections the following definitions and concepts will be useful. Rockets used to launch spacecraft into orbit are referred to as space vehicles or launch vehicles. Most of them are multi-stage vehicles. It is customary to refer to the first stage as the booster stage and to the remaining stages as upper stages.

The term spacecraft is used in a general sense to describe the assemblage and its payload that is lifted into space and injected into an orbit by a launch vehicle. If the launching rocket with its payload follows a vertical or nearly vertical trajectory, and remains within one Earth radius of the ground, then the combination is normally referred to as a sounding rocket. When the spacecraft is injected into a closed orbit about the Earth, it becomes an artificial satellite of the Earth. The orbit is described by four factors: apogee and perigee, the points farthest from and nearest to Earth, given in miles or kilometres; the inclination of the plane of the orbit to the plane of the Earth's equator, given in degrees; and the period, or time required for one circuit of the orbit. Spacecraft that are injected into trajectories leading to escape from the Earth are called space probes. These may be classified as lunar, Venus, Mars, deep-space, etc.

The sounding rocket, artificial satellite and deep-space probe provide the scientist with a powerful means for research in a number of areas. Among the important subjects of investigation by space techniques are: the space environment; Sun-Earth relations; investigation of the Moon; investigation of the planets; determination of astronomical constants; investigation into the nature of the universe; and life in space.

Each of these areas can be investigated along lines that were impossible prior to the advent of the rocket. Of particular importance is the opportunity in astronomy to observe celestial objects and phenomena in the gamma ray, X-ray, ultraviolet, infrared and radio wavelengths that cannot penetrate the Earth's atmos-

TABLE II.—*Regions of Space: Orbital Space*

Region	Distance*	Environmental conditions	Astronautic aspects
Terrestrial space (Ehricke) Magnetosphere	1–10 $r_\oplus$	Atmospheric effects on orbit in inner terrestrial space. Electrical and magnetic effects caused by Earth; especially the high-intensity radiation belt with maximum at 1.5 and 3.5 $r_\oplus$, outer limit 10–15 $r_\oplus$ on sun side Gravitational dominance of terrestrial perturbations over lunar perturbations	Manned space stations appear safe in region between relevant atmosphere (450 km.) and beginning of intense corpuscular radiation (650 km.); or beyond radiation belt (>15 $r_\oplus$) Instrumented satellites: meteorological: 1.1–6.5 $r_\oplus$ observational: 3-6.5 $r_\oplus$ (6.5 $r_\oplus$ = 24 hr. orbit) communication: 3-6.5 $r_\oplus$
Cislunar space (Ehricke) Magnetosphere (on antisolar side of Earth)	10–50 $r_\oplus$	Equivalence and finally dominance of lunar perturbations over terrestrial perturbations. Gravitational pull of Moon alternately joins and opposes Earth's pull Solar perturbation may be effective on cislunar transfer orbits Tail of Earth's magnetosphere on antisolar side	Essentially research satellites to explore environmental conditions in cislunar space Region is traversed by lunar and interplanetary vehicles Region of chemical and nuclear heat-exchanger rocket drives
Translunar space (Ehricke)	70 to ~ 300 $r_\oplus$	Outer region of geocentric space Lunar and terrestrial gravitational forces never act in opposite direction	No firm technical concepts yet on utilization of translunar space
Region in which the Moon can hold a satellite	40–45 $r_\leftmoon$	Lunar gravity field dominant Special environmental conditions near the Moon, not yet known in detail	Lunar satellites Lunar base
Inner solar system	0–1.6 A.U.†	Central force field of the Sun Intense radiation Zodiacal dust cloud	Region of nuclear heat-exchanger drives for fast missions, low-thrust drives for large payload missions
Ecosphere (Strughold)	~ 0.5–1.5 A.U.†	Liquid water belt Oxygen belt Biotemperature belt	Natural biotechnical region of solar system
Outer solar system	1.6–38 A.U.†	Central force field of the Sun Strong perturbation potential of major planets	Low-thrust drives (electrostatic or electrodynamic)
Region in which respective planet can hold a satellite	Generally greater than 150 planet radii (except Mercury)	Gravitationally, a local anomaly in heliocentric space Probably special electrical, magnetic and radiation conditions not yet known in detail	Region in which artificial satellites can be established
Region in which the planet rather than the Sun should be considered as centre body for an arriving or departing body	Venus: ~ 100 r_φ Earth: ~ 144 $r_\oplus$ Moon: ~ 34 $r_\leftmoon$ Mars: 170 r_σ Jupiter: 690 $r_{2\!\!4}$	Gravitationally defined as the region in which the Sun is a lesser perturbative force in the planetocentric co-ordinate system than the planet is in the heliocentric co-ordinate system. A co-ordinate system is the frame of reference in which the orbit of the spacecraft (or comet, etc.) is measured. Its centre is in the centre of the planet or the Sun, respectively, for the two above mentioned co-ordinate systems	Region in which transfer from heliocentric to planetocentric orbits (capture) or the reverse (escape) takes place. The transfer orbit between a closed planetocentric (satellite) orbit and a heliocentric (elliptic) orbit is a hyperbola. The limit of the activity sphere is reached when the hyperbolic orbit merges with the heliocentric orbit

*$r_\oplus$ = Earth radius; $r_\leftmoon$ = Moon radius; r_φ = Venus radius; r_σ = Mars radius; $r_{2\!\!4}$ = Jupiter radius; each measured from the centre of the respective planet.
†1 A.U. = 1 astronomical unit = mean distance from sun to earth, measured from the centre of the sun.

phere to reach observatories on the ground. Especially exciting is the possibility that space biology may contribute significantly to efforts to develop a comprehensive biological theory. This collection of researches carried out in space or by means of space techniques is referred to as space science.

The scientific applications and technological advances made during the first five years of the space age were truly remarkable, but perhaps the most important aspect of this early phase of the space age was the establishment of a new capability in human endeavour. Out of these early activities man could develop the knowledge and techniques; trained and experienced personnel; assembly, launch, tracking and telemetering facilities; industrial complexes and teams; flight "hardware," and the ability to operate in space that go to make up a total space capability. With this capability, man can extend his domain beyond the earth as he once did over the land and the seas, into the air and into the ocean depths. (H. E. NL.)

I. ASTRODYNAMICS

Astrodynamics is the branch of astronautics that deals with the mathematical theory of the motions of spacecraft. These motions may be separated into two groups: the translations of the centre of mass of the spacecraft, and the rotations of the spacecraft about its centre of mass. The translations describe the flight path or trajectory or orbit of the spacecraft; the rotations describe its wobbling or angular motion.

In each of the two groups there are three motions. The translational motions are (1) surge, a motion generally along the velocity vector or direction of travel of the spacecraft; (2) heave, a motion perpendicular to the surge and generally in the direction of the gravity vector; and (3) swerve, a motion perpendicular to the surge and the heave. The rotational motions are (1) roll, a spinning motion around the velocity vector of the spacecraft, *i.e.*, spin about the surge direction; (2) pitch, an angular motion about the swerve direction; and (3) yaw, an angular motion about the heave direction. Thus, a spacecraft has a total of six motions which must be completely accounted for if its performance is to be known accurately, or if its position and attitude in space are critical to accomplishing its basic mission.

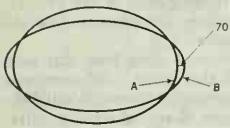


FIG. 1.—EQUATORIAL BULGE IS 21.4 KM. DISCREPANCY BETWEEN OBSERVED SHAPE (A) AND SHAPE OF HYDROSTATIC EQUILIBRIUM (B) IS 70 M.

The theory and applied mathematical methods of astrodynamics depend heavily on the methods of celestial mechanics, exterior ballistics and the flight dynamics of aircraft and missiles.

A. Two-Body Motion

Sir Isaac Newton stated that a sphere, such as the Earth, which is homogeneous in concentric layers, attracts an exterior point mass, such as a spacecraft, as if all of the mass of the sphere were concentrated at its centre. Furthermore, his universal law of gravitation states that any two particles of mass in the universe attract each other with a force that varies directly as the product of their masses and inversely as the square of the distance between them. Thus, a statement of a problem concerning the translational motions of a spacecraft takes the form of a set of equations of motion, which are ordinary differential equations of the second order. A proper understanding of astrodynamics, therefore, requires a knowledge of this branch of mathematics.

For the case of two bodies, such as a spacecraft and a simple spherical Earth, a solution to the differential equations of motion is known. Two types of translational motion are possible. The first is an elliptical or circular orbit of the spacecraft around the Earth. The second is a flight path away from the Earth, a path that may be hyperbolic or parabolic in shape.

While these simple solutions are quite useful in many cases, they are not correct, strictly speaking, since the Earth is not a simple spherical body. It is known, for example, that the Earth has a bulge at its equator. (*See* fig. 1.) This bulge, although amounting to only 21.4 km., seriously affects the motion of a spacecraft. For example, a spacecraft in a nominally circular polar orbit experiences a greater attraction as it passes over the equator and thus it is pulled toward the Earth. Also, its velocity is increased. When the spacecraft passes over the poles, it feels a lesser attraction and therefore it moves out farther from the Earth and its velocity decreases. The bulge at the equator, therefore, has caused a surging and heaving motion of the spacecraft as well as a variation in its altitude. Also, for a spacecraft in a 500-mi. orbit, the perigee altitude variation is approximately 5 mi. in two months.

For a spacecraft in an inclined orbit the bulge also causes the major axis of the orbit to rotate in the orbital plane and the orbital plane to rotate in space about the Earth's spin axis. The size of these two rotations depends upon the exact inclination of the orbital plane to the equator of the Earth. At an inclination of 63.4° there is no rotation of the major axis. For angles less than 63.4° the major axis rotates in the same direction as the satellite motion around the earth. For angles greater than 63.4° the major axis rotates in the opposite sense. There is no rotation or precession of the orbital plane when the inclination is 90° (polar orbit). Maximum precession, however, occurs for an equatorial orbit.

1. Effect of Asymmetrical Bodies.—Previous work in celestial mechanics on the motion of the innermost satellite of Jupiter and the innermost satellite of Saturn has been helpful to the astrodynamicist in understanding the effect of the bulge on the orbits of near-Earth spacecraft. By actually observing the variations in the orbits of near-Earth satellites the astrodynamicist has discovered that the Earth has, in fact, an extremely complicated shape. Not only is there a lack of symmetry between the northern and southern hemispheres but there is also an asymmetrical distribution of mass about the Earth's spin axis. The latter means that the equator of the Earth is not circular but is elliptical in shape (approximately 100 m.). It is therefore necessary to represent the gravity field of the earth in terms of a potential expressed in spherical harmonics, which include the zonal harmonics (wrinkles in latitude), the sectorial harmonics (wrinkles in longitude), and the tesseral harmonics (wrinkles in both latitude and longitude). This potential may be expressed as

$$U = -\frac{G\,M}{r}\sum_{n=0}^{\infty}\sum_{m=0}^{n}\left(\frac{a_e}{r}\right)^n J_{nm}\,P_{nm}\,(\sin\phi)\,\cos m\,(\lambda-\lambda_{nm})$$

where

$$P_n(\sin\phi) = \frac{1}{2^n\,n!}\,\frac{d^n\,(\sin^2\phi-1)^n}{d\,(\sin\phi)^n}$$

$$P_{nm}(\sin\phi) = \cos^m\phi\,\frac{d^m\,P_n\,(\sin\phi)}{d\,(\sin\phi)^m}$$

In these equations, G = universal gravitational constant; M = mass of Earth; r = distance from origin of co-ordinates to computation point; a_e = equatorial radius of Earth; ϕ = geocentric latitude; and λ = geocentric longitude.

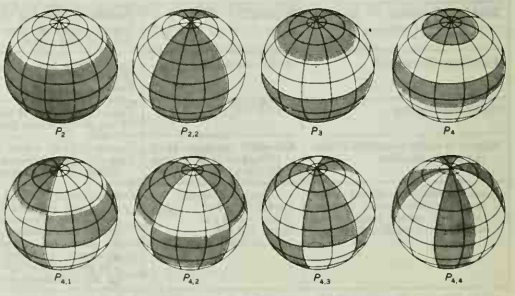


FIG. 2.—EFFECT OF SPHERICAL HARMONICS ON ELEVATION OR DEPRESSION OF THE GEOID; SHADED AREAS ARE ELEVATED, LIGHT AREAS DEPRESSED

Astronaut Edward H. White as he maneuvered in space more than 100 mi. above the surface of the earth. The 25-ft. "umbilical" line which provided him with oxygen and tethered him to the spacecraft may be seen in this photograph

White photographed from inside the spacecraft by McDivitt. The visor of White's helmet was gold-coated to protect him from unfiltered solar rays; in one hand he held the self-propulsion device which enabled him to maneuver while floating in space

BOTH PHOTOGRAPHS BY COURTESY OF NASA

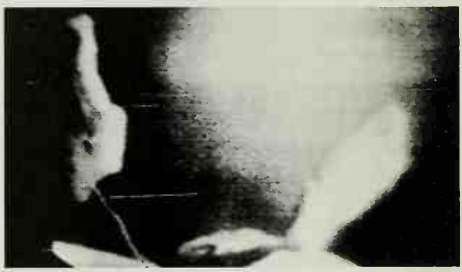
THE FLIGHT OF GEMINI 4

The flight of Gemini 4 in June 1965 lasted 97 hours and 56 minutes, during which the spacecraft containing United States astronauts James A. McDivitt and Edward H. White II made 66 orbits of the earth. During the third orbit, as Gemini 4 swept toward the Pacific coast of the United States, astronaut White emerged from the spacecraft and floated in space for about 22 minutes. Photographs of the first man to float in space, Soviet cosmonaut Alexei Leonov, may be found on the following plate

SPACE EXPLORATION

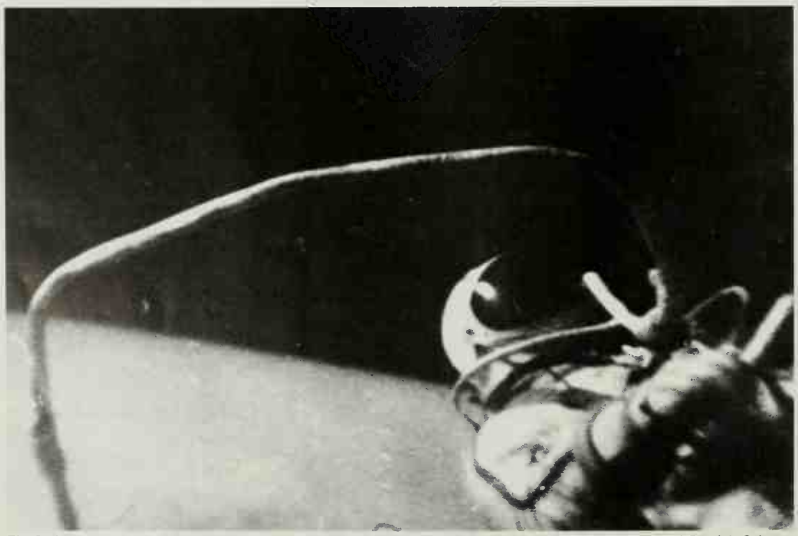
Soviet cosmonaut Alexei A. Leonov emerging from the two-man Voskhod 2 capsule. The Russian spacecraft transmitted to earth live television pictures of Leonov's venture into space. This photograph and that at right above were made directly from a television screen on the ground

Cosmonaut Leonov pushing off into space, attached to the capsule by a 16-ft. "umbilical" line which supplied him with oxygen. Unlike U.S. astronaut White, Leonov did not have a self-propulsion device



Leonov, free of the capsule, floating upside down at the end of his tether; at right may be seen the open hatch of Voskhod 2. Leonov remained in space for about ten minutes

THE FLIGHT OF VOSKHOD 2, MARCH 1965



The Soviet cosmonaut floating in space 107 mi. above the earth. During their flight, cosmonauts Leonov and Pavel I. Belyayev completed 17 orbits of the earth

ALL PHOTOGRAPHS FROM SOVFOTO

In cases where $m = 0$ the harmonic is called a zonal harmonic, and its values are constant along parallels of latitude. The second zonal harmonic is the term (P_2 term) which describes the equatorial bulge. As indicated in fig. 2, this harmonic adds elevation to the equatorial regions of the earth and subtracts it from the poles. The fourth zonal harmonic (P_4 term) continues this work, adding elevation both at the equator and at the poles but slightly depressing the middle latitudes. The proper combination of the second and fourth zonal harmonics will produce a figure that coincides with an ellipsoid within 1 m. Almost all common representations of the earth's figure consider only the second and fourth zonal harmonics. Higher-order even zonal harmonics add to the wrinkles on the earth, remaining symmetric with respect to the equator. The third order zonal harmonic (P_3 term) is the so-called pear-shape term. This term, as with any odd order harmonic, has an opposite effect in the northern and southern hemispheres.

Where $m = n$ the harmonic is called a sectorial harmonic and has values of consistent sign along lines of constant longitude; however, these values change in magnitude as the latitude changes. In addition, for any given latitude the value varies in magnitude and direction as the longitude changes. The $P_{2,2}$ term is the major term indicating that the earth does not have rotational symmetry, and is referred to as the elliptic equator term.

In those cases where $m \neq 0$ and $m \neq n$, the harmonic is called a tesseral harmonic and maintains a value of consistent sign within a tessera, changing sign and magnitude as functions of both latitude and longitude. A tessera is a curvilinear rectangle, in this case bounded by lines of constant longitude and latitude.

Although a satellite may be tracked to an accuracy of 0.1 mi., the lack of knowledge of the gravity field permits prediction of position to an accuracy of only 0.25 mi. per 6 hr. or 1 mi. per day. Thus, even for two relatively simple bodies like a spacecraft and the Earth, the translational motions are complicated because of the complexity of the gravitational force causing them. This same situation also exists for other bodies in space. The Moon, for example, is known to be lopsided, and therefore a spacecraft orbiting the Moon will experience similar variations in its orbit.

2. Additional Forces.—In addition to the gravitational force, a spacecraft in its flight around the Earth experiences a drag force due to impact with molecules of air. Furthermore, radiation pressure and a particle wind, originating in the Sun, exert some pressure on the spacecraft and affect the orbit. Additional forces can arise due to the magnetic field of the Earth, as its lines of force are cut by the satellite. All of these forces also must be included in the differential equations for the motion of a spacecraft.

The basic effects of some of these forces are known. The air drag, for example, causes an elliptic orbit eventually to become circular and a circular orbit to decrease in size, with the result that the satellite reenters the atmosphere and either burns up or strikes the Earth. Solar radiation pressures can cause large changes in the altitude of perigee.

B. Three-Body Problem

The simple two-body theory when applied by Newton to the motion of the planets about the Sun verified the excellent observations of Tycho Brahe and the generalizations of Johannes Kepler, except that small variations or perturbations in the orbits were noted. These perturbations in the orbit of one planet about the Sun occurred whenever another planet came near. In a similar manner the orbit of the Moon about the Earth demonstrated appreciable perturbations due to the gravity field of the Sun. It was clear, therefore, that all bodies in space influence all other bodies to some extent, and thus many complicated gravity fields must be considered simultaneously. Even for the relatively simple case of only three bodies, there is no general solution for the equations of motion (*see* Celestial Mechanics).

The situation for a spacecraft is particularly complicated since during some portions of its flight it may be under the strong influence of different bodies and during other portions under weak but simultaneous influences of many bodies. Although the various external forces and gravity fields acting upon a spacecraft may be known approximately and the appropriate equations for the translational motion of the satellite may be set down, no general solutions to the equations are available. Thus the translational motions of the spacecraft would be unknown, were it not for the development of the high-speed computer. The application of the computer by the ballistics scientist to solve projectile trajectory problems has provided the astrodynamicist with an extremely valuable tool that enables him to integrate numerically the complicated equations of motion and, therefore, to arrive at a knowledge of the complete translational motions of a spacecraft. This direct numerical integration is called the special perturbations technique.

If one has a considerable knowledge of the position of the spacecraft as a function of time, it is possible to use the high-speed computer to find special curves which can best represent the observed motions. This process is known as the general perturbations technique.

C. Spacecraft Rotations

Spacecraft injected into orbit without rotation and subject to no external torques will not rotate; the roll, pitch and yaw will all be zero. For spacecraft injected with rotation, the total angular momentum remains a constant if no external torques are present. This characteristic of conservation of angular momentum for spacecraft is fundamental to understanding their rotational performance.

A rotating spacecraft is often subject to small external damping torques, such as magnetic torques arising from the magnetic field of the Earth and air damping torques. There may also be small internal damping torques arising from the movement of components, such as wires, fluids and other masses. The effect of these torques is to cause the spacecraft to change its axis of rotation until it is finally rotating about its axis of maximum moment of inertia. The angular momentum is unchanged when only internal damping is involved. Therefore, spacecraft which are spin-stabilized are designed not as elongated bodies like classical projectiles but rather as somewhat stubby shapes that will fly stably about their initial rolling axis.

Often it is necessary to control actively the angular motion of spacecraft. Again, the fundamental underlying theory is the conservation of angular momentum. The primary control devices are generally small jets to provide torques about the three spacecraft axes. Sometimes internal inertia wheels are used to change the distribution of the angular momentum about each axis.

A special case of the passive control of the angular position of a spacecraft is the use of the gravity gradient torque. This technique is based on the fact that an elongated body in a near-Earth orbit experiences a significant torque due to the difference in the gravity force acting on different parts of spacecraft located at

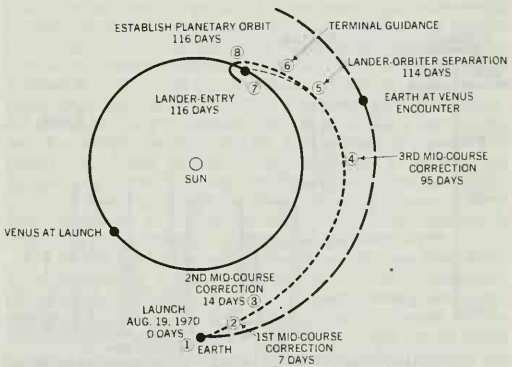


FIG. 3.—TYPICAL TRAJECTORY TO VENUS WITH MANEUVERING POINTS INDICATED

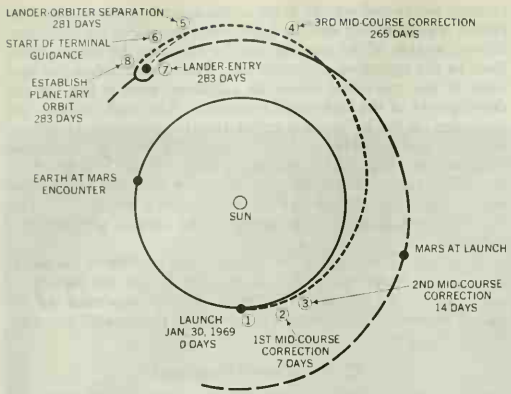


FIG. 4.—TYPICAL TRAJECTORY TO MARS WITH MANEUVERING POINTS INDICATED

different distances from the centre of the Earth. This torque causes the spacecraft to point its long axis toward the Earth. If, in addition, there exists some damping torque, the spacecraft will be stably Earth-oriented. The gravity stabilization technique is quite reliable because of its simplicity, and there is no limit to its lifetime. It is of special interest for spacecraft used in communications, meteorology, reconnaissance and mapping of the Earth and other planets.

D. Applications

Because the translational motions of a near-Earth satellite are greatly influenced by the complex gravity field of the Earth, and also because there is a need to know the exact locations of satellite tracking stations with respect to each other and with respect to the centre of mass of the Earth, it was recognized that a satellite could be used to determine these quantities. This knowledge, in turn, would make possible both a better understanding of the translational motions of a satellite and an accurate prediction of the orbits of future satellites. Thus, a special geodetic satellite, ANNA (army, navy, NASA and air force), was conceived and placed into orbit. Various precision tracking techniques, such as optical tracking, doppler tracking, transponder tracking, interferometric tracking and laser tracking, were used to determine: (1) the gravity field of the Earth; and (2) the location of the various tracking stations in geocentric co-ordinates. This was a case, therefore, of a special satellite being used to determine basic astrodynamical information which in turn could be employed to enable a better prediction of spacecraft motion.

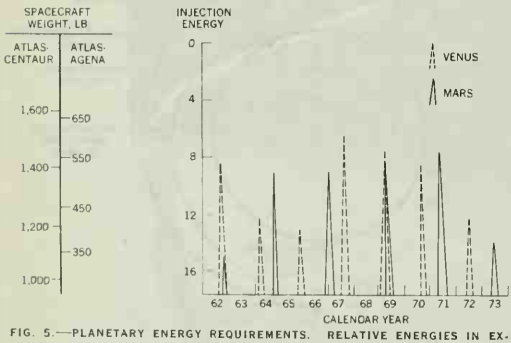


FIG. 5.—PLANETARY ENERGY REQUIREMENTS. RELATIVE ENERGIES IN EXCESS OF ESCAPE ENERGY ARE SHOWN FOR MARS AND VENUS FLIGHTS AT MOST OPPORTUNE TIMES

Not only is the gravity field of the Earth of interest but also there is considerable interest in determining the gravity fields of the Moon and other planets to enable accurate satellite flight operations in their vicinities. Consider, for example, the motion of a spacecraft traveling from the Earth past Venus. Relative to the Earth the spacecraft travels on a hyperbolic orbit and initially the Earth's gravitational field predominates. However, as the spacecraft moves outward, the Moon's gravitational field becomes dominant over that of the Earth. After passing beyond the Earth-Moon system, the influence of the gravity field of the Sun becomes the primary controlling factor. During its long voyage to Venus (approximately 100 days), the spacecraft follows an elliptical orbit relative to the Sun, subject to small perturbations produced by other planets in the solar system. However, as the spacecraft approaches Venus the gravity field of Venus takes on major importance. Although there is no solution for the translational motions of a spacecraft on such a voyage, the equations of motion may be numerically integrated on a computer and thus the flight path may be obtained (*see* fig. 3 and 4).

In addition, it is necessary to consider the launching energy required to reach a planet when a favourable opportunity is available. As indicated in fig. 5, the favourable opportunities for Mars occur about every 26 months and for Venus about every 19 months. The launching energy, however, is different at each opportunity because of the difference in the geometrical situation. Even at a favourable opportunity there is only a short period (called a launch window) when the energy remains within feasible bounds. It has been shown mathematically that the lowest energy requirement results when it is possible to use what is known as a Hohmann transfer flight path. This simple space maneuver is illustrated in fig. 6. To transfer from the smaller circular orbit to the larger, sufficient velocity is first added in a direction along the orbit to produce a new elliptical orbit with an apogee equal to the radius of the larger orbit. After the satellite has passed along this elliptical orbit and has reached the larger orbit, it is again necessary to add velocity along the orbit to produce the circular orbit at that altitude.

There are also minimum energy transfers for non-coplanar orbits. However, the additional energies are generally quite large. Some of the energy limitations in launching out of the ecliptic plane are shown in fig. 7. The rays represent inclination to ecliptic plane. The circles represent distance from the Sun in astronomical units (A.U.). For example, a 250-lb. spacecraft launched in the ecliptic plane (0°) can almost reach the Sun (0.08 A.U.). However, if this same spacecraft is launched 37° out of the ecliptic plane it only travels halfway to the Sun.

A voyage of a spacecraft represents a significant breakthrough in the science of astrodynamics. Heretofore, astronomers had only the planets and their motions as a basis for understanding the various gravity fields and the fundamental distances in the solar system. Now man is able to use the spacecraft and their motions in conjunction with those of the planets to obtain a better knowledge of many of the physical constants governing the motion of the planets. For example, a spacecraft flying past the Moon can be used to evaluate both the distance to the Moon and the lunar mass. Similarly, a spacecraft passing near Venus gives trajectory data that can be used to obtain an entirely new estimate of the mass of Venus and the distance from the Earth to Venus. This information, in turn, enables us to obtain a new value for the astronomical unit, which is the fundamental factor for understanding all distances in the solar system and beyond. As a result of the flight of Mariner 2 to Venus and radar observations of the planet, the astronomical unit is estimated to be 149,599,244

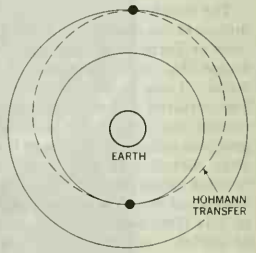


FIG. 6.—SIMPLE SPACE MANEUVER, THE HOHMANN TRANSFER, IN WHICH AN ELLIPTICAL TRANSFER PATH IS EMPLOYED, WITH ENERGY IMPULSES AT PERIGEE AND APOGEE

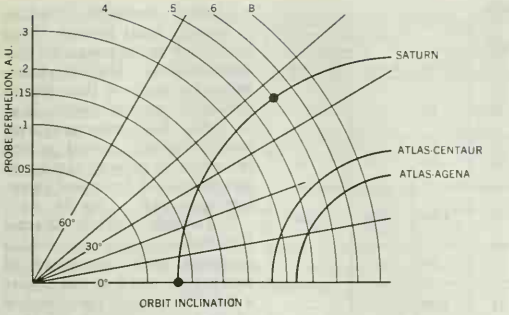


FIG. 7.—NECESSARY BOOSTER PERFORMANCE FOR VARIOUS LAUNCHING VEHICLES TO LAUNCH A 250-LB. SPACECRAFT TOWARD THE SUN AT DIFFERENT ANGLES OUT OF THE EARTH'S ECLIPTIC PLANE

km., accurate to within ± 278 km. (*see* Table III).

1. Angular Motion and Controls.—Some examples of the angular motion of spacecraft are as follows: In the case of the Transit navigation satellite it was desired that the satellite not spin in its orbit. However, in order to launch the satellite it was necessary to spin-stabilize the last stage of the launch vehicle. As a result the satellite was placed into orbit spinning. To stop the spin small weights were placed at the satellite's equator. When released, these weights payed out on strings and applied a counter-rotating torque to the satellite. In this way most of the angular momentum of the spacecraft was transferred to the weights. The weights were then released into space carrying their share of the spacecraft's angular momentum with them, and the satellite was left with a very small spin.

TABLE III.—*Values for the Astronomical Unit*

Mariner 2	149,599,244 ± 278 km.
Goldstone*	149,598,845 ± 250
Millstone*	149,597,850 ± 400
Jodrell Bank*	149,599,800 ± 5,000
U.S.S.R.*	149,599,500 ± 800

*Radar determinations

The specifications for the Tiros weather satellite called for a spacecraft which was spin-stabilized and from which pictures of the clouds could be taken, but there could be no pitching or yawing motion. Therefore, internal viscous damping was provided; the viscous torque caused the wobbling motion to damp out and gave a stable rolling motion about the axis of maximum moment of inertia. This type of passive angular control has been extremely successful.

Tiros also furnishes an excellent example of a simple technique for changing the orientation of the spin axis of a spacecraft. When a change is desired, an electromagnet in the satellite is activated. This magnet then tries to align itself with the local magnetic lines of force of the Earth and in doing so applies a torque that causes the spacecraft to precess, thereby changing the orientation in space of its spin axis. The electromagnet is turned off once the spacecraft has reached the desired new orientation.

In the case of the Orbiting Astronomical observatory (OAO), inertia wheels will be used to vary the distribution of the angular momentum of the spacecraft about each of its three principal axes. By redistributing the angular momentum from the spacecraft to and from these spinning wheels, it was possible to control the pointing accuracy of the telescope to fractions of a second of arc.

2. In-Space Maneuvering.—Maneuvers in space are quite different from those on earth. In space a body moves in a curved line, the curvature depending upon the instantaneous location vector and velocity vector of the spacecraft. For example, to apply a trajectory correction (*i.e.*, a mid-course correction) to the Mariner 2 spacecraft, it was necessary to know the future time at which the correction would be made. Furthermore, it was necessary to know accurately the location and velocity of the spacecraft at that time. The change in the velocity vector to place the spacecraft in the new trajectory had to be predicted so that the spacecraft would just miss Venus, which was also moving on a curved path. In addition, the angular orientation of the spacecraft in space had to be known and had to be changed to a new orientation so that the corrective force would be applied in the proper direction. Without going into the details of computation, timing and execution, the complexity of such a maneuver is evident (*see* fig. 5 and 6). Mariner 2 originally was on a trajectory that would have missed Venus by 233,000 mi. After the mid-course correction it missed Venus by 21,600 mi., which was on the safe side and only 11,600 mi. off the intended "bull's-eye miss" of 10,000 mi. *See* also AERODYNAMICS; BALLISTICS: *Exterior Ballistics;* GEODESY. (Jo. D. N.)

II. LAUNCH VEHICLES

Launch vehicles are powered heavier-than-air craft that are used to place a secondary craft or payload into ballistic trajectory. They are closely akin to their predecessors, the guided missiles or military rockets, which were fashioned to deliver warheads against military targets. Launch vehicles, however, must be capable of greater flexibility to perform a variety of tasks.

1. History.—Scientific uses for rockets were suggested by Konstantin Tsiolkovski of Russia in his treatise on space travel (1903). Robert Goddard of the United States launched the first rocket vehicle employing a liquid-fueled rocket engine on March 16, 1926. The fast technological growth in aeronautics during the 1930s and 1940s was effectively adapted to rocket technology to produce the V-2 in Germany (1938–45), the first supersonic aircraft, X-1 in the United States (1947), and a series of ballistic missiles such as Corporal, Thor, Atlas and Titan. The perfected state of guidance and control systems was a most essential part of this rapid utilization of rocket power. (A detailed history of rocketry will be found in the article ROCKETS AND GUIDED MISSILES.)

2. Configurations.—Launch vehicles usually consist of one or more rocket-powered stages utilizing precise guidance systems in conjunction with control systems that actuate steering mechanisms. Small sounding rockets and upper stages of larger vehicles are sometimes stabilized in a fixed orientation by spin imparted from auxiliary power sources or by means of canted aerodynamic fins.

The airframe of a launch vehicle is closely integrated with the propellant tanks, which make up most of the vehicle's volume. Propulsion is most commonly supplied from rocket motors operating on the reaction forces produced from the high-pressure combustion of liquid fuel and liquid oxidizer combinations or from solid combustibles. Propulsion produced from the combustion of a fuel with the oxygen of atmospheric air, such as the ramjet, is not suitable for application to space launch vehicles.

3. Guidance.—Guidance systems are composed of several basic elements which provide: (1) continuous flight information for the determination of vehicle position, velocity and angular orientation; and (2) computations based on this information for appropriate control commands. Guidance elements can be located both in the vehicle and on the ground. When all guidance elements are in the vehicle the system is referred to as "all-inertial," and guidance accuracy depends upon the precision of inertial sensors that provide the reference for measurement of position and orientation. These systems, being insensitive to external influences (*e.g.*, radio jamming), are usually employed in military missiles. Nonmilitary vehicles more commonly employ ground-based equipment in conjunction with the airborne elements in order to increase flexibility, performance and accuracy.

4. Launch Facilities.—Final launch vehicle flight preparation is accomplished at a range which provides checkout, propellant, photographic, calibration, tracking, telemetry, timing and safety services. Ranges in use in the 1960s included: (United States) Atlantic missile range, Cape Kennedy, Fla.; Pacific missile range, Lompoc, Calif.; White Sands missile range, N.M.; Wallops station, Wallops island, Va.; (France) missile centre and test range near Arcachon, southwest of Bordeaux; Colomb-Béchar, Alg.; Le-

TABLE IV.—*Summary of U.S. Launch Vehicles*

Vehicle	Stage or Engines	Propellant	Thrust (1,000 lb.)	Diameter* (ft.)	Height† (ft.)	Payload (lb.)		
						300 NM‡	Escape	Mars/Venus
Scout	1. Algol II-A 2. Castor I-ES 3. Antares X-259 4. Altair X-248	Solid Solid Solid Solid	86§ 64 23 3	3.3	72	240	—	—
Delta	1. Thor (DM-21) 2. AJ10-118 3. Altair X-248	LOX/RP‖ WIFNA/UDMH¶ Solid	170§ 7.7 3	8	90	800	105	90
TAD (Thrust Augmented Delta)	1. Thor (DM-21) plus three XM-33♀ 2. AJ10-118 3. Altair X-248	LOX/RP‖ Solid WIFNA/UDMH¶ Solid	170§ 162§ 7.7 3	8	90	1,000	150	120
Thor-Agena B	1. Thor (DM-21) 2. Agena B	LOX/RP‖ IRFNA/UDMHδ	170§ 16	8	76	1,600	—	—
TAT (Thrust Augmented Thor-Agena)	1. Thor (DM-21) plus three XM-33♀ 2. Agena B	LOX/RP‖ Solid IRFNA/UDMHδ	170§ 162§ 16	8	76	2,200	—	—
Atlas-Agena D	1. Atlas D 2. Atlas D (sustainer) 3. Agena D	LOX/RP‖ LOX/RP‖ IRFNA/UDMHδ	368§ 80 16	10	91	5,950	950	600
Centaur	1. Atlas D 2. Atlas D (sustainer) 3. Centaur (Two RL-10)	LOX/RP‖ LOX/RP‖ LOX/LH○	368§ 80 30	10	100	8,500	2,300	1,300
Atlas D		LOX/RP‖	368§	10	72	—	—	—
Titan II	1. XLR-87 2. XLR-91	Storable Storable	430§ 100	10	90	approx. 7,000	—	—
Saturn I	1. S-1 (Eight H-1) 2. S-IV (Six RL-10)	LOX/RP‖ LOX/LH○	1,500§ 90	21.6	120	15,000	—	—
Saturn 1B	1. S-1B (Eight H-1) 2. S-IVB (One J-2)	LOX/RP‖ LOX/LH○	1,500§ 200	21.6	142	35,000	—	—
Saturn V	1. S-IC (Five F-1) 2. S-II (Five J-2) 3. S-IVB (One J-2)	LOX/RP‖ LOX/LH○ LOX/LH○	7,500§ 1,000 200	33	281	220,000	95,000	70,000

*Maximum, excluding fins. †Less spacecraft. ‡Ballistic trajectory; range, 300 naut. mi. §At sea level. ‖Liquid oxygen and kerosene. ¶White inhibited fuming nitric acid and unsymmetrical dimethylhydrazine. ♀"Strapon" units. δInhibited red fuming nitric acid and unsymmetrical dimethylhydrazine. ○Liquid oxygen and liquid hydrogen.

Manned spacecraft tracking is done by radar because of the usually low orbits associated with these missions. The Deep Space network consists of three instrumentation stations with large steerable "dish" antennas approximately equally spaced around the world to provide continuous surveillance of lunar and planetary spacecraft. (*See* fig. 8.)

7. Launch Vehicle Characteristics.—A single-stage vehicle operating under conditions of zero gravity and drag will acquire a velocity gain in the direction of thrust according to the classic rocket equation, $\Delta v = Ig \cdot \ln (m_i/m)$, where Δv is the ideal stage velocity increase; I is the specific impulse of the propellant combination used in the rocket motor and is equal to rocket thrust divided by propellant flow; g is gravitational acceleration (standard); m_i is the stage mass at initiation of rocket thrust; and m is the stage mass at rocket thrust termination. The quantity Ig, possessing the units of velocity, is the "effective exhaust velocity" of the rocket. It serves as a measure of propellant effectiveness. Characteristic values for Ig extend from 7,000 to 14,000 ft. per second; values in the 9,000–10,000 range are commonly attained in practice. The masses m_i and m contain the inert or tare weights of the payload or spacecraft. Elimination of the latter results in m_i/m being the mass ratio of the bare stage, this term then being a measure of the structural efficiency of the stage designs. Values for m_i/m of 5 to 10 are common. Lower values are associated with rigid propellant tank designs and the higher numbers with pressure-stabilized designs.

Because of the logarithmic character of the rocket equation, it is convenient to use a logarithmic scale for the payload on these performance charts. For multi-stage vehicles the ideal velocity is the sum of the ideal velocities of the stages. Thus,

$$V_i = \{Ig \cdot \ln (m_i/m)\}_1 + \{Ig \cdot \ln (m_i/m)\}_2 + \ldots$$

In actual practice, velocity is lost because of gravitational and aerodynamic forces. These losses are heavily dependent upon the characteristics of the vehicle and the particular trajectory flown, and the computation of them is complicated. (R. B. M.)

vant Island range, south of Toulon; Papeete range in the South Pacific, Tuamotu archipelago; (U.S.S.R.) Kapustin Yar, southeast of Volgograd; Aral'sk on the Aral sea; Franz Josef Land in the Arctic; Mirny in Antarctica; (Canada) Fort Churchill; (Japan) Akita rocket range, in northwestern Honshu; Kagoshima space centre, southern Kyushu; (Australia) Woomera.

5. Operation.—Vehicles are usually launched vertically and so maneuvered in roll, pitch and yaw as to produce the proper final trajectory. Multi-stage vehicles can conveniently make use of coasting periods or parking orbits to change the flight direction in pitch. Trajectory analysis is a highly specialized art requiring extensive use of trained personnel and automatic computation.

6. Vehicle and Spacecraft Communication.—Range-based facilities usually provide for the required vehicle tracking and data acquisition (telemetry) associated with launches. Vehicles spending long periods of time in powered flight or in parking orbits frequently need tracking and data services from bases located away from the launch sites. These bases are linked to a common information centre by a radio and landline network.

The demands for data and tracking information over extended periods of time are met by several established networks. NASA organized three such networks for nonmilitary missions: the Earth Satellite network, the Manned Space Flight network and the Deep Space network. Responsibility for the development, implementation and operation of the major tracking and data acquisition networks for the manned and unmanned satellite programs was assigned to the Goddard Space Flight centre, Greenbelt, Md. The Jet Propulsion laboratory, Pasadena, Calif., is responsible for the Deep Space network. Stations for the Earth Satellite network were established at 14 locations on 5 continents. The Smithsonian Astrophysical observatory, Cambridge, Mass., operates a worldwide optical network in support of certain scientific satellites.

Primary unmanned satellite tracking is based to a large extent upon interferometric systems operating in a radio frequency band of 136–137 mc. Telemetry data are acquired by means of high-gain antenna systems located at data readout stations. The telemetry information is usually recorded on magnetic tape.

III. SPACECRAFT

During the first eight years of the space age more than 300 spacecraft were successfully orbited by the United States and the U.S.S.R. Some of these are listed in Tables V and IX. The number of successful launchings increased steadily. The percentage of successes for the U.S. rose from 29% in 1958 to about 90%.

A. Function

The purpose of the spacecraft is to carry a cargo into space, and sometimes to bring the cargo back to Earth. The cargo may be instruments or it may be man with his instruments. In this section only automated spacecraft will be discussed; *see* also *Manned Space Flight*, below.

The spacecraft must provide the environment and services required for the instrumentation ("instrumentation" here refers to the equipment associated with experiments, measurements, etc., to be made by the spacecraft; *e.g.*, radiation sensors, television cameras and micrometeor counters). With respect to environ-

ment, it is necessary to maintain the proper temperature, pressure, vibration level, rigidity and thermoconductivity. With respect to services, it is necessary to supply: (1) a structural platform; (2) power, both electrical and propulsive; (3) communications; (4) a diagnostic monitoring system for determining the condition of the spacecraft; (5) a command centre for channeling various command signals to the various types of instrumentation aboard; and (6) a control centre for guidance and navigation of the spacecraft.

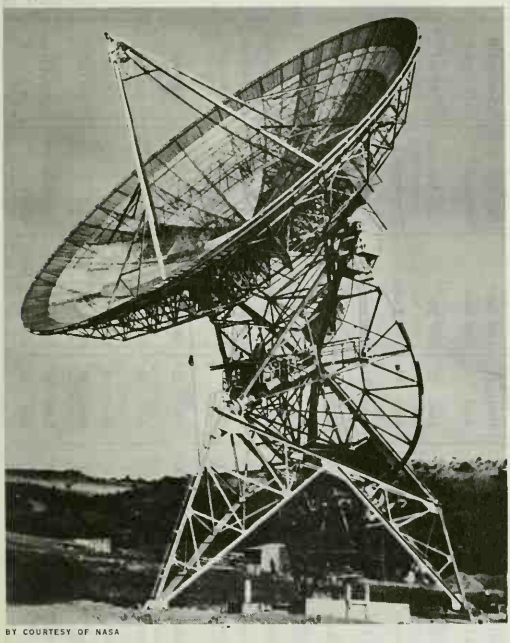
In general the environment and services required by instrumentation are quite similar from one spacecraft to another. However, the instrumentation itself may differ widely in character and function. At one extreme may be a scientific satellite carrying delicate and sophisticated instrumentation for precision measurement and at the other extreme may be a simple transmitter, as in the first Vanguard satellite.

B. Typical Spacecraft

Some typical spacecraft will be described in order to illustrate their diversity and some of their design details.

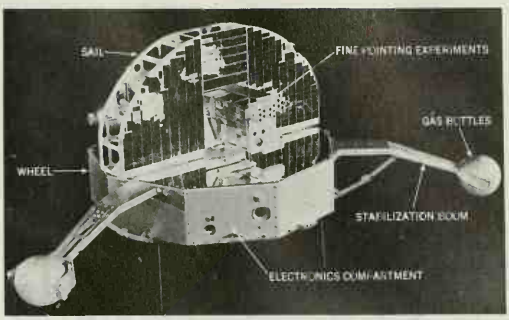
1. Atmospheric Structure Satellite.—The Atmospheric Structure satellite (Explorer 17) was a 405-lb. spacecraft carrying instruments to measure the density, composition, pressure and temperature of the atmosphere. The spacecraft was a hermetically sealed sphere designed to be nearly leakproof, thus minimizing contamination of the atmosphere of space. For example, at apogee the internal satellite pressure was 10^{13} times greater than the "hard vacuum" of space on the outside; any gas leakage from the spacecraft would interfere with measurements of the tenuous space environment.

Electrical energy to the spacecraft was supplied by approximately 150 lb. of silver-zinc chemical batteries. With all instrumentation operating, the satellite consumed 110 w. of power and could operate at this load for about 70 hr. Since there was no provision for data storage aboard the spacecraft, and since satellite "turn-ons" were to last about 5 min., several hundred data



BY COURTESY OF NASA

FIG. 8.—DEEP-SPACE NETWORK TRACKING STATION AT JOHANNESBURG, S.AF., HAS "DISH" ANTENNA 85 FT. IN DIAMETER. SIMILAR STATIONS ARE AT GOLDSTONE, CALIF., AND WOOMERA, AUSTR.



BY COURTESY OF NASA

FIG. 9.—ORBITING SOLAR OBSERVATORY, FIRST LAUNCHED SUCCESSFULLY IN MARCH 1962

"passes" could be obtained. A reasonable useful lifetime of the satellite was computed at between two and three months (the actual lifetime was 99 days).

When injected into orbit, the satellite was spin-stabilized at 1.5 cycles per second. An inclination of 58° from the equator gave it a coverage of the Earth's surface to within several degrees of the arctic and antarctic regions. The orbital period was 90 min. The spacecraft carried an aspect system to determine the orientation of its spin axis in relation to a reference plane; this information was needed to allow the experimenters to interpret their data correctly. A four component aspect system was used, comprised of a digital-fan sensor, a slit-fan sensor, a Sun-Moon switch and two infrared Earth sensors. The system was designed to use the Earth and either the Sun or the Moon as references against which to locate the satellite spin axis. A digital-fan sensor uses an optical masking technique and six photodiodes to measure the angle of a light source viewed by the sensor. The slit-fan sensor, located in the same housing as the digital fan, is used to measure azimuth angles accurately.

The Sun-Moon switch reduced the gain of the amplifier when sunlight was viewed and restored full gain when the Moon was the reference object. The two Earth sensors had a pencil beam of approximately 2° solid angle and were mounted in the same housing. One sensor had an elevation of 22° north of the equator of the satellite, and the other pointed 22° south of the equator. These Earth sensors operated in the 7–15 micron infrared wavelength region and detected the discontinuity at the Earth's horizon. Prior to launch, steps were taken to insure that the satellite was stable about its predetermined axis of rotation. The battery and electronics components were distributed to ensure maximum moment of inertia about this axis.

The Atmospheric Structure satellite was the first scientific Earth satellite containing a pulse-code modulation (PCM/FM) telemetry system. Previous satellites used the pulse-frequency modulation (PFM/AM) system. The solid-state PCM system provided an output power of 500 mw. and was capable of supplying 40 separate channels of information in digital form.

Provisions were made to permit selection of experiments while the satellite was in orbit. Two command receivers were used, each capable of two independent commands. The number of allowable combinations permitted resolution of possible interference between the various detectors. While equipment "turn-on" was controlled from the ground, "turn-off" of all equipment was automatic and produced from on-board instrumentation, ensuring that power would not be wasted when the satellite passed out of the range of the command receiver.

2. Interplanetary Explorer.—The Interplanetary Explorer (Explorer 18) is a 138-lb. spacecraft placed into a highly elliptical orbit: an apogee altitude of 122,800 mi., a perigee of 120 mi. and an inclination to the equator of 33°, corresponding to an orbital period of approximately four days. The spacecraft is designed to study interplanetary radiation and magnetic fields in the vicinity

TABLE V.—*Advances in Space Exploration by Unmanned Artificial Satellites and Space Probes*

Name	Nation	Launched	International code	Launch vehicle	Perigee/Apogee (mi.)	Period (min.)	Lifetime*	Weight in orbit† (lb.)	Remarks
		1957							
Sputnik 1	U.S.S.R.	Oct. 4	1957 Alpha	n.a.‡	143.5/584	96.17	1958	(184)	Establishment of first artificial satellite
Sputnik 2	U.S.S.R.	Nov. 3	1957 Beta	n.a.‡	128/1,056	103.7	April 14, 1958	1,120	First inhabited space capsule (dog, "Laika")
		1958							
Explorer 1	U.S.	Jan. 31	1958 Alpha	Jupiter-C	224/1,585	114.7	4–10 yr.	30.8	Discovery of inner (Van Allen) radiation belt
Vanguard 1	U.S.	March 17	1958 Beta	Vanguard	405/2,462	134.3	Indefinite	3	First proof that Earth is pear-shaped
Sputnik 3	U.S.S.R.	May 15	1958 Delta	n.a.‡	133/1,178	105.8	April 6, 1960	(2,926)	First instrumented multipurpose space laboratory (atmospheric pressure, composition; positive ion concentration; primary cosmic ray data, etc.)
Score	U.S.	Dec. 18	1958 Zeta	Atlas	115/915	101.5	Jan. 21, 1959	8,800	First transmission of voice messages from space
		1959							
Lunik 1	U.S.S.R.	Jan. 2	1959 A.P. 1§	n.a.‡	91.5/214.7‖	450¶ (about Sun)	...	3,245	First space probe to become artificial planetoid; radio contact to 371,000 mi.
Vanguard 2	U.S.	Feb. 17	1959 Alpha	Vanguard	347/2,064	125.9	...	22	Returned first cloud-cover photographs; transmitted for 18 days
Explorer 6	U.S.	Aug. 7	1959 Delta	Thor-Able	157/26,366	768	July 1961	143	Discovered Earth's electrical current system; returned first television photograph of Earth; first use of solar-cell paddles for power
Lunik 2	U.S.S.R.	Sept. 12	1959 ...§	n.a.‡	(Moon impact)	...	33 hr. 32 min.	3,324	First spacecraft on Moon; discovery that Moon is enveloped by layer of low-energy ionized gas
Vanguard 3	U.S.	Sept. 18	1959 Epsilon	Vanguard	317/2,329	130.2	...	100	Mapped Earth's magnetic field, returned micrometeorite and radiation data
Lunik 3	U.S.S.R.	Oct. 4	1959 Theta	n.a.‡	25,500/294,250	15.3¶ (about Earth)	April 1960	958.7	First circumnavigation of Moon; first pictures of opposite side of Moon
Explorer 7	U.S.	Oct. 13	1959 Iota	Juno II	346/676	101.2	...	92	Transmitted data on magnetic fields and storms, solar flares, radiation belt
		1960							
Pioneer 5	U.S.	March 11	1960 Alpha	Thor-Able	(Solar orbit)	312¶	...	95	Interplanetary probe; measured size of solar system, returned data on magnetic field and solar wind out to 22,500,000 mi.
Tiros 1	U.S.	April 1	1960 Beta	Thor-Able	428.7/465.9	99.19	75 yr.	320	Study of TV systems for weather satellites
Transit 2A	U.S.	June 22	1960 Eta	Thor-Able-Star	382/657	101.66	50 yr.	1,587	First successful dual-satellite launch
Discoverer 13	U.S.	Aug. 10	1960 Theta	Thor-Agena A	161/436	94.1	Nov. 14, 1960	(300)	First successful recovery of capsule from orbit
Echo 1	U.S.	Aug. 12	1960 Iota	Thor-Delta	945/1,049	118.24	...	241.4	(Aluminum-coated plastic sphere.) First successful voice and picture transmission by reflection off satellite; first measurements of satellite perturbations caused by solar radiation pressure
Discoverer 14	U.S.	Aug. 18	1960 Kappa	Thor-Agena A	116/502	94.54	Sept. 16, 1960	(300)	First successful air-recovery of orbital capsule
Sputnik 5	U.S.S.R.	Aug. 19	1960 Lambda	n.a.‡	190/211	90.68	Aug. 20, 1960	10,120	First controlled landing of animals in space vehicle (two dogs, six mice, insects); animal behaviour in orbit monitored by TV and radio
Explorer 8	U.S.	Nov. 3	1960 Xi	Juno 2	258/1,423	112.75	19 yr.	101	Collection of data for ionospheric studies; static electricity accumulation on satellite surface; impact of micrometeorites
Tiros 2	U.S.	Nov. 23	1960 Pi	Thor-Delta	387/453	98.27	75 yr.	330	Radiation and optical mapping of Earth; TV pictures received but, wide-angle pictures of poor quality
Discoverer 18	U.S.	Dec. 7	1960 Sigma	Thor-Agena B	154/459	94.1	April 2, 1961	(300)	Re-entry test; exposure of human eye tissue, bone marrow and blood cells, algae, spores, etc., to corpuscular radiation; air-recovery successful
		1961							
Sputnik 8/Venus Probe 1	U.S.S.R.	Feb. 12	1961 Gamma	n.a.‡	123/198 66.7/94.6‖	89.7 3009	Feb. 25, 1961 (Sputnik 8)	(1,419) (V.P. 1)	First hyperbolic escape of probe to measure interplanetary magnetic fields and solar radiation in space and near Venus; radio contact lost about 4,700,000 mi. from Earth
Explorer 9	U.S.	Feb. 16	1961 Delta	Scout	395/1,605	118.3	...	15	Atmospheric density data provided by optical tracking of 12-ft. balloon
Sputnik 9	U.S.S.R.	March 9	1961 Theta	n.a.‡	115/155	...	March 9, 1961	10,340	Habitable vehicle with dog, guinea pigs, mice, insects and plant seeds; re-entry and recovery successful
Sputnik 10	U.S.S.R.	March 25	1961 Iota	n.a.‡	111/153	88.4	March 26, 1961	10,330	Cabin with dog "Zvezdochka" recovered after 17 orbits
Transit 4A	U.S.	June 29	1961 Omicron 1	Thor-Able Star	534/623	103.7	...	175	All-weather global navigation satellite; first to use nuclear auxiliary power system
MA-5	U.S.	Nov. 29	1961 Alpha Iota	Atlas D	99.6/147.5	88.5	Nov. 29, 1961	2,900	First U.S. orbital flight carrying animal (chimpanzee "Enos")

TABLE V.—*Advances in Space Exploration by Unmanned Artificial Satellites and Space Probes (continued)*

Name	Nation	Launched	International code	Launch vehicle	Perigee/Apogee (mi.)	Period (min.)	Lifetime*	Weight in orbit† (lb.)	Remarks
		1962							
Tiros 4	U.S.	Feb. 8	1962 Beta	Delta	441/525	100.4	Indefinite	285	Returned total of 32,593 cloud-cover photographs up to June 10, 1962
OSO-1	U.S.	March 7	1962 Zeta	Delta	344/370	96.2	...	458	First orbiting solar observatory, transmitting solar flare, radiation data
Cosmos 1	U.S.S.R.	March 16	1962 Theta	n.a.‡	135/609	96.35	May 25, 1962	n.a.‡	Unmanned spacecraft to study conditions of prolonged flight and effects of Van Allen belts on long-range radio communications
Ariel	U.K.-U.S.	April 26	1962 Omicron	Delta	242/754	100.9	...	32	First international co-operative satellite
Tiros 5	U.S.	June 19	1962 Alpha Alpha	Delta	367/604	100.5	...	286	Weather satellite; returned 57,875 cloud-cover photographs
Telstar 1	U.S.	July 10	1962 Alpha Epsilon	Delta	593.4/3,501.8	157.8	...	(170)	First transatlantic relay of television signals (July 10); first colour TV relay (July 16); tests of broad-band microwave communications in space
Mariner 2	U.S.	Aug. 27	1962 Alpha Rho	Atlas-Agena B	65.5/113.8‖	345.9¶	...	(447)	Geophysical studies of Venus, passing planet at 21,594 mi.; measurements of atmospheric and surface temperatures, magnetic field, planetary density; radio contact to 51,300,000 mi.
Alouette	Can.-U.S.	Sept. 28	1962 Beta Alpha	Thor-Agena B	620/638	105.4	...	320	Canadian research satellite for ionospheric sounding
Mars 1	U.S.S.R.	Nov. 1	1962 Beta Nu	n.a.‡	(Solar orbit)	519¶	...	1,965	Mars probe; transmissions ceased after 66,000,000 mi.
		1963							
Telstar 2	U.S.	May 7	1963 13A	Delta	604/6,713	225	.	175	Active-repeater communications satellite, transmitting on 136,050 mc.
(None)	U.S.	May 9	1963 14A	Atlas-Agena B	2,254/2,295	166.6	Indefinite	n.a.‡	Ejected 50-lb. package of copper dipole needles for experiments in reflecting high-frequency radio signals over long Earth distances
Syncom 2	U.S.	July 26	1963 31A	Delta	22,132/22,823	1,436	...	86	Active-repeater communications satellite in synchronous orbit, initially over Brazil; communications tests successful
		1964							
(None)	U.S.	Jan. 29	1964 5A	Saturn I	164/471	94.8	< 1 yr.	37,700	Heaviest weight in orbit to date; thrust at lift-off, 1,500,000 lb.
Elektron 1	U.S.S.R.	Jan. 30	1964 6A	n.a.‡	252/4,410	169	...	n.a.‡	Both spacecraft launched into different orbits by one launch vehicle to study separated regions of space
Elektron 2	U.S.S.R.	Jan. 30	1964 6B	n.a.‡	285/42,352	1,360	...	n.a.‡	
Ranger 7	U.S.	July 28	1964 41A	Atlas-Agena B	(Moon impact)	...	July 31, 1964	806	Transmitted more than 4,000 still television photographs of Moon's surface; last photographs taken at about 1,000 ft. above surface
OGO-1	U.S.	Sept. 5	1964 54A	Atlas-Agena B	173/92,721	3,839	•••	1,073	Carries 20 co-ordinated space experiments
Explorer 22	U.S.	Oct. 9	1964 64A	Scout	551/1,673	104.8	.	116	First satellite to carry successful laser experiment
Explorer 24	U.S.	Nov. 21	1964 76A	Scout	329/1,551	116	...	19	With Explorer 25, first simultaneous launch of two spacecraft by NASA
Explorer 25	U.S.	Nov. 21	1964 76B	Scout	329/1,551	116	.	90	
Mariner 4	U.S.	Nov. 28	1964 77A	Atlas-Agena D	(Mars probe)	..	Indefinite	585	Photographed surface of Mars and studied the Martian atmosphere
Zond 2	U.S.S.R.	Nov. 30	1964 78C	n.a.‡	(Mars probe)	...	...	n.a.‡	Scheduled to pass Mars about July 1965, but communication equipment failed
San Marco	Italy-U.S.	Dec. 15	1964 84A	Scout	128/510	95	...	254	First launch of a satellite by a team of nationals from a country other than U.S. or U.S.S.R.
		1965							
Pegasus A	U.S.	Feb. 16	1965 9A	Saturn I	308/463	97.1	...	33,200	2,000 sq.ft. micrometeoroid sensor area
Ranger 8	U.S.	Feb. 17	1965 10A	Atlas-Agena B	(Moon impact)	...	Feb. 20, 1965	...	Transmitted 7,137 still photographs of Moon's surface; covered much wider area than Ranger 7
Ranger 9	U.S.	March 21	1965	Atlas-Agena B	(Moon impact)	...	March 24, 1965	...	Photographs of Moon transmitted over commercial television
Early Bird 1	U.S.	April 6	1965 28A	thrust-augmented Delta	22,251/22,270	1,437	.	87	First commercial satellite launched by NASA for the Comsat Corp.
Zond 3	U.S.S.R.	July 18	1965	n.a.‡	(Space probe)	...	...	n.a.‡	Photographed far side of Moon while on unspecified deep-space mission

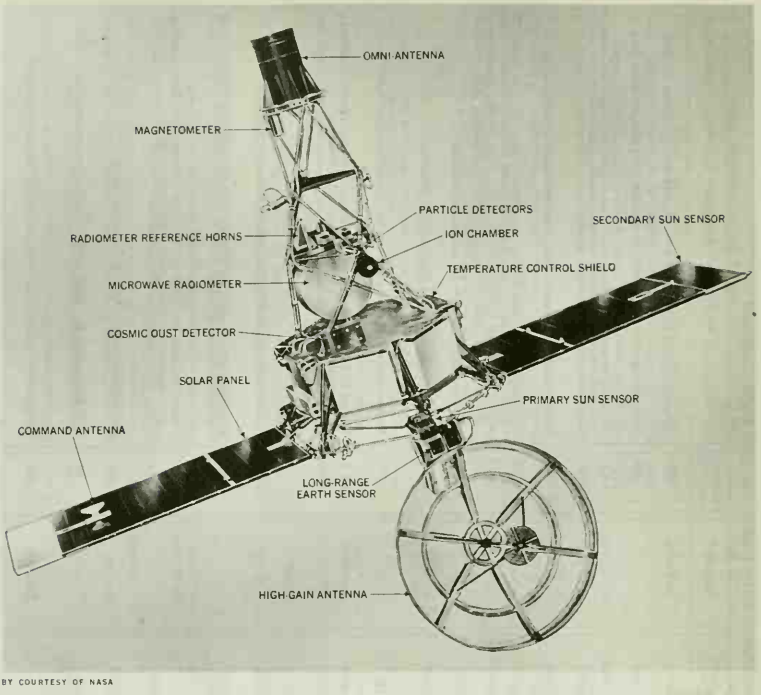
*Dates indicate time of destruction during re-entry or successful re-entry of satellite; other lifetimes are estimated.
†Figures in parentheses indicate weight of spacecraft or capsule only; other figures show total weight in orbit, including booster stage, etc.
‡n.a.—not announced. §Prior to Oct. 1, 1959, only Earth satellites were included in the international code; A.P.=artificial planet.
‖Perihelion and aphelion of artificial planet in miles × 1,000,000. ¶Days. ♀Period, Venus Probe 1, in days.
Source: Based on data compiled by National Aeronautics and Space Administration.

of the Earth and Moon. Explorers 21 and 26 are also Interplanetary Explorers.

Explorer 18 is an octagon-walled platform, 28 in. across and 12 in. deep, fabricated from nylon honeycomb and fibreglass. A rubidium-vapour magnetometer is located on a 71-in. boom on top of the platform and two fluxgate magnetometers on 14-ft. booms at the sides. Four spring-loaded solar-cell paddles extend from the main structure to supply the required power.

The satellite transmits data continuously during its operating time in orbit on the 136.110 mc. frequency; signals from the 1.5-lb. 4-w. transmitter can be received by NASA stations at distances of 123,000 mi. Data are recorded 100% of the time for approximately the first six months of flight, and then for 25% during the remainder of the satellite's lifetime. The telemetry system is designed so that the recorded data can be processed 16 times faster than real time.

3. Orbiting Solar Observatory.—The Orbiting Solar observatory (OSO) is a 500-lb. spacecraft designed to accommodate interchangeable experiments for studying the Sun. A weight of about 250 lb. is allotted for the total experiment payload.

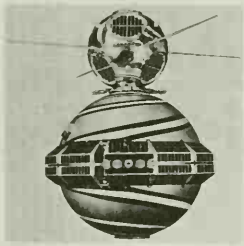


BY COURTESY OF NASA

FIG. 11.—A MODEL OF THE MARINER 2 SPACECRAFT WHICH WAS LAUNCHED AUG. 27, 1962, ON A COURSE WHICH CARRIED IT TO WITHIN 21,600 MI. OF VENUS. ITS MISSION WAS SUCCESSFULLY COMPLETED DEC. 14, 1962, WITH THE RADIOING OF MESSAGES BACK TO THE EARTH. THE MODEL IS SHOWN WITH ITS SOLAR PANELS EXTENDED IN CRUISE POSITION

The OSO consists of two sections: a nine-sided wheel composed of wedge-shaped compartments and a semi-circular section, mounted above the wheel, called the sail. When the spacecraft is in orbit, three arms extend from the wheel and slow its rotation to about 30 r.p.m.; the arms can be used as antennas. The wheel diameter is 44 in. Five of the nine wedge sections, each with a volume of approximately 1,000 cu.in. and a maximum weight capacity of 30 lb., are available for experiments. When in orbit, the wheel rotates much like a gyroscope, each section facing the Sun every two seconds.

The sail, joined to the wheel by a rotating shaft, is designed so that it aims always at the Sun. The sail carries an array of solar cells to supply power and recharge batteries that produce about 16 w.; of this, 9 w. is available for power experiments. In addition, the sail carries two compartments for experiments which must point at the Sun; each is 4 × 8 × 36 in. and is able to carry approximately 50 lb. of equipment, but the two may be combined to give double the space and weight capacity.



BY COURTESY OF U.S. NAVY

FIG. 10.—TRANSIT 2A NRL SATELLITE LAUNCHED JUNE 22, 1960, FROM CAPE CANAVERAL (NOW CAPE KENNEDY), FLA.

Data from each experiment are stored sequentially in digital form on a tape recorder operating continuously throughout each 95-min. orbit. The information from the recorder is read out at high speed on command from a ground station. The first OSO was launched in 1962, the second early in 1965.

4. Transit.—Transit (fig. 10) is a satellite weighing approximately 135 lb. It is designed to provide navigational data to military and civilian ships. The spacecraft has an octagonal body of fibreglass and honeycomb plastic, 12 in. high and 18 in. across. Four panels containing 18,000 solar cells are mounted on the sides, and the power these produce is stored in nickel-cadmium batteries. The spacecraft employs gravity-gradient stabilization, utilizing a single self-erecting boom equipped with a long fine-damping spring with a small mass on the end. The primary instrumentation consists of two ultrastable oscillators operating at 150 and 400 mc.

5. Mariner.—The Mariner spacecraft (Mariner 2) weighed 447 lb. and was launched by an Atlas-Agena vehicle in a 180,000,000 mi., 109-day trajectory past Venus. The spacecraft (fig. 11) was 5 ft. in diameter at the base and 9 ft., 11 in. high. In the cruise position, with solar panels and high-gain antenna extended, it was 16.5 ft. across and 11 ft., 11 in. high.

The design was a variation of the hexagonal concept used for the Ranger series. The base housed a liquid-fuel rocket motor for trajectory correction and six modules containing the attitude control system, electronic circuitry for the scientific experiments, power supply, battery and charger, data encoder and command subsystem, digital computer and sequencer, and radio transmitter and receiver. Sun sensors and attitude control jets were mounted on the exterior of the base.

A tubular superstructure extended upward from the base and scientific experiments were attached to this framework. An omnidirectional antenna was mounted at the peak of the superstructure. A parabolic, high-gain antenna was hinge-mounted below the base, and two solar panels also were hinged to the base; these folded up alongside the spacecraft during launch, parking orbit and injection phases, and unfolded like butterfly wings when the craft was in space. A command antenna for receiving transmissions from

Earth was mounted on one of the panels.

The solar panels contained 9,800 solar cells occupying an area of 27 sq.ft. They collected energy from the Sun and converted it into electrical power at a minimum of 148 w. and a maximum of 222 w. The amount of power available from the panels increased slightly during the mission as the spacecraft moved nearer to the Sun. Each solar cell had a protective glass filter to reduce the Sun's ultraviolet and infrared radiation, which would produce heat but not electrical energy. Prior to deployment of the solar panels, power was supplied by a 33.3-lb., silver-zinc rechargeable battery with a capacity of 1,000 watt-hours, and this battery supplied the long-term power requirements of the Venus mission (109 days), both directly for switching and sharing peak loads with the solar panels and also during trajectory correction when the panels were not directed at the Sun.

Two-way communication aboard the Mariner 2 was supplied by the receiver/transmitter, two transmitting antennas (the omnidirectional and high-gain antennas) and the command antenna for receiving instructions from Earth. The spacecraft transmitting power was only 3 w. An Earth sensor was mounted on the antenna yoke near the rim of the high-gain, dish-shaped antenna to search for and keep the antenna pointed at the Earth.

Stabilization of the spacecraft for yaw, pitch and roll was provided by ten cold-gas jets mounted in four locations and fed from two titanium bottles containing 4.3 lb. of nitrogen at 3,500 p.s.i. The jets were linked by logic circuitry to three gyroscopes in the attitude-control subsystem, to the Earth sensor on the parabolic antenna, and to six Sun sensors mounted on the spacecraft frame and on the backs of the two solar panels. The Sun sensors were light-sensitive diodes that informed the attitude-control system when they "saw" the Sun. The attitude-control system responded to these signals by turning the spacecraft and pointing the longitudinal or roll axis toward the Sun. Power for these maneuvers was provided by the cold-gas jets. Enough nitrogen was provided to operate the gas jets for a minimum of 200 days.

A second probe, Mariner 4, was successfully launched toward Mars on Nov. 28, 1964, passing that planet in July 1965. While the general design was like that of the probe toward Venus, the scientific payload, power supply and telemetry were modified because of the greater distance of Mars from the Earth and the Sun. A television camera on the spacecraft took 22 pictures of the surface of the planet which were transmitted to Earth; other instruments gathered data on the atmosphere and searched unsuccessfully for trapped radiation and a magnetic field. Along the way Mariner 4 made measurements in interplanetary space.

6. Orbiting Geophysical Observatory.—In 1964, the first Orbiting Geophysical observatory, OGO-1, was launched into an elliptical orbit about the Earth. This type of spacecraft is larger than the Explorer class, weighing somewhat more than 1,000 lb. and designed to accommodate up to 50 individual experiments.

It is designed to be stabilized by motor-driven inertial flywheels and gas jets to keep one face toward the Earth, maintained so that one axis in the spacecraft is always along the line to the centre of the Earth. The solar paddles were designed to face continuously toward the Sun to ensure maximum power. The payload from an OGO consists primarily of experiments selected to make co-ordinated measurements and observations of many aspects of the Earth environment simultaneously in order to supply a clearer picture of the relationships between them and their dependence upon solar activity. Two types of OGO's are scheduled to be used. The EGO, or Eccentric Orbiting Geophysical observatory, such as OGO-1, is planned to have a highly elliptical orbit reaching beyond the Earth's magnetic field or magnetosphere into the near-Earth interplanetary space.

The POGO, or Polar Orbiting Geophysical observatory, will be placed in a low-altitude orbit over the poles for studies primarily of the upper atmosphere and the ionosphere, and for monitoring incoming radiations.

The OGO is essentially a boxlike structure, 6 × 3 × 3 ft. in size, of aluminum sandwich panels, that houses the scientific instrumentation and the equipment needed for power, telemetry, communications and attitude control. Projecting from the box antennas for communications, booms for experiments requiring separation from the spacecraft and the other experiments, a boom and containers for experiments that require alignment in the orbital direction, two paddles of 75 sq.ft. combined area that carry the solar cells, an antenna for the radio astronomy experiment and experiments that must point to the Sun. In orbit OGO-1 resembles a giant insect with a 20-ft. wing spread.

7. Ranger.—Ranger 7, an 806-lb. spacecraft, was launched toward the moon by an Atlas-Agena B rocket on July 28, 1964. The mission of Ranger was to take photographs of the moon before impacting on its surface. The tubular-shaped craft was equipped with two "wings" that extended from the bottom of the main tower structure at approximately right angles; these wings, constructed so that they would face the sun during flight, contained panels of solar cells that converted the sunlight into the electricity needed to power the craft's equipment. Six television cameras mounted in the tower of Ranger focused on the moon through an aperture in the centre of the tower. At the top of the tower was an omnidirectional antenna used during launch and in a mid-course maneuver. Facing away from the moon and toward the earth was a dish-shaped antenna that transmitted the photographs.

During the last 17 minutes of its flight Ranger returned more than 4,000 high-resolution photographs of the moon's surface, the last at an altitude of about 1,600 ft. Rangers 8 and 9 later successfully performed similar missions.

C. Spacecraft Design

The paramount requirements in spacecraft design are reliability and long life; the spacecraft must work, and it must work for a long time. The total cost of a particular development program depends directly on the ability of the designer to fulfill these requirements. A satellite that works for a long time collecting desired data clearly obviates the need for additional spacecraft to collect the same data, and therefore saves additional fabrication and launching costs. Reliable long life, therefore, is the keystone to success in spacecraft technology, not only for near-Earth orbits but especially for spacecraft journeying to other planets. A trip to Mars, for example, takes almost a year, and during that time the spacecraft must be capable of performing guidance and navigation functions as well as scientific measurements. If it is to land when it finally reaches Mars, it must then be capable of operating on the surface. A Martian year is about two Earth years. Therefore, the requirement for long life, completely unattended by human hands, is the great challenge that faces the designer of automated spacecraft.

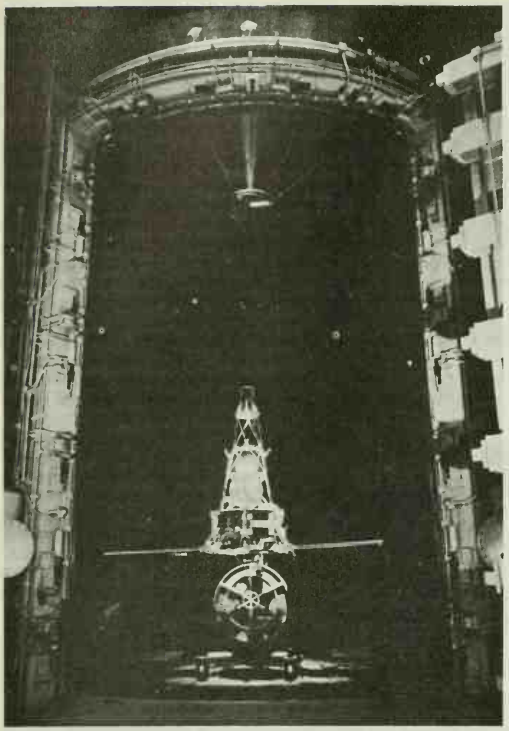
The first essential for success is simplicity, a feature exhibited by some of the early satellites. They tended to be spherical so that the drag forces would be nearly constant and the resultant orbits could be more readily computed and tracked. They had simple power supplies such as storage batteries or small banks of solar cells to utilize the energy coming from the Sun. Their transmitters and associated electronic circuitry were elementary. That this simplicity of design was successful is evident in the case of the Vanguard satellite, which operated continuously in orbit for over five years.

D. Spacecraft Testing

The importance of skilled workmanship and strict quality control in spacecraft construction is obvious. However, it is well known that reliability greatly decreases as the number of components and complexity of design increase. To achieve success in complex satellites, it has been necessary to develop extensive facilities for testing spacecraft before they are launched.

Fundamentally, successful spacecraft performance depends upon: (1) simple, sound design; (2) highly skilled workmanship; (3) strict quality control; (4) mechanical acceleration and vibration tests; (5) thermal and vacuum tests; (6) radiation tests; (7) guidance and control tests; and (8) complete in-flight monitoring of performance.

1. Acceleration and Vibration Tests.—Every satellite must endure and survive the rugged accelerations and vibrations im-



BY COURTESY OF NASA

FIG. 12.—MARINER SPACECRAFT IN AN ENVIRONMENTAL FACILITY THAT SIMULATES APPROXIMATE TEMPERATURES AND PRESSURES IN SPACE

posed during launch, so testing facilities that simulate these stresses have been developed. The launching accelerations are simulated by a centrifuge. The accelerations during launch can be programmed, and the spacecraft can be mechanically, structurally and elastically examined in detail under simulated conditions.

A vibration machine is used to pinpoint those resonances that may produce structural failure or distortion. Just as wind-induced resonances can destroy a large suspension bridge, so small vibrations can cause solar paddles, antennas and the main structural components of a satellite to fail. Vibration tests have been extremely important in uncovering weak points in mechanical design that otherwise might not be discovered. About 30% of the spacecraft failures occur during the vibration test.

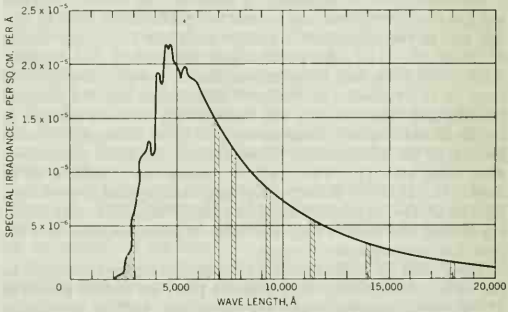
2. Pressure and Temperature Tests.—An important hazard to spacecraft performance is the low pressure in outer space. Some components, for example, might blow up when placed in a vacuum; some materials might simply evaporate or lose their normal characteristics. The vacuum of space is considerably lower than vacuums readily achievable on Earth, and, therefore, it has been necessary to improve vacuum apparatus. In the mid-1960s vacuum chambers capable of housing complete spacecraft could attain pressures of 10^{-6} mm. of mercury and sometimes as low as 10^{-8}. For small components lower pressures were available.

Temperature presents another problem, for a spacecraft receives heat by radiation directly from the Sun, without the shielding effect provided by the Earth's atmosphere. The amount of heat depends upon the thermoconductivity and the radiation characteristics of the satellite. The temperature of the spacecraft depends upon the heat it absorbs and, in particular, on the heat it radiates away. It is necessary to achieve a good thermal balance in order to minimize extremes of temperature, and it is also desirable to achieve slow changes in temperatures, when changes are unavoidable. Temperature has a direct effect on the lifetime of electronic components; in general, the higher the temperature the poorer their reliability and the shorter their lifetime. Sudden changes in temperature are also serious because they introduce thermoworking and thermofatigue, phenomena that can lead to failure in components and circuitry. The temperature, the radiation and the thermal balance characteristics of spacecraft are thoroughly explored on Earth in thermal chambers, usually combined with vacuum chambers. Heat from simulated "suns" is directed at the satellite. The walls of the test chambers are cooled to the blackbody temperatures of space by refrigerants. The satellite itself may be oriented in a systematic manner with respect to the "sun" in order to simulate its flight plan in space. Thermocouples distributed throughout the satellite measure the transient flow of heat and internal distribution of temperatures under various working conditions. In this way "hot spots" may be detected and eliminated.

3. Radiation Testing.—Another problem is that of the intense radiation in space. Radiations coming from the Sun, as well as from the cosmos, pose a special hazard to the performance and life of electronic components, as well as to the characteristics of some structural materials. Solar cells, transistors, resistors, etc., are all subject to serious damage from radiation, particularly during severe solar storms. Artificial radiation also occurs in space as a result of high-altitude explosions of nuclear bombs. Facilities were being constructed in the 1960s to test satellite components and complete spacecraft in strong radiation environments.

4. Guidance and Control Testing.—If spacecraft are to be properly controlled in space, there must be testing facilities on Earth to first prove out the entire guidance and control servomechanism. An excellent approach is to mount the complete spacecraft so that it is free in all three modes of angular motion (roll, pitch and yaw). In this way the servomechanism can be programmed and the spacecraft can be put through various maneuvers. At the same time the system can be placed in a thermovacuum chamber and can be operated for long periods to determine its lifetime in the space environment. In this way any deficiencies in the various components of the control system can be discovered and corrected. Complex guidance systems present a major challenge with respect to reliability and long life. Passive attitude-control systems are preferred because of their simplicity. However, in many cases performance requirements dictate direct and immediate control in all three modes of motion or require successive positioning in precise fashion.

For example, the OSO spacecraft must be capable of being pointed in a given direction with an accuracy of one minute of arc; for the OAO satellite the requirement is measured in seconds of arc.



REPRINTED FROM "SATELLITE ENVIRONMENT HANDBOOK," ED. BY FRANCIS S. JOHNSON, WITH THE PERMISSION OF THE PUBLISHERS, STANFORD UNIVERSITY PRESS. © 1961 BY LOCKHEED AIRCRAFT CORPORATION

FIG. 13.—DISTRIBUTION OF ENERGY IN THE SOLAR RADIATION INCIDENT ON THE EARTH'S UPPER ATMOSPHERE. NARROW SHADED LINES ARE REGIONS OF STRONG ATMOSPHERIC ABSORPTION; SHADED AREA BELOW 3000 Å IS PORTION OF THE ULTRAVIOLET SPECTRUM ABSORBED BY THE ATMOSPHERE

In such cases, complex command and control systems are required and these systems, therefore, are of vital concern in determining satellite reliability and performance. The best approach to an answer lies in the long-life guidance and control tests carried out on the ground.

Of equal importance, however, are diagnostic instrumentation and telemetry during the actual space flight, because if reliability and lifetime are to be improved it is necessary to know exactly why past satellites have failed. For this reason some weight and power must be allocated for diagnostic instrumentation which will continuously monitor the performance of the spacecraft, its circuitry, temperature, voltages and currents, and general performance. Thus, in the actual space environment the design and performance of the spacecraft can be monitored and analyzed, and corrective modifications can be made before future launchings.

(Jo. D. N.)

IV. SPACE SCIENCE: PHYSICAL SCIENCES

Science is the expression of man's desire to explore and understand his natural environment. Space science is the extension of that exploration to those new regions of space and to those new phenomena which have become accessible to man as a result of the development of sounding rockets, satellites and deep-space probes. Before this, man was limited in his explorations to the Earth and its atmosphere up to an altitude of about 40 km. He was limited to the study of the natural phenomena which either occurred in this region of space or which emitted electromagnetic or particulate radiation which could penetrate the Earth's atmosphere to this region. He could observe the stars and galaxies in those colours of light which were not absorbed by the atmosphere, and could only speculate about the appearance of the planets and the stars in the ultraviolet and infrared light which is absorbed by the atmosphere. The development of the sounding rocket, the satellite, and the deep-space probe enormously increased the region of space which man can explore and the phenomena he can observe.

In order to understand the factors that must be taken into account in space science and to provide a proper perspective for this article it is necessary to remember that the Earth is immersed in the "atmosphere" of a small star, the Sun, and that the Sun controls the properties of the space occupied by the planets and provides the energy for the processes which take place on the Earth and planets.

A. Solar Phenomena

The physical properties of the Sun are given in the article Sun. This section will cover primarily those solar phenomena of importance to space science or those which can only be studied by space techniques.

The Sun emits electromagnetic radiation in all wavelengths, from gamma rays to long wavelength radio waves. It also emits charged particles—protons, alpha particles, the positively charged nuclei of heavier atoms, and presumably electrons, although these have not yet been detected. The total power radiated by the Sun, 3.8×10^{33} ergs per second, is nearly constant, varying by less than 1%. However, certain regions of the electromagnetic spectrum, as well as the energies and intensities of the charged particles, are highly variable, both in time and in the location on the Sun from which they are emitted.

Fig. 13 is a graph of the amount of energy that the Sun radiates at each wavelength. The Sun radiates 99% of its energy in wavelengths between 3000 and 40,000 Å. The atmosphere is transparent to this portion of the solar spectrum (with the exceptions of the various absorption bands at 6884, 7621, 9419, 11,350, 14,000 and 18,000 Å, shown cross-hatched in fig. 13) and consequently it has been studied extensively at ground-based observatories. The spectral distribution in this portion of the spectrum is similar to that of a 6,000° K. black body, and the radiation comes almost entirely from the photosphere of the Sun. This region of the solar spectrum is of importance to space science because it is primarily responsible for the temperature of objects in space. The solar energy incident on a planet determines the temperature of the planet. Mercury, being closest to the Sun, has a maximum temperature of about 410° C. (770° F.) whereas Jupiter, which is 5 astronomical units from the Sun, has a temperature of −138° C. (−216° F.). As a spacecraft moves closer to the Sun, more solar energy falls on it and its temperature rises.

The shorter wavelengths, below 2900 Å, which are absorbed by the atmosphere can only be studied with instruments on sounding rockets and satellites. Although the amount of energy involved is small, it is an interesting and important region of the solar spectrum. Below 2000 Å the character of the spectral-irradiance curve changes. It is no longer similar to that from a 6,000° K. black body; instead, there is a continuum similar to that of a 4,000° K. black body with a number of very intense emission lines. The intensity of these lines varies with solar activity. The

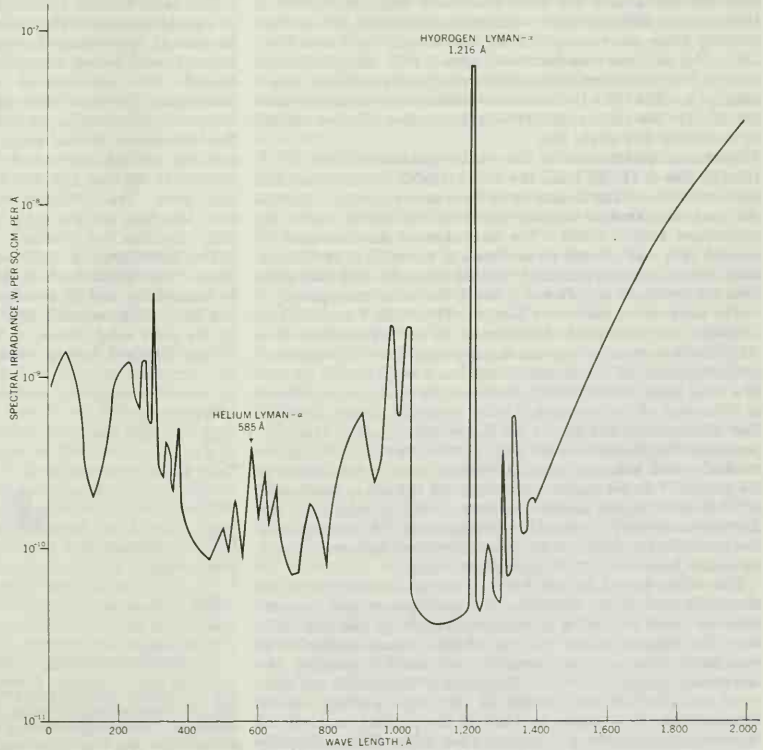
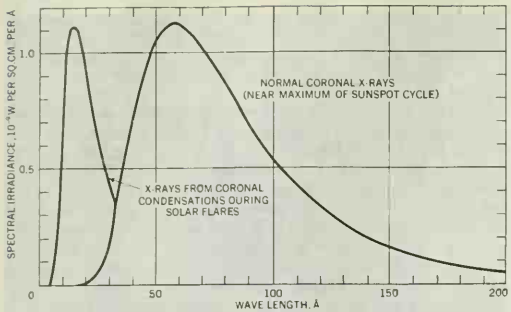


FIG. 14.—SOLAR ULTRAVIOLET SPECTRAL IRRADIANCE ABOVE THE EARTH'S ATMOSPHERE (*see* TEXT)



REPRINTED FROM "SATELLITE ENVIRONMENT HANDBOOK," ED. BY FRANCIS S. JOHNSON, WITH THE PERMISSION OF THE PUBLISHERS, STANFORD UNIVERSITY PRESS. © 1961 BY LOCKHEED AIRCRAFT CORPORATION

FIG. 15.—SOLAR X-RAY SPECTRAL IRRADIANCE ABOVE THE EARTH'S ATMOSPHERE (*see* TEXT)

radiation in this part of the spectrum comes not from the photosphere but from the chromosphere and corona of the Sun. Fig. 14 is an enlargement of that portion of the spectral-irradiance curve from 0 to 2000 Å, showing the more important emission lines and the continuum.

The region from about 1000 to 2000 Å is sometimes referred to as the vacuum ultraviolet, the region from 100 to 1000 Å as the extreme ultraviolet (EUV) and the region below 1000 Å as the X-ray portion of the spectrum.

The Lyman alpha line of hydrogen (1216 Å) is the most prominent emission line in this region of the spectrum. The first measurements of this line were made with sounding rockets in 1949, and its intensity was continuously monitored on the first Orbiting Solar observatory (OSO-1) in March, April and May, 1962. All of these measurements indicate that the intensity of 6×10^{-7} w. per square centimetre is very nearly constant over a solar cycle. The OSO-1 observations indicated a systematic variation of a few per cent correlated with the number of active regions on the visible disk of the Sun.

Systematic observations of the relative intensities of the 304 Å (HeII), 284 Å (FeXV) and the 335 Å (FeXVI) lines were also made on OSO-1. The intensities of these lines varied in a systematic way with sunspot number and with the flux of 2,800 mc. radio noise from the Sun. The intensities of the iron lines increased by a much larger factor than the intensity of the helium line, the differences presumably arising from the fact that these lines are produced at different levels in the solar atmosphere.

The solar corona also emits X-rays in the 0–100 Å region. Fig. 15 shows the solar spectral irradiance in the region from 0 to 100 Å for the normal corona and during a solar flare. The spectral-irradiance curve of the normal corona has a shape similar to that of a black body at 500,000° K.; however, the total energy radiated in this region of the spectrum is many orders of magnitude below that which a body the size ot the Sun would radiate at this temperature. At shorter wavelengths, 8–20 Å, the intensity varies markedly with solar activity. A typical value in this region is $0.1–2 \times 10^{-8}$ w. per square centimetre; for the region below 8 Å, 10^{-13} to 10^{-15} w. per square centimetre. During solar flares the shortest wavelength observed is normally about 1 Å although on a few occasions 0.1 Å X-rays have been observed, and on very rare occasions X-rays of 0.02 Å have been noted.

The values quoted for the fluxes of the various regions in the ultraviolet and X-ray spectrums are preliminary and in some cases are based on fleeting measurements made by sounding rockets. The absolute values will undoubtedly change somewhat as systematic, long-term measurements are made in orbiting observatories similar to OSO-1. The study of the spatial and temporal variations of this portion of the solar spectrum should contribute to the solutions to some of the puzzling problems of the chromosphere, the corona and solar flares discussed in the article SUN.

In addition to their importance in leading to an understanding of solar phenomena, the radiations emitted in the ultraviolet and X-ray region of the spectrum are equally important in the understanding of the physics of the upper atmosphere. Fig. 16 shows the altitudes at which the various portions of the spectrum are absorbed by the atmosphere. The properties of a particular altitude region of the atmosphere are determined by the radiation which is absorbed there. The interaction between solar radiation and tbe atmosphere will be discussed below under *Interplanetary Phenomena.*

The Sun also continuously emits a flux of low-energy charged particles generally referred to as the "solar wind." During large solar flares the Sun emits large numbers of high-energy protons or solar cosmic rays. The solar wind and solar cosmic rays are best observed in interplanetary space, and the solar wind in particular controls many of the properties of interplanetary space.

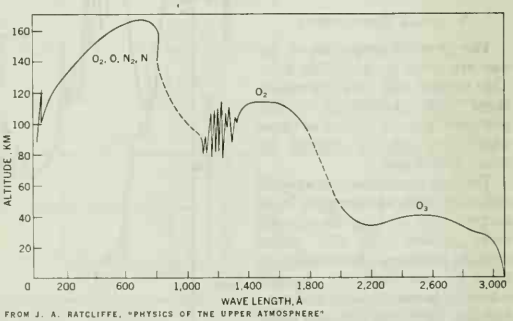
B. Interplanetary Phenomena

Interplanetary space is defined as that region in the solar system whose properties are primarily determined by solar phenomena rather than by the presence of other bodies such as the planets. The Earth and its magnetic field, for instance, influence a region of space extending out about 70,000 km. on the daylight side of the Earth and about 300,000 km. on the night side. Outside this region the properties of space are determined by solar phenomena.

Four interplanetary phenomena will be discussed: the solar wind; the interplanetary magnetic field; solar and galactic cosmic rays; and micrometeorites. These phenomena determine the conditions in interplanetary space and vary appreciably with both time and position in space. In order to understand and describe their behaviour it is necessary to make measurements in interplanetary space.

The solar wind is a moving solar plasma (a plasma consists of a gas of electrically charged particles). Although the individual particles of a plasma are charged, the charges are both positive and negative, and the net charge on any finite volume of plasma is neutral. The solar plasma consists primarily of protons and, presumably, electrons, although the electrons have not yet been measured. The protons in this plasma have energies of about 1 Kev (thousand electron volts), which corresponds to a velocity of about 500 km. per second. Occasionally there is a period of activity on the Sun followed by changes in the behaviour of the solar wind. The number of the particles increases (for example, from less than one per cubic centimetre to more than ten) and their velocities also increase.

The interplanetary field has a value of about 5×10^{-5} gauss. The direction of the field lies approximately in the plane to the ecliptic and at an angle of 40° to 50° with the Earth-Sun line. The magnetic field is carried outward from the Sun by the solar wind; hence, it fluctuates with solar activity. Although the field is very weak, it influences the trajectories of



FROM J. A. RATCLIFFE, "PHYSICS OF THE UPPER ATMOSPHERE"

FIG. 16.—PENETRATION OF SOLAR ULTRAVIOLET RADIATION INTO THE ATMOSPHERE. ABSORPTION ABOVE 2,000 Å IS DUE PRINCIPALLY TO OZONE, BETWEEN 850 AND 2,000 Å TO MOLECULAR OXYGEN AND BELOW 850 Å TO ALL CONSTITUENTS

cosmic rays since it exerts a continuous force on any moving charged particle.

Galactic cosmic rays are thought to originate outside the solar system. Their intensity shows only a very small variation with solar activity and does not vary with direction. Their properties are discussed in detail in the article Cosmic Rays. After a major solar flare the intensity of protons in the energy range from about 1 to several hundred Mev (million electron volts) increases by several orders of magnitude in the first few hours and then gradually returns to normal over a period of several days. During the first hour after a flare, while the intensity is increasing, the number of particles observed depends upon the direction of observation. It must also depend upon the position of observation in space, although no direct observation of such a variation has been made. The protons responsible for the increase in intensity are called solar cosmic rays. Except for their large variation in intensity and position they are identical with galactic cosmic rays. In addition to protons there are the positively charged nuclei of heavier atoms—helium, nitrogen, oxygen, and so on up the atomic scale. The relative abundances of the various nuclei in the solar cosmic rays differ from those in the galactic cosmic rays and reflect the relative abundances of the elements in the Sun.

About two particles will strike a sphere of 1 sq.cm. cross-sectional area in 1 sec. from the normal galactic cosmic-ray flux. About 100,000 particles will strike a similar sphere in 1 sec. at the peak of a major solar flare.

Meteoroids are a class of astronomical objects which travel in eccentric orbits about the Sun. A meteoroid may strike the Earth's atmosphere and produce an observable effect called a meteor or "shooting star." The solid body which constitutes a meteoroid is called a meteorite and may vary in mass from 10^{-14} g. to several thousand kilograms. The very light meteorites, with mass of less than about 10^{-6} g., are called micrometeorites or "cosmic dust" and appear to be concentrated in the vicinity of planets. Measurements made by Pioneer 1 and Mariner 2, when compared with measurements made by satellites in the vicinity of the Earth, showed that the flux of dust at great distances from the Earth was almost 1,000 times smaller. The material in space is dominated by particles with a mass of about 10^{-9} g. About one particle of mass 10^{-10} g. strikes an area of 1 sq.m. in 1 sec. near the Earth. The integrated accretion rate over the entire surface of the Earth is about 10,000 tons per day.

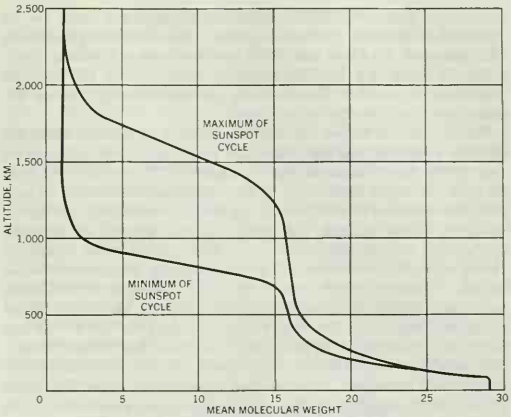
C. The Earth

The Earth, of course, is the planet which has been studied most extensively. It is convenient to divide material in the realm of space science under four major headings: Geodesy, the study of the solid earth; the Upper Atmosphere; the Ionosphere; and the Magnetosphere.

1. Geodesy.—A satellite is a powerful tool for studying the gravitational field of the Earth and for locating points on the surface of the Earth. The position of a satellite in orbit can be determined within limits of about 10 or 20 m. By photographing a passive satellite or a satellite carrying a flashing light it is possible to determine its location relative to the fixed stars. Then, by determining the orbit precisely, studying variations in the orbit and correcting for those variations resulting from atmospheric drag and solar radiation pressure, it is possible to determine the gravitational potential and the exact shape of the Earth. From a knowledge of the gravitational potential it is possible to determine the variation of density within the Earth and to draw some inferences about the internal strength of the Earth.

Simultaneous observations of a satellite can be used to determine precisely the relative locations of the observing stations on the surface of the earth. In this way triangulation measurements can be extended over great distances. (*See* also Geodesy.)

2. Upper Atmosphere.—As altitude increases, the first significant departure from the sea-level composition of the atmosphere occurs at about 30 km., where significant amounts of ozone (O_3) are observed. The upper atmosphere will be considered here as beginning at this altitude and extending up to about 200 km.



REPRINTED FROM "SATELLITE ENVIRONMENT HANDBOOK," ED. BY FRANCIS S. JOHNSON, WITH THE PERMISSION OF THE PUBLISHERS, STANFORD UNIVERSITY PRESS. © 1961 BY LOCKHEED AIRCRAFT CORPORATION

FIG. 17.—VARIATION OF MEAN MOLECULAR WEIGHT OF THE ATMOSPHERE WITH ALTITUDE, SHOWING VALUES AT THE TWO EXTREMES OF THE SUNSPOT CYCLE

Above this point, the properties of the atmosphere are determined by ions and free electrons and will be discussed in the section entitled *Ionosphere.* The upper atmosphere is studied primarily with sounding rockets; balloons are not feasible above 40 km. and satellites cannot be flown below about 200 km.

In order to describe and understand the atmosphere at a particular point, it is necessary to know its composition, temperature, pressure and density. All of these quantities vary with altitude and latitude, the time of day, the season and the solar cycle. An official publication, *A United States Standard Atmosphere,* gives these quantities as a function of altitude for an idealized middle-latitude atmosphere averaged over the yearly and solar-cycle variations.

The variation in temperature with altitude is largely controlled by the absorption of solar radiation. The rapid increase in temperature starting at about 100 km. and extending to about 300 km. marks the region in which ultraviolet radiation is absorbed. No temperature increase is noted above about 300 km. because no large amount of energy is absorbed in this region. Fig. 16 is a plot of the altitude at which about $\frac{2}{3}$ of the photons of a particular wavelength have been absorbed. The longer wavelengths (1000 to 2000 Å) are absorbed below about 50 km. by the ozone in the Earth's atmosphere. The wavelengths between 1000 and 2000 Å are absorbed by the oxygen in the atmosphere in the region between 50 and 150 km. The extreme ultraviolet radiation (500 to 1000 Å) is absorbed in the region between about 140 and 160 km.

Solar radiation is absorbed in the atmosphere by means of several processes, the two most important of which are photoionization and dissociation. In photoionization a neutral atom or molecule absorbs a photon and emits an electron, thereby forming an electron-ion pair. The electron will have an energy determined by the amount of energy required to liberate it from the atom and the amount of photon energy absorbed. In general, the electron will be emitted with an energy greater than the thermal energy of the ambient neutral particles. The electron then loses this energy by one of three processes: (1) inelastic collision with neutral particles; (2) elastic collision with ambient electrons; or (3) elastic collision with neutral particles and ions. The electron may remain a free electron or it may attach itself to a neutral molecule, thereby forming a negative ion. At altitudes below 60 km. electrons rapidly attach themselves to air molecules to form negative ions, the primary effect being an increase in conductivity of the air. At altitudes above 60 km. during the day, the density of the air is so low that the probability of attachment is small;

the rate of production of electrons is high and thus there are an appreciable number of free electrons. At night the production of electrons is very low, and those that are produced attach themselves to molecules, forming negative ions. Above 120 km. the density is so low that the electrons produced during the day remain free electrons during the night.

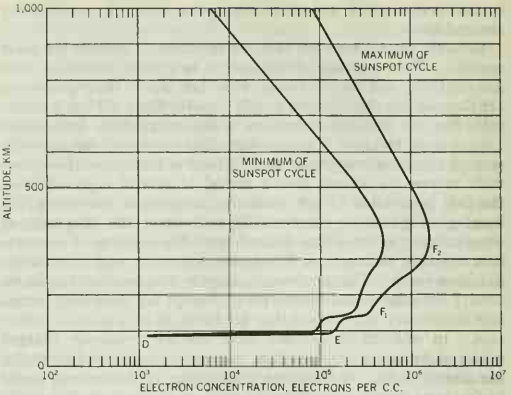
In the second major process, dissociation, a neutral molecule absorbs a photon and separates into two atoms. An example of this is the dissociation of the O_2 molecule into atomic oxygen. At very low altitudes, around 30 km., the neutral atoms of oxygen attach themselves to an O_2 molecule forming ozone, O_3. At higher altitudes, above 90 km., the density is low enough so that the probability of attachment to an O_2 molecule is small compared to the rate of dissociation and there is a large amount of atomic oxygen. The net effect is that there is a variation of the mean molecular weight of air with altitude. Fig. 17 shows the mean molecular weight for the two extremes of the solar cycle. At low altitudes, below about 90 km., the atmosphere has a mean molecular weight corresponding to the normal proportions of molecular oxygen and nitrogen. From 90 km. upward, the mean molecular weight drops rapidly as the oxygen molecules dissociate. The sharp drop in the mean molecular weight at 800 km. at solar minimum and 1,600 km. at solar maximum is due to the presence of the lighter constituents in the atmosphere. In this region the atmosphere is in diffusive equilibrium; therefore, the lighter components tend to separate out and become the dominant constituents at higher altitudes. Measurements on Explorer 8 showed that there is a large amount of helium ions present at about 1,600 km. At the very high altitudes, above 2,500 km., the mean molecular weight corresponds to that of atomic hydrogen.

3. Ionosphere.—One of the major energy-absorption processes in the atmosphere is the production of electron-ion pairs. At altitudes below about 90 km. electrons freed by this process are promptly attached to another molecule, forming a negative ion. Above 90 km. the electrons persist for a considerable length of time: in the regions around 200 km. the high concentrations of electrons persist even through the night. The ionosphere was one of the first regions of the upper atmosphere to be studied, because it reflects low-frequency radio waves that can be produced and detected with elementary equipment. A radio wave propagating vertically from the surface of the earth will be returned from a level in the atmosphere where the electron density and the frequency of the radio wave are related by the formula $N_e = 1.24 \times 10^4 f^2$ electrons per cubic centimetre, where f is the frequency in megacycles per second.

Experimentally it is found that the electron density increases with altitude up to about 350 km. Therefore, the higher the frequency used to probe the ionosphere, the higher the altitude from which the signal is returned. The ionosphere has been studied by sending a variable-frequency radio signal vertically upward and measuring the length of time required for the echo to return. This time is then a measure of the altitude at which the wave is reflected. At altitudes above about 300 km. and a frequency above 10 mc. no echo is returned. At higher frequencies the wave is propagated upward; that is, the ionosphere becomes "transparent."

During the early studies of the ionosphere it was observed that as the probing frequency was increased it would be reflected at a particular level up to a certain limit; then the altitude of reflection would change discontinuously to a higher level, where a wave of higher frequency would be reflected. Thus, it was thought that there were distinct layers, each of higher electron density than the one below. These observations led to the characterization of certain distinct ionospheric layers or regions, the so-called D, E, F_1 and F_2 regions. However, this concept of sharp layers of ionization with little or no free electrons in the intervening regions proved to be wrong as shown in fig. 18. The regions or layers are really locations where the electron concentration reaches a peak, remains relatively constant for some distance, and then again increases.

The D Region.—The D region extends from 60 km. to about 90 km. Photoionization is produced by the Lyman alpha radiation



REPRINTED FROM "SATELLITE ENVIRONMENT HANDBOOK," ED BY FRANCIS S. JOHNSON, WITH THE PERMISSION OF THE PUBLISHERS, STANFORD UNIVERSITY PRESS. © 1961 BY LOCKHEED AIRCRAFT CORPORATION

FIG. 18.—VARIATION OF ELECTRON DENSITY WITH ALTITUDE, SHOWN AT THE TWO EXTREMES OF THE SUNSPOT CYCLE

of hydrogen and the primary ions formed are NO^+. Fig. 16 shows that there is a "window" at the frequency of the Lyman alpha line which makes it possible for this radiation to penetrate to the 60 km. altitude. In addition to the normal D region ionization produced in this way, high-energy solar X-rays emitted at times of solar flares can increase the electron concentration in the D region, and can extend appreciably downward the lowest altitude levels of ionization. Also, protons from solar flares can cause an enhancement of the electron concentrations over the polar regions of the Earth. When such an event as this takes place the increased ionization in the D region completely blacks out radio transmission. It has been shown that the increased ionization is produced by protons in the energy range from a few Mev up to several hundred Mev.

The E Region.—The altitude range from about 90 to 140 km. is designated the E region. It has been established that the "soft" solar X-rays are responsible for the photoionization which occurs in this range. The dominant ionic species in this region are O_2^+ and NO^+. The electron density in the E region is of the order of 10^5 electrons per cubic centimetre at noon during sunspot minimum and is increased by about 50% during sunspot maximum. Electron concentration is highest near local noon and falls off roughly symmetrically with time on either side of noon.

The F_1 Region.—There is no clear resolution between the F_1 region and the E and F_2 regions. The F_1 region is generally and rather arbitrarily defined as the altitude range from 140 to 200 km. The helium 304 line in the solar spectrum is responsible for the photoionization in this region. The photons of the helium 304 line have an energy of 41 electron volts. The energy required to free an electron is of the order of 13 electron volts, and therefore the electrons released in this region have a high energy and share their energy, through the mechanism of elastic collision, with the ambient electrons rather than with the ambient ions. Therefore, the electron temperature is several hundred degrees higher than the ion temperature during the daytime. Electron density in this region is of the order of 5×10^5 electrons per cubic centimetre at noon during sunspot minimum and 2×10^6 electrons per cubic centimetre during sunspot maximum. This region is not usually detected at night, when the electron density falls below 10^4 electrons per cubic centimetre.

The F_2 Region.—The F_2 region is not too well understood. It is thought that the helium 304 line from the Sun is also responsible for ionization there. The reason a second ionization peak is formed in this region is that the recombination rate decreases with altitude more rapidly than does the ionization rate, so that larger ion concentrations occur at higher altitudes.

Above the region of the second peak the electron concentration

decreases as shown in fig. 18. The ionosphere in this region is studied by instruments carried on satellites, the first direct measurements being made by the early Sputniks. Alouette 1, a Canadian satellite, carried a swept-frequency "topside" sounder, an instrument similar to the conventional Ionosonde used on the ground. This satellite transmitted a radio signal in the 1–30 mc. range which was reflected at a depth in the ionosphere given by the formula quoted earlier. In this case, since the transmitter was above the ionosphere, the higher the frequency of the transmitted signal the deeper it penetrated, until finally, at a certain critical frequency, it passed completely through the ionosphere and was reflected from the ground.

4. Magnetosphere.—The magnetosphere is defined as the region of space surrounding the Earth, influenced by the Earth's magnetic field, which begins at an altitude of about 400 km. and extends to about 10 earth radii (64,000 km.) on the sunlit side of the Earth and to perhaps 40 earth radii (256,000 km.) on the dark side. It is the region in space in which charged particles oscillating from the northern to the southern hemisphere have a lifetime that is long compared to the time required for a single oscillation. It extends out to a boundary which is sharp and well defined but whose location in space varies with solar activity and with the angle to the Earth-Sun line.

At the boundary the magnetic field changes direction sharply and decreases in strength. The intensity of the trapped radiation also decreases sharply. The transition region or boundary is only about 70 km. thick. Inside the boundary the magnetic field agrees with that predicted for a dipole magnetic field. Outside, the field is turbulent for a distance of several Earth radii until the conditions of interplanetary space are reached.

A charged particle moving in a magnetic field experiences a force at right angles to its direction of motion and to the magnetic field. As a result of this force the particle will spiral along a magnetic line of force. In a uniform magnetic field the particle would maintain a constant velocity and a constant "pitch angle" between its velocity vector and the magnetic field. If, however, it is in a region in which the strength of the field is changing, the pitch angle and the size of the spiral will change.

It is believed that the protons in the inner region arise from the decay of neutrons produced by the interaction of solar and galactic cosmic rays with air nuclei.

D. The Moon and Planets

The discussions in the articles Moon and Planet are based for the most part upon data from ground observations and theoretical analyses. The program of systematic exploration of the Moon and planets with unmanned and manned spacecraft did not begin until the 1960s. Soviet scientists measured the strength of the lunar magnetic field at the Moon's surface with Lunik 2 and photographs of the far side of the Moon were obtained with Lunik 3 (1959) and Zond 3 (1965). Ranger 7 in July 1964, together with Ranger 8 in Feb. 1965 and Ranger 9 in March 1965, provided more than 15,000 high-resolution photographs of the lunar surface. These photographs showed the lunar surface in the *maria* to be peppered with craters of various sizes and to have a smooth, undulating contour.

The first measurements in the vicinity of a planet were made by Mariner 2, which passed Venus at a distance of 34,760 km. Magnetometers and radiation detectors on board the spacecraft showed no change in the magnetic field and no increase in the radiation level. This implies that Venus has no radiation belt, at least none extending out to 35,000 km., and that any field that it may have does not exceed the interplanetary magnetic field at that altitude. Two experiments to measure the planet's temperature gave a surface temperature of about 427° C. (801° F.).

In July 1965 Mariner 4 passed Mars at a distance of 9,868 km. (6,118 mi.) and transmitted 22 photographs (19 of them useful) of about 1% of the planet's surface to the Earth. The pictures revealed a barren, crater-pocked surface that resembled that of the Moon. Mariner 4 also discovered that the Martian atmosphere was thinner than had been expected, and found no evidence of a magnetic field or of a surrounding belt of radiation.

E. Astronomical Phenomena

The atmosphere absorbs the ultraviolet light from the stars just as it does from the Sun. However, while the Sun emits the bulk of its energy in the visible region of the spectrum, there are stars which emit most of their energy in the ultraviolet portion. In addition to absorbing a major portion of the electromagnetic spectrum, the atmosphere also distorts and scatters a large fraction of the light. For these reasons extremely accurate pointing of large telescopes for long periods of time is required.

The ultraviolet spectra of stars have been measured with instruments on sounding rockets. Preliminary results indicate that intensity in the ultraviolet is below that which would be predicted from our present understanding of stellar processes.

Explorer 11 carried an experiment to measure the flux of high-energy gamma rays. This experiment gave a very low flux of gamma rays and gave no evidence of any discrete sources.

Experiments with sounding rockets have given evidence for discrete sources of X-rays. If subsequent experiments confirm these early results the measurements of the X-ray flux will help to understand stellar processes. (J. E. N.)

V. SPACE SCIENCE: LIFE SCIENCES

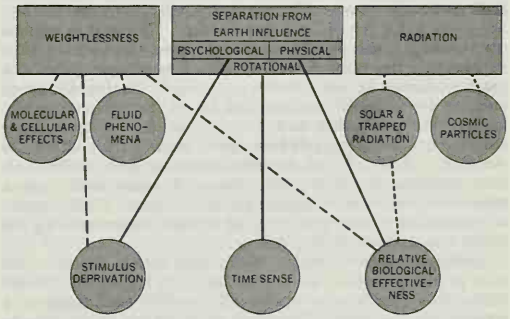
Bioastronautics, a term used to describe the space life sciences, includes acquisition of knowledge in space biology and application of this knowledge to manned space exploration.

Of the technologies which transform basic scientific knowledge into practical use, biologistics is perhaps the broadest. It includes all knowledge, techniques, equipment and supplies required for life support and crew protection, such as environmental control systems, space suits, food, water and waste management devices. This article will treat bioastronautics in two broad categories. The first, concerned with the search for fundamental scientific knowledge, will be called space biology. The second will describe applied bioastronautics in terms of major unsolved technical problems and the special knowledge, procedures and equipment needed to support astronauts in space.

A. Origin of Bioastronautics

Although men have for many centuries speculated on the probable biological effects of the space environment, the first serious investigations of space hazards resulted from a near-disaster in 1862 when two British balloonists, having reached an altitude of about 29,000 ft., narrowly escaped death from acute oxygen deprivation or hypoxia. The French physiologist Paul Bert, stimulated by reports of this experience, began studying respiratory physiology and the biological effects of barometric pressure.

Bert had the first low-pressure chamber suitable for studies on human subjects. His classic observations formed the basis for aviation medicine, a biomedical specialty which developed slowly until the fourth and fifth decades of the 20th century, when revolutionary advances in aeronautics resulted in a broad and vigorous program of research. One of the first units devoted to biomedical



BY COURTESY OF NASA

FIG. 19.—BIOLOGICAL EFFECTS OF THE SPACE ENVIRONMENT, SHOWING INTERRELATIONSHIP OF THE VARIOUS EFFECTS (*see* TEXT AND TABLE VI)

research exclusively in space problems was the department of space medicine, established at the U.S. air force School of Aviation Medicine in 1949 by Harry G. Armstrong, an aviation medicine research pioneer. The rapidly growing body of knowledge in aviation medicine, together with technical developments in aircrew protective equipment, by the time of Sputnik 1 and Sputnik 2 formed the scientific and technological base for bioastronautics in the space age.

B. Space Biology

1. Environmental Biology.—Underlying the many problems in the adaptation of man to space flight are fundamental considerations of the relationship between any living organism and its environment. In a broad sense, the definition of life and its fundamental characteristics are established by this relationship.

Space flight offers an unprecedented opportunity for studying the influence of environment on life characteristics and processes; namely, to determine what effects may proceed from the absence of the normal gravitational force and rotation of the Earth and the presence of ionizing radiation in levels much greater than those on Earth. The absence of the normal gravitational force to which Earth organisms are adapted through evolution may be expected to have effects on the systems in Table VI, with interactions as shown in fig. 19.

Table VI.—*Biological Processes for Study in Space*

A. The role of gravity in:
(1) Plant morphogenesis
(2) Animal morphogenesis
(3) General physiology of cells:
(a) Cell nutrition
(b) Protein and nucleic acid syntheses
B. Effects of weightlessness (lack of gravitational stimulus) on:
(1) Animal co-ordinative processes (*i.e.*, orienting reflexes)
(2) Plant and animal fluid transport (*i.e.*, blood and lymph flow)
(3) Mammalian physiological processes (*i.e.*, homeostatic mechanisms)
C. Space radiation effects:
(1) Biological effectiveness of particles having energy not available from accelerators
(2) Possible synergism between ionizing radiation and other factors in orbital flight (*i.e.*, weightlessness, altered rhythmic processes)
D. Effects of removal from the Earth's rotation on diurnal cycles, time-dependent phenomena, growth of spiral structures

Experimental systems were under development in the 1960s in the biosatellite project of NASA; it was planned to put recoverable spacecraft in orbit for periods of up to 30 days, carrying a wide range of biological experiments involving organisms from bacteria to subhuman primates, to seek answers to the problems outlined in Table VI and fig. 19.

2. Exobiology.—Exobiology is the science of extraterrestrial life. The most revolutionary finding anticipated in the exploration of space would be the discovery of extraterrestrial life or evidence of its preexistence. Aside from its popular impact, this discovery and subsequent investigation would have deep bearing on such topics as the origin of life, organic evolution and even the evolution of the universe.

One of the most widely accepted characteristics of life is the ability to derive energy and matter from the environment and to utilize them for growth and reproduction. Chemical studies on the synthesis of carbon compounds associated with terrestrial life suggest that the synthesis of such substances must occur with great frequency in the universe; therefore, the probability of multiple formation of carbon-based life in the universe is very high. (The probability that life may be derived from other chemical systems is not ruled out, but it is much more difficult to evaluate.) Knowledge of terrestrial evolution leads to the fruitful proposition that life, once it appeared, would evolve into a wide variety of forms to exploit the environment. This proposition raises the distinct possibility of sentient, communicating life forms elsewhere in the universe.

Within the narrow confines of the solar system, however, the most probable forms of life outside the earth are microorganisms (the extant theories of evolution all postulate such forms as a starting point of life). Therefore, exobiology is primarily concerned with the recognition of carbon-based microorganisms on planets of the solar system. Of these planets, Mercury may have too high a temperature for ordinary organic syntheses to occur. Venus, from data derived from Mariner 2, appears to be too hot for conventional biochemistry based on an aqueous medium. Jupiter is too remote for study in the present state of technology.

Despite the apparently negative evidence obtained by Mariner 4, Mars is by far the most attractive prospect for the presence of life processes. The seasonal appearance of dark areas associated with the apparent thaw of polar caps may mean that some life forms, perhaps vegetation, revive as warmth and water become seasonally available.

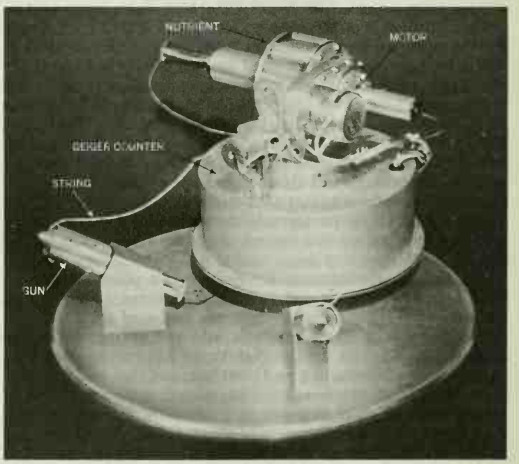
The principal U.S. programs directed toward the investigation of life on Mars are the NASA Mariner and Voyager projects. These include "fly-by" and orbiter missions to gather and record data from electromagnetic radiation either emitted or reflected from the planet by means of spectroscopic analysis, television photography, thermal mapping, etc. Landing capsules were planned that would detect life by identifying some of its most universal characteristics (Table VII); physical-chemical methods were proposed to identify organic compounds. A typical life detection device is illustrated in fig. 20.

Table VII.—*Characteristics or Techniques Useful in the Search for Extraterrestrial Life*

1. Biologically sigoificant elements and diatomic combinations
2. Carbon compounds; organic compounds; optically active compounds
3. Water in liquid form in sufficieot amount to permit activity; theoretical substitutes for water
4. Temperatures less than 100° C.
5. Proteins, nucleic acids and amino acids
6. Special staining or dye reactions typical of macromolecules
7. Growth and metabolism
8. Carbon-dioxide assimilation (photosynthetic and autotrophic bacteria)
9. Enzymes or eozyme activity
10. Structure and organization typical of living systems
11. Gross observations by means of near-scan, high-resolution television
12. Bioelectrical phenomena

C. Problems of Man in Space

1. Biological Space.—A general scientific definition of space appears in the introduction to this article. In bioastronautics, the definition of space is necessarily associated with its relationship to man. For instance, the unadapted human being becomes partially incapacitated from hypoxia when exposed to altitudes above 10,000 ft. Equipped with pressure-breathing oxygen apparatus, an aviator can function effectively up to altitudes of about 45,-



BY COURTESY OF NASA

FIG. 20.—RADIOISOTOPE BIOCHEMICAL PROBE KNOWN AS "GULLIVER." LIFE-DETECTION DEVICE IS 5½ INCHES IN DIAMETER AT THE BASE AND HAS THREE "STICKY STRINGS." FIRED IN DIFFERENT DIRECTIONS BY BULLETS, THEN REELED BACK, CARRYING SOIL PARTICLES AND MICROORGANISMS INTO A NUTRIENT BROTH. BACTERIA, IF PRESENT, WILL GROW IN BROTH, AND APPROPRIATE SIGNALS WILL BE SENT TO EARTH

000 ft. However, at any altitude above 18,000 to 20,000 ft., and particularly above 30,000 ft., there is some risk of an attack of the "bends," or decompression sickness. At altitudes above 45,000 ft. an artificial pressure environment is essential to prevent severe decompression effects and to permit adequate oxygenation.

Where space begins is not determinable from a bioastronautics point of view. From a physiological standpoint, space equivalence in terms of atmospheric pressure may be said to begin at about 65,000 ft. The radiobiologist interested in manned space flight recognizes the vicinity of 90,000 ft. as an altitude at which space radiation begins to be significant.

2. Hazards and Stresses of Space Flight.—The environmental hazards of manned space flight include weightlessness, penetrating radiations, "hard" vacuum, temperature extremes, meteorites, and intense, though transient, accelerative and vibratory forces of launch, orbital injection, in-flight maneuvers, reentry and landing. Other factors which may pose problems of safety or effective performance for space travelers include reduced magnetic fields; empty visual fields; unattenuated solar radiation in the ultraviolet, visible and thermal zones; the absence of the customary terrestrial day-night cycle; and the vastness, desolation and silence of space.

Weightlessness.—The accumulating human experience in the "zero-g" weightless state typical of orbital space flight will show whether or not exposure to this state, prolonged for many days, weeks or months, will produce detrimental biological effects. The Mercury and Vostok flights showed that astronauts were able to carry on normally and efficiently as spacecraft operators. Some evidence has accumulated of adverse effects on arterial blood pressure, seen immediately after landing from orbital flights. A very practical question concerns the possibility that prolonged exposure to weightlessness may so de-condition the astronaut that his cardiovascular system may not be capable of compensating for the increased accelerative stresses of reentry, nor for the circulatory demands of standing or moving about during the first few hours or days back on Earth. Moreover, there is lively scientific interest in the possible influence of prolonged weightlessness on various metabolic and regulatory processes, particularly on calcium metabolism and the possibility of skeletal atrophy, kidney stone formation, and predilection for arterial blood clots as a result of elevated blood calcium levels. Similarly, the possible adverse effects on muscular size and strength must be elucidated. Since it is impossible to simulate prolonged weightlessness in Earth laboratories, the answers to numerous questions about its biological and behavioural effects must come from experience in actual space flight, whether from orbiting animal-carrying satellites or spacecraft with human crews.

Radiation.—It was estimated that the Apollo manned lunar landing and return mission could be accomplished without exposing the crew to radiation of more than 25 rad (radiation absorbed dose; *see* RADIOLOGY: *Therapeutic Radiology: Dosage*). This includes radiation from all sources—the artificial electron belt, the Van Allen Belt, galactic sources and high-magnitude solar flares. An emergency total dose of 150 rad is regarded as a maximum consistent with mission completion and minimum crew safety. This much radiation, however, if absorbed in as short a time as one hour, might produce some symptoms of acute radiation sickness, such as nausea and malaise; however, control of these symptoms with standard drugs is considered feasible.

Since knowledge of the biological effects of high-energy protons and other high-energy space radiations is insufficient to make precise estimates of the hazard to space travelers, laboratory proton accelerators are being used, in part, for proton biological effects studies.

There is a modest hope that at least two types of radiation treatment may evolve to a practical state: antiradiation prophylactic drugs and the use, after radiation exposure, of autologous bone marrow.

More powerful space boosters, capable of orbiting large payloads, will allow for the inclusion in the spacecraft design of considerably more inert radiation shielding than can be carried in, for example, the Gemini spacecraft. In the mid-1960s there was optimism regarding the development and perfection of a magnetic means of deflecting charged, high-energy particles from spacecraft.

High-Intensity Light.—Visible solar radiation in space is so intense that, without protective filters in his glasses, helmet visor or spacecraft windows, the astronaut viewing the Sun directly could suffer temporary blindness or even permanent retinal burns. The form of objects viewed in such unscattered solar light may be difficult to discern accurately because of the high contrast between the illuminated and nonilluminated parts.

Ultraviolet Light.—Unattenuated solar ultraviolet light is another hazard to the space traveler. This is not a serious danger because the space suit will efficiently shield against ultraviolet radiation. There is some concern, however, about the deteriorating effects of ultraviolet radiation on materials used in the construction of spacecraft and space suits.

Temperature Extremes.—The range of temperatures to which an astronaut might be exposed is well illustrated by temperature estimates of the Moon's surface. On the dark side at lunar midnight, which is about 160 hr. after lunar sunset, the surface temperature drops to an estimated −150° C. (−238° F.). Conversely, at lunar midday, it would reach a scorching 180° C. (356° F.). Since these temperature extremes are well beyond the tolerance of an unprotected man, the lunar explorer and the astronaut operating in free space outside the spacecraft require insulated pressure suits with reflective outer layers and built-in, temperature-controlled ventilation.

The Vacuum of Space.—Interplanetary space is an extreme vacuum (with a pressure in the order of 10^{-14} mm. of mercury). Exposed to such a condition, the unprotected human being would have less than 15 sec. of consciousness before the onset of acute hypoxia and immediate, catastrophic decompression symptoms including bends, chokes, palsies and ebullism (the profuse evaporation or "boiling" of body fluids). To protect the astronaut against the vacuum of space, U.S. spacecraft are designed to contain a normal operating pressure of 5 p.s.i.a. (pounds per square inch, absolute) of 100% oxygen, and a minimum emergency pressure of 3.5 p.s.i.a. In addition, the astronaut wears a pressure suit which will maintain an internal pressure of 3.5 p.s.i.a. for as long as may be necessary to correct any environmental control system malfunction or to complete the mission and return to Earth.

Isolation and Sensory Deprivation.—Space flights subject the astronauts to varying degrees of physical and social isolation. Laboratory experiments, as well as the experiences of castaways and explorers, indicate that man, when deprived of the customary sights, sounds, smells, vibrations and thermal and kinesthetic sensations, as well as social interactions, may experience profound psychological disturbances to the point of severely impaired judgment and noneffectiveness. Studies and experience show that communication, as by radio, reduces somewhat the sense of isolation. Space missions with multiple crew members, good communications with Earth stations and demanding work schedules will greatly relieve the relative isolation and reduced level of sensory experience.

Empty Visual Fields.—The so-called empty visual field has been important in aviation because it causes problems of accommodation and visual effectiveness. It has been shown that in the presence of an empty visual field, such as a cloudless sky in broad daylight, subjects cannot relax their visual accommodation sufficiently to assure a visual focus at infinity. Instead, the subject experiences a temporary nearsightedness. This situation interferes radically with effective visual search for targets located at visual infinity. Whether or not this impairment will be important in space operations remains to be proved. It is probable that the stars, planets and Moon will afford some of the needed detail in the distant visual fields to permit accommodation of the eye.

Acceleration.—Throughout a typical space mission the astronaut is subjected to an acceleration environment that differs radically from the customary terrestrial situation. The principal features of this unusual experience are: the intense, short-dura-

tion, positive accelerations of the launch and staging phases; the long period of weightlessness in orbit; and the various accelerative forces of orbital maneuvers; all culminating in acute deceleration during retrofire, atmospheric reentry and the landing impact.

Aerospace medical research has determined that, with the subject semireclined and well supported, the predicted accelerations in all proposed manned flights are not only tolerable but compatible with an acceptable degree of mental and visual effectiveness and competence in using the fingers and hands for control manipulations. Atmospheric reentry of a Gemini-type spacecraft from a low Earth orbit, such as 120 mi., is marked by a period of decelerative force which lasts for several minutes and may reach as high as 8 to 10 g. Fig. 21 shows two reentry deceleration curves which have been proved by centrifuge testing to be within human tolerance when the subject is properly positioned and supported.

3. Biologistics.—This subject includes the characteristic problems that arise in designing and equipping manned spacecraft. Spacecrew requirements, as listed in Table VIII, form the basis for many aspects of spacecraft design and are essential specifications for biologistics. From these requirements are evolved the designs for environmental control systems to regulate cabin temperature, pressure, atmospheric composition and humidity; space suits; acceleration couches and restraining devices; food- and waste-handling equipment; and other equipment needed to sustain the space crew throughout the mission.

Manned space flights can be provided with oxygen, food and water for missions of six weeks or more. Partial recovery of water from human wastes and wash water, and regeneration of oxygen from carbon dioxide, would permit somewhat longer missions. However, for long-duration space flights of many weeks or months, in order to keep within the weight and space limitations imposed by the energy requirements for over-all mission completion, it will be necessary to eliminate all but an initial store of consumable supplies and to depend upon a closed ecological system for waste conversion and recovery of life-supporting essentials. In long-duration, earth-orbiting missions it might be feasible to resupply the spacecraft periodically by logistics vehicles. One of the most fascinating, if not the most difficult, technological challenges of the space age is the development of practical closed ecological system concepts in which essentially all waste products, both man- and machine-generated, could be converted for reuse. (O. E. Re.; J. M. Ta.)

TABLE VIII.—*Spacecrew Requirements*

Area	Requirements	Item
Metabolic	Oxygen consumption	2.0 lb./man/day
	Carbon-dioxide output	2.3 lb./man/day
	Heat output	11,300 B.T.U./man/day
	Water consumption	6.0 lb./man/day
	Food consumption	2,800 kg.cal./man/day
Environmental	Cabin pressure	5 p.s.i.a. and 100% oxygen or 15 p.s.i.a. mixed oxygen-nitrogen
	Space suit pressure	3.5 p.s.i.a.
	Cabin temperature	75° F. ± 5° (rel. hum. 40%–70%)
	In-flight acceleration	(examples, "eyeballs in" or back to front) 10 g for 1.5 min., 9 g for 2 min., 7.5 g for 3 min.
	Noise levels	80 decibels over-all; 55 db in 600–4,800 cps range

VI. MANNED SPACE FLIGHT

Initial Programs.—A goal of both the U.S. and Soviet space programs was the achievement of manned orbital flight. Both nations based their projects on the use of existing ballistic-missile launch vehicles. Because the Soviet ballistic missile had been designed to carry a heavier warhead, their manned spacecraft could also be heavier than its U.S. counterpart; the U.S.S.R. Vostok spacecraft weighed approximately 10,400 lb., while the U.S. Mercury weighed about 3,000 lb.

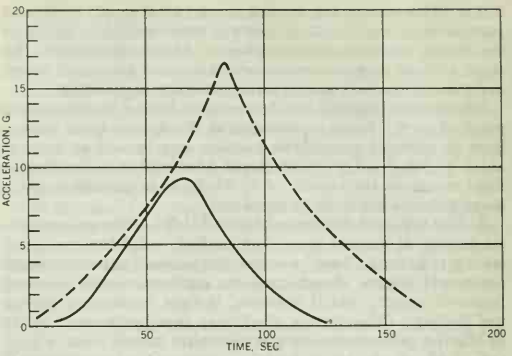


FIG. 21.—TWO EXPERIMENTAL REENTRY DECELERATION CURVES, SHOWING VALUES THAT ARE WITHIN RANGE OF HUMAN TOLERANCE

Soviet Flights.—On April 12, 1961, Yuri Gagarin became the first man to orbit the Earth. In his spacecraft, Vostok 1 (East 1), he circled the Earth once, after a launching near Baikonour, U.S.S.R. On Aug. 6, 1961, Gherman Titov circled the Earth 17 times in Vostok 2, in a 25-hr. flight.

The duration of weightless flight was further extended in a "group flight" by Andrian G. Nikolayev and Pavel Popovich. Nikolayev was launched Aug. 11, 1962, and Popovich was injected into a nearly identical orbit one day later. Both remained in orbit until Aug. 15, Nikolayev completing 64 orbits and Popovich 48. At one point the craft came within 5 mi. of each other but were more than 1,000 mi. apart before reentry.

A second group flight took place in June 1963. On June 14 Valery Bykovsky was launched in Vostok 5. Two days later Valentina Tereshkova, the first woman cosmonaut, was launched in Vostok 6. Both spacecraft returned on June 19, Bykovsky completing 82 orbits and Miss Tereshkova completing 49. In neither group flight was there an attempt at rendezvous.

United States Flights.—The first U.S. manned space flight took place on May 5, 1961, when Alan B. Shepard flew a short-range ballistic trajectory in a Mercury spacecraft boosted by a Redstone launch vehicle. He reached an altitude of 116.5 mi., a speed of 5,100 m.p.h., and covered a distance of 300 mi. The period of weightless flight was about 5 min. On July 21, 1961, Virgil I. Grissom made a nearly identical flight.

The first U.S. orbital flight took place on Feb. 20, 1962, when John H. Glenn, Jr., completed three orbits in his Mercury spacecraft "Friendship 7." M. Scott Carpenter completed a similar flight on May 24, 1962, and Walter M. Schirra, Jr., made a six-orbit flight on Oct. 3, 1962. The last flight in the Mercury series took place on May 15–16, 1963, when L. Gordon Cooper, Jr., circled the Earth 22 times.

Effects of Space Flight.—Results of the first 12 U.S. and Soviet manned flights did not indicate any deleterious effects that could be attributed to space flight, at least for the periods of weightlessness encountered. Titov did report ". . . unpleasant sensations resembling seasickness . . ." during the course of his one-day flight. However, none of the other pilots reported such sensations on even longer flights.

The U.S. astronauts demonstrated repeatedly that a well-trained pilot can greatly enhance the probability of mission success. All U.S. pilots manually controlled the attitude of their craft while in orbit; in at least two cases the lack of this capability would have caused the mission to fail. A decisive demonstration of man's capabilities in space was given by Cooper after a complete failure of his automatic control system. In spite of this failure Cooper controlled his retrofire maneuver with such precision that he landed less than 5 mi. from the primary recovery ship.

2. Gemini.—While the Mercury and Vostok programs demonstrated man's capabilities in space, their scope was too limited to permit the development of operational proficiency in manned space flight. In the United States the Gemini project was designed to continue where Mercury left off.

Objectives.—The Gemini project had two prime objectives: to extend the duration of flight to as long as two weeks, and to develop techniques for performing a rendezvous in space. The two-man Gemini spacecraft resembles the Mercury spacecraft in shape but is somewhat larger; it weighs more than twice as much as Mercury. The Gemini spacecraft is not fitted with the Mercury-type escape rocket; instead, ejection seats are used for emergencies. This is possible since the Gemini booster—the Titan II—uses storable hypergolic propellants which create a smaller explosion hazard than the liquid oxygen and kerosene used in the Atlas.

TABLE IX.—*Manned Space Flights*

Name	Nation	Pilot	Launch date	Weight (lb.)	Perigee/ Apogee (mi.)	Number of orbits	Time (hr.: min.)
Vostok 1	U.S.S.R.	Yuri A. Gagarin	April 12, 1961	10,418	112/203	1	1:48
"Freedom 7"	U.S.	A. B. Shepard	May 5, 1961	2,855	116*	302†	:15
"Liberty Bell 7"	U.S.	Virgil I. Grissom	July 21, 1961	2,836	118*	303†	:16
Vostok 2	U.S.S.R.	Gherman S. Titov	Aug. 6, 1961	10,430	111/160	17	25:18
"Friendship 7"	U.S.	John H. Glenn, Jr.	Feb. 20, 1962	2,987	100/162	3	4:56
"Aurora 7"	U.S.	M. S. Carpenter	May 24, 1962	2,975	99/167	3	4:56
Vostok 3	U.S.S.R.	Andrian G. Nikolayev	Aug. 11, 1962	10,160	114/156	64	94:35
Vostok 4	U.S.S.R.	Pavel R. Popovich	Aug. 12, 1962	10,160	112/158	48	70:57
"Sigma 7"	U.S.	W. M. Schirra, Jr.	Oct. 3, 1962	3,029	100/176	6	9:13
"Faith 7"	U.S.	L. Gordon Cooper, Jr.	May 15, 1963	3,033	100/166	22	34:20
Vostok 5	U.S.S.R.	Valery F. Bykovsky	June 14, 1963	10,160	109/138	82	119:10
Vostok 6	U.S.S.R.	Valentina Tereshkova	June 16, 1963	10,160	114/145	49	70:50
Voskhod 1	U.S.S.R.	Vladimir M. Komarov, Konstantin P. Feoktistov, Boris B. Yegorov	Oct. 12, 1964	11,729	110/254	16	24:17
Voskhod 2	U.S.S.R.	Pavel Belyayev, Alexei Leonov	March 18, 1965	n.a.‡	108/308	17	26:2
Gemini 3	U.S.	Virgil I. Grissom, John W. Young	March 23, 1965	7,031	100/139	3	4:53
Gemini 4	U.S.	James A. McDivitt, Edward H. White	June 3, 1965	7,875	100/175	66	97:56
Gemini 5	U.S.	L. Gordon Cooper, Jr., Charles Conrad, Jr.	Aug. 21, 1965	7,000	100/219	120	190:56

*Maximum altitude, statute miles (ballistic trajectory). †Range, statute miles. ‡n.a.=not announced.

Rendezvous Maneuvers.—In a typical Gemini rendezvous mission, the "target" vehicle is placed into orbit first. The target is an Agena stage, placed into orbit by an Atlas launch vehicle. (The Atlas-Agena used for the Gemini target vehicle is essentially the same as the launch vehicle used in the Ranger and Mariner projects.) If the Agena's orbital period is 90 min., then it again passes over the launch area 16 orbits (or 24 hr.) later. If the Gemini spacecraft is launched at the correct time, it can be inserted into an orbit that is identical to that of the Agena, with the spacecraft and its target flying in close proximity to each other. The final docking, or joining of the spacecraft and its target, is performed by the pilots in the spacecraft, using small rockets to maneuver their craft into position.

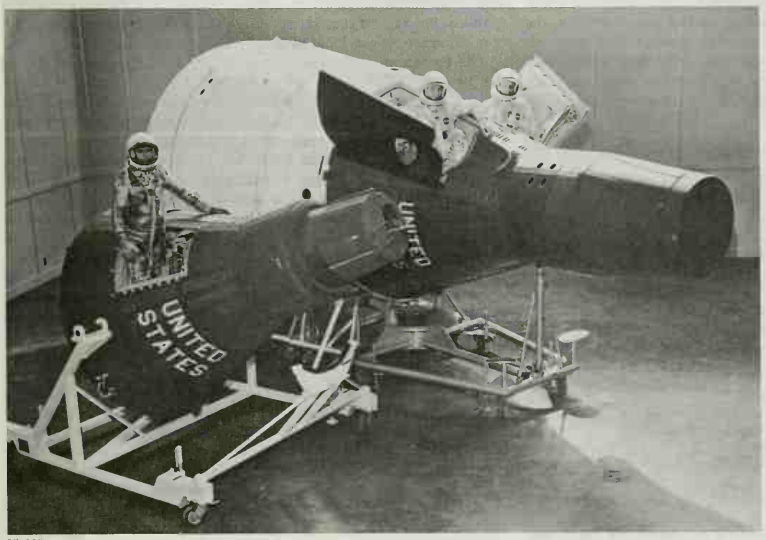
For many practical reasons it may not always be possible to launch the spacecraft at the correct moment to achieve an immediate rendezvous. Under normal conditions this moment can be extended to a "launch window," several minutes long, (by providing sufficient fuel to maneuver the spacecraft) to permit docking even though the target is now farther from the spacecraft.

If the launch window has to be extended even farther, then gross maneuvers are required to complete a rendezvous and docking mission. Normally, the Agena's propulsion system will be used to assist in these maneuvers. If the two craft are in the same orbital plane but are out of phase (*i.e.,* if they pass the same point at different times), the required corrections can be made with a minimal expenditure of fuel. In this case, one of the craft is placed into a more elongated orbit with a longer period; the second craft, in its original orbit with the shorter period, will then overtake the first. The orbit of the first is again shortened after the distance between the two craft has been decreased. These changes of orbits require relatively little propulsion.

If the two spacecraft are injected into different orbital planes, the corrective maneuver becomes more difficult. For a given plane change α, a velocity increase equal to $2V\sin\frac{\alpha}{2}$ is required, where V is satellite velocity. Thus, for a 5° plane change, a velocity increment of about 2,200 ft. per second would be required, while for a 60° change an increment equal to satellite velocity (about 25,500 ft. per second) is needed. Even with Agena propulsion, only minor plane changes are possible in Gemini. A rendezvous maneuver, therefore, contains the following: launch within a specified "window," plane change, phase changes and docking.

On March 23, 1965, astronauts Virgil ("Gus") Grissom and John W. Young made the first manned flight in a Gemini spacecraft, a three-orbit mission. The second manned Gemini flight, a four-day mission, took place in June 1965, with astronauts James A. McDivitt and Edward H. White. White left the spacecraft for about 22 min. during the third orbit, and maneuvered freely in space at the end of a 25-ft. tether. In Aug. 1965 L. Gordon Cooper and Charles Conrad circled the Earth for eight days in the third manned Gemini flight.

3. Voskhod.—The first Soviet multi-manned space flight took place on Oct. 12–13, 1964, when Voskhod 1 ("Sunrise 1") orbited the Earth 16 times. Piloted by Vladimir M. Komarov, the craft also carried Konstantin P. Feok-



BY COURTESY OF NASA

FIG. 22.—MERCURY AND GEMINI (TWO-MAN) SPACECRAFT, SHOWING COMPARATIVE SIZES

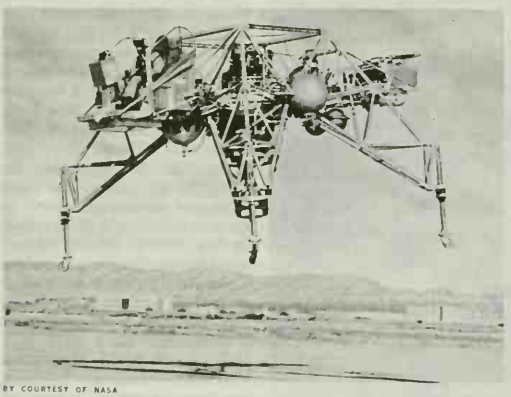
tistov, an engineer-scientist, and Boris B. Yegorov, a physician. The crew wore standard clothing rather than pressure suits during this flight; a landing system comprised of parachutes and braking rockets permitted the craft and its occupants to land on the ground. Voskhod 2, a two-man vehicle piloted by Pavel Belyayev, was launched on March 18, 1965. During this flight Alexei Leonov, the co-pilot, left his spacecraft for a short period of time, becoming the first man to float freely in space. Voskhod 2 apparently developed difficulties at the end of the 16th orbit and safely completed an emergency landing one orbit later.

4. Lunar Landing Program.—In May 1961 the United States announced its plans to land a manned spacecraft on the Moon by the end of the 1960s. The name given to this program was Apollo.

Lunar Landing Techniques.—Several techniques for the lunar mission appeared to be feasible and were studied in detail. These included the "direct method," wherein a single giant launch vehicle (Nova) would launch a spacecraft toward the Moon. Near the Moon's surface the entire spacecraft would be slowed down by its own propulsion system, called the lunar landing stage. Another propulsion system, the lunar takeoff stage, would launch the spacecraft from the Moon and send it on its return trajectory toward Earth. After several mid-course corrections, the spacecraft would reenter the Earth's atmosphere at a speed of 36,000 ft. per second, nearly 1½ times satellite velocity.

A second proposal, called the "earth-orbit rendezvous" method, differed from the direct method in that two somewhat smaller launch vehicles (Saturn V) would separately transport two parts of the spacecraft (*i.e.*, the lunar landing stage as one part, and the lunar takeoff stage and spacecraft proper as the second part) into Earth orbit. There they would be joined and then propelled as a single craft toward the Moon. The remainder of the mission would be the same as with the "direct method."

A third method, called the "lunar orbit rendezvous" method, would employ a single Saturn V launch vehicle to send the three-man Apollo spacecraft, weighing 90,000 lb., on a trajectory toward the Moon. Approximately 2½ days later, as the craft approached the Moon, its own propulsion system would place it into an orbit around the Moon. In lunar orbit one element of the spacecraft, called the lunar excursion module (LEM), with two men aboard, would separate from the mother ship. The two men would then fly the LEM toward the Moon, while the third man circled the Moon in the other section of the spacecraft, the "command module." After executing a "soft" landing on the lunar surface, the two lunar explorers would make scientific observations and collect geological samples. After about one day, they would take off from the surface and fly toward a rendezvous with the command module in lunar orbit. After the rendezvous, spacecraft propulsion would eject the spacecraft from lunar orbit and send it back toward the Earth. Precise guidance and control on the return trip would allow the spacecraft to enter within the limits of a rather narrow "reentry corridor." This corridor, about 25 mi. high, is limited at the lower side by too thick an atmosphere, which would cause the craft to exceed acceptable deceleration limits, and at the upper by so thin an atmosphere that the craft would not be "captured" and therefore would not reenter. Within the atmosphere the Apollo command module will have a limited degree of maneuverability. Although its lift-to-drag ratio is only about 0.34, the reentry path will be 1,000 mi. long and 200 mi. wide as a result of the high entry velocity.



BY COURTESY OF NASA

FIG. 23.—LUNAR LANDING RESEARCH VEHICLE BEGINNING ITS FIRST TEST FLIGHT AT THE NASA FLIGHT RESEARCH CENTRE, EDWARDS AIR FORCE BASE, CALIF., OCTOBER 1964

5. Manned Spacecraft Design.—Two factors are of major importance in the design of manned spacecraft: weight and reliability. The need for minimum weight becomes apparent when it is recognized that, for a manned lunar mission, 750 lb. of launch vehicle weight is required for every pound returned to Earth at the end of the mission; or, since the takeoff thrust is about 25% higher than the takeoff weight, an additional 1,000 lb. of thrust is needed for each added pound of spacecraft weight. Reliability is generally assured by providing redundant elements where a failure would critically affect the success of the mission. In manned missions the pilot can significantly enhance reliability by detailed failure analyses and corrective actions.

External Characteristics.—The external shape and materials of a manned spacecraft are governed by the necessity of reentering the Earth's atmosphere at speeds approaching satellite or escape velocity, depending on the type of mission. The tremendous heat generated during reentry is dissipated either by radiation from the craft's surface, or by ablating (evaporating or boiling away) some of the surface material. For example, in the Mercury and Gemini spacecraft (fig. 22), the major portion of the heat is dissipated by ablation from the blunt forward-facing heat shield; heat is also rejected from the conical afterbody, where thin (0.016-in.), high-temperature alloy "shingles" quickly reach an equilibrium temperature and give off heat by radiation.

The blunt shape of the forward face of the Mercury, Gemini and Apollo spacecraft also helps in the heat rejection process. This shape creates a strong shock wave which in turn dissipates much of the craft's kinetic energy; therefore, only a small part of the total energy is given off as heat at the vehicle's surface.

Major Subsystems.—Subsystems for providing a habitable atmosphere, for generating electric power, for determining and maintaining the correct attitude, flight path and velocity, and for communications are located inside the spacecraft.

The environmental control system provides oxygen (or air) at the proper temperature, pressure and humidity, and removes carbon dioxide and other contaminants. In all U.S. spacecraft, oxygen at a pressure of 5 p.s.i. is used for breathing purposes; in the U.S.S.R. Vostok spacecraft, a standard mixture of oxygen and nitrogen at a pressure of 14.7 p.s.i. is employed. Although the cabin atmosphere itself is habitable, the pilot is further protected by a space or pressure suit which allows him to survive in case the cabin is punctured by a meteoroid or is otherwise damaged.

Electric power sources include ordinary batteries, solar cells and hydrogen-oxygen fuel cells. For short missions, up to two or three days in duration, ordinary batteries are acceptable; for longer missions, solar cells or fuel cells can generate greater amounts of power with lower weight requirements. Fuel cells have the added advantage that they produce potable water as a by-product.

The spacecraft's attitude is stabilized with small rocket motors, or thrusters. The thrusters can be fired automatically or manually by the pilot through the use of a control similar to an airplane control stick. Whichever method is chosen, the necessary information concerning the craft's attitude comes from an autopilot, which uses gyroscopes to measure attitudes and attitude changes. In Earth satellite missions it is important to stabilize the craft's attitude precisely about all axes of rotation just prior to retrofire, so that reentry will be made in the proper direction. In rendezvous missions and in lunar missions, attitude control also

must be maintained while making course corrections and during docking maneuvers.

Guidance and navigation systems are employed to find the spacecraft's position and velocity in space, to determine course corrections that must be applied to follow the desired trajectory, and to measure changes in speed and direction while maneuvers are made. A typical guidance system consists of an inertial measurement unit (an array of gyroscopes and accelerometers) used to measure the acceleration vectors relative to a fixed point in space, and a guidance computer to translate acceleration measurements into trajectory data and to determine required course corrections. With this information the pilot then applies thrust in the proper direction.

Communications systems in manned spacecraft include voice receivers and transmitters, telemetry receivers and transmitters for sending measurement data, and radar beacons for transmitting tracking information to ground network stations. (G. M. Lo.)

VII. APPLICATIONS

Satellites have two properties that make them uniquely useful: a satellite provides a view of a large portion of the Earth's surface at any instant, and can also view all of the surface over a period of time if it is in a near-polar orbit; and, conversely, a satellite is mutually visible from a large number of widely separated points on the Earth's surface.

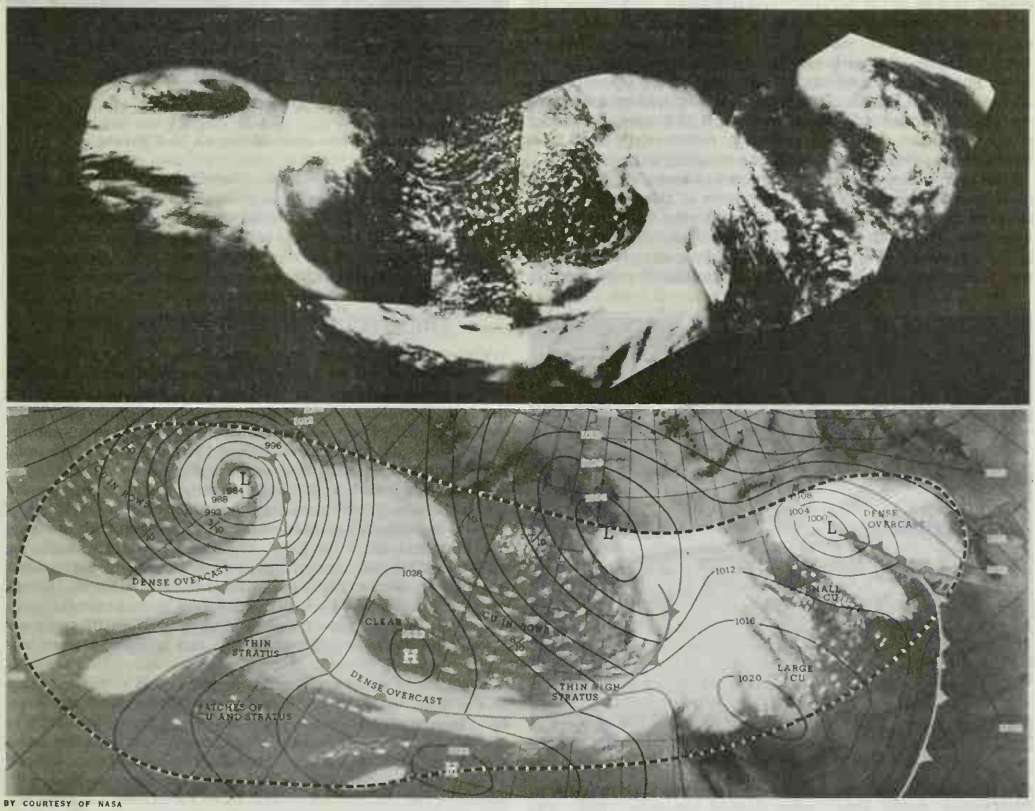
1. Meteorological Satellites.—Satellites such as the Tiros series provide the means to monitor the Earth's cloud cover and temperature distribution. This information is telemetered to data-acquisition stations on the Earth in the forms of television pictures and temperatures as measured with infrared-sensitive radiation detectors. (*See* fig. 24.)

Since weather is affected by conditions over a large portion of the Earth and by the solar energy input to the Earth, it is important that meteorological data be obtained over the entire surface of the Earth. Complete global monitoring was previously impossible because of the large number of observation stations required and inability to place such stations in many areas, particularly in the ocean areas where severe weather disturbances originate.

Satellites such as Tiros provided coverage of some of these meteorological events for the first time. Satellite observations have led to the discovery of severe storms, hurricanes and typhoons long before they could be detected by conventional means.

They have also led to a new understanding of the tremendously large area of influence of certain weather conditions. There was little doubt that this new capability to observe from above, and over the entire surface of the Earth, would result in better information on which to base weather forecasting and lead to a more thorough understanding of the nature of the weather processes, all of which would make the interpretation of observations more accurate.

2. Communications Satellites.—Another promising application of satellites, which takes advantage of the property that satel-



BY COURTESY OF NASA

FIG. 24.—STORMS AND FRONTS AS SEEN BY TIROS. (TOP) MOSAIC OF TIROS PHOTOGRAPHS; (BOTTOM) WEATHER MAP FOR MAY 20, 1960, PREPARED BY USING TIROS CLOUD DATA

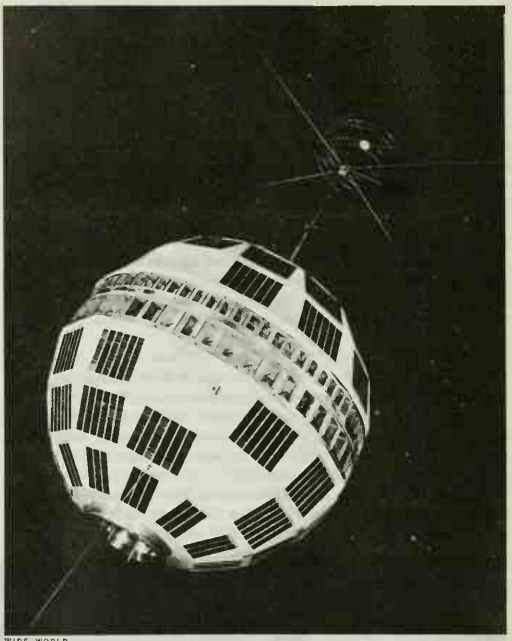
lites can be seen simultaneously from widely separated points on the earth, was first proposed in 1945 by Arthur C. Clark. He suggested the use of a manned satellite station for relaying communications over large distances. Although they were unmanned, the Echo, Telstar, Relay and Syncom satellites demonstrated the utility of this means of communication.

Satellite Coverage.—Satellites such as Telstar and Relay are in intermediate-altitude orbits (3,000 to 6,000 mi. high) and therefore move with respect to the Earth. Thus, to provide continued communications between two widely separated points would require that satellites passing out of range be replaced by other satellites just coming into view. From 30 to 50 satellites would be required at these altitudes to provide essentially continuous communications, if the satellites were distributed randomly with respect to each other in their orbits.

The Syncom satellite was placed in an orbit at an altitude of 22,245 mi. At this altitude the period of rotation of the satellite about the Earth is exactly 24 hr. By selecting an equatorial orbit and a period that is the same as the Earth's rotational period, the satellite remains stationary or fixed relative to the surface of the Earth. Thus a single "synchronous" satellite of this type would be available continuously for communications between any two stations within its area of visibility, and three such satellites, suitably spaced, could provide global coverage with the exception of the extreme polar regions.

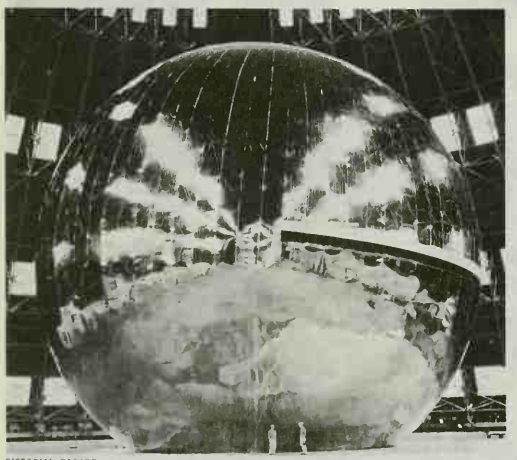
Types of Communications Satellites.—Telstar, Relay and Syncom are all "active" satellites; *i.e.*, they carry transmitter-receiver sets to receive, amplify and retransmit signals from the ground, as well as the necessary power supplies. There are also "passive" communications satellites which carry no electronics equipment but merely reflect a signal directed at them back toward the Earth. Examples are the 100-ft. inflatable sphere Echo 1 which was visible as a rapidly moving "star" in the evening skies, and its successor, Echo 2. Passive satellites require larger receiving antennas on the ground and larger amounts of ground transmitter power than active satellites because they reflect only a small fraction of the incident energy back to the receiving antenna.

Commercial Exploitation of Communications Satellites.—The successful development and demonstration of these experimental satellites provided the technical basis for the Communications Satellite corporation (Comsat), formed in 1962. This corporation is in the process of establishing an interim commercial satellite communication capability across the North Atlantic with the first launch of Early Bird, a synchronous satellite, in April 1965. Starting in 1966 the corporation plans to establish a truly global system of satellites to provide worldwide communications service on a commercial basis. The system will be operated by an international consortium of common carriers and represents the first large-scale commercial exploitation of the space technology that has been developed by the U.S. space program. (L. Ja.)



WIDE WORLD

FIG. 26.—TELSTAR 1, FIRST PRIVATELY OWNED COMMUNICATIONS SATELLITE, WAS LAUNCHED IN 1962, MAKING POSSIBLE THE TRANSMISSION OF LIVE TELEVISION PROGRAMS BETWEEN THE U.S. AND EUROPE



PICTORIAL PARADE

FIG. 25.—ECHO 1, A HIGH-ALTITUDE REFLECTOR FOR LONG-RANGE COMMUNICATIONS, SHOWN HERE DURING GROUND INFLATION TESTS. THE ALUMINUM-COATED BALLOON, WHICH INFLATES TO 100 FT. IN DIAMETER, WAS LAUNCHED AUG. 12, 1960

VIII. INTERNATIONAL SPACE ACTIVITIES

The space age began in an atmosphere of international cooperation. Following suggestions made in Rome in 1954 by the central planning body, Comité Spéciale de l'Année Géophysique Internationale (CSAGI), for the International Geophysical Year, the U.S. and the U.S.S.R. announced plans for artificial scientific satellites. Both countries, plus the United Kingdom, Canada, Japan, Australia and France (unofficially), also organized programs for the launching of small, scientific sounding rockets.

CSAGI subsequently considered resolutions for the exchange of satellite and sounding-rocket information consistent with the open, scientific character of the IGY. The western members sought the widest possible exchange of information relating to experiments and results. Soviet scientists preferred to restrict the resolutions essentially to published scientific results. The fact that the tools, techniques and certain of the research objectives of space research were common to both scientific and military interests was clearly a limiting factor.

Scientists of countries concerned with satellites or sounding-rocket work, including tracking, met with delegates from interested international scientific unions in London in Nov. 1958 to draft a charter for a permanent International Committee for Space Research (COSPAR) to function under the International Council of Scientific Unions. The meeting was convened by Homer E. Newell of the United States. The group formulated a charter and selected H. C. Van de Hulst of the Netherlands as president. Representatives of the U.S.S.R. Academy of Sciences participated

in the drafting of the charter, but at the next meeting of COSPAR, in March 1959 in The Hague, they alleged that the charter gave inadequate "representation" to the eastern bloc.

Other delegates strongly opposed overt consideration of political factors in the organization of a nongovernmental scientific body. A compromise provided for two vice-presidents, one each from the U.S. and the U.S.S.R., with additional members of COSPAR's executive committee to be selected from slates proposed by the two countries.

COSPAR subsequently met amicably in Nice, France (1960), Florence, Italy (1961 and 1964), Washington, D.C. (1962), and Warsaw, Pol. (1963). Little change was made in the original IGY patterns for exchange of data and notification of launchings. While other nations went well beyond the IGY-COSPAR conventions for co-operation, the Soviet Union in general continued to publish scientific results often lacking sufficient detail for independent evaluation, listed "look-angles" for satellites at specific locations but failed to provide their orbital elements, and furnished only general information on numbers of sounding rockets launched in various areas, without listing experiments or results. Catalogues of rocket-satellite data deposited in the IGY World Data centre in Moscow were to have been provided at six-month intervals but were not made available after 1960. Only an approximate compliance with the fundamental requirement for information of scientific results can be said to have been obtained.

Beginning in 1962 COSPAR attempted to modify its structure so as to make a positive contribution to the organization of useful space research. Significant progress became evident by 1963 in the planning of synoptic sounding-rocket programs in meteorology, ionospherics and aeronomy.

Another forum, principally for annual symposia on engineering aspects of space research, exists in the International Astronautical federation, which brings together a considerable number of national rocket societies with broad and long-standing interests in the promotion of space exploration.

1. U.S. Programs of Co-operation.—The charter of the National Aeronautics and Space administration provided, among its objectives, for co-operation with other nations in space activities. Co-operation was offered by NASA with officially sponsored, central civilian agencies in other countries, with literal sharing of the requirements for specific projects of scientific validity and mutual interest, without exchange of funds. The program included:

Joint Satellite Projects.—The first international satellites were Ariel and Alouette, the former instrumented in the U.K. and engineered by NASA, the latter prepared entirely in Canada, and both launched successfully in 1962 by NASA. By the end of 1964 a second British and a first Italian satellite had been successfully launched, the latter by an Italian team using U.S. launch facilities. Additional co-operative satellites were planned by British, French and Italian national space councils. Under a 1964 agreement, NASA undertook the launching, in 1967, of two satellites to be prepared by the European Space Research organization (*see* below).

Individual Satellite Experiments.—A number of experiments proposed by scientists outside of the U.S. and judged in competition with U.S. proposals were selected by NASA for incorporation in large observatory satellites to be launched over a period of several years.

Sounding Rocket Experiments.—Australia, Canada, India, Italy, Japan, New Zealand, Pakistan, Denmark, Norway, Sweden, France, the Netherlands and Argentina co-operated with NASA in the launching of scientific sounding rockets, from sites in several countries. Responsibility was divided for the essential components: sounding rockets, payloads and ground equipment.

Ground-Based Experimental Participation.—More than 40 countries carried out weather observations made with aircraft, balloons or ground instruments, synchronized for correlation purposes with cloud photographs obtained by NASA's Tiros weather satellites; 10 countries established major ground terminals for experimental communications by means of NASA communications satellites. Scientists in many other countries co-ordinated important observations with satellites investigating the ionosphere.

Tracking and Data-Acquisition Operations.—About two-thirds of NASA stations outside the U.S. are operated wholly or in part by citizens of the host countries, which, in a number of cases, also provide and support the technical staff.

Training.—Three NASA programs offer (1) graduate fellowships at U.S. universities for students sponsored by their central space agencies; (2) training arrangements at NASA centres in support of agreed co-operative projects; and (3) associateships at NASA centres for gifted senior scientists.

2. U.S.–U.S.S.R. Relationships.—United States representatives attempted during the IGY to broaden exchanges between the two major space powers but met with little success. Other bilateral overtures were made privately, and in 1960 the administrator of NASA publicly offered tracking services by Project Mercury stations for future Soviet manned flights. The offer was not accepted.

In Feb. and March 1962 Pres. John F. Kennedy responded to routine congratulations from Chairman Nikita S. Khrushchev, on the occasion of the first U.S. manned orbital flight, by proposing specific co-operative projects and designating Hugh L. Dryden, deputy administrator of NASA, as chief negotiator. Khrushchev designated Soviet academician A. A. Blagonravov to meet with Dryden. By June 1962 their discussions produced an agreement calling for co-ordinated weather satellite launchings and data exchange, a complementary satellite survey of the Earth's magnetic field, and demonstration of experimental space communications by means of the U.S. satellite Echo 2. The agreement was confirmed by the Soviet foreign office and the U.S. department of state. By Aug. 1963 a detailed implementing agreement became effective.

With increasing strength in the U.S. space program and the easing of international tension following the nuclear test ban treaty of 1963, President Kennedy was encouraged to suggest, in his speech to the general assembly of the United Nations in Sept. 1963, that the two countries explore co-operation in their manned lunar programs. U.S. Pres. Lyndon B. Johnson reaffirmed this proposal in Dec. 1963.

The bilateral agreement of 1962 led, early in 1964, to communication tests by means of Echo 2, conducted through transmissions from a U.K. station to a Soviet station. The Soviet side provided data on the tests. By Oct. 1964 a conventional communications link for exchanging weather data was operating between Washington and Moscow on a shared cost basis. Conventional weather data was being exchanged twice daily but satellite data was not yet exchanged since it was not yet available from the Soviet side.

3. European Regional Co-operation.—In the period 1961–63, 9 countries established a European Space Research organization (ESRO) to develop scientific satellites and employ sounding rockets. By late 1963 plans had been made for a budget of $300,000,000 over an eight-year period, and for the establishment of a data centre in Darmstadt, Ger., a technical centre in Delft, Neth., and a test centre in Italy, as well as administrative headquarters in Paris. ESRO member countries were the United Kingdom, France, West Germany, Spain, Switzerland, Denmark, Sweden, Belgium and the Netherlands.

A second agency, the European Launcher Development organization (ELDO), was organized during the same period. Consisting of seven countries, including Australia, ELDO sought to develop a European satellite booster using the British Blue Streak vehicle as a first stage, a French second stage and a German third stage. Belgium, the Netherlands and Italy were the other participants. ELDO's budget called for expenditures of about $200,000,000 over five years. Its launching range was planned at Woomera, Austr. The first stage of the Blue Streak was successfully flight-tested in 1964.

4. UN Activities.—Following a U.S. proposal, the United Nations established an *ad hoc* committee on the peaceful uses of outer space in 1959. The Soviet Union, the United Arab Republic and India abstained, essentially on issues of political representation and procedure. The committee prepared a report describing the opportunities for useful activity by UN specialized agencies such as the World Meteorological organization and the International Telecommunications union, stating opportunities for the

development of space law and seeking to encourage co-operation among nations in peaceful space research. By late 1961 a formula was arrived at which expanded the committee, increasing the representation of the eastern bloc, and provided for the committee to attempt to do its business on the basis of unanimity. Designated the UN Committee on the Peaceful Uses of Outer Space, it established two subcommittees, one on scientific and technical matters, the other on legal matters. It was agreed that the committee's function would be to facilitate co-operation among nations rather than engage in operational activities. The Scientific and Technical subcommittee met in Geneva, Switz., in mid-1962 and recommended the publication of information on national program activities, requested support for specific programs of the international scientific community and provided for possible UN sponsorship of an international sounding-rocket range to be established on the geomagnetic equator. India offered to establish such a range, then in preparation through bilateral arrangements with the United States. The parent committee unanimously approved the Scientific and Technical subcommittee's report in September, and the general assembly added its approval later in the fall of 1962. Although the subcommittee reconvened in Geneva in the spring of 1963 and again in 1964, little more was accomplished other than the receipt and preparation of various reports. The international rocket sounding range at Thumba in India did, however, progress well, receiving U.S., French and Soviet contributions in the form of equipment. While reflecting Soviet political inhibitions in part, the absence of further substantive achievement also could be ascribed to the limited scope available at the time for UN activity within its prescribed nonoperational function.

5. Prospects.—Despite political and security difficulties, unexpected progress was made in the first six years of the space age in bringing nations together in useful co-operation of more than token character. The bulk of substantive activity followed from the international co-operative projects of NASA. However, other centres of international collaboration began to appear. The regional organizations in Europe, ELDO and ESRO, mapped ambitious programs. The International Committee on Space Research (COSPAR) began to generate useful operational activities pursuant to recommendations made by certain of its working groups. Most promising in potential were the first steps taken toward significant co-ordination of effort in this field by the Soviet Union and the United States. (A. W. Fr.)

IX. SPACE LAW

It was apparent from the outset of the era of space flight that the legal status of outer space must differ essentially from that of air space. It is a universally recognized principle of international law that every nation has complete and exclusive sovereignty over the air space above its territory and therefore possesses the right to exclude aircraft from its air space and to determine unilaterally the conditions of entry and use, unless a different rule has been prescribed by an international agreement to which the nation is a party. The vertical extent of the air space has never been defined by international agreement and there is no agreed limit upon the upward reach of the area within which the underlying nation may exercise exclusive control.

If the rule of exclusive sovereignty which prevails in the air space were to extend to altitudes at which orbital flight can be accomplished, it would be necessary to secure the prior consent of nations over which a satellite would pass. In apparent reliance, however, on a basic difference between the legal status of air space and the legal status of outer space, neither the United States nor the Soviet Union sought the prior consent of other nations before sending its satellites aloft. No objection was registered by any other nation, and no nation contended that the flight of such spacecraft over its territory required its prior permission.

In addition to the question of the legal status of outer space, the related question of the legal status of celestial bodies was raised by the prospect of exploration and exploitation of the Moon and the planets of the solar system.

Both of these questions were the subject of Resolution 1721 (XVI) adopted unanimously by the general assembly of the United Nations on Dec. 20, 1961. The resolution, which was the first authoritative statement of legal principles applicable to space activities, "commends to States for their guidance" the principles that "international law, including the Charter of the United Nations, applies to outer space and celestial bodies" and that "outer space and celestial bodies are free for exploration and use by all States in conformity with international law and are not subject to national appropriation."

Resolution 1721 (XVI) also requested the United Nations Committee on the Peaceful Uses of Outer Space "to study and report on the legal problems which may arise from the exploration and use of outer space." During the two years after the adoption of the resolution, a consensus on certain general principles emerged from the deliberations of the committee and its Legal subcommittee.

These principles were given formal expression in Resolution 1962 (XVIII), entitled "Declaration of Legal Principles Governing the Activities of States in the Exploration and Use of Outer Space," which was adopted unanimously by the general assembly on Dec. 13, 1963.

In addition to restating the principles enunciated in Resolution 1721 (XVI), Resolution 1962 (XVIII) declared that states bear international responsibility for national activities in outer space, whether carried on by governmental agencies or by nongovernmental entities; that the activities of nongovernmental entities in outer space require authorization and continuing supervision by the state concerned; that each state which launches or procures the launching of an object into outer space, and each state from whose territory or facility an object is launched, is internationally liable for damage caused by such object to a foreign state or to its natural or juridical persons on the Earth, in air space or in outer space; that the state of registry of an object launched into outer space retains jurisdiction and control over the object and its personnel while in outer space; that ownership of objects launched into outer space is not affected by their passage through outer space or by their return to Earth, and there is a duty to return any such object to the state of registry regardless of the place where found; and that personnel engaging in space flight who make a landing on the territory of a foreign state or on the high seas because of accident or emergency shall be safely and promptly returned to the state of registry.

Resolution 1962 (XVIII) recognized that activities in outer space might be carried on by international organizations, and it provided that in such cases responsibility for compliance with the principles set forth in the resolution shall be borne by the international organization and by the states participating in it. Taking account of the possibility of harmful interference being caused by a particular space activity with the activities of other states in the peaceful exploration and use of outer space, the resolution declared that if a state has reason to believe that an outer space activity planned by it or its nationals would have such an adverse effect, it shall undertake appropriate international consultations before proceeding with any such activity.

Although both Resolution 1721 (XVI) and Resolution 1962 (XVIII) declared the principle that outer space is free and is not subject to national appropriation, neither resolution attempted to define at what altitude outer space begins or air space ends. The problem presented by the apparent juxtaposition of two realms governed by antithetical legal regimes was left by the general assembly for future consideration. It appeared that the future development of manned space flight, involving extensive flight during return to earth at altitudes below that at which orbital flight can occur but above that at which the flight of conventional aircraft has taken place, eventually would require a more precise treatment of this question.

See also references under "Space Exploration" in the Index.

(J. A. Jo.)

Bibliography.—Arthur Clarke, *The Exploration of Space* (1959); H. E. Newell, *Express to the Stars* (1961); S. Ramo (ed.), *Peacetime Uses of Outer Space* (1961); Krafft A. Ehricke, *Space Flight* (1962); R. Jastrow (ed.), *Exploration of Space* (1960); R. M. L. Baker and M. W. Makemson, *An Introduction to Astrodynamics* (1960); G. P. Sutton, *Rocket Propulsion Elements* (1956); H. S.

Seifert (ed.), *Space Technology* (1959); H. H. Koelle, *Handbook of Astronautical Engineering* (1961); R. B. Morrison and M. J. Ingle (eds.), *Design Data for Aeronautics and Astronautics* (1962); Francis S. Johnson, *Satellite Environment Handbook* (1961); John A. Ratcliffe (ed.), *Physics of the Upper Atmosphere* (1960); John C. Houbolt, "Lunar Rendezvous," *International Science and Technology*, No. 14, p. 62 (Feb. 1963).

SPACE-TIME. The theory of relativity has brought about a fundamental change in the scientific conception of space and time, described in a famous saying of Hermann Minkowski—"From henceforth space in itself and time in itself sink to mere shadows, and only a kind of union of the two preserves an independent existence." This union is called space-time. In this article, the concept itself is discussed first. The method of determining the structure of space-time is described in the article RELATIVITY. A summary of the results is given below, under *Recent Developments*.

All our thoughts and concepts are called up by sense-experiences and have a meaning only in reference to these sense-experiences. On the other hand, however, they are products of the spontaneous activity of our minds; they are thus in no wise logical consequences of the contents of these sense-experiences. If, therefore, we wish to grasp the essence of a complex of abstract notions we must for the one part investigate the mutual relationships between the concepts and the assertions made about them; for the other, we must investigate how they are related to the experiences.

So far as the way is concerned in which concepts are connected with one another and with the experiences there is no difference of principle between the concept-systems of science and those of daily life. The concept-systems of science have grown out of those of daily life and have been modified and completed according to the objects and purposes of the science in question.

The more universal a concept is, the more frequently it enters into our thinking; and the more indirect its relation to sense-experience, the more difficult it is for us to comprehend its meaning; this is particularly the case with prescientific concepts that we have been accustomed to use since childhood. Consider the concepts referred to in the words "where," "when," "why," "being," to the elucidation of which innumerable volumes of philosophy have been devoted. We fare no better in our speculations than a fish which should strive to become clear as to what is water.

SPACE

In the present article we are concerned with the meaning of "where," that is, of space. It appears that there is no quality contained in our individual primitive sense-experiences that may be designated as spatial. Rather, what is spatial appears to be a sort of order of the material objects of experience. The concept "material object" must therefore be available if concepts concerning space are to be possible. It is the logically primary concept. This is easily seen if we analyze the spatial concepts for example, "next to," "touch," and so forth, that is, if we strive to become aware of their equivalents in experience. The concept "object" is a means of taking into account the persistence in time or the continuity, respectively, of certain groups of experience-complexes. The existence of objects is thus of a conceptual nature, and the meaning of the concepts of objects depends wholly on their being connected (intuitively) with groups of elementary sense-experiences. This connection is the basis of the illusion which makes primitive experience appear to inform us directly about the relation of material bodies (which exist, after all, only in so far as they are thought).

In the sense thus indicated we have (the indirect) experience of the contact of two bodies. We need do no more than call attention to this, as we gain nothing for our present purpose by singling out the individual experiences to which this assertion alludes. Many bodies can be brought into permanent contact with one another in manifold ways. We speak in this sense of the position-relationships of bodies (*Lagenbeziehungen*). The general laws of such position-relationships are essentially the concern of geometry. This holds, at least, if we do not wish to restrict ourselves to regarding the propositions that occur in this branch of knowledge merely as relationships between empty words that have been set up according to certain principles.

Prescientific Thought.—Now, what is the meaning of the concept "space" which we also encounter in prescientific thought? The concept of space in prescientific thought is characterized by the sentence: "We can think away things but not the space which they occupy." It is as if, without having had experience of any sort, we had a concept, nay, even a presentation, of space; and as if we ordered our sense-experiences with the help of this concept, present *a priori*. On the other hand, space appears as a physical reality, as a thing which exists independently of our thought, like material objects. Under the influence of this view of space the fundamental concepts of geometry, the point, the straight line, the plane, were even regarded as having a self-evident character. The fundamental principles that deal with these configurations were regarded as being necessarily valid and as having at the same time an objective content. No scruples were felt about ascribing an objective meaning to such statements as "three empirically given bodies (practically infinitely small) lie on one straight line," without demanding a physical definition for such an assertion. This blind faith in evidence and in the immediately real meaning of the concepts and propositions of geometry became uncertain only after non-Euclidean geometry had been introduced.

Reference to the Earth.—If we start from the view that all spatial concepts are related to contact-experiences of solid bodies, it is easy to understand how the concept "space" originated, namely, how a thing independent of bodies and yet embodying their position-possibilities (*Lagerungsmöglichkeiten*) was posited. If we have a system of bodies in contact and at rest relatively to one another, some can be replaced by others. This property of allowing substitution is interpreted as "available space." Space denotes the property in virtue of which rigid bodies can occupy different positions. The view that space is something with a unity of its own is perhaps due to the circumstance that in prescientific thought all positions of bodies were referred to one body (reference body), namely the earth. In scientific thought the earth is represented by the co-ordinate system. The assertion that it would be possible to place an unlimited number of bodies next to one another denotes that space is infinite. In prescientific thought the concepts "space" and "time" and "body of reference" are scarcely differentiated at all. A place or point in space is always taken to mean a material point on a body of reference.

Euclidean Geometry.—If we consider Euclidean geometry we clearly discern that it refers to the laws regulating the positions of rigid bodies. It turns to account the ingenious thought of tracing back all relations concerning bodies and their relative positions to the very simple concept "distance" (*Strecke*). Distance denotes a rigid body on which two material points (marks) have been specified. The concept of the equality of distances (and angles) refers to experiments involving coincidences; the same remarks apply to the theorems on congruence. Now, Euclidean geometry, in the form in which it has been handed down to us from Euclid, uses the fundamental concepts "straight line" and "plane" which do not appear to correspond, or at any rate, not so directly, with experiences concerning the position of rigid bodies. (On this it must be remarked that the concept of the straight line may be reduced to that of the distance. A hint of this is contained in the theorem: "The straight line is the shortest connection between two points." This theorem served well as a definition of the straight line, although the definition played no part in the logical texture of the deductions.) Moreover, geometricians were less concerned with bringing out the relation of their fundamental concepts to experience than with deducing logically the geometrical propositions from a few axioms enunciated at the outset.

Let us outline briefly how perhaps the basis of Euclidean geometry may be gained from the concepts of distance. We start from the equality of distances (axiom of the equality of distances). Suppose that of two unequal distances one is always greater than the other. The same axioms are to hold for the inequality of distances as hold for the inequality of numbers. Three distances $\overline{AB'}$, $\overline{BC'}$, $\overline{CA'}$ may, if CA' be suitably chosen, have their marks

BB', CC', AA' superposed on one another in such a way that a triangle ABC results. The distance CA' has an upper limit for which this construction is still just possible. The points A, (BB') and C then lie in a "straight line" (definition). This leads to the concepts: producing a distance by an amount equal to itself; dividing a distance into equal parts; expressing a distance in terms of a number by means of a measuring-rod (definition of the space-interval between two points).

When the concept of the interval between two points or the length of a distance has been gained in this way we require only the following axiom (Pythagoras' theorem) in order to arrive at Euclidean geometry analytically. To every point of space (body of reference) three numbers (co-ordinates) x, y, z may be assigned—and conversely—in such a way that for each pair of points A (x_1, y_1, z_1) and B (x_2, y_2, z_2) the theorem holds:

$$\text{measure-number } AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

All further concepts and propositions of Euclidean geometry can then be built up purely logically on this basis, in particular also the propositions about the straight line and the plane. These remarks are not, of course, intended to replace the strictly axiomatic construction of Euclidean geometry. We merely wish to indicate plausibly how all conceptions of geometry may be traced back to that of distance. We might equally well have epitomized the whole basis of Euclidean geometry in the last theorem above. The relation to the foundations of experience would then be furnished by means of a supplementary theorem. The co-ordinate may and *must* be chosen so that two pairs of points separated by equal intervals, as calculated by the help of Pythagoras' theorem, may be made to coincide with one and the same suitably chosen distance (on a solid). The concepts and propositions of Euclidean geometry may be derived from Pythagoras' proposition without the introduction of rigid bodies; but these concepts and propositions would not then have contents that could be tested. They are not "true" propositions but only logically correct propositions of purely formal content.

Difficulties.—A serious difficulty is encountered in the above represented interpretation of geometry in that the rigid body of experience does not correspond *exactly* with the geometrical body. There are no absolutely definite marks, and, moreover, temperature, pressure and other circumstances modify the laws relating to position. It is also to be recollected that the structural constituents of matter (such as atom and electron, *q.v.*) assumed by physics are not in principle commensurate with rigid bodies, but that nevertheless the concepts of geometry are applied to them and to their parts. For this reason consistent thinkers have been disinclined to allow real contents of facts (*reale Tatsachenbestände*) to correspond to geometry alone. They considered it preferable to allow the content of experience (*Erfahrungsbestände*) to correspond to geometry and physics conjointly.

This view is certainly less open to attack than the one represented above; as opposed to the atomic theory it is the only one that can be consistently carried through. Nevertheless it would not be advisable to give up the first view, from which geometry derives its origin. This connection is essentially founded on the belief that the ideal rigid body is an abstraction that is well rooted in the laws of nature.

Foundations of Geometry.—We come now to the question: what is a priori certain or necessary, respectively in geometry (doctrine of space) or its foundations? Formerly we thought everything; nowadays we think—nothing. Already the distance-concept is logically arbitrary; there need be no things that correspond to it, even approximately. Something similar may be said of the concepts straight line, plane, of three-dimensionality and of the validity of Pythagoras' theorem. Even the continuum-doctrine is in no wise given with the nature of human thought, so that from the epistemological point of view no greater authority attaches to the purely topological relations than to the others.

Earlier Physical Concepts.—We have yet to deal with those modifications in the space-concept which have accompanied the advent of the theory of relativity. For this purpose we must consider the space-concept of the earlier physics from a point of view different from that above. If we apply the theorem of Pythagoras to infinitely near points, it reads

$$ds^2 = dx^2 + dy^2 + dz^2$$

where ds denotes the measurable interval between them. For an empirically given ds the co-ordinate system is not yet fully determined for every combination of points by this equation. Besides being translated, a co-ordinate system may also be rotated. This signifies analytically: The relations of Euclidean geometry are covariant with respect to linear orthogonal transformations of the co-ordinates.

In applying Euclidean geometry to prerelativistic mechanics a further indeterminateness enters through the choice of the co-ordinate system: the state of motion of the co-ordinate system is arbitrary to a certain degree, namely, in that substitutions of the co-ordinates of the form

$$\begin{aligned} x' &= x - vt \\ y' &= y \\ z' &= z \end{aligned}$$

also appear possible. On the other hand, earlier mechanics did not allow co-ordinate systems to be applied of which the states of motion were different from those expressed in these equations. In this sense we speak of "inertial systems." In these favoured-inertial systems we are confronted with a new property of space so far as geometrical relations are concerned. Regarded more accurately, this is not a property of space alone but of the four-dimensional continuum consisting of time and space conjointly.

Appearance of Time.—At this point time enters explicitly into our discussion for the first time. In their applications space (place) and time always occur together. Every event that happens in the world is determined by the space co-ordinates x, y, z, and the time co-ordinate t. Thus the physical description was four-dimensional right from the beginning. But this four-dimensional continuum seemed to resolve itself into the three-dimensional continuum of space and the one-dimensional continuum of time. This apparent resolution owed its origin to the illusion that the meaning of the concept "simultaneity" is self-evident, and this illusion arises from the fact that we receive news of near events almost instantaneously owing to the agency of light.

This faith in the absolute significance of simultaneity was destroyed by the law regulating the propagation of light in empty space or, respectively, by the Maxwell-Lorentz electrodynamics. Two infinitely near points can be connected by means of a light-signal if the relation

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 = 0$$

holds for them. It further follows that ds has a value which, for arbitrarily chosen infinitely near space-time points, is independent of the particular inertial system selected. In agreement with this we find that for passing from one inertial system to another, linear equations of transformation hold which do not in general leave the time-values of the events unchanged. It thus became manifest that the four-dimensional continuum of space cannot be split up into a time continuum and a space continuum except in an arbitrary way. This invariant quantity ds may be measured by means of measuring-rods and clocks.

Four-dimensional Geometry.—On the invariant ds a four-dimensional geometry may be built up which is in a large measure analogous to Euclidean geometry in three dimensions. In this way physics becomes a sort of statics in a four-dimensional continuum. Apart from the difference in the number of dimensions the latter continuum is distinguished from that of Euclidean geometry in that ds^2 may be greater or less than zero. Corresponding to this we differentiate between timelike and spacelike line-elements. The boundary between them is marked out by the element of the "light-cone" $ds^2 = 0$ which starts out from every point. If we consider only elements which belong to the same time-value, we have

$$-ds^2 = dx^2 + dy^2 + dz^2$$

These elements ds may have real counterparts in distances at

rest and, as before, Euclidean geometry holds for these elements.

Effect of Relativity, Special and General.—This is the modification which the doctrine of space and time has undergone through the restricted theory of relativity. The doctrine of space has been still further modified by the general theory of relativity, because this theory denies that the three-dimensional spatial section of the space-time continuum is Euclidean in character. Therefore it asserts that Euclidean geometry does not hold for the relative positions of bodies that are continuously in contact.

For the empirical law of the equality of inertial and gravitational mass led us to interpret the state of the continuum, in so far as it manifests itself with reference to a noninertial system, as a gravitational field, and to treat noninertial systems as equivalent to inertial systems. Referred to such a system, which is connected with the inertial system by a nonlinear transformation of the co-ordinates, the metrical invariant ds^2 assumes the general form:

$$ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx_\mu dx_\nu$$

where the $g_{\mu\nu}$'s are functions of the co-ordinates and where the sum is to be taken over the indices for all combinations 11, 12, . . . 44. The variability of the $g_{\mu\nu}$'s is equivalent to the existence of a gravitational field. If the gravitational field is sufficiently general it is not possible at all to find an inertial system, that is, a co-ordinate system with reference to which ds^2 may be expressed in the simple form given above:

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$$

but in this case, too, there is in the infinitesimal neighbourhood of a space-time point a local system of reference for which the last-mentioned simple form for ds holds. This state of the facts leads to a type of geometry which Riemann's genius created more than half a century before the advent of the General Theory of Relativity of which Riemann divined the high importance for physics.

Riemann's Geometry.—Riemann's geometry of an n-dimensional space bears the same relation to Euclidean geometry of an n-dimensional space as the general geometry of curved surfaces bears to the geometry of the plane. For the infinitesimal neighbourhood of a point on a curved surface there is a local co-ordinate system in which the distance ds between two infinitely near points is given by the equation

$$ds^2 = dx^2 + dy^2$$

For any arbitrary (Gaussian) co-ordinate system, however, an expression of the form

$$ds^2 = g_{11}dx^2 + 2g_{12}dx_1dx_2 = g_{22}dx_2^2$$

holds in a finite region of the curved surface. If the $g_{\mu\nu}$'s are given as functions of x_1 and x_2 the surface is then fully determined geometrically. For from this formula we can calculate for every combination of two infinitely near points on the surface the length ds of the minute rod connecting them; and with the help of this formula all networks that can be constructed on the surface with these little rods can be calculated. In particular, the "curvature" at every point of the surface can be calculated; this is the quantity that expresses to what extent and in what way the laws regulating the positions of the minute rods in the immediate vicinity of the point under consideration deviate from those of the geometry of the plane.

This theory of surfaces by Gauss has been extended by Riemann to continua of any arbitrary number of dimensions and has thus paved the way for the general theory of relativity. For it was shown above that corresponding to two infinitely near space-time points there is a number ds which can be obtained by measurement with rigid measuring-rods and clocks (in the case of timelike elements, indeed, with a clock alone). This quantity occurs in the mathematical theory in place of the length of the minute rods in three-dimensional geometry. The curves for which $\int ds$ has stationary values determine the paths of material points and rays of light in the gravitational field, and the "curvature" of space is dependent on the matter distributed over space.

Just as in Euclidean geometry the space-concept refers to the position-possibilities of rigid bodies, so in the General Theory of Relativity the space-time concept refers to the behaviour of rigid bodies and clocks. The space-time continuum, however, differs from the space continuum in that the laws regulating the behaviour of these objects (clocks and measuring-rods) depend on where they happen to be. The continuum (or the quantities that describe it) enters explicitly into the laws of nature, and conversely these properties of the continuum are determined by physical factors. The relations that connect space and time can no longer be kept distinct from physics proper. Nothing certain is known of what the properties of the space-time continuum may be as a whole. Through the general theory of relativity, however, the view that the continuum is infinite in its timelike extent but finite in its spacelike extent has gained in probability.

Time

The physical time-concept answers to the time-concept of the extrascientific mind. Now, the latter has its root in the time-order of the experiences of the individual, and this order we must accept as something primarily given. One experiences the moment "now," or, expressed more accurately, the present sense-experience (*Sinnen-Erlebnis*) combined with the recollection of (earlier) sense-experiences. That is why the sense-experiences seem to form a series, namely the time-series indicated by "earlier" and "later." The experience-series is thought of as a one-dimensional continuum. Experience-series can repeat themselves and can then be recognized. They can also be repeated inexactly, wherein some events are replaced by others without the character of the repetition becoming lost for us. In this way we form the time-concept as a one-dimensional frame which can be filled in by experiences in various ways. The same series of experiences answer to the same subjective time-intervals.

The transition from this "subjective" time (*Ich-Zeit*) to the time-concept of prescientific thought is connected with the formation of the idea that there is a real external world independent of the subject. In this sense the (objective) event is made to correspond with the subjective experience. In the same sense there is attributed to the "subjective" time of the experience a "time" of the corresponding "objective" event. In contrast with experiences external events and their order in time claim validity for all subjects.

This process of objectification would encounter no difficulties were the time-order of the experiences corresponding to a series of external events the same for all individuals. In the case of the immediate visual perceptions of our daily lives, this correspondence is exact. That is why the idea that there is an objective time-order became established to an extraordinary extent. In working out the idea of an objective world of external events in greater detail, it was found necessary to make events and experiences depend on each other in a more complicated way. This was at first done by means of rules and modes of thought instinctively gained, in which the conception of space plays a particularly prominent part. This process of refinement leads ultimately to natural science.

The measurement of time is effected by means of clocks. A clock is a thing which automatically passes in succession through a (practically) equal series of events (period). The number of periods (clock-time) elapsed serves as a measure of time. The meaning of this definition is at once clear if the event occurs in the immediate vicinity of the clock in space; for all observers then observe the same clock-time simultaneously with the event (by means of the eye) independently of their position. Until the theory of relativity was propounded it was assumed that the conception of simultaneity had an absolute objective meaning also for events separated in space.

This assumption was demolished by the discovery of the law of propagation of light. For if the velocity of light in empty space is to be a quantity that is independent of the choice (or, respectively, of the state of motion) of the inertial system to which it is referred, no absolute meaning can be assigned to the conception

of the simultaneity of events that occur at points separated by a distance in space. Rather, a special time must be allocated to every inertial system. If no co-ordinate system (inertial system) is used as a basis of reference there is no sense in asserting that events at different points in space occur simultaneously. It is in consequence of this that space and time are welded together into a uniform four-dimensional continuum. *See* RELATIVITY.

(A. EI.)

RECENT DEVELOPMENTS

The more recent developments in the structure of space-time were in three different directions. First, cosmological theories attempted to describe the properties of space-time in the large. By this we mean that instead of asking what the structure of space-time is, for instance, in the neighbourhood of a star or in the solar system or even in the neighbourhood of a galaxy, we attempt to describe the gross properties of the universe as a whole in terms of the gross geometrical structure of space-time. Second, the unified field theories were developed, the aim of which was to show that not only gravitational forces, but also electromagnetic (and nuclear) forces are manifestations of the geometry of space-time. Third, modern developments in the quantum theory of the electromagnetic field suggested that a certain structure should be attributed to those portions of space-time which in the past were considered to be empty. The detailed description of this structure forms the modern quantum theory of the "vacuum."

A brief description of the new results must necessarily be incomplete. According to the theory of relativity, the geometrical structure of space-time determines the physical processes; hence, a detailed description of this structure would be equivalent to the characterizing of all the physical processes which can take place. This would entail rather more than any short discussion.

Cosmological Theories.—Let us review now what cosmological theories tell us about space-time. Most cosmological considerations are based on the so-called cosmological principle, which states that on a large scale all positions in space are equivalent; irregularities are only local. Thus on a large scale the universe is homogeneous. This still tells us nothing about the development of the universe in time. With this postulate we will find two classes of theories. The steady-state theory assumes that the universe should be stationary; *i.e.*, its broad aspects should not change in time. The other theories assume the existence of a cosmic time which measures the evolution of the universe. The limiting case of both classes is the static universe. However, about 1930, E. Hubble at the Mount Wilson observatory discovered that the nebulae in the observable part of the universe are receding from us. This observation automatically rules out the possibility that space-time should be static as was originally proposed by Albert Einstein (1917). The evolving universes can be of the following kinds: (1) In the very early initial stages of the evolution, the universe was exceedingly small and contained exceedingly high densities of matter; that is to say, it has developed from a point source. (2) The universe was originally a static Einstein universe; from this stage it is now developing by expansion (Lemaître-Eddington universe). (3) The universe originally contracted from an extremely large initial size, reached a minimum size and is expanding again.

From these different theories, the second was the most commonly accepted. However, since 1940 interest in the other theories revived. The point-source model 1 was invoked by George Gamow, Edward Teller (1939) and others to explain not only the origin of the universe but the generation of the elements composing the universe as well.

The steady-state theory was put forward by Hermann Bondi and T. Gold (1948) and by Fred Hoyle (1948). According to this theory, the large-scale structure of space-time is independent of the position of the observer and of the time of observation. This postulate is called the "perfect cosmological principle," in contradistinction to the above-mentioned "cosmological principle" which postulates homogeneity only as far as space is concerned. However, with this assumption a difficulty arises. Notwithstanding the homogeneity in space and time, we must incorporate the observed fact that the universe is expanding and carrying matter away from the neighbourhood of any observer. Thus, on the one hand an observer should see on the average the same amount of matter around him, clustered into nebulae; and on the other hand he should see these nebulae, with the matter contained in them, moving away. If this be true, the thinning out of matter by expansion must be replaced by a continuous creation of matter out of nothing. This process cannot be observed per se (the amounts are too small) but it gives rise to a number of predictions of things that are observable. Among others is the distribution of age among nebulae. We do not know yet how far the theory is confirmed by observations.

Unified Field Theories.—In the General Theory of Relativity all forces which produce acceleration are gravitational forces. Moreover, these gravitational forces stem from the geometry of space-time. This, however, does not mean that the only interactions which can be described by the theory are the forces commonly called gravitational. In the General Theory all the usual forces, such as electromagnetic, elastic, etc., produce accelerations by modifying the geometry of space-time which in turn produces a gravitational field. Mathematically this arises in the following fashion: For every interaction, whether it be pure gravitational, electromagnetic, etc., there is a corresponding energy-momentum tensor; and the geometry of space-time is determined by this energy-momentum tensor. (*See* RELATIVITY.) However, the General Theory alone cannot give the laws for these tensors, *i.e.*, how the electromagnetic fields modify the geometry of space-time. The unified field theories attempt to accomplish this as far as electromagnetic forces are concerned, conceiving these laws to be geometrical restrictions on space-time.

This modifies the geometry of space-time profoundly. It makes the geometry either higher dimensional or non-Riemannian. The modifications brought about this way are too abstract and technical to be discussed here. (For details, *see* RELATIVITY: *Unified Theories.*)

Modern Quantum Theory of Fields.—This area of study has also left its mark on the development of space-time concepts. Quantum theory denies the possibility of the simultaneous measurement of certain quantities with arbitrary accuracy. For example, it asserts that the momentum (mass times velocity) and position of a particle cannot be measured simultaneously with arbitrary accuracy. These rules are expressed by the Heisenberg uncertainty relations. In general, three main directions can be pursued.

1. The question can arise of how far the uncertainty relations will reflect themselves in our specification of space-time. First, is it possible to attach labels to points in space-time which would tell with complete accuracy what we mean by "here" and "now"? Second, even if it were possible to do this labeling locally, are we able to spread the labeling over the whole of space-time while preserving complete accuracy? The answers to these questions are not yet known.

2. Suppose we were able to answer the first question in the affirmative, thus making possible the use of the concept of space-time as it stands. How would the uncertainty relations effect the General Theory? It is immediately apparent that they will affect the theory. For, according to the Heisenberg relations, it is not possible to measure the position and momentum of a particle simultaneously and precisely. On the other hand, if the gravitational field of the body can be measured, the General Theory permits just that. Thus, the General Theory must be modified to make it consistent with quantum theory. This can be accomplished for weak gravitational fields, with the following results. Space-time is still characterized by the $g_{\mu\nu}$ tensor (*see* above: *Effect of Relativity, Special and General*) but this $g_{\mu\nu}$ tensor cannot be known in general with perfect precision; rather we know the probability that in a given physical situation it will have certain values. Thus the discussion returns again to the questions raised in point 1 above.

3. In the General Theory, if no gravitational fields are present we consider space-time empty and its metric to be flat. If there is any physical process going on there will be a gravitational field

generated by it, and space-time will become curved. According to the modern quantum theory of fields, however, even empty space is not completely void. This empty space, called the vacuum, is endowed with potentialities. Electrons, positrons, photons and other elementary particles can be spontaneously created there even if only to be annihilated immediately. These virtual physical processes have an observable influence on certain physical phenomena, for example, on the actual emission of light by actual atoms. Now we may ask: How will the existence of these virtual processes influence our description of space-time? Is there any gravitational field associated with these processes? Does a gravitational field influence these virtual processes? It is hard to tell what the answers will be.

See also QUANTUM MECHANICS; COSMOGONY; COSMOLOGY.

BIBLIOGRAPHY.—*General Discussions:* A. Einstein, *The Meaning of Relativity* (1953); E. Schrödinger, *Space-Time Structure* (1950) (*see* also for the unified theories). *Cosmological Discussions:* H. Bondi, *Cosmology* (1952); E. Schrödinger, *Expanding Universes* (1956). *Unified Theories:* P. G. Bergmann, *Introduction to the Theory of Relativity* (1942). *Space-Time and Quantum Field Theories:* There are no monographs on the subject. For a review of the present situation, *see Rev. Mod. Phys.*, vol. 29, no. 3 (1957). For the research in Russia *see* D. Iwanenko and A. Sokolow, *Klassische Feldtheorie*, Akademie Verlag (1953). (N. L. B.)

SPADEFISH (*Chaetodipterus faber*), a very deep-bodied, almost circular fish of the family Ephippidae, 2 to 3 ft. long, common in Atlantic waters from Massachusetts to Brazil and much prized for food. It has a small mouth, with brushlike teeth. The lateral line follows the humped curve of the body. In colour it is grayish, with four dark bands running from the top of the back down the body. It is also called angelfish.

SPADIX, in botany, a form of flowering spike with a fleshy or succulent axis, usually subtended by, or enclosed in, a bract called a spathe (*q.v.*). It is the characteristic inflorescence of the arum family. *See* ARACEAE. (J. M. Bl.)

SPAGHETTI: *see* MACARONI.

SPAGNA, LO (*c.* 1450–1528), the usual designation (because of his Spanish origin) of the Italian painter GIOVANNI DI PIETRO, a pupil and assistant of Perugino. Nothing is known of his early life. The chief of his numerous panel paintings are the "Nativity," in the Vatican, and the "Adoration of the Magi," in the Kaiser Friedrich museum, Berlin. Lo Spagna's frescoes reach a higher standard of merit than his panel pictures. The museum of the Capitol in Rome possesses a series of life-sized fresco figures by him, representing Apollo and the nine muses, formerly ascribed to Raphael.

SPAHIS, originally the holders of fiefs in central Asia who yielded personal military service to their superior chief (in Persian *siparhi*, meaning "warriors," and synonymous with sepoy). In time the term came to be applied to the soldiery furnished in their own stead. A similar institution existed in Turkey, and the spahis were the light irregular cavalry which from the time of Sultan Murad I (1362) down to the beginning of the 19th century formed the flower of the Turkish army; at one period they are estimated to have numbered 130,000.

SPAIN, in its own language *España*, a state in the southwest of Europe, comprising eleven-thirteenths (190,115 sq.mi.) of the Iberian peninsula and in addition the Balearic Islands (1,936 sq.mi.), the Canary Islands (2,894 sq.mi.) and certain towns on the Mediterranean coast of Morocco (Ceuta, Peñón de Vélez, Alhucemas and Melilla), with the Chafarinas Islands. The possessions of Spain are: (1) Spanish Sahara, comprising the provinces of Rio de Oro (area 73,362 sq.mi.; chief town Villa Cisneros) and Sekia el Hamra (area 32,047 sq.mi.; chief town El Aainn); (2) Ifni (area 741 sq.mi.; chief town Sidi Ifni); (3) Spanish Guinea (10,832 sq.mi.), comprising Rio Muni or continental Guinea (6,734 sq.mi., chief town Bata) and the islands of Fernando Póo, Elobey, Annobón and Corisco (819 sq.mi., chief town Santa Isabel, on Fernando Póo). The former Spanish protectorate of Morocco extended over a northern and a southern zone (*see* MOROCCO). This article is concerned with Spain proper.

PHYSIOGRAPHY

The Iberian peninsula covers the major portion of that section of the earth's surface which is included between latitudes 43° 47′ N. (Estaca de Vares) and 36° 0′ N. (Isleta de Tarifa) and between the meridians of 3° 19′ E. (Cape de Creus) and 9° 30′ W. (Cape da Roca). Of its area of about 223,000 sq.mi., approximately 85% is Spanish. It has been thought proper, therefore, to treat here of the physical geography and geology of the peninsula as a whole rather than to confine the treatment to Spain.

The peninsula is the least European in appearance of the three great Mediterranean peninsulas. Its massif character, the great central Meseta, the narrow Strait of Gibraltar by which it is separated from Africa and the high barrier of the Pyrenees by which it is separated from the rest of Europe all give it the physical appearance of a disconnected outlier of the Moroccan Meseta rather than that of a part of Europe. Moreover, it is separated from the Mediterranean by coastal ranges which leave between themselves and the sea only narrow and discontinuous strips of coastal plain that alone, of all the peninsula, have a Mediterranean history.

At first glance it would seem that the peninsula should serve to unite Africa and Europe; instead, its ranges of east-and-west mountains serve to separate them. Moreover, the peninsula furnishes no natural traffic routes between the Mediterranean and the Atlantic or between the Mediterranean and northwestern Europe. The closer physical connection with Africa than with the rest of Europe had the consequence of the Islamic invasion and the 500-year struggle with Islam which is accountable in large part for the slow development of Spain along European lines. The physical character of the peninsula likewise resulted in the separation of Portugal from Spain and the 700-year separation of the kingdoms of Castile and Aragon. So sharply defined are the natural regions occupied by these two states that even now their peoples differ sharply in physical type, customs and social organization.

The Coasts.—Seven-eighths of the border of the peninsula are washed by the sea; yet the peninsula as a whole has relatively little access to the sea and comparatively few good harbours. The Portuguese seaboard is 500 mi. long and, as compared with the rest of the peninsula, relatively low. On the southern half of the coast, however, mountain masses and spurs terminate in a number of high rocky headlands like Capes Mondego, Carvoeiro, da Roca, Espichel, de Sines and St. Vincent. The lagoon of Aveiro and the estuaries of the Minho, Douro, Mondego, Tagus, Sado and Guadiana rivers deeply indent the coast and furnish the principal harbours. The high continental mass which forms the greater part of Spain has practically no navigable rivers and a coast line that is for the most part steep and rocky, with relatively few indentations that are easily approached from the interior. The northern coast has numerous indentations that form convenient harbours for small craft, although the current that flows eastward along it frequently builds bars at their mouths. The best harbours are found in the *rias* or fjordlike indentations of the coast of Galicia. There are the fine natural harbours of Pontevedra, Vigo, Corunna and Ferrol. Bilbao, on the Nervión river, has been since mediaeval times the principal maritime outlet of the Basque provinces. Santander and San Sebastián, likewise on the northern coast, are secondary centres of overseas shipping.

Along the Gulf of Cádiz is the broadest coastal plain of the whole peninsula. Eastward as far as the marshes at the mouth of the Guadalquivir (Las Marismas) it is lined by a series of sand dunes and bordered in part by a line of low, wave-built islands. Eastward from the mouth of the Guadalquivir the coast is more varied and includes the excellent harbour of the Bay of Cádiz.

The Mediterranean coast is bold and rocky from the Strait of Gibraltar almost to Cape Palos but is broken by the British-owned harbour of Gibraltar and the fine harbour of Cartagena. North of Cape Palos the coastal ranges again recede, leaving a narrow coastal plain; but in the northern part of the province of Alicante the mountains again advance to the sea and form the lofty headland of Cape de la Nao. The whole coast of the Bay of Valencia is low and swampy and lined with lagoons, furnishing little in the way of good harbours. All this coast is built of sediments brought down from the Meseta. This sedimentary character

culminates in the great delta of the Ebro river. North of the mouth of the Ebro the coast alternates between playas and rocky headlands. The Llobregat river also forms a considerable delta just south of the harbour of Barcelona. From Barcelona a long monotonous sand beach extends to the mouth of the Tordera river. There begins the famous Costa Brava, divided into two sections by the Gulf of Rosas and presenting a rough varied coast line of deep coves and steep, rocky bluffs.

Inland Configuration.—The surface of the peninsula is noted for its contrasts and its vast expanses of dreary uniformity. There are mountains rising with alpine grandeur above the snow line, but often sheltering rich and beautiful valleys at their base. Naked walls of white limestone tower above dark woods of cork and olive. In other parts, as in the Basque country, in Galicia, in the Serranía de Cuenca (between the headwaters of the Tagus and those of the Júcar), in the Sierra de Albarracín (between the headwaters of the Tagus and those of the Guadalquivir), there are extensive tracts of undulating forest-clad hill country. Almost contiguous to these there are broad plains of level tableland, some almost uninhabitable and some with a network of irrigation canals and richly cultivated like the Requena of Valencia. Continuous mountain ranges and broad tablelands give the prevailing character to the scenery; but there are, in addition, lofty isolated mountain peaks such as Monseny, Montserrat and Mont Sant in Catalonia, the Peña Golosa in Valencia and Moncayo on the border of Aragon and Old Castile, as well as many small secluded valleys such as those of Vich and Olot among the Catalan Pyrenees.

The Meseta and Its Borders.—More than half the peninsula is occupied by the Meseta, an ancient earth block consisting mostly of flat-lying strata, broken and eroded and reduced in large part to a state of peneplanation. The Meseta slopes gradually from north to south and from west to east, its average altitude being about 2,000 ft. On the north it is bordered by the Cantabrian mountains and on the northeast by the Iberian ranges, which separate the Meseta from the deep tectonic depression in which flows the Ebro river and terminate near the Gulf of Valencia. On the south the Meseta is bordered by the Sierra Morena, which is actually only the steep southern edge of the Meseta along the Baetic depression, and the Sierra de Alcaraz between the headwaters of the Guadalquivir and Segura rivers.

The Cantabrian mountains are divided into two sections—the Cantabro-Asturian section on the northern border of the Meseta and the Galician section which extends into Portugal as far as the Douro river. The Cantabro-Asturian section runs sensibly parallel to the north coast and culminates in the Peña de Cerredo (8,786 ft.), the highest point in the ranges of the Meseta. A characteristic of these mountains is the many *parameras* or isolated plateaus surrounded by steep mountains or walls of sheer cliff. On account of the diversity of their origin and their many interruptions the Iberian ranges can be characterized better as the upturned northeastern corner of the Meseta than as a true mountain system. This corner reaches its highest point in the Sierra de Moncayo (7,595 ft.). South of the Jalón river it divides into two sections: one, which borders on the Meseta, forms the broken and picturesque Serranía de Cuenca; the other, which lies nearer the Ebro depression, forms the coast ranges that border the low, lagoon-strewn Bay of Valencia.

The Meseta is divided by a central cordillera into northern and southern submesetas, representing the ancient regions of Old Castile and New Castile. ' This central cordillera consists of a series of crust-block mountains formed by successive fractures and vertical movements and extending *en échelon* northeastward from the Atlantic coast near Lisbon to the Nudo de Albarracín, where it connects with the Iberian ranges. It divides the basin of the Douro river from that of the Tagus. There are lofty peaks reaching above the line of perpetual snow in the sierras of this central cordillera. The highest peak, the Plaza del Moro Almanzor (8,504 ft.) in the Sierra de Gredos, is only a little lower than the highest of the Cantabrian peaks, while the Pico de Peñalara in the Sierra de Guadarrama reaches 7,972 ft. Even these highest crests are, however, only about 5,000 ft. above the general level of the Meseta. Between the Sierra de Guadarrama and the Sierra de Gredos and between the latter and the Sierra de Gata are high irregular surfaces filled with hills and *parameras*.

Parallel to the western part of the central cordillera between the basins of the Tagus and Guadiana rivers are the Montes de Toledo, a shorter series of sierras that form a denuded highland of relatively low elevation and are separated from the Iberian ranges by the section of the Meseta known as La Mancha. Their highest point is Las Villuercas (4,734 ft.) in the Sierra de Guadalupe. Their westward extension into Portugal is known as the Montanhas do Alentejo. The Sierra Morena is an even less continuous range than the central cordillera or the Montes de Toledo. Its westerly extension into Portugal ends at the rocky headland of Cape St. Vincent.

Geologically, geographically and politically the Meseta is the true Spain. On all sides it is in general remarkably isolated from the coast so that the passes on its border and the river valleys that lead from it to the coast and to the Ebro and Guadalquivir depressions have been from earliest time geographical features of the highest importance. On the northwest, communication between León and Galicia is furnished by the Sil river which at an early date was followed by military roads and in modern times became part of the route of the railway to the port of Corunna. In the Cantabrian mountains the passes are fairly numerous and over them lie the routes of several railways connecting the interior with the ports of the Bay of Biscay. The two most remarkable passes in these mountains are the Pass of Pajares, which carries the railway from León to Oviedo and the port of Gijón, and that of Reinosa which leads down to the deep valley of the Besaya and is crossed by the railway from Valladolid to Santander. The eastern part of the range is crossed by railways from Burgos to Bilbao and San Sebastián, the latter following the picturesque gorge of Pancorvo and crossing the range at Idiazábal.

On the northeast the valley of the Jalón affords the most important communication between Madrid and the Ebro depression. At an early date an important military road followed this depression. It is now traversed by the Madrid-Saragossa railway. Farther south the southerly extension of the Iberian ranges long made direct communication between Madrid and Valencia difficult, and even now, although a branch line connects the Jalón valley with the coastal railway a short distance north of Valencia, the main communications between Madrid and the eastern and southeastern coasts are across the southeastern part of the Meseta and the hill country between the Meseta and the Mediterranean.

The descent from the Meseta to the depression of the Guadalquivir is comparatively gradual; but passes through the eastern part of the Sierra Morena are few, the principal one being the Puerto de Despeñaperros where the Magaña river, a tributary of the Guadalimar, has cut a deep gorge that the railway from Andalusia to Madrid follows in its ascent to the Meseta.

The central cordillera also offers considerable obstruction to easy communication between the northern and southern submesetas, although the hilly intervals between the sierras afford comparatively easy routes. Carriage roads cross the three main passes in the Sierra de Guadarrama (the passes of Somosierra in the northeast, Novacerrada near the Pico de Peñalara and Guadarrama a few miles south and west), while the railway from Madrid to Segovia passes through a tunnel close to the Guadarrama pass and that from Madrid to Avila crosses the southwestern portion of the same sierra by a remarkable series of tunnels and cuts. The Sierra de Gredos has a road across it connecting Ávila with Talavera de la Reina by way of the Puerto del Pico; but for the most part there are only bridle paths across this sierra and the Sierra de Gata and no railway crosses either.

The Meseta is bordered on the northeast and on the south by two great transverse depressions. That on the south is known as the Baetic depression, now deeply eroded by the bed of the Guadalquivir. It is a gigantic fracture that has cut like a knife across all the strata of the Meseta and stands out with remarkable clearness from Cape St. Vincent to the eastern end of the Sierra Morena. The present Baetic depression is believed to be the remains of a strait which, up to the Tertiary period, connected the Mediter-

ranean with the Atlantic.

Later, on account of movements that produced the Cordillera Penibética by which the present depression is bounded on the south and southeast, this strait was reduced to a gulf opening only to the Atlantic and was in time converted, whether by further uplift or by sedimentation alone, to the present valley. The sedimentary deposits of the Guadalquivir, which flows along the most abrupt border of the depression at the foot of the Meseta, have been built up on the materials worn off from the Meseta during the period of submergence and the sediments of marine origin to form the present fertile valley of Andalusia.

The Ebro depression forms a deep ravine between the northeastern corner of the Meseta and the Pyrenees. Its bordering walls converge to form an angle at the point where the depression is shut off from the Mediterranean by the Catalan coastal range, through which the Ebro river has cut a tortuous gorge. At one time the depression was occupied by the sea, which penetrated by way of the Gulf of Rosas but did not extend to the Bay of Biscay. When the Catalan coastal range emerged the depression was converted into a closed lake. The definition of the Ebro river and the successive terraces which characterize both sides of the depression were brought about by the later uplift from which resulted the present architecture of the Pyrenees.

The Ebro valley is divided into two sections by spurs from the Pyrenees on the one side and from the Sierra del Moncayo of the Iberian ranges on the other. The uppermost of these, a plateau of between 1,000 and 1,300 ft. above sea level, is only about one-fourth the size of the lower portion. The valley of the Guadalquivir is also divided into an upper and a lower section as indicated by the change in the course of the river from a due west to a more southwest direction. The small upper valley is of considerable elevation while the much larger lower valley is mainly lowland and, from Seville to the sea, is composed of a perfectly level and marshy alluvium (Las Marismas).

Mountains Outside the Meseta Border.—The Pyrenees (*q.v.*) form a wall between Spain and France with an average altitude higher than that of the Alps. Their snow-capped crests reach their highest point in the Pico de Aneto (11,168 ft.) in the central massif of granitic rocks that forms the range's core. Railroads pierce the Pyrenees over two routes (Jaca-Pau and Ripoll-Foix). The Catalan coastal range that borders the Ebro depression is the youngest range of the peninsula and all evidence indicates that the region suffered movements in the Quaternary period. It represents the upper portions of a land mass now in large part submerged beneath the Mediterranean. The peaks Monseny (5,713 ft.) and Montserrat (4,016 ft.) are the outstanding crests. Erosion has moulded them into picturesque forms. Nearer the coast a series of hills of low elevation borders the Catalan range throughout its whole length.

The Baetic system of the southeastern part of the peninsula has been said to be the result of a series of thrusts that had their focus farther south and forced up against the rigid mass of the Meseta the land mass located in front of it. It is generally accepted that a long longitudinal fault which extends from the lower course of the Segura river to the Genil river divides the system into two groups. The mountains to the south of this fracture are known as the Cordillera Penibética. They lie near the Mediterranean coast and are considered to be prolonged across the Strait of Gibraltar by the mountains of the Rif. The great block of the Sierra Nevada of the Cordillera Penibética contains the highest crests of the peninsula and reaches its greatest altitude in Cerro Mulhacén (11,414 ft.). It is completely alpine in character with glacial cirques, lakes and moraines. The Cordillera Penibética includes at its eastern end the volcanic zone that forms the Cape de Gata. The extension westward from the Sierra Nevada to Point Marroquí appears to be cut by a series of transverse faults. The mountains north of the Penibaetic fault, though not exhibiting the high crests of the Cordillera Penibética, have a much broader area of high altitudes. They close the eastern end of the Baetic depression and extend eastward to form the Cape de la Nao. The dividing line between them and the Iberian ranges is generally placed at the Júcar river.

Rivers and Lakes.—The main water divide of the peninsula follows rather closely, in the north, east and south, the border of the Meseta and the greater part of the drainage is westward by four large rivers, the Tagus (Spanish *Tajo,* Portuguese *Tejo*), Douro (Spanish *Duero*), Guadiana and Guadalquivir. Only a limited part of the northeastern corner of the Meseta is drained by the Ebro on account of the barrier of the Iberian ranges, and all the remaining rivers which drain to the Mediterranean are short as compared with those draining to the Atlantic. All the large rivers of the Atlantic slope rise in Spanish territory. The Tagus and the Douro have their lower reaches in Portugal and the lower Guadiana is partly in Portugal and partly on the boundary. Only the Guadalquivir and the Ebro, of the larger rivers, are entirely in Spanish territory. The longest of the rivers is the Ebro (575 mi.). The rest in order of their length are the Tagus (565 mi.), the Guadiana (510 mi.), the Douro (480 mi.) and the Guadalquivir (350 mi.). The Minho, which divides the Cantabrian from the Galician mountains, and its tributary, the Sil, which rises in the northwestern corner of the Meseta, are much shorter.

The mouths of the Minho, Douro and Tagus form estuaries which provide good harbours for shipping, but all rapidly become unnavigable toward the interior. In fact only the Guadalquivir can really be classed as navigable for any appreciable distance from the sea, ships being able to ascend it as far as Seville. The others flow in deep rocky valleys which they have cut across the Meseta. Their water supply is scanty, their deeply cut valleys are difficult to cross and the river beds so far below the level as to make them useless for irrigation. The Ebro makes its way to the Mediterranean by way of a tortuous gorge and so affords no means of river transportation to the sea. Its great importance lies in its plentiful waters for irrigation purposes derived from both the Pyrenees and the Cantabrian mountains. Lakes of any considerable size are few, the only important ones being two coastal lagoons near the Mediterranean—the Albufera de Valencia, the Mar Menor in Murcia—and the Laguna de Janda in Cádiz behind Cape Trafalgar. There are many small alpine lakes and small salt lakes are to be found in every steppe region.

GEOLOGY

Geologically the Iberian peninsula consists of a great massif (the Meseta) which has been composed by Archaean, Palaeozoic and eruptive rocks partly concealed by a covering of Tertiary strata but characterized by the absence, excepting in its margins, of any marine deposits of Mesozoic age and bordered on the north, east and south by zones of folding in which the Mesozoic and early Tertiary beds are involved.

The Meseta is a fragment only of the great Hercynian mountain system which was formed across Europe at the close of the Carboniferous period. The earth block which forms the Meseta was individualized at the close of the Palaeozoic era. During the Mesozoic era the Hercynian system of which it is a part was shattered and large portions of it sank beneath the sea and were covered by Mesozoic and Tertiary strata. But other portions remained above the sea and of these the Meseta was one. Around it the deposits of the Jurassic and Cretaceous seas were laid down. During the Tertiary era these deposits, together with the earlier Tertiary beds, were crushed against the old massif, thus forming the folded zones of the Baetic ranges on the south, the hills of southern Aragon in the east and the Pyrenees on the north.

There are clear evidences of earlier (Huronian and Caledonian) movements on the Meseta, but worn folds of Carboniferous formations embedded in the Hercynian structure may represent in places an older discordant structure resulting from earlier movements.

Tertiary movements soldered to the Meseta the formations of the northeast, where the folding action lay within the early (Palaeogene) Tertiary period, and of the southeast, where the folding was most intense, in the later Neogene period before the Tortonian epoch. The Tertiary pressures also caused many important movements in the Hercynian massif itself; ancient fractures were rejuvenated, large interior basins were formed and the strata were folded and crushed with great violence in the borders

of the massif. During the Pliocene period the Meseta was given a general tilt westward. Later differential Tertiary movements modified the central cordillera system and accentuated the separation of the northern from the southern submeseta.

The areas of regional metamorphism on the Meseta are formed of plutonic rocks, broadly granites, which pass upward into an Archaean or strata-crystalline series in Galicia, Portugal and the central cordillera system. They appear elsewhere aligned with the Palaeozoic series, principally along a line from Alcántara to Andújar. This series shows these levels: the lowest, augen-gneiss; the intermediate, micaceous gneiss with schists and crystalline limestones intercalated; and the uppermost, lustrous schists and phyllites. These materials appear now at the surface among granitic rocks in the massif of Galicia and northern Portugal and the central sierras, in the Montes de Toledo and in southern Portugal on the plateau of Évora and Beja which extends eastward by the Sierra de Aracena to the Sierra Morena at a point north of Seville.

Palaeozoic.—The oldest Palaeozoic strata are referred, from their included fossils, to the Cambrian, Ordovician and Silurian systems. They range through a vast region of Andalusia, Estremadura, Castile, Salamanca, León and Asturias and along the peaks of the Pyrenees and the Cantabrian mountains. In eastern Galicia and Asturias and in western León the Cambrian and Silurian trend lines swing round a quarter circle following the outline of Galicia and northern Portugal. The Silurian of the sierras (Ocejón, Alto Rey) lying to the northeast of the Sierra de Guadarrama trends south-southwest, south and, finally, southwest. Round the gneissic vestiges of the Toledan massif the Cambrian and Silurian, partly metamorphosed in the neighbourhood of the granites, trend westward and then northwestward in the Montes de Toledo and the Sierra de Altamira. From the Sierra Morena front the trend lines, running at first parallel to the line of front, swing to the west and west-northwest round the strata-crystalline of southern Portugal. Farther east they run freely to the northwest in the great belt of Palaeozoic which stretches from La Mancha to beyond the Portuguese frontier. The older Palaeozoic of the north culminates in the Teleno (7,181 ft.) belonging to the Montañas de León. In the south the Sierra Morena is the scarped southern edge of the main mass of the Palaeozoic, cut by a series of faults partly of Permian and partly of Pliocene date. The Palaeozoic also formed the two plateaus to the south of the Sierra Morena and a low plateau in the east notable for the complex lead-silver zone of Linares and La Carolina; and a broad plateau in the west extending from the Andévalo district south of the Sierra de Cuacena to the western shores of Portugal.

In the hinterland the Palaeozoic sierras trend finally to the northwest, washing the Hercynian axial lines which curve into this direction from both the Sierra de Toledo and the Sierra Morena. At Montoro the course of the Guadalquivir has been imposed on the Palaeozoic by the removal of overlying Tertiary formations. East of Almadén the Palaeozoic is pierced by Quaternary basalts of the Campo de Calatrava beyond which tongues of Mesozoic projecting from La Mancha bring the Guadiana drainage into the Palaeozoic area.

Mesozoic.—In Portugal the Mesozoic series consists of deposits laid down in the geosyncline which existed to the west of the Meseta during that era and which disappeared at its close after retreating north of the present Tagus river draining the Upper Cretaceous. Geosynclinal conditions were most marked during the Jurassic whose "Tusitanian" stage has a thickness of 5,000 ft. in the Tôrres Vedras area. Jurassic limestones sketch the structural skeleton of the Mesozoic fringe of the Meseta. In the Mondego region a Jurassic upfold, flanked on two sides by Turmian formations, runs west-southwest to Soure and then west-northwest to Cape Mondego by Verride and Buarcos.

This axis delimits on the south an area of worn Cretaceous reliefs above which stands out the Jurassic of Cantanhede, separated from the sea by a broad belt of high dunes. The Cretaceous, which appears in depressions along the principal axis as well as on the flanks of the Jurassic zone, is widely developed in the Cintra region where the Jurassic scarcely emerges. South of the Tagus river movements of Alpine date have brought up the Jurassic to form the higher parts of the complicated Serra da Arrábida which runs west-southwest to Cape Espichel. In Algarve the Lias and Jurassic, with a fringe of Cretaceous toward the littoral, are folded against the southern edge of the plateau and contribute granite and limestones to the complexity of the zone.

The Lower Carboniferous rocks of Spain consist partly of limestones and partly of sandstones and conglomerates like the culm of Devonshire. It is in the culm of the province of Huelva that the celebrated copper mines of Río Tinto are worked. The Upper Carboniferous is formed, to a large extent, of sandstones and shales, with seams of coal; but beds of massive limestones are often intercalated, and some of these contain *Fusulina* and other fossils like those of the Russian *Fusulina* limestone. In the Cambrian and Silurian areas of the northwest, Tertiary movements have probably compressed the strata in the arc of the western Asturias, but not in the Carboniferous area to the east where the Carboniferous has its most extensive development in Spain, covering a considerable section of the eastern Asturias and stretching more or less continuously through the provinces of León, Valencia and Santander.

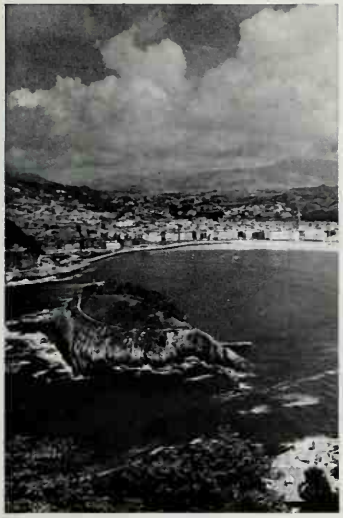
The blue-gray mountain limestone of the Lower Carboniferous forms the triple massif of the Picos de Europa and the crests of many sierras. The culm, which also forms mountains, especially along the upper Deva, marks the transition to the continental conditions in which originated the sandstones and slates of the middle Carboniferous, with their intercalated beds of coal, principally the beds of the Tangres and Mieres basins. Tertiary movements raised the Carboniferous along an east-west axis to form the Cantabrian chain.

To the north of the Carboniferous of the central Asturias lies an area of Mesozoic marine invasion with deposits in a belt trending east and west behind the Palaeozoic "island" by the Cape de Peñas district. From Peña Prieta eastward the altitude of the summits of the Cantabrian chain declines and the Cretaceous is rapidly substituted for the Carboniferous. In the half circle between Peña Tabra and the Sierra de Alcaraz the Meseta is bordered by highlands principally of Mesozoic formations. The tectonic relations between these highlands and the Cantabrian system, which continues eastward in the Cretaceous to the Pyrenean front, are not clear.

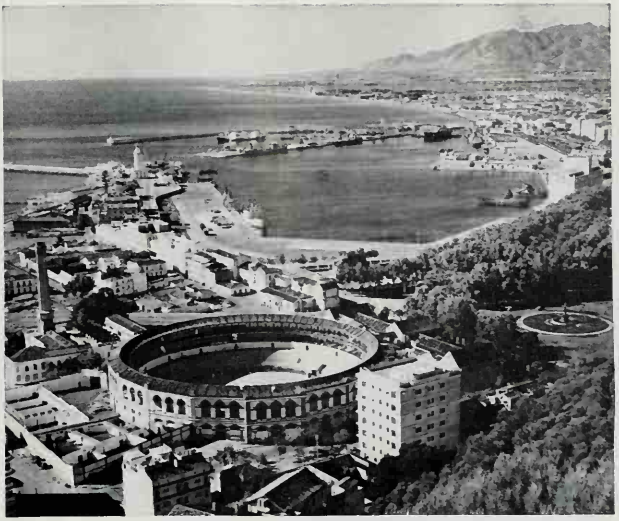
The Permian is probably represented by some of the red sandstones, conglomerates and slates in the Pyrenees, in the Serranía de Cuenca and in Andalusia. The Triassic system is well developed in the north of the peninsula along the Cantabrian chain and eastward to the Mediterranean. It is composed of red and variegated sandstones, dolomites and marl, and contains deposits of gypsum, aragonite and rock salt. The Jurassic overlies these strata and is especially important in the eastern part of the peninsula between Castile and Aragon, along the Mediterranean border, in Andalusia and on the flanks of the Pyrenees. The early Jurassic conglomerates and grits are followed by limestones, some dolomitic variegated marl and overlying spongy limestones or dolomites. The neritic Lias and Jurassic are important only on borders of New Castile.

The Cretaceous forms the broad plateau of La Lora on the southern front of the Cantabrian mountains northwest of Burgos and extends southward and westward into a Cretaceous isthmus in the northern side of the Miocene "Straits of Burgos." An upfold brings it to the surface again in the islandlike Sierra de Atapuerca on the south side of the "straits" toward the Montes de Oca. From the scarp of the Urbión the Cretaceous dips gently southward and Upper Cretaceous limestones form small, bare, fault-bounded plateaus sharply contrasting with the pine-clad slopes of the Lower Cretaceous and appearing at the surface not only on the margin of the zone but in the upfolds of small radius among the Tertiaries of the Douro, at El Burgo de Osma, Soria and elsewhere.

South of the Jalón river beyond the Parameras de Molina the border of the Meseta is formed principally of Jurassic limestones and marl commonly capped by horizontal limestones of the Cretaceous and forming a series of high plains rather than mountains in which the *muela* or molar-tooth summit is typical. The dif-



San Sebastián, major tourist centre of Spain and seat of the government during summer months, showing La Concha bay and Santa Clara Island



View of Málaga, one of the oldest Mediterranean seaports, and the harbour. In the foreground is the bull ring

Valle de los Caídos (Valley of the Fallen), Spain's magnificent memorial to the dead of the Civil War; the crypt, hewn out of rock, is surmounted by 492-ft. cross

Entrance of the picadors to a bull ring

Salmon fishing in the Deva river near La Hermida, northern Spain

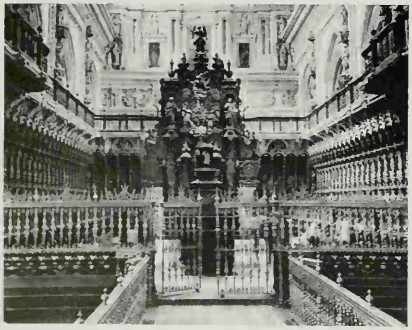
VIEWS OF SPAIN

PHOTOGRAPHS, (CENTRE RIGHT) INGE MORATH FROM MAGNUM, (OTHERS) J. ALLAN CASH—RAPHO GUILLUMETTE

SPAIN

Central nave of the Gothic cathedral at Burgos, begun in the 13th century

The high altar in the Church of San Francisco el Grande (1761–85) in Madrid



The 16th-century choir of Córdoba cathedral, originally a mosque founded in 785 but consecrated in 1236 to honour the Assumption of Mary

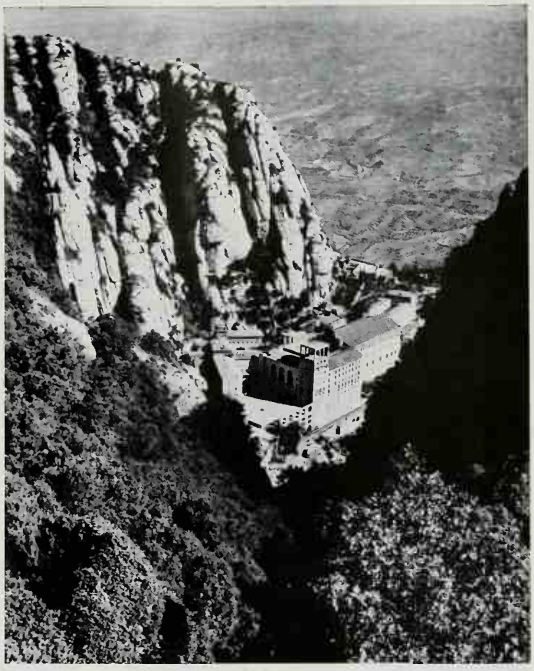
Interior of the Córdoba cathedral which contains a forest of 850 pillars connected by open Moorish arches

INTERIORS OF NOTABLE SPANISH CHURCHES

PHOTOGRAPHS (TOP LEFT) AMPLIACIONES Y REPRODUCCIONES "MAS" (TOP RIGHT) LATHAM FROM MONKMEYER, (BOTTOM LEFT, BOTTOM RIGHT) J. ALLAN CASH—RAPHO GUILLUMETTE

Santiago de Compostela, a place of pilgrimage since the middle ages. Spires of the Romanesque cathedral (begun 1078) are in the background

Thread spinner in a factory at Roda del Ter, Barcelona province



Montserrat monastery (founded in the 9th century), in the huge fissure called Valle Malo above the Llobregat river. Eroded rock formations appear to be stone figures when seen from a distance

A cobbled street in Santiago de Compostela

Edificio de España, a modern office block in Madrid. In front of the building is the Cervantes memorial, with figures of Don Quixote and Sancho Panza (1928)

SPANISH SCENES

BY COURTESY OF (TOP LEFT) SWEDISH-AMERICAN LINE, (CENTRE LEFT) MINISTRY OF TOURISTS AND INFORMATION, SPAIN; PHOTOGRAPHS, (TOP RIGHT, BOTTOM LEFT) J. ALLAN CASH—RAPHO GUILLUMETTE, (BOTTOM RIGHT) PIERO SAPORITI FROM BLACK STAR

SPAIN

Cave dwellings in the hills at Lumbreras

El Puente aqueduct at Segovia built of granite blocks, a late Roman structure

Harvesting olives in the rolling hill country near Úbeda

Steel plant at Avilés, Oviedo

Waterfront at La Coruña, an important seaport of Galicia region. Buildings in the background have *miradores* or glazed balconies, characteristic of the city

ANCIENT AND MODERN SPAIN

BY COURTESY OF (CENTRE LEFT, CENTRE RIGHT) MINISTRY OF TOURISTS AND INFORMATION, SPAIN; PHOTOGRAPHS, (TOP LEFT) INGE MORATH FROM MAGNUM, (TOP RIGHT) AMPLIACIONES Y REPRODUCCIONES "MAS," (BOTTOM RIGHT) J. ALLAN CASH FROM RAPHO GUILLUMETTE

ferent alineations, however, are known as the Serranía de Cuenca, the Montes Universales and the Sierra de Albarracín.

Tertiary.—Deposits of Tertiary age cover rather more than a third of Spain. On the Meseta the Tertiary formations were laid down in the basins of Old and New Castile and in the minor systems of Castelo Branco (Beira Baixa), Badajoz and others of the Tagus and Guadiana valleys. In the west arkose grits of Lutetian date form a broad, monotonous plain in the provinces of Zamora and Salamanca, and the old Tertiary (Palaeogene) beds appear also in the east near Lerma and in the plateau of Soria. Elsewhere the outer deposits of Palaeogene are insignificant. The continental Miocene, unconformable on the Palaeogene, consists of fine-grained materials laid down within the basins of the Palaeogene; and the Cretaceous to which it is unconformable had been folded in the margins of the basins, by rivers (Tortonian), brackish ponds (Sarmatian) and the marshy ponds (Pontian). In spite of stratigraphic discontinuity the forms of the Miocene relief are constant and characteristic.

The type form is the tableland (*mesa*), with its flat upper surface of the limestone cap (*páramo*), its slope of marl (*cuesta*) and its sands or clays of the lower slopes forming plains (*llanuras, campiñas, campos*). As the erosion advances the worn tablelands appear as truncated pyramids or form low mesas capped by the beds of harder marl. Finally great stretches of plain are formed on the level of the Tortonian on lower Sarmatian. Elsewhere, the Eocene marine strata are developed in the basin of the Ebro, and Miocene deposits occupy some small tracts especially on the coast of Valencia, although most of the sandy Tertiary rocks of the latter district are Pliocene. In the Baetic depression the marls of the Tertiary rocks of whatever level are of great importance, supplying the rich soil of the Cijarafe olive groves west of Seville and of the vineyards of Jerez.

Other Features.—Quaternary deposits spread over about a tenth of the country. The largest tract of them is to be seen to the south of the Cantabrian mountains. Another flanks the Sierra de Guadarrama and spreads out over the great plain from Madrid to Cáceres.

From the rivers Douro, Tagus, Guadalquivir and Ebro, E. Hernández-Pacheco described four fundamental terraces at heights of 30, 100, 200 and 330 ft., which are fairly constant within limits of 4-33 ft. from the higher terraces, and he commented on the absence of terraces on the Spanish section of the Guadiana river. The highest terrace is of late Pliocene date, and to it are referred the famous *rañas*, extensive platforms of coarse detritus high on the northern front of the Montes de Toledo. The others seem to be related to the terminal moraines of the Pleistocene glaciation with which, in some cases at least, they are continuous. These moraines are found as low as 2,300 ft. on the Serra da Estrêla (one of the westernmost sierras of the central cordillera); and the topography of the Picos de Europa (of the Cantabrian mountains), the Sierra de Guadarrama and the Sierra de Gredos (of the central cordilleras), the Sierra Nevada and the Iberian ranges from the Demanda to Moncayo has been modified to some degree by glaciation. But the centres of glaciation in each of these cases were relatively small, and the action local. Only in the Pyrenees are the glacial phenomena of real topographical importance. The terraces, however, are of first importance, making irrigation possible in areas otherwise arid.

CLIMATE AND SOILS

Looking at Spain, as a whole, for the year as a whole, we can distinguish, in respect of temperature, the very different climates of the equable Biscayan and northern Atlantic coasts; of the interior, with extremes of temperature always, but with a winter that varies from the traditional nine months of the higher parts of Old Castile to the short two months' winter of the lowlands of Estremadura; of the Mediterranean coast with its very short winter and rather hot summer; and of the subtropical south and southeast, where winter, in the popular sense, hardly exists.

More important is the distinction in respect of rainfall between Spain of the northwest and north (roughly as far south as the line León-Pamplona-Huesca), with a mean annual rainfall of 24 in. and more, fairly well distributed over the year; and the rest of the country, arid Spain, with rains insufficient in quantity or badly distributed or, more commonly, both. There is a general decline in the mean annual rainfall from northwest (Santiago, 66 in.) to southeast (Almería, *c.* 10 in.), but this is interrupted by the minimal rainfall of the basin of Old Castile, where Salamanca and Zamora, like Saragossa in the Ebro basin, lie within an isohyet of 12 in., and by the orographic rainfalls, which are sometimes remarkable. Thus, in the Baetic calcareous zone, in spite of a rainless summer season, the annual rainfall may be 80 in. on the outskirts of the Serranía de Ronda.

The rains of arid Spain are spasmodic, Mediterranean rains falling in large drops for a few hours of a few days of the year. They are also irregular in amount from year to year; the rainfall records of 95 years for San Fernando, near Cádiz, for example, show a mean deviation, from the mean, of 26%. (Annual records from 48 stations are practically complete from 1920 on, and a few are much older.)

Irregularity of the seasonal rain supply and the entrenched courses of the rivers of the interior prevent much use being made of their waters for irrigation. To overcome both of these difficulties for the rivers of the Ebro system immense engineering works were planned. (*See* ARAGON.) On the lower courses of the rivers of the short eastern slope the relief is, by exception, highly favourable to the establishment of important irrigation systems. (*See* VALENCIA.)

Properly conserved, arid soils, although they are inevitably exposed to the sterility resulting from sequences of years of drought, have, nevertheless, certain advantages of chemical composition because of the smaller degree to which plant foods are lost by leaching, which enable them to maintain a standard of productivity lower than that of more humid regions and requiring a greater output of labour, but more independent of artificial fertilization, especially if natural manure is supplied by pasturing. Dry farming, the *cultivo de secano*, is the third and most widespread of the geographical institutions of arid Spain.

Subsoils include the Andalusian black earths and new soils, chestnut or red in colour. An extreme type is the *calvero* soils. In the *calveros*, which arise commonly from attempts to extend cultivation beyond its proper limits, the soil cap is discontinuous and the native rock conspicuous on the surface; tufts of permanent vegetation depend on and protect the small patches of soil; in the open spaces lichens are the chief covering, with occasionally transitory grasses.

Saline soils occur locally, in the more arid parts of Old and, especially, of New Castile and of the Ebro basin and elsewhere; but the extension of the saline area is greatly modified by human agency. In eastern Spain, in general, and in Andalusia, red soils, the Mediterranean red earths, are commoner than chestnut, and in the dry southeastern corner tawny or gray soils are commoner than red; the precise significance of the colour is not known.

FLORA AND FAUNA

Flora.—Because of the extraordinary variety of climates, altitudes, exposures and soils, there are more botanic species in Spain than in any other European district of the same size. The number of endemic species, particularly on the tablelands and the mountain ranges, is exceptional. Spain may be divided botanically into four zones.

In the humid north and northwest the flora is mesophytic, of the same type as that found in central and western Europe. A great part of the country once covered with forest is now barren. Deciduous trees flourish only in the north and northwest or on damp mountain slopes and along watercourses. Typical of these are oaks, beech, birch, ash and sweet chestnut. In the Pyrenees they ascend to about 5,000 ft., above which level grow birches and conifers such as silver fir (*Abies alba*), mountain pine (*Pinus mugo*) and Scotch pine (*Pinus sylvestris*). Higher up are alpine meadows and heaths, rich in alpine flora. Westward from the Pyrenees, through the Basque country and the Cantabrian mountains to Galicia, the lower slopes are well wooded. Higher up, heath moors contain box, blackberry, gorse and heather; while

apples are cultivated in the open meadows and maize, oranges, lemons and other fruits near the coast.

The central tableland or Meseta and the steppe lands carry xerophytic types. Typical trees are pines and evergreen oaks (*Quercus ilex* and *Q. suber*), though poplar and ash grow by the streams. On the deforested areas above 7,000 ft. thin *matorral*, mainly broom heath, is found. *Matorral* consists of a scrub of evergreen bushes and herbaceous plants belonging mainly to the Cistaceae, Labiatae, Leguminosae and Ericaceae families. It flourishes where there is plenty of sunshine, low rainfall and high evaporation. Myrtle (*Myrtus communis*), mastic bush (*Pistacia lentiscus*), sage (*Salvia officinalis*), sweet marjoram (*Origanum majorana*), wormwood (*Artemisia absinthium*) and tree heath (*Erica arborea*) are common in *mattoral*. As the soil becomes more arid this changes into steppe which consists chiefly of bare ground sparsely covered with small bushes and tufts of grasses. Spanish broom (*Spartium junceum*) and gorse are common northeast of the Sierra de Gredos, on the southern side of which olives, figs and vines are cultivated, while on the rocky western slopes are huge areas of woody labiates.

On the Montes de Toledo and Sierra Morena are *matorrales* of broom heath and cistus. Those dominated by cistus are known as *jarales*. On the vast plains of La Mancha the *matorral* is of aromatic labiates such as thyme, lavender and rosemary (these are known as *tomillares*). Sometimes the scrub is very dense and may be 10–15 ft. high, though typically it is about 6 ft. high. In the Sierra Morena evergreen oaks are the dominant trees with ash, alder, black poplar and others along the river banks. There is a great deal of cork oak (*Quercus suber*) in the southwest and sweet chestnut (*Castanea sativa*) in the west. Steppe lands occur mainly in the high southeastern plains where the climate is dry and the soil sandy or salt. There is a large variety of plants, but esparto grass (*Stipa tenacissima*) is dominant.

In the dry Andalusian plain of the southeast, African and semi-desert plants are found. Much land is under cereals and legumes, though there are many vineyards and groves of oranges and olives. Along the streams grow patches of dwarf palms (*Chamaerops humilis*), prickly pears and oleanders. In the Andalusian mountains evergreen oaks grow everywhere, pinewoods in the hollows, and olives are common. The *matorral* is chiefly of cistus and labiates. On the Sierra Nevada typical Mediterranean fruits grow about 5,000 ft. up, below the *matorral* of gorse and aromatic shrubs. In irrigated parts cultivated fruit trees form gardens called *vegas* or *huertas*. The flat marshy expanse of Las Marismas, near the mouth of the Guadalquivir, supports reeds and salt-marsh plants.

The lush, green, almost tropical vegetation of the southern coastal strip is the result of irrigation. Date palms, bananas, sugar, rice and cotton are grown in *vegas*. On the hills grow pines, oaks and carob, along the coastal areas jujube, almonds, figs, olives and pomegranates and on the lowlands mastic bush and sweet laurel.

On Spain's Mediterranean coast line Catalonia has the densest vegetation. South of Barcelona are woods of pine and cork oak as well as typical *matorral*. Oleanders grow in the Ebro valley. North of Barcelona cork flourishes. The rainfall is higher there and so eucalyptus, pines, cypresses and poplars are common. Farther north the deciduous trees are found and the Mediterranean flora merges with the Pyrenean.

Fauna.—Cut off from Europe by the Pyrenees and formerly joined to Africa, Spain has two clearly distinguished groups of animals. The first group includes both species found in the temperate forests of central Europe and occurring in the north of Spain and also isolated colonies of alpine forms occurring in the higher mountain ranges—for instance, deer, ibex, eagle, vulture. The second group is typically Mediterranean and is found in Spain in the dry southeastern regions; there are many indigenous forms and many North African species.

Almost everywhere are found the rabbit, various dormice and squirrels, the rat (*Rattus norvegicus*) which was probably brought into Spain by Napoleon's armies in the 19th century, bats (21 kinds, including the only tropical species found in Europe) and red deer. Indigenous to the country are the Spanish hare (*Lepus granatensis*), the water vole (*Arvicola terrestris sapidus*), the squirrel (*Sciurus vulgaris segurae*) found in the Sierra de Segura, the red deer (*Cervus elaphus hispanicus*) and the Spanish ibex (*Capra pyrenaica*).

In the region of the Pyrenees the fauna is similar to that of France. There live the squirrel, mole, chamois, hare (*Lepus europaeus*) and bear (*Ursus arctos*). In the Cantabrian mountains the fauna is of the European type and includes the chamois, gray dormouse, squirrel, bear, yellow marten and ermine. The typical animals of Galicia are the Galician hare and the ibex.

In the central plateau of Spain indigenous species of squirrel, mole, muskrat and roebuck are found. In the Sierra de Guadarrama live species of wolf, lynx, genet (an African animal), stone marten, badger, weasel and fox. In the central strip between the Tagus and Guadiana rivers there is a predominance of African species. Among the indigenous animals of the southeastern regions are peculiar species of jackal, wolf, squirrel and rabbit. In the lower Guadalquivir valley lives the wild boar.

In Andalusia are a species of dormouse, the Egyptian mongoose (*Herpestes ichneumon widdringtoni*), which is not found in Europe except in the Iberian peninsula, the Andalusian weasel and the hedgehog, while the ibex inhabits the mountains inland from the Mediterranean. At Gibraltar are the famous Barbary apes, probably introduced from Africa. There are no squirrels in the Mediterranean regions.

As Spain is on the route of bird migration between northern Europe and Africa, many species, such as ducks, wild geese and wading birds, winter there on their way south. Others such as flamingos, snipe (by the Guadalquivir), two sorts of quail and the southern shearwater (in the south) stop in the spring, on their way north from Africa.

Birds of prey are numerous and varied, the most common being the buzzard. There are many kinds of vultures, hawks and eagles. The imperial eagle (*Aquila heliaca adalberti*), also found in North Africa, is indigenous and so are some owls, the royal cuckoo (*Cuculus canorus minor*) and, in central Spain and the Sierra Morena, the azure-winged magpie (*Cyanopolius cooki*), otherwise only found in northeastern Asia. Common on the steppes is the red-legged partridge, and the great bustard feeds in flocks in the cornfields of the Meseta.

Amphibians and reptiles are particularly numerous in the south. There are many lizards, chameleons in the Guadalquivir valley, salamanders in the mountains, tortoises and, everywhere, frogs and toads. Trout abound in the mountain streams and lakes, and barbel, carp and tench are common fresh-water fish.

The insect life is quite varied and contains African and European species. Those in the rainy northwest are different from those in the arid zones, where the praying mantis and locusts are found.

BIBLIOGRAPHY.—Maps published by the Instituto Geográfico y Catastral, by the Depósito de la Guerra, by the Instituto Geológico de España and by the Portuguese Direcção Geral dos Trabalhos Geodésicos e Topográphicos and Commissão do Serviço Geológico; Max Sorre, *Espagne—Portugal*, vol. 7 of *Géographie universelle*, ed. by L. Gallois and P. Vidal de La Blache (Paris, 1934); L. Martín Echeverría, *Geografía de España*, 3 vol., 3rd ed. (Barcelona, 1937–38); E. H. del Villar, *El Valor geográfico de España* (Madrid, 1921); J. Dantín Cereceda, *Ensayo acerca de las regiones naturales de España* (Madrid, 1922); L. Doporto, *Geografía general de España*, 3rd ed. (Seville, 1930); L. Mallada, *Explicacion del mapa geológico de España*, 7 vol. (Madrid, 1895–1911); Museo Nacional de Ciencias Naturales, *Serie geológica: Trabajos* (Madrid, 1912 *ff.*); Instituto Geográfico y Catastral, Oficina Central Meteorológica, *Resumen* (annual), and *Memorias*; E. H. del Villar, *España en el mapa internacional de suelos* (Madrid, 1927); Comisión de Investigaciones Paleontológicas y Prehistóricas, *Memorias* (Madrid, 1915 *ff.*); M. Willkomm and J. Lange, *Prodromus Florae Hispanicae*, 3 vol. (Stuttgart, 1861–80, with supplement, 1893); E. Fodor (ed.), *Spain and Portugal in 1952* (1952); Doré Ogrizek (ed.), *L'Espagne* (Paris, 1951). (X.)

HISTORY

It is impossible to introduce any accurate chronology into the prehistory of Spain. Before the Roman pacification in the 2nd century B.C. with its certain dates and events lie a couple of centuries of Punic penetration from Carthage; still earlier we know

of the coasting voyages of Greek traders along the Mediterranean shore and catch some glimpse of the Iberian civilization with which these Greeks came into contact; but all this brings us only into the 7th century B.C. Perhaps for several hundred years before the advent of the Greek ships there had been Phoenicians, mainly from Tyre and Sidon, who had maintained trading posts along the southern shore of Spain and had sailed through the Gibraltar straits to Cádiz on the Atlantic. Then the perspective of years lengthens out immeasurably and we step back, not by centuries but by thousands of years, through neolithic into palaeolithic times, when some type of men—we cannot guess to what race they belonged or what form of language they knew—hunted the beasts of those early days and decorated with marvellously fine likenesses of these animals the inner walls and roofs of deep caverns.

Palaeolithic Spain.—These cave paintings rank among the most impressive discoveries of the past. There can be no doubt of their vast antiquity, since they certainly represent animals of the late Quaternary geological period and are engraved upon bones which are identifiable as those of animals now long extinct. Nor can there be much dispute about the extraordinary veracity and vivacity with which wild horse and bison, hind and stag, ibex and ox and boar are delineated with little save their outline and with no visible hesitation of stroke.

The best-known of the Spanish caverns and one of the first to be discovered is Altamira near Santander. The publication of its contents by Edouard Cartailhac and the abbé Henri Breuil (Monaco, 1906) is the best introduction to the study of palaeolithic Spanish archaeology. Other important sites are Hornos de la Peña, Pasiega and Castillo in the province of Santander; Basondo in the province of Biscay; and Pindal, Buxu and La Peña in Asturias. Palaeolithic decorated caverns are largely confined to the small mountainous area in the north which extends from the westernmost Pyrenees through the coastal hills beyond Santander. They may therefore be justly termed Cantabrian. Elsewhere, however, beyond the Spanish border, the same type recurs, chiefly on the northern slopes of the Pyrenees and in the region of the Dordogne in southwestern France. These three districts—the one in northern Spain, the two others in southern France—are inseparably allied and were products of contemporary cultures. In other parts of Spain, where caves are little more than open rock shelters, a distinctively different art has been found.

Prehistorians therefore speak of "cave art" and "rock-shelter art," and distinguish sharply between Cantabrian and east Spanish palaeolithic. Beginning in the early 1920s there was great activity in the exploration and elucidation of this second group. Except that the fauna of the paintings there implies a slightly warmer climate, much the same series of animals are rendered in much the same style; but intimately combined with them appear weirdly represented men and women—a subject almost completely neglected by the Cantabrian cave artists. Often the men are armed with bow and arrow and shown in pursuit of their quarry; but whereas the beasts are naturalistically correct in the surprising palaeolithic manner, some of the huntsmen with threadlike limbs more nearly resemble the praying mantis, while others with sudden protuberances of calf or thigh suggest unevenly inflated rubber toys. The women, in long flounced petticoats very remotely suggestive of the elegant ladies of Minoan Crete, seem to take part in dances and conversations, poorly visualized and crudely rendered. This striking combination of a naturalistic animal art with geometrically conventionalized, highly primitive human representations is a phenomenon of great significance, since the groupings and actions of the men and women are almost as clearly allied to palaeolithic rock drawings in North Africa as the ibexes and stags, horses and oxen and bulls are allied to the cave paintings of Cantabria. As in historic, so in prehistoric times, the destiny of Spain must have been a succession of invasions, now from the north, pouring round or over the Pyrenees, now from the south out of Morocco and Algeria. The most important sites of the east Spanish group are Cogul by Lérida, Valltorta in Castellón, Albarracín by Teruel, Alpera and Minateda in Albacete and Cantos de la Visera in Murcia—all characteristically within two or three days' journey afoot from the sea.

Neolithic and Bronze-Age Spain.—The Neolithic Age in Spain testifies to a wholly different condition of mankind. Perhaps it is too ingenuously simple to ascribe the naturalism of palaeolithic art to a race of hunters and the geometric conventionalism of the neolithic to a race that tilled the soil. A fairly uniform culture spread over the entire peninsula; for the rock shelters with neolithic drawings have a very wide distribution and show a striking similarity. They have almost no pictorial attraction and often resemble picture script rather than actual illustration. The pottery finds are the great archaeological aid to neolithic chronology; and the occurrence of some of the conventionalized devices of the rock paintings upon pottery which can with certainty be assigned to the transition from the Neolithic to the early Bronze Age puts the general epoch of this culture beyond doubt.

A closer comparison of ceramic types shows that beneath this superficial uniformity there lies sufficient diversity to warrant many subdivisions. Chief among these is the clear distinction between the more indigenous central and the southeastern or Almería culture which was probably under African influence and may actually have been the nucleus from which the true Iberian culture developed. In late neolithic times the more extensive central culture exhibits a distinctive and striking bell-shaped (the so-called campaniform) pottery. At this time it is likely that there was some exportation of copper and silver from Spain to the nearest of the Mediterranean islands, since there are ceramic affinities with finds of the Aëneolithic Age (or Stone and Copper Age, lasting in Spain from about 3000 to 2500 B.C.) in the Balearics and in Sardinia and Sicily. The peculiar megalithic structures in these parts, such as the talayots of the Balearics, the nuraghi of Sardinia, the underground chambers of Malta, are fairly well paralleled by megalithic grave structures in Portugal and Spain. But there is as yet no certain proof of intercourse between the eastern and western basins of the Mediterranean for these early times.

Somewhat after 2500 B.C. the Bronze Age displaced the Aëneolithic without apparent interruption; but still the Aegean influence, which can be proved to have reached Sicily and southern Italy, failed to extend as far as Spain. If anything, the Bronze Age marks a retrogression in the civilization of Spain, where at midcentury there had been no archaeological discovery to indicate a powerful or wealthy Bronze Age culture even remotely comparable to that of Crete or the eastern Mediterranean lands.

Tartessus.—Herodotus (iv, 152) narrates the adventurous voyage of a Greek sailing ship from Samos, which was driven by storm past the Pillars of Hercules and "at last reached Tartessus. This trading-town was in those days a virgin port unfrequented by merchants, and the Samians in consequence made a greater profit than any Greeks before their day." This event supposedly took place about 630 B.C. During the century thereafter, Tartessus was certainly frequented by trading vessels of the Ionian Greeks of Phocaea in the Bay of Smyrna. It has been maintained that this rich Spanish city on the Atlantic was the Tarshish of the Old Testament, which "with silver, iron, tin and lead" traded in the fairs of Tyre (Ezekiel xxvii, 12).

Yet the modern search for Tarshish-Tarsis-Tartessus failed. The German archaeologist Adolf Schulten after several campaigns regretfully admitted in 1926 that the mysterious city must lie deep underriver and below the ground-water level.

THE IBERIANS

The whole of the Hispano-Lusitanian peninsula was sometimes called Iberia by ancient writers; but the true focus of the Iberian civilization was the southeastern corner of the land. There, a stock out of northern Africa (Oran?) and therefore probably Hamitic (Berber?) may have been infiltrating ever since the aëneolithic days of the Almería culture (*c.* 3000 B.C.). They were probably racially akin to the Tartessians, who would therefore have also been African invaders. We become completely certain of their presence toward the end of the Bronze Age and, several centuries later, can trace the diffusion and inland penetration of their culture until it was finally absorbed by the Roman civilization almost as late as the time of Christ. Roman writers describe the typical Iberians as dark complexioned, with unkempt hair, small

of face but with the cheekbones emphasized and the lower lip prominent, small framed, alert and wiry. Iberian bronzes display them as riders of horses, and we know what their swords and daggers, spears and other weapons were like. The women are represented sometimes as wearing a single-piece cloak drawn up over the head from the shoulders like a mantilla or shawl, and sometimes with a short, hooded garment which flares out below the waist and often has swallow-tailed sleeves. We may judge of their jewellery and their headdresses from the famous sculptured bust, the "Lady of Elche." Iberian art was rude but vigorous, with a leaning toward sculptured animals of stone and human figurines in bronze. The pottery was decorated with simple linear themes, artistically of no particular distinction, until Greek examples offered more sophisticated decorative motives to copy. The best Iberian art comes from the provinces of Murcia and Albacete in the southeast, where the Greek influence was strongest. The Iberians lived in walled towns. The cyclopean masonry in the bottom courses of the great walls of Tarragona is indubitably a remnant of an old Iberian stronghold. Tartessus and Massia (now Cartagena) were similarly enclosed; but the Carthaginians completely destroyed both of these cities, making Gades (now Cádiz) take the commercial place of the former and refounding the latter as Carthago Nova. Behind this seaboard fringe of Tartesso-Iberians the high-lying interior was in the possession of a variety of indigenous peoples of whom we have no clear knowledge except that, along with Catalonia, they passed through an early Iron Age (or Hallstatt phase) associated with the invasion of Celtic tribes who, in the 7th and 6th centuries B.C., poured through the Pyrenees by the western passes and gained possession of large areas of the peninsula, submerging the previous inhabitants. The chief Celtic zones today are Galicia and Portugal. This Celtic invasion may justly be called a historical event. We can date its successive waves within a few decades, define its sources, follow its spread and perceive its results. In the extreme northern mountain lands the earlier inhabitants managed to preserve their individuality in the face of the irruption. It has even been argued that the Basques, who maintain customs and a language whose isolation is a mark of extreme antiquity, are ultimate descendants of the palaeolithic Cantabrian folk which produced the cave art so many thousand years earlier. Elsewhere, during the pre-Roman times, we find Celtic tribes (Beribraces, Sefes, Cempsos) in possession of the land.

Such is the ethnic picture during the 6th and 5th centuries B.C. With such a Spain the Ionic Greeks now came into contact, in rivalry with the Carthaginians who had fallen heirs to the centuries-old Phoenician exploitation of these western marts. Neither the Phoenicians nor the Carthaginians left any very permanent mark upon the land, while the Greeks influenced it profoundly. Ships from Tyre and Sidon may have traded beyond the straits and in Cádiz at least as early as the 9th century B.C.; yet modern archaeology, which has located and excavated Greek, Iberian and Roman towns, has not laid bare a single Phoenician settlement or found more important Phoenician remains than the odds and ends of trinkets and jewels and similar articles of barter. The inference is clear that, except perhaps at Cádiz, the Phoenicians built no towns, but had mere trading posts and points of call.

The Phocaean Greeks, on the other hand, whose western focus was Massilia (Marseilles; founded 600 B.C.), founded true colonies along the east and south coasts where unmistakable traces of their settlements have survived. The sites of Hemeroskopeion (near the modern Denia) and Mainake (at the mouth of the Río de Vélez near Málaga) have been identified, though not yet excavated; and Emporion (now Ampurias, on the Pyrenean east coast near the French frontier) has been systematically dug, revealing a fortified town with strong gates and walls, streets more or less at right angles, remains of houses and a shrine of Aesculapius with a cult statue of the god.

While there is almost nothing in Iberian art ascribable to Phoenician or Punic influence, the art of the Greeks had indisputably a revolutionary effect. A collection of sculptures found at Cerro de los Santos in the province of Murcia is a hybrid of archaic Greek and Iberian art. Fragments of ibericized Ionic Greek architectural mouldings were discovered in the same region; and the pottery from Elche and the surrounding districts is full of unmistakable borrowings from the ornamental repertory of Greek vases. The earliest Iberian money was coined in direct imitation of Greek types of the 4th century B.C. Archaic Greek bronzes discovered in many parts of Murcia and Alicante and the actual material traces of the Phocaean colonies tell us that the Greeks were once in this part of Spain.

Elsewhere in the Mediterranean this period is covered by written chronicles of the ancient historians; but the records for Spain are extraordinarily meagre. We hear no more about Tartessus after the 6th century B.C. Instead, Greek sources of the 5th century begin to speak of the whole Tartesso-Iberian region as Iberia; and recent archaeological exploration has confirmed this emphasis upon the Iberians and established them as the great civilizing force of that time in ancient Spain.

During the 4th century B.C. the Carthaginian encroachment reached most of the Iberian homeland, the Greek contact was broken and the Graeco-Iberian culture in Murcia waned. But in recompense, and probably under the pressure of this Punic expansion, the Iberian penetration of the interior gathered great headway. During the 3rd century B.C. the most flourishing Iberian centre in Spain was the valley of the Ebro, where in the 5th century there had been no Iberian settlements whatever; and Greek writers of the time no longer refer to the Castilian uplands as Celtic but as Celtiberian. Excavation confirmed this gradual iberianization of almost the whole of pre-Roman Spain. Largely as a result of the careful and systematic work of Schulten in laying bare the sites of the Roman and Celtiberian encampments there, an extensive insight into this iberianized Celtic culture of the 3rd and 2nd centuries B.C., in particular that of Numantia, has been gained.

Thus, when the Romans began their efforts to latinize the land after the close of the Second Punic War (201 B.C.), they found not a Celtic civilization like that of Gaul but an Iberian culture of the general type which the Greeks had encountered and influenced before them. Archaeological evidence shows that the Iberian and the Roman civilizations must have co-existed side by side for more than a century. The latest sculptures from Cerro de los Santos, though still Iberian in style, borrow Roman statuary motives. In other regions true Iberian pottery is often found mixed with thoroughly Roman ware (*terra sigillata*). Iberian jewellery and gold have been found amid Roman surroundings; and most of the coinage bearing in Iberian script the names of Iberian towns dates from after the Roman conquest.

(R. C.; X.)

THE ROMANS

Carthage, defeated by Rome in the First Punic War, turned to Spain as a natural treasure house to furnish reparations, to finance new offensives and to supply auxiliary troops. Hamilcar's activities after 237 B.C., followed by those of Hasdrubal (including the founding of the fortress at Carthago Nova) and Hannibal, forced upon Rome a difficult, distant and costly war for which it had no desire and which ushered in many decades of stubborn effort to subdue a rich, proud and brave country. Victory in the Second Punic War left Rome with two embryonic Spanish provinces: Hispania Citerior, the eastern coastal strip based on Tarraco (Tarragona) and Carthago Nova; and Hispania Ulterior, comprising the southern coast and the valley of the Baetis (Guadalquivir). Thereafter the task of deepening the provinces by advance inland was bitterly pursued. Year by year, with each province under its proconsul, Roman soldiers and Roman money were poured in, until Scipio Africanus' epic capture of Numantia in 133 B.C. allowed advance northward to the Durius (Douro) and westward to the middle Tagus. Administration, though usually opportunist, was not always merely oppressive or self-interested; and although annual tribute was imposed on native communities they received certain advantages—security, peace, safe communications, new opportunities of production and commerce, employment under Roman arms and acquaintance with Roman civil law. Agriculture, viticulture and fisheries flourished in the south and east, as the local coin types attest, even if these activities were controlled by

Roman capital; trunk roads were constructed (*e.g.*, that from the Pyrenees to Carthago Nova) and coinage developed from the proceeds of the innumerable mines, of which Rome claimed the gold while leasing out the silver, lead and (probably) copper.

When in 80–70 B.C. the Roman Quintus Sertorius used Spanish supporters against Roman political rivals, he did so as the leader of an already half-romanized army. Julius Caesar's governorship of Hispania Ulterior (61–60 B.C.) saw the frontiers extended up to the mouth of the Durius. Later, as dictator, he initiated a new colonial and municipal policy, bringing Italian settlers to Spain and granting Latin rights to certain indigenous communities. Gades became a fully fledged Roman municipium. Roman peace was swiftly followed by Roman political consciousness; language, religion and public life were henceforth moulded and unified by the central Roman model.

Augustus' first task was to undertake, by bloody and savage warfare, the conquest of the northwestern (Asturio-Cantabrian) corner of Spain—not complete until 19 B.C. Now, with the whole peninsula tamed, he effected its total reorganization, though the precise date (in any case earlier than 2 B.C.) is uncertain. Baetica (roughly the Guadalquivir valley with the southern coast) became a "public" province, without regular troops, under senatorial proconsular administration; Lusitania (a new province roughly coextensive with modern Portugal south of the Durius) and Tarraconensis (comprising all the rest of the peninsula) were imperially administered under propraetors with three legions as a permanent garrison in the latter. Major regional towns or cities became centres of government and local jurisdiction. Emerita Augusta (Mérida), a pattern of Augustus' many newly founded watchdog colonies of veteran soldiers, typifies, by the wealth of its remains today, their imposing character. Colonies and full municipia, formed *ex hypothesi* of Roman citizens, were self-governing and subject only to citizens' taxes. Unprivileged communities were administered by the provincial governor and subject still to payment of tribute. Although Augustus was chary in extending the Roman franchise, his work for Spain was enormously beneficial and widely recognized as such. The building of roads and bridges, the encouragement of local enterprise, the unification of sentiment and the establishment of profound peace, while raising the standard of living everywhere, were not imposed at the cost of rigid conformity. Political, religious and financial anomalies were tolerated. Rome's principal reward consisted of the vast output (perhaps unequalled elsewhere) of gold, silver, copper, iron, tin and lead from Spanish mines.

The Augustan model was faithfully preserved. Claudius reduced the garrison to two legions, Vespasian to one. The road complex was developed swiftly. Spanish wine, by its cheapness, and oil, by its quality, disturbed the previous pattern of world markets. Vespasian imposed stricter regulations for the control of all the principal Spanish mines, the revenue of which afterward flowed into the imperial treasury. But early in his reign he recognized a century of astonishing progress by simultaneously promoting about 350 unprivileged communities to Latin rights. Spain could not now deny the presence of the golden age, marked by security, a high and well-diversified culture and a balanced and critical civilization, which, if it produced only few statesmen, gave Rome the younger Seneca, Lucan, Martial, Quintilian, Columella and Pomponius Mela as literary or academic ornaments. The 3rd and 4th centuries saw initiative, if not prosperity, declining. City walls crumbled in neglect, through excessive dependence on Rome; and when the barbarians burst the Rhine frontier in A.D. 406–407 and flooded Gaul and Spain, there could be neither resistance nor hope of recovery. (C. H. V. S.)

With the irruption of the Vandals, the Suebi (Sueves, Suevi) and the Alani (Alans), the history of Spain enters on a long period of division and confusion. The Vandals and their associates, who plundered far and wide, were not numerous enough to establish a rule of their own. When in 428 their king Gaiseric passed out of Spain to found the Vandal kingdom of Carthage, his whole horde numbered only 80,000 persons, including old men, women and children and runaway slaves. Salvian says that many of the subjects of the empire preferred poverty among the barbarians to the tyranny of the imperial tax collectors. The great landowners had, moreover, almost as much to fear from the agrarian insurgents known as Bagaudae, who are found acting with the Suebi, as from the barbarians. In the north the Asturians and Basques, the least romanized part of the population, appear from the beginning of the age of barbarization as acting for themselves. In the mountain country of Cuenca, Albacete and the Sierra Nevada the native Orospedans were entirely independent in the middle of the 6th century. Spain was thrown back into the state of division from which it had been drawn by the Romans—with the vital difference that the country now possessed the tradition of the Roman law, the municipalities and one great organization in the Christian Church.

No help was to be expected from the empire. Unable to aid itself it had recourse to the Visigoths (*see* GOTHS). Ataulphus (*q.v.*), the successor of Alaric I and husband of Placidia, whom he had married against the wish of her brother the emperor Honorius, entered Spain in 414 but was murdered in 415. After the speedily ensuing murder of his murderer and successor Sigeric, Wallia (415–419), who was elected to the kingdom, continued Ataulphus' work. He destroyed the Alani and drove the Vandals and Suebi into the northwest. But then he handed Spain back to the imperial officials, that is to say, to weakness and corruption, and marched with all his people into Aquitania Secunda, which had been assigned to them by Honorius. (X.)

THE VISIGOTHS

The Visigothic rulers of Aquitaine began to intervene in Spanish affairs early in the 5th century but it was not until 494, during the reign of Alaric II (484–507), that large-scale migration to Spain began. The migrants, who certainly did not number more than 200,000, included peasants as well as warriors and clergy and established their overlordship without serious opposition (497). In the areas where they elected to settle, the Hispano-Roman landowners (*hospites*) were compelled to surrender to the newcomers two-thirds of their cultivated land (*sortes Gothicae*), but there was no general expropriation even in Old Castile, where Visigothic settlement was densest. The political capital of the Visigothic monarchy was transferred from France to Toledo in the reign of Athanagild (551–567).

Though the invaders were both politically weak and already highly romanized, powerful influences at first prevented any rapid assimilation to the native population. The Visigoths were followers of the Arian heresy and remained subject to their own legal codes, while the Hispano-Romans were Catholics and continued to be governed by Roman law. Intermarriage between the two races was prohibited. Instability was to be the outstanding feature of the state set up by the Visigoths in Spain. Because of the ambitious Visigothic nobility's strict insistence on the principle of elective monarchy, the crown was constantly menaced by political intrigue, and the removal of the sovereign by deposition or assassination was frequent. Leovigild (568–586)—one of the few effective Visigothic kings—finally destroyed (585) the remnant of the Germanic kingdom set up by the Suebi in northwestern Spain about 175 years before and managed to bring the ever-rebellious Basques under his control; he also dealt successfully with a Catholic rebellion against the state religion. But the next king, Reccared (586–601), was baptized a Catholic soon after his accession, and the third council of Toledo then proclaimed the conversion of the whole kingdom (589).

The removal of the religious barrier allowed the assimilation of the two races to progress more rapidly, and the councils of Toledo, where nobles as well as clergy were represented, now began to dominate the crown. The reign of Swintila (621–631) saw the final expulsion of the Byzantine prefecture set up in southern Spain in Athanagild's time. In the time of Sisenand (631–636), the fourth council of Toledo—presided over by St. Isidore—completed the subjection of king and government to ecclesiastical authority. An event of great significance for the subsequent history of mediaeval Spain was the promulgation by Recceswinth (649–672) of the *Liber judiciorum* as the sole legal code for both races. This was a new compilation, drawn up by order of the king *c.* 654,

in which Germanic customary law predominated over Roman legal traditions.

It is probable, however, that the extent of the assimilation between the two races brought about by Reccared and Recceswinth has been somewhat exaggerated. Even after their reforms only nobles of pure Visigothic descent could be elected to the throne, and the Germanic nobility was thus encouraged to maintain its racial purity up to (and after) the final fall of the monarchy.

Wamba (672–680), the last king of any stature, attempted to reorganize the decadent military institutions of his kingdom but was deposed by a trick; and the energies of the next three kings were chiefly concerned with holding their thrones against the intrigues of their political enemies. By the time of Roderick (710–711) the disintegration of the state was complete. Apparently as a result of an invitation from Roderick's opponents, Tarik, the Arab governor of Tangier, crossed the Strait of Gibraltar with an invading Arab and Berber army in 711. Roderick called for national unity against this grave threat, but even then the Visigothic nobles declined to abandon their suicidal partisanship. Defeatism was deliberately encouraged by some of them in the army that Roderick had assembled; in consequence it was routed by Tarik south of Medina-Sidonia on July 19, 711, the king himself disappearing from history.

The collapse of the Visigothic monarchy in Spain followed at once. A substantial number of members of the large slave population abjured Christianity to improve their social position. The Spanish Jews, subjected to fierce persecution since the reign of Sisebut (612–621), welcomed the tolerant invaders as liberators. A number of high Visigothic nobles accepted Arab suzerainty. Toledo fell without resistance and the invaders speedily completed the occupation of almost the whole peninsula. Though the Arabs took one-fifth of the land for themselves the Christian communities were left with their religious and legal independence intact and were protected by Koranic law.

The social system of the Visigothic monarchy had consisted broadly of an aristocracy of nobles (*primates*) and clergy; a large class of freemen (*ingenui*) of mixed racial origin; and a numerous slave class (*servi*). The *aula regia* was made up of palace officials (*palatini*), who held the title of count by virtue of their offices. Provinces were governed by *duces* or *comites* and subdivided into areas ruled by *judices*. This stratified society without clear political ideals proved incapable of creating any sense of national cohesion during the two centuries of Visigothic rule in Spain. Though they were by no means without creative capacity in letters and the visual arts, it was as lawgivers that the Visigoths exercised their greatest influence on the history of Spain. Recceswinth's *Liber judiciorum* remained the basic legal code of Christian Spain for centuries after the disappearance of the last Visigothic king; and, by a curious paradox, memories of the feeble Visigothic state were now to prove strong enough to sustain the small Christian states of the north in their long and desperate resistance to Islamic military and cultural power. (P. E. R.)

MOORISH SPAIN TO 1031

During the reign in Spain of the Visigothic king Witiza (701–709) Arab forces of the caliphate (*q.v.*) had conquered northern Morocco and laid siege to Ceuta, the last remnant of the Byzantine possessions. It seems certain that Julian, the imperial count or governor of Ceuta, incited the Arab viceroy in North Africa, Musa ibn Nusair, to attack Spain, possibly by arrangement with the sons of Witiza. After a successful preliminary raid by the Berber Tarif in 710, Musa ordered his governor at Tangier, Tarik ibn Ziyad, to make a descent in force. In 711 Tarik crossed with 7,000 men, mostly Berbers, and, after reinforcement by 5,000 more, awaited Roderick's advance near the Laguna de Janda, south of Vejer de la Frontera. There, on July 19, Roderick was defeated, and probably killed. The sons and partisans of Witiza, who had withdrawn during the battle, now joined Tarik and encouraged him to advance northward to seize Toledo, while a detached force occupied Córdova. In June 712 Musa crossed with an army of 18,000, mostly Arabs, captured Seville and Mérida (after a siege lasting over the winter) and dispatched his son 'Abd al-Aziz to the southwest, before himself joining Tarik at Talavera. After wintering at Toledo, he resumed his advance in 714, captured Saragossa and made with Tarik a two-pronged expedition into León and Galicia before returning to Damascus at the caliph's command. His son 'Abd al-Aziz, after occupying Portugal, completed the conquest of Granada and Murcia, where the duke Theodomir (called Tudmir by the Arabs) became tributary. Other Arab forces pushed northeast into Septimania; and later governors continued to raid into Gaul, where, although defeated by Charles Martel at Poitiers in 732, they maintained an advanced base at Narbonne until its capture by Pepin in 751.

The invaders met little opposition on the whole. The sons of Witiza and other great Visigothic families, whether converted to Islam or not, compounded by payment of tribute for extensive domains. The Jews, freed from persecution, were eager allies, and the serfs gained a measure of freedom. Conversions were on a large scale, and the Spanish converts (*muwallads*) became an active and turbulent element in the general Moorish population. The unconverted, called *mustaribs* (Mozarabs), did not suffer much from interference and, like the Jews, formed prosperous communities in the Moslem cities. The Arabs, in spite of continuous immigration, were too few to colonize the country; they formed the administrative and military cadres, maintained by the allotment of fiefs in the region of Saragossa, the eastern and southern coasts and the Guadalquivir valley. The Berbers settled mainly in the centre and in mountainous regions which resembled their native land and favoured their anarchical tendencies. The Berber revolt in Morocco in 740 produced a parallel outbreak in Spain, and Berber risings, large and small, continued for the next two centuries. Among the Arabs, also, the intertribal feud between Kais and Kalb (*see* CALIPHATE) was reflected in Spain, complicated by rivalries between the early immigrants (*baladis*) and the Syrian contingents (*shamis*) which came in from 741 onward. In this patchwork of racial and tribal discords, rebellion was almost endemic, and the resulting lack of unity remained a permanent characteristic of al-Andalus, as Moorish Spain was called.

The Omayyad Dynasty.—In 755 the Omayyad prince 'Abd-ar-Rahman ibn Mu'awiya, escaping from the massacre of his relatives in Syria, landed at Almuñécar, and with Kalbite support captured Córdova (756). He waged pitiless warfare against internal disorder for more than 30 years and established a centralized power based upon an imperial guard of Berbers and European slaves (Slavonians). The pious Hisham I (788–796) tried to stem the expansion of the Asturian kingdom. A series of *muwallad* revolts, at Saragossa, at Toledo, at Mérida and in Córdova, threatened to dislodge his successor, al-Hakam I (796–822), who, however, by his implacable resolution reaffirmed the Omayyad power and laid the foundations of the material and literary culture which developed under the urbane 'Abd-ar-Rahman II (822–852). The peace was again broken under Mohammed I (852–886) by successive *muwallad* and Berber revolts, often with Asturian support, notably those of the "Visigothic" Beni Kasi at Tudela and Saragossa and of Omar ibn Hafsun in the Sierra de Ronda. General disintegration and civil war among and between Arabs, Berbers and *muwallads* in all quarters forced his son 'Abdallah (888–912) to adopt a temporizing and defensive strategy.

'Abdallah's grandson 'Abd-ar-Rahman III (912–961) by resolute and far-sighted leadership lifted the kingdom from this morass to an apogee of power, culture and magnificence. Within ten years he mastered Ibn Hafsun (d. 917) and his sons, and in ten more he reunited Moorish Spain. He proclaimed himself caliph in 929, taking the title of *al-Nasir* (the Victorious). In spite of occasional reverses, he forced the Christian princes to pay tribute to Córdova, now a metropolis enjoying immense reputation and prestige in all Europe. He greatly enlarged the Slavonian guard, whose officers acquired increasing power at the expense of the old Arab aristocracy. The reign of his son al-Hakam II (961–976) was no less brilliant. The weakness of Hakam's son Hisham II (976–1013) favoured the rise of a military dictator, the Arab Ibn Abi Amir, later entitled al-Mansur (Almanzor) (981–1002), under whom, as under his son 'Abd al-Malik al-Muzaffar (1002–08), the military

power of the caliphate, now recruited mainly from Morocco, was at its height. But beneath the surface the old antagonisms and social discontents were still alive. In 1009 they broke out in a political crisis which temporarily reduced Moorish Spain to chaos and from which it never fully recovered. While Slavonians and Berbers turn by turn occupied Córdova and made and unmade caliphs, the rest of al-Andalus threw off all allegiance to the Omayyads and was dissected among a host of petty rulers (*muluk al-tawaif*). After Córdova itself had been occupied by the Moroccan Arab house of Hammuda (1016–27), the last Omayyad prince was dethroned in 1031 and replaced by a republic.

(H. A. R. G.)

CHRISTIAN SPAIN TO 1479

The history of Christian Spain during the period 711–1479 is dominated by the interplay of two opposed forces: the idea of an ultimate Christian reconquest of the whole peninsula; and the inescapable fact that, both culturally and economically, the Islamic and Christian peoples of Spain had become irretrievably associated with each other. The urge to reconquer, based partly on religious belief, partly on the Visigothic tradition of Spanish unity and partly, too, on economic pressures, did not express itself predominantly through the forms of militant Christianity until four centuries after the Moorish occupation. Then, when religious fanaticism had made its appearance, among both Moors and Christians, at the end of the 11th century, the Christian conquest of Andalusia was undertaken in a spirit of intolerance that brought ruin to southern Spain and severely damaged the economy of the conquering power, Castile. Once again, however, the fact of the *convivencia* between the two Spains asserted itself, and fanaticism had to be moderated. Thus, in 1238, Ferdinand III, the conqueror of Andalusia, was compelled to acquiesce in the foundation of the new Moorish kingdom of Granada and to give it his protection because it was recognized that without Moorish tribute, the economic structure of Castile could not survive while it digested the newly reconquered territories. For the same reason the Christian kings had to continue to accept the existence within their frontiers of communities of Islamic vassals (Mudéjares) to whom they also gave protection.

The Kingdom of Asturias.—Before the tide of Arab and Berber invasion a number of Visigothic nobles and their retinues withdrew to the mountains of Asturias and there elected a Visigothic prince, Pelayo (?718–737), as their king. The Berber revolt in Morocco and Spain and the consequent withdrawal of the Berber garrison in Galicia (750) led to the attachment of Galicia to the new kingdom by Alphonso I (Alfonso; 739–757). Under Alphonso II (792–842) the Asturian capital was moved from Cangas de Onís to Oviedo. Later kings pushed forward from the mountains to the central plain and their armies operated with success far to the south. The supposed discovery of the tomb of St. James at Compostela in Galicia (*c.* 830) made the Asturian kingdom guardian of a shrine of European significance and gave it a symbol of national unity. Alphonso III (866–910), the greatest of the Asturian kings, expanded his domains as far as Coimbra in the southwest and Burgos in the southeast. By now the situation was stable enough for his successor, García I (909–914), to move the capital from the Asturian mountains to the city of León.

The Asturian kings, like their Leonese successors, regarded themselves as direct heirs to the traditions of the fallen Visigothic monarchy; and, for this reason, the historical writings of Alphonso III's reign emphasized their obligation to undertake the reconquest of all Spain. The Asturian administration followed Toledan models as closely as circumstances permitted. Yet the organization, social and political, inherited from the Visigoths was already being changed by the immigration of a free Mozarabic peasantry from the south, while Frankish political ideas also exerted some influence on Asturian society. In general, however, it can be said that the Asturian kings handed on to the kings of León a tradition of isolation, not only from Moorish Spain but also from the rest of Christian Europe, which experience was to show could not be sustained.

The Kingdom of León.—The transfer of the Christian capital from Oviedo to León coincided with the establishment of the caliphate of Córdova and, until the early 11th century, the balance of power in the peninsula swung strongly in favour of the Moors. That the Leonese kingdom survived at all was the result of the generalship of Ordoño II (914–924) and Ramiro II (932–950) who, in unfavourable circumstances, secured important victories against the caliph's armies. Ramiro was, however, unable to prevent the breakaway from his kingdom of the Christian county of Castile. This district, partly recolonized by immigrants from the north, had rejected the Visigothic *Forum judicum* in favour of Castilian customary law and showed itself hostile in other ways to the Visigothic traditions of León until, under Count Fernán González (923–970), it achieved *de facto* independence. Henceforth clashes between the innovating tendencies of Castile and the conservative traditions of León were to weaken the Christians. Ordoño III of León (951–956) acknowledged the hegemony of the caliphate, and the later attempts of Bermudo II (984–999) to assert his independence of Córdova brought disaster on the kingdom. Almanzor's armies occupied and destroyed León (988) and sacked Santiago de Compostela itself (997), sparing only the Apostle's tomb. Under Alphonso V (999–1027) the capital was repopulated and recovery began. However, early in the reign of Bermudo III (1027–37), a new threat to the independence of León materialized when Sancho III the Great of Navarre, after seizing the county of Castile (1028), occupied the eastern part of the Leonese kingdom. Later he entered the capital (1034) and there assumed the title of "emperor of Spain," Bermudo fleeing to Galicia. León was never wholly to recover its supremacy again.

The Rise of Catalonia and Navarre.—During the reign of Charlemagne the Franks succeeded in capturing Barcelona (801) and in driving the Moors out of the northeastern corner of the peninsula. This area, later known as the Marca Hispanica, was then occupied by a number of Visigothic counts under Frankish suzerainty. In the time of Count Wilfred (865–898) the county of Barcelona became independent and established its hegemony over the other counties of the Marca Hispanica. Under Borrell II (954–992) Almanzor's armies devastated Catalonia and burned Barcelona (985); but the region recovered during the rule of Raymund Borrell (Ramón; 992–1018). Under Berengar-Raymund I (Berenguer Ramón, 1018–35) it began to flourish. Because of the influence of the Franks it had already developed political and social institutions different from those of the rest of Christian Spain. The 10th century also saw the dramatic rise of the Basque kingdom of Navarre, which has little history before this period. Sancho I Garcés (905–926) added the small county of Aragon to his dominions but suffered severely at the hands of 'Abd-ar-Rahman III's armies. A period of subordination to the caliphate followed, but Sancho II Garcés (970–995) was spared the worst of Almanzor's attacks and devoted himself to the political organization of his kingdom. There was, however, nothing to indicate that, under Sancho III the Great (1000–35), Navarre was to become, for a brief period, the dominant power in Christian Spain.

The Mediaeval Spanish Empire.—Though Sancho the Great had assumed the imperial title in León, his political ideas were hostile to the old Visigothic-Leonese tradition of peninsular unity, and his empire was largely built on a simple desire for personal aggrandisement. It included Navarre, Castile, León, Sobrarbe and Ribagorza while Berengar-Raymund I of Barcelona became his vassal. Sancho deliberately encouraged contacts between his dominions and the rest of Christian Europe and probably did much to make the Spanish nobles aware of the feudal aspirations of their class north of the Pyrenees. He regarded his empire as his personal property and distributed it among his four sons in his will. The eldest, García Sánchez III (1035–54), inherited an enlarged Navarre. Castile was made a kingdom and given to his second son, Ferdinand; Sobrarbe and Ribagorza were united in a separate kingdom and given to another son, while the small county of Aragon was also erected into a kingdom for his fourth son, Ramiro I (1035–65). Ramiro soon incorporated Sobrarbe and Ribagorza into his new kingdom (1037), and Aragon thus began its independent history.

Sancho's will provided for the return of León to Bermudo III on his death, but this restoration was overthrown by Ferdinand I of Castile (1037–65), who speedily annexed León and reoccupied (1054) the Castilian territories given in his father's will to his eldest brother, García of Navarre. Ferdinand also took the imperial title and by virtue of this claimed suzerainty over his brothers. He achieved great successes against the Moors, pushing his frontiers to the Tagus in the south and the Mondego in the west and making the *taifa* kings of Saragossa, Seville and Toledo his tributaries. On his death Ferdinand, as his father had done, divided his dominions among his sons, and there was a period of violent strife before the second of them, the brilliant Alphonso VI (1065–1109), succeeded to the whole inheritance. His reign brought great changes to Spain. Toledo was permanently reoccupied (1085); and soon afterward the Cid Campeador took possession of Valencia (1094). All the *taifa* kingdoms paid tribute to Alphonso. There was no attempt to expel the newly conquered Moorish populations, who were promised royal protection and the preservation of their religion and customs. To give formal expression to his tolerant policy, Alphonso changed the imperial title by styling himself "emperor of the two religions." It seemed for a brief time as if these ideas might produce a genuine Hispano-Moorish civilization under Christian political control. Unfortunately the Almorávides' invasion of Andalusia (1086) soon removed the *taifa* kingdoms from Alphonso's control and replaced the tolerant spirit of the Spanish Moors by one of bigoted fanaticism. Meanwhile, under the influence of his wife Constance of Burgundy, the emperor allowed French Cluniac monks to get complete control of the Spanish church. The Spanish tradition of *convivencia* was wholly repugnant to these ardent disciples of Gregory VII's doctrine of the universal church. They forced Alphonso to modify his promises of toleration toward the conquered Mudéjares, compelled him to replace the old Mozarabic missal by the Roman and contrived to have the ancient Visigothic script of Spain replaced by the Carolingian hand in use in the rest of western Europe. These innovations brought into Christian relations with the Moors something of the crusading zeal which the French monks had found deplorably lacking and ensured that mediaeval Spain would in the future belong much more definitely to the common political and cultural pattern of mediaeval Europe. This uprooting of the traditions of the past was, however, partly responsible for the backwardness of Spanish culture for the next 150 years, while the new determination to regard the reconquest as a crusade rather than a recovery of territorial dominion in the end wrecked the Spanish economy, both Moorish and Christian.

Alphonso was followed by his daughter Urraca (1109–25), widow of Count Raymund of Burgundy; her turbulent reign cannot be followed here. She was succeeded by her son, Alphonso VII (1126–57), who gave new meaning to the imperial title by extending his influence over the county of Barcelona and by forcing Ramiro II of Aragon and García Ramírez of Navarre to accept his suzerainty. The counts of Gascony and Provence also accepted him as overlord and, at a ceremony at León (1135), attended by his new Christian and Moorish vassals, he assumed the title "emperor of all Spain." He also achieved resounding, if inconclusive, successes against the Moors, and the European prestige of the Spanish imperial court was recognized when Louis VII visited Alphonso at Toledo (1150). Nevertheless separatist tendencies in the peninsula secured another ominous victory when the count of Portugal, Alphonso Henriques, extracted from the emperor recognition of Portugal's status as an independent kingdom (1139). On his death, Alphonso VII, too, divided his realm among his sons, leaving Castile to the elder, Sancho III, and León to Ferdinand II. This new partition proved the deathblow to any future revival of the imperial idea originally inherited from the Visigoths. Castile and León now remained separated for 73 years, during which the magnates began to display in a marked degree that anarchical partisanship and resistance to royal authority which were to plague Spanish life for the next three centuries. Some important successes were nevertheless achieved against the Moors during this time. In 1212 Alphonso VIII of Castile (1158–1214), Sancho's son, supported by the armies of Aragon, Navarre and Portugal, routed the Almohade amir of Morocco, Mohammed III, at Navas de Tolosa and so removed the last really serious Islamic threat to Christian hegemony in Spain. In an atmosphere of ever-mounting religious fanaticism the way was now open to the conquest of Andalusia.

The Rise of Aragon.—The kingdom of Aragon was brought into existence (1035) by a testamentary act of Sancho the Great of Navarre. During the rest of the 11th century it gradually encroached on Moorish territory north of the Ebro. Saragossa was captured (1118) by Alphonso I (1104–34) who, in spite of his entanglements in the internal affairs of Castile, was able to follow up this important success by occupying a large area southwest of the former Moorish capital. In 1137 Petronilla, the infant daughter of Ramiro II (accession 1134; death 1154), was betrothed to Raymund-Berengar IV, count of Barcelona, to whom Ramiro thereupon handed over the government of Aragon (1137–62). The son of the marriage, which was solemnized in 1150, Alphonso II (1162–96), formally initiated the personal union of both countries under the house of Barcelona which was to last until 1410. In 1179 an important step for the future of Aragon was taken when Alphonso made an agreement with Castile whereby the task of reconquering the Moorish kingdom of Valencia was reserved to the Aragonese crown. In exchange Aragon relinquished any claims to any other Moorish-held territory in the peninsula. During this reign, too, Aragon developed political interests north of the Pyrenees; Provence and Roussillon fell to Alphonso by inheritance. This preoccupation with French affairs proved fatal to Alphonso's son, Peter II (1196–1213) who, having added Montpellier to his dominions by marriage, became involved in the Albigensian war and was killed by Simon de Montfort at the battle of Muret (1213). His death put an end to major Aragonese political pretensions north of the Pyrenees. It is to be noted that, from 1137 to 1412, Catalonia was the dominant partner in the association of the two countries. Economically and culturally Catalonia was much in advance of isolationist and conservative Aragon, whose nobility often found itself more in sympathy with the politics of the Castilian magnates than with the bourgeois ideas of the nationalistic Catalans. Catalan was the language of the court of Aragon during the whole of this period.

During the reign of Ramiro II Navarre seceded (1134) from the Aragonese crown to which it had been united since 1076. Sancho VII of Navarre (1194–1234) tried to nominate James I of Aragon as his successor and to reunite the two kingdoms, but the Navarrese preferred Thibaut (Theobald) IV of Champagne as king (1234–53), and Navarre was thereafter ruled mostly by a succession of French princes. The little mountain kingdom nevertheless retained most of its distinctively Spanish institutions.

End of the Reconquest.—The last king of León, Alphonso IX (1188–1230), was succeeded, on his death, by his son, who was already king of Castile. Castile and León were thus finally united. The new sovereign, Ferdinand III the Saint (1217–52), at once embarked on a great series of campaigns to subdue Andalusia. These began with the capture of Córdova (1236) and culminated in the surrender of Seville (1248). Influenced by the crusading zeal instilled into the Spanish church by the Cluniac and Cistercian orders, Ferdinand at first expelled the Moorish inhabitants of the Andalusian cities en masse but was later forced to modify his policy by the collapse of the Andalusian economy that inevitably ensued. He also assented, chiefly for financial reasons, to the establishment of the new Moorish kingdom of Granada under Castilian suzerainty. The Granadine Moors were forced to create, by hard work, a high degree of productivity in their new state in order to pay to Castile a large annual tribute, which probably became essential to the well-being of Castile's finances. This reviving tendency toward a renewal of the older tolerance toward the Moors was also reflected in the cultural field; Toledo now became famous throughout Europe as a centre where Moorish philosophy and science were, under the stimulus of successive archbishops, made available in translation to European scholars. During the same period James I of Aragon (1213–76), Peter II's son, who was legislator and *littérateur* as well as warrior, completed Aragon's part in the reconquest. After occupying the Balearics (1235) he captured Valencia (1238) and so established the final peninsular

frontiers of Aragon. By these Castilian and Aragonese achievements the reconquest was, for all practical purposes, brought to an end. Both countries now gave themselves up to internal social and constitutional struggles.

Castile and León From 1252 to 1479.—The reign of the cultured Alphonso X the Wise (1252–84) was politically ill-fated. Alphonso involved himself in wars with Portugal, Aragon and Navarre and, having got himself elected to the German throne (1257), devoted much effort and treasure to an unsuccessful attempt to make good his claims against papal opposition. At home a constitutional struggle developed over the rival claims to the succession of Ferdinand of La Cerda and of Sancho, respectively the grandson and the second son of the king. This ended in an attempt by a junta of prelates and nobles, supporters of Sancho, to depose Alphonso (1282) and saw the beginning of an acute phase in the violent clash between the magnates and the crown which was to dominate Castilian politics for 200 years. Alphonso's legal works, based on the renascent doctrines of Roman and canon law, laid the theoretical foundations for monarchical supremacy; but the king was too inept a politician to succeed in enforcing them. The troubled reign of Sancho IV (1284–95) was complicated by war with Aragon, which supported the succession of the house of La Cerda; and Castile did not emerge from continuous civil strife until the end of the minority of Sancho's grandson Alphonso XI (1312–50). This Alphonso, aided by the armies of the autonomous municipalities (*concejos*), sternly repressed the magnates and succeeded in establishing a measure of absolutism which found expression in the legal and political reforms promulgated at Alcalá (1348). He also took up arms successfully against a threat from Morocco, defeating the invaders at Salado (1340).

The trend toward strong central govérnment would probably have continued under his son Peter I the Cruel (1350–69) without major trouble, had not Alphonso left a number of powerful bastard sons to contest his legitimate heir's inheritance. The eldest, Henry of Trastamara, posing as the defender of the magnates against royal attempts to limit their ancient privileges, began a struggle to seize his half-brother's throne which lasted about 19 years. Meanwhile Aragonese designs on Murcia—Castile's only outlet to the Mediterranean—and Castilian ambitions in western Aragon plunged the two countries into a ferocious war (1356–66) which became even more embittered when Peter IV of Aragon espoused the Trastamaran cause in Castile. It was not, however, until France intervened actively in support of Peter IV and Henry of Trastamara that Peter of Castile's victories in Aragon were checked. Both France and England believed that the support of the powerful Castilian fleet might prove decisive in settling the outcome of the Hundred Years' War. Peter I's alliance with England (1362) therefore led to the dispatch of the free companies to Castile, financed by France and the pope (1365). Peter fled to Gascony and secured the military intervention of the Black Prince, who routed Henry's supporters at Nájera (1367). Peter was restored but quarrelled with his English ally, who then began negotiations with Aragon, Navarre and Portugal for a quadripartite partition of Castile. In 1369 Henry of Trastamara, again with French help, defeated and killed the Castilian king at Montiel and finally seized the kingdom (1371).

To counteract this French success John of Gaunt married Peter's daughter and heiress Constance (1371) and thus became pretender to the Castilian throne (1372–87). He was, however, unable to invade Castile during the reign of his rival Henry II (1369–79), who succeeded in consolidating his position against strong loyalist opposition. For several decades to come a close political and military alliance with France dominated Castilian foreign policy, and Castilian sea power menaced the English south coast. In spite of the conservative and class-conscious ideas of the Trastamaran monarchy, the constant wars in which they were involved compelled Henry and his son John I (1379–90) repeatedly to seek financial aid from the third estate, which thus in the latter's reign secured a considerable measure of control over the government. John brought military and financial ruin to his kingdom, partly because of new Lancastrian attempts to conquer Castile and partly because of his own ambitions to add Portugal to his crown. His crushing defeat by the Portuguese at Aljubarrota (1385) was followed by the landing of John of Gaunt in Galicia with a large army (1386). French support and John of Gaunt's military ineptitude, however, saved the Trastamaran dynasty and the dynastic dispute was finally settled (1388) by the marriage of Catherine of Lancaster, Peter I's granddaughter, to John I's heir, Henry III (1390–1406). Violent outbreaks against the Jews marked Henry's accession. In his short reign the crown reverted with success to the policies of Alphonso XI and Peter I and reduced the magnates to order by stern measures. At this time, too, the occupation and colonization of the Canary Islands was begun. The long reign of the weak John II (1406–54) saw a continuance of the political disintegration of Castile as a result of a renewal of the struggle of the magnates with the crown. Civil war was now frequently accompanied by armed Aragonese and Navarrese interventions. John, a politically inept lover of the arts, left the government in the hands of his favourite, Alvaro de Luna. Luna's power was not broken until 1453, when the king, acceding to the demands of the nobles and of the queen, ordered his execution.

Under the impotent Henry IV (1454–74) the Trastamaran dynasty reached its lowest ebb. Civil strife continued unabated and was now accompanied by a collapse of the moral standards of the court. The king had married Joan (Juana) of Portugal (1455), but their daughter and heiress, also called Joan, was alleged to be the child of an adulterous relationship of the queen's. Opponents of the court accordingly favoured the adoption of the king's sister Isabella (Isabel) as heiress to the throne. The magnates formed a league in support of Isabella though popular opinion tended, as usual in such circumstances, to support the king. Joan's marriage to Alphonso V of Portugal and his active support of her claims led, on Henry's death, to five years of civil war before Isabella triumphed and was acknowledged to be queen of Castile by the peace of Trujillo (1479). Ten years before she had married Ferdinand, son and heir of John II of Aragon. Thus, on John II's death (also in 1479), the crowns of Castile and Aragon became associated though not yet formally united. The 15th century had so far been an age of political, financial and moral decay in Castile. There was nothing to hint, in 1474, that the central kingdom would, within a few years, become the heart of a world state.

Aragon and Catalonia, 1276–1479.—The conquest of Valencia having ended Aragonese participation in the reconquest, it became necessary for the eastern kingdom to expand outside the peninsula if it was to avoid becoming subordinate to the hegemony of Castile. Under Peter III (1276–85) Aragon therefore began political and military intervention in Italy. As husband of Constance, daughter of the emperor Frederick II's bastard son Manfred (*see* SICILY), Peter accepted the throne of Sicily and promised to free his new subjects from the rule of Charles of Anjou. In pursuit of this aim Aragonese troops and Catalan ships secured substantial successes against French military power and papal diplomacy and made Aragon a major factor in Italian politics. Papal pressure on Peter III's sons Alphonso III (1285–91) and James II (1291–1327) resulted in Aragonese undertakings to abandon Sicily; but these were not accepted by the Sicilians, and James's brother Frederick (Fadrique) was eventually recognized as king of the island of Sicily for life (*see* FREDERICK III). James also secured the grant of Sardinia to Aragon as a papal fief (1297), though the island was not completely subjugated until the 15th century and Aragonese claims to it led to prolonged hostilities with Genoa. Aragonese power in the Mediterranean was further strengthened by the activities of the Catalan company in Asia Minor and Greece, which eventually brought about the establishment of the Catalan duchy of Athens, first under Sicilian and then under Aragonese suzerainty. The Aragonese dynasty in Sicily remained closely associated with the Aragonese crown, and both countries were united when Martin I of Aragon (1395–1410) succeeded his son Martin I of Sicily in 1409. The reign of Alphonso V (1416–58) saw a further dramatic extension of Aragonese power in Italy when, after a long struggle, Alphonso finally became the undisputed ruler of the kingdom of Naples (1443) and transferred his capital there.

These events reacted unfavourably on the political situation in the parent kingdom, where Aragonese and Catalan particularism grew apace and reached its climax under John II (1458–79), when Catalonia was in revolt against the crown for 11 years. The discontent of the Catalans dated back to the compromise of Caspe (1412), which had placed Alphonso V's father, the Trastamaran prince Ferdinand I (1412–16), on the Aragonese throne when the male line of the house of Barcelona had become extinct. Since 1164 Catalonia had been the dominant partner in the Aragonese union. Now the Catalans feared, with some reason, that the Castilian dynasty would favour Aragonese interests at their expense.

The preoccupation of successive kings with Italian affairs weakened the crown at home. In 1283 the nobles and municipalities of Aragon formed an *unión* and extorted a General Privilege granting them special privileges and immunities. These were increased substantially in 1287 and remained in force throughout the reigns of James II (1291–1327) and Alphonso IV (1327–35). But in 1348 Peter IV (1336–87), who had strong absolutist leanings, routed the army of the *unión* at Epila and put an end to an intolerable constitutional situation. Peter's reign represents something of a break with the general policy of Aragon during this period. Having annexed the kingdom of Majorca (1344), he challenged the peninsular hegemony of Castile (1356); but, in the 11 years' war which followed, a large part of his kingdom was overrun and occupied by the armies of Peter I of Castile. He saved himself by joining in the French effort to dethrone his Castilian rival and replace him by Henry of Trastamara. Henry promised to cede the whole of eastern Castile to Aragon in exchange for Aragonese support but declined to fulfil his promise when he had secured the Castilian throne. Long years of negotiation for an Anglo-Aragonese alliance against Castile followed, during which Peter IV sought to play off French and English interests in the peninsula for his own ends without finally committing himself to either side. In the last years of his reign the common problems with which the schism in the papacy faced the peninsular kingdoms led him to try to form a peninsular bloc that would be neutral in both religious and military affairs; but French pressure prevented the success of this far-sighted plan. Peter's son John I (1387–95), elder brother and predecessor of Martin I, left the conduct of diplomacy largely to his French wife Yolande of Bar, and Aragonese foreign policy became largely subservient to that of France. The accession of a Castilian prince to the Aragonese throne as Ferdinand I in 1412 had already foreshadowed the union of the two largest peninsular states, and this was made certain in 1469, during the reign of Ferdinand I's second son John II, when Ferdinand, the heir to the throne (born 1452), married the Castilian heiress Isabella (born 1451). The same year which saw Isabella undisputed ruler of Castile (1479) brought her husband to the Aragonese throne as Ferdinand II. The destinies of the two countries were thereafter bound together.

Mediaeval Institutions: Asturias, León, Castile.—The Asturo-Leonese kingdom inherited from the Visigoths the principle of elective monarchy; but in practice its sovereigns were always chosen from the same family. By the 11th century primogeniture was accepted, though the elective tradition was never formally abandoned. In accordance with their declared policy, the first Asturian kings sought to re-establish at Oviedo the civil and ecclesiastical administration which had existed at Toledo. Outlying regions were governed by *comites* in the king's name, and the officers of the royal household were the direct successors of the Visigothic *palatini*. The steward (*maiordomus*) and the standard-bearer (*armiger regis* or *dapifer*) eventually emerged as the chief civil and military officers of the court. By the 13th century the governorship of the larger provinces was usually held by an *adelantado*, who exercised regional military and judicial authority on behalf of the king. In some areas similar duties were carried out by the *merino mayor*. The *merino menor*, who was appointed by either of these two officers, acted as their deputy in matters of criminal jurisdiction reserved to the crown (*voz de rey*). The traditional military organization of the country was overhauled by John I of Castile in 1383 when, following French practice, a constable aided by two marshals took over the duties previously performed in the king's name by his *alférez*. The office of admiral of Castile was created by Ferdinand III. The *cancellarius* rarely appears before the 12th century. By the time of Alphonso X, however, the chancellor was recognized to be the chief civil officer of the household. The *cancillería*, like the court itself, remained ambulatory until the end of the mediaeval period. The employment by the king of a privy seal (*sello de la poridad*), independent of the chancery and held in the custody of its own *canciller*, was well established by the end of the 13th century. As elsewhere in Europe the privy seal was evidently used by Alphonso XI and Peter I to further the direct control of the king over the government since, in the political reaction brought about by the accession of Henry of Trastamara, attempts were made to restrict its use. The treasury, whose officials were largely Jewish, was headed by the *almojarife mayor*, later known as the *tesorero mayor*. Peter I created several *contadores reales* to supervise the work of local tax collectors.

In theory feudalism did not exist in Castile or León, the magnates (*ricoshombres*) holding their territorial possessions, from which their rank derived, by virtue of an individual grant for life only from the king. Vassalage was not tenurial, and lands were not held for military service. In practice, however, feudal grants did occur and, from the 10th century onward, the magnates constantly attempted to usurp feudal rights for themselves. The ambiguity of their status may, indeed, have been partly responsible for their determination to control the crown. The lesser nobles (*infanzones* and *hijos d'algo*) could be vassals of the crown or of a lord. The free peasants paid tribute for permission to till the soil either to the king on royal lands (*realengos*) or to a lord on seigneurial lands (*señoríos*). On the *behetrías* the peasants were semifree, having the theoretical right to change their lord at will. This institution was chiefly found in those areas of Castile and León proper which had been repopulated by Mozarabic immigration during the reconquest. The serfs (*solariegos*) were chiefly found on the lands of the magnates, of the military orders and of the church (*abadengos*). The status of the Mudejares (Moslems living under Christian protection) varied according to political conditions and to locality. The Jewish communities in the towns (*aljamas*) enjoyed royal protection until the end of the 14th century, when an outburst of popular fanaticism led to great massacres of Jews in Seville, Córdova, Toledo and other cities (1391) and Jewish communal life was irreparably damaged.

The most notable institution of mediaeval Castile was the semi-autonomous municipality (*concejo*) comprising a town (*villa*) and its surrounding *comarca*. The rise of the *concejos* dates from the 11th century, and they were largely brought into being by the need to attract settlers, by special privileges, to newly reconquered areas. This was achieved by the royal grant of a franchise (*fuero*) giving the *concejo* in perpetuity a large measure of control over its own affairs and its own legal code. The *concejos* raised their own militia and formed military associations with each other (*hermandades*) in defense of their interests. They were invaluable allies of the crown in its struggles with the nobles, particularly during the 14th century, when their deputies (*procuradores*) to the Castilian parliament (*Cortes*) dominated that body. The appearance of the third estate in the *Cortes* dates from the end of the 12th century.

In the later middle ages the king depended on the vote of the *procuradores* for the grant of financial aid (*servicio*) necessary to cover civil and military expenditure. The heavy costs of the wars of Henry II and John I increased the dependence of the king on the third estate. John I was forced to meet the severe criticisms of his policy voiced by the *procuradores* by creating the royal council (1386). At the beginning four representatives of each estate were appointed to the new body, and the king delegated to it (theoretically at least) many of his powers. In the 15th century the *concejos* themselves fell victims to the growth of the royal power and had to accept royal officials known as *corregidores*. A hint of impending decay had already appeared during the reign of Henry III (1390–1406) when, against the wishes of the *Cortes*, the king succeeded in making permanent the *alcabala*, a highly

unpopular sales tax of one-tenth, which then became one of the main sources of royal revenue.

Mediaeval Institutions: Aragon, Catalonia, Valencia.—In spite of their union under the Aragonese crown, these three regions retained their own separate institutions and parliaments. There was a strong tendency toward feudalism in Aragon from the beginning and grants of fiefs for military service were frequent. After the union with Catalonia, where feudalism was fully developed, the Aragonese nobility intensified its efforts to obtain full feudal rights. The smaller municipalities (*universidades*) were less important than in Castile partly because of the dominant position of Saragossa, Barcelona and Valencia respectively in the parliamentary representation of the popular arm. The Aragonese nobility consisted of four classes, *ricoshombres de natura, ricos-hombres de la mesnada* (created by James I), *caballeros* and *infanzones*. These, like the representatives of the *universidades,* attended parliamentary sessions by right. In Catalonia there were three classes of magnates (*comtes, vescomtes* and *valvassors*). The Aragonese *Cortes,* which consisted of four estates (*brazos*) because of the separate representation of the greater and lesser nobles, was governed by complicated rules of procedure. Redress preceded supply, and unanimity was in theory required for the approval of any measure. The *Cortes* also claimed the right to declare war. The most important political office in the country was that of the *justicia de Aragón,* whose presence at sessions of the *Cortes,* to which he was responsible, was indispensable. He was arbitrator between king and nobles or people when disputes arose with the crown, and his decisions were virtually final. A permanent parliamentary commission (*diputación del reyno*) existed to deal with finance, observance of the *fueros* and matters pertaining to the peace of the realm. The organization of the Catalan and Valencian *corts* was somewhat similar, though the former asserted a right to control all legislation. A *diputació general* in Catalonia performed functions like those of the Aragonese permanent parliamentary commission. In spite of the many restrictions on their powers it has been suggested that the Aragonese kings had greater *de facto* authority than the kings of Castile, except perhaps during the domination of the *unión* of 1283–1348. The appointment of the *justicia* was made by the crown, and only the crown could initiate legislation.

A peculiar feature of Aragonese administration was the absence of intermediate authorities between the central government and purely local officials. In Catalonia, land belonging to the king was ruled on his behalf by the *veguers,* with subordinates known as *sots-veguers* or *batlles;* both were crown appointments. The importance of Barcelona gave that city a dominant place in the life of Catalonia. It was ruled by the *concell de cent* though, in practice, this body's authority was normally delegated to a small *concell* of five persons with whom were associated the *veguer* and the *batlle*. Catalan law was based on the *Usatges* (*c.* 1060), the oldest written feudal code. The kingdom of Valencia was governed by the constitution (*Furs*) granted to it by James I in 1239. There were wide variations in the conditions of the unprivileged classes under the Aragonese crown. The legal and economic position of the Catalan serfs (*pagesos de remença*) was worse than that of serfs anywhere else in the peninsula. On the other hand the Mudéjares were comparatively well treated in Aragon proper where, until the 15th century, crown and nobles appreciated their economic importance and allowed them freedom of worship and their own tribunals. The Jews, vassals of the king, were favourably treated until the 15th century when, as in Castile, their position rapidly deteriorated. (P. E. R.)

THE "CATHOLIC KINGS"

The accession of Isabella has long been traditionally regarded as the final stage of Spanish mediaeval history. When Ferdinand succeeded to the crown of Aragon (1479), which possessed also Catalonia and Valencia as well as the Balearic Islands, the married pair ruled a larger area of the peninsula than had been united for many centuries. Though Castile remained a sovereign and independent state completely separated from the states of the crown of Aragon, the combined power of the devout and energetic Isabella and the subtle and unscrupulous Ferdinand (V of Castile and II of Aragon) soon sufficed to make Spain one of the leading powers of Europe.

In Castile aristocratic anarchy, rapacity and violence had reduced the whole land to poverty and taught the people to look to the crown for strong government. Ferdinand and Isabella, to whom Innocent VIII gave and Alexander VI confirmed the title of "Catholic kings," responded to the demands of their subjects by a series of measures directed to the weakening of the overmighty nobility. The rebellious nobles were first defeated in a series of battles in many parts of Castile, and all castles not entirely necessary for the defense of the kingdom were destroyed. Soon the private jurisdictions of the nobility were considerably limited and their most pretentious imitations of royal customs forbidden. The administration of the great military orders, hitherto the exclusive preserve of the nobility, was progressively taken over, with their revenues, by the crown. The feature of the royal policy that contributed most toward establishing an efficient bureaucracy was, however, the principle of depriving the nobles of their influence on the royal council, replacing them by *letrados* (trained civil servants steeped in the absolutist traditions of Roman law) or by ecclesiastics (owing their position to royal patronage). The nobles were also attracted to the court, where they tended to lose their local influence, to dissipate their wealth in ostentatious living and to change from powerful chieftains to obsequious courtiers under the jealous vigilance of the crown, which was indeed their only source of employment and honours. They continued too to be excluded from the *Cortes,* in accordance with the 15th-century custom of regarding the superior estates as no longer essential thereto. Meanwhile the number of towns represented in the *Cortes,* which had once been as many as 50, sank to 18; and any further enfranchisement was opposed by the privileged towns (*Cortes* of Valladolid, 1505). Henceforth a meeting of the *Cortes* of Castile consisted merely of 36 *procuradores,* two from each town—a number conveniently small for the exercise of royal influence. These *procuradores* never succeeded in asserting the principle that redress precedes supply; and the old custom whereby every *procurador* was bound by detailed instructions (*poderes*) from his town was abrogated at least as early as 1505 (probably much earlier). Henceforth they were plenipotentiaries, free to consent to any proposal of the crown. Even their legislative power was not complete, since one of the 83 laws to which the *Cortes* of Toledo of 1502 agreed recognized that kings possessed the power "to make *fueros* and laws and to interpret and amend them whereinsoever they shall think it convenient to do so." The Laws of Toro (1505) did much to remedy the diversity of laws from which Castile suffered.

In the states of the crown of Aragon the Catholic kings were unable to make any large direct increase in their power. The greater nobles of Aragon enjoyed feudal liberties and privileges almost unequalled elsewhere in Europe and kept the peasantry in abject servitude. The Aragonese *Cortes* with its four estates was so powerful as to reduce royal power to a mere shadow—a shadow still further diminished by the overwhelming power of the *justicia mayor,* who, sitting in a chair, crowned the kneeling king with the formula: "We who are as good as you swear to you who are no better than we, to accept you as our king and sovereign lord, provided you accept all our liberties and laws; but if not, not." The *Corts* of Catalonia and that of Valencia were only a degree less powerful than that of Aragon. It was only by skilful use of the church and of the Inquisition that royal power could be increased in these eastern regions of the peninsula.

The Spanish Inquisition.—The Spanish church was wealthy and powerful because the people were intensely religious and because it was largely a national institution in which no foreigner might hold office and in which the crown was supreme (papal power having been reduced almost to the vanishing point). It was, consequently, a fact of serious political importance that during the anarchy of Henry IV's reign (1454–75) the Jews gained great power and influence. They might compel—sometimes by means of their usury—their debtors to renounce the Christian religion; and Marranos (baptized Jews) often preserved their old religious faith

in secret. At the same time the power of the Moriscos (baptized Moors) had increased, and they were reviving ancient heresies such as the half-forgotten Manichaeism. The Catholic kings consequently consulted Pope Sixtus IV, who thereupon issued a bull (Nov. 1, 1478) authorizing them to choose two or three inquisitors notable for their virtue and learning, to whom he granted jurisdiction. The bull was put into force by a royal *cédula* (decree) issued in Medina del Campo (Sept. 17, 1480) ordering the establishment of the Holy Office in Castile.

As a royal instrument for the strengthening of the monarchy and the unification of the peninsula the value of the Inquisition would be difficult to exaggerate. The whole organization was completely under royal control, papal sanction being little more than a matter of form. The *suprema*, which controlled it, was merely one of the royal councils; and the officials of the Inquisition, from the inquisitor general downward, were the paid servants of the crown, who could appoint them or dismiss them at its pleasure. Also, except for one brief period (1507–17), there was to be but one Inquisition and one inquisitor general for the whole of Spain. As far as the Inquisition was concerned the political divisions between Castile and Aragon, Valencia and Catalonia soon ceased to exist.

Furthermore the Inquisition took no account of those privileges of the nobility and those extensive liberties and franchises that had been a ceaseless embarrassment to the crown. Indeed, it was used by the crown to achieve secular aims in cases where the ordinary courts were difficult to move, or where the laws and customs of the country were used to place obstacles in the path of the advancing monarchy. The familiars of the Inquisition, exercising ceaseless vigilance in the remotest corners of Spain, may be fittingly compared with the justices of the peace who did so much to uphold the throne of Tudor England. Though frequently popular among Spaniards of pure Old Christian descent, the new inquisitorial activities met with strong protest in Aragon, where it was established in 1484. The inquisitor of Saragossa, Pedro de Arbués (1441–85), was murdered—probably at the instigation of rich Marranos (baptized Jews). Throughout the centuries the chief victims of the Inquisition were men of Morisco or Marrano ancestry. It cannot be denied that the Inquisition was guilty of abuses and cruelties in the course of its long history, but it was no more unjust or inhumane than most other courts of the Europe of its day. The traditional exaggerations about it were derived from propaganda against Spain at the time of its greatness and from the works of Juan Antonio Llorente (1756–1823), 19th-century liberals and a number of historical novelists and dramatists.[1]

The Inquisition was abolished by Joseph Bonaparte in 1808, restored by Ferdinand VII in 1814, abolished and restored again (1820 and 1823) and finally abolished by Maria Cristina in 1834.

The Conquest of Granada.—After a hard-fought war of more than ten years (1481–92) the last of the Mohammedan kingdoms of the peninsula fell (Jan. 2, 1492) to the Catholic kings. The long process of reconquest had occupied the energies of the many provinces of Spain and thus made itself a unifying influence. It also created the Spanish military tradition which was to dominate the battlefields of Europe for the subsequent 150 years. At the conquest of Granada the Moors were granted by treaty the most liberal rights for the exercise of their religion. But the efforts of Christian missionaries, prompted by Francisco Jiménez (Ximénez) de Cisneros, the queen's confessor, stirred the Moors to a revolt (1501) which was put down after fierce fighting and was regarded as having invalidated their treaty. They were, consequently, ordered by a *pragmática* (Feb. 12, 1502) to leave the peninsula under conditions so difficult that they accepted the only feasible alternative, Christian baptism. Thus all the Moors of Granada became Moriscos and therefore subject, as Christians, to the Inquisition. But their Christianity became only nominal, since the overwhelming majority had neither the opportunity nor the desire to understand the Catholic faith. Attempts to instruct them failed for lack of money and for lack of priests and other missionaries able to speak to them in the Arabic language. They consequently remained an anti-Catholic and, therefore, anti-Spanish population in possession of the southeastern corner of the peninsula—of the region, that is to say, most liable to invasion from the African Moors or from the Turk.

The Expulsion of the Jews.—The tide of national enthusiasm, religious fanaticism and indignation at Jewish financial operations reached its high-water mark about three months after the fall of Granada, in a decree of March 31, 1492, which offered all the Jews in Spain the alternatives of Christian baptism or exile within three months. The emigrants were allowed to sell or take with them their goods, except that a law forbade the export of gold or silver (an exception which use of the credit system made less cruel than it would otherwise have been). The number of those who preferred exile to the renunciation of their religion has been credibly estimated at 165,000. They spread over Portugal, Italy, Greece, Turkey and North Africa, where they still speak the Spanish current in Castile at the time of their expulsion. By the middle of the 20th century their descendants (Sephardim), more than 2,000,000 in number, were spread widely over both hemispheres.

Spain in the New World.—The discovery of America (Oct. 12, 1492) was the result largely of the enterprise of Isabella, and upon her fell the responsibilities inseparable from her position as proprietress of half the world. She immediately laid down certain principles of policy that endured for centuries. The most important of these was the limitation of American trade to a single port—at first Cádiz, but from 1503 Seville—which maintained a virtual monopoly, in spite of its inaccessibility, till 1717. Secondly, the control of all American trade was exercised by the Casa de Contratación (House of Trade), which rapidly developed into a huge organization for the fulfilment of three important functions: as a ministry of commerce it carried out all the ordinances about the lading of ships, the registering of cargoes, smuggling and unlicensed emigration to America; as a school of navigation it formed itself into a kind of nautical university, one of the greatest centres of mathematical and scientific research in Europe (its books were translated into many languages and used by navigators throughout the world); and as a court of law it had an extensive jurisdiction over everything that happened on its ships and over most matters connected with American trade.

The exploration of America (known always to Spaniards as *las Indias*, "the Indies") was pushed on with great speed immediately after its discovery. Diego Velázquez conquered Venezuela (1499–1500) and Cuba (1511). The famous conquest of Mexico by Hernan Cortes took place between 1517 and 1524. During the opening decades of the 16th century settlements were founded in Paraguay and Colombia. In 1513 Vasco de Balboa reached the Pacific. The Philippines were discovered in 1521 and formally occupied (1565) during the reign of Philip II, after whom they were named. The expeditions of Pánfilo de Narváez and Ferdinando de Soto (1528–42) brought the Spaniards to Florida. In 1539 the Mississippi was discovered, and during the same year Ecuador, Bolivia and Chile were added to the possessions of the Spanish crown. Buenos Aires was founded in 1535. After nearly 20 years of war Francisco Pizarro won his victory over the Inca state of Peru (1543).

As the conquest went on, it followed the traditions of the reconquest of Spain from the Moors in the middle ages. As a consequence the natives were granted a certain amount of protection and self-government. But the grant of great estates (*encomiendas*) to Spaniards by the crown led to the use of native labour under conditions that often amounted to cruel exploitation, which the Spanish priest Bartolomé de las Casas exerted himself to the utmost to check. Among other measures designed to mitigate the hardships of the natives he introduced the importation (1512) of Negro slaves from Africa. The policy of Isabella, Charles V and

[1]Thus, for example, Llorente states that the tribunal of Seville alone, in 1481, tried 21,000 accused persons and burned 2,000 of them alive and as many in effigy. He obtained these fantastic figures by adding those pardoned under an edict of grace to those accused of heresy. Careful investigation points to the conclusion that the number of heretics burned during the whole lifetime of the Inquisition was not as large as that attributed by Llorente to the Seville tribunal in this one year. Llorente's *Histoire critique de l'inquisition d'Espagne*, 4 vol. (Paris, 1817-18) was widely used by anti-Catholic propagandists and translated into several languages. The author's life was such that his work may well be supposed to be extremely tendentious; he wrote to please those who happened to be in power. He used many documents taken from the archives of the Inquisition but carefully selected them to support the case he wished to advocate.

Philip II toward the natives was distinguished by its humanitarianism and enlightenment. Yet it must be admitted that the good intentions of Spanish kings rarely did much more than pave the colonial hell.

The energies of the first colonists were largely absorbed in one industry: the production of precious metals. Though all mines were regarded as royal property, Spaniards were allowed quite freely to discover and to operate them, provided that they went through certain formalities such as the registering of their claims before royal officials and the swearing of an oath to bring all the precious metals they obtained to be taxed and stamped at the royal offices. The royal share, at first a large one, was fixed in 1504 at one-fifth. This royal *quinto*, which was rarely varied in certain parts of America, amounted to 22,000,000 maravedis (about £28,000) in 1505 and had risen to more than double that amount by 1518. These modest sums were the earnest of that vast and dazzling wealth which was soon to lure Spain on to attempt world domination. The precious metals imported into Spain before 1530 were mostly gold; after that date they were from 85% to more than 99% silver. They rose in quantity steeply till the closing years of the 16th century, reaching a yearly average of about £4,000,000. In the earlier half of the 17th century they declined almost as steeply as they had arisen.

Joanna the Mad and Philip I.—On the death of Isabella (Nov. 29, 1504) her eldest surviving daughter Joanna (Juana; 1479–1555) succeeded to the throne of Castile, her father Ferdinand continuing to rule the states of the crown of Aragon till his death in 1516. Since Joanna's mental derangement (she is known as *la Loca*, "the Mad") had become apparent even before her accession, her Habsburg-Burgundian husband Philip the Handsome, son of the emperor Maximilian I and archduke of Austria, ruled in her stead as Philip I of Castile till his sudden death (Sept. 25, 1506). Thereafter Castile was governed by Jiménez (who became a cardinal and inquisitor general in 1507) and by Ferdinand, Joanna having been persuaded to retire to a life of seclusion at Tordesillas. On the death of Ferdinand (1516) the whole of Spain, which now included Ferdinand's recent acquisitions—Cerdaña (Cerdagne) and Roussillon (1493) and Spanish Navarre (invaded 1512; incorporated with Castile 1515)—came under the governance of Charles I, better known as the emperor Charles V. This eldest son of Joanna the Mad, who ruled in her name, inherited not only the Aragonese and Italian possessions of his grandfather Ferdinand but also the great possessions in the Netherlands and central Europe of his Habsburg grandfather Maximilian.

THE HOUSE OF AUSTRIA

Charles V (1516–56).—Charles arrived in Spain an unprepossessing youth (Sept. 19, 1517) surrounded by grasping Flemish favourites and unable to speak Spanish. By his reckless promotion of Flemings he offended the Spanish nobility; and by his demands for money he offended the common people, who were also alarmed at his obvious intention of ruling Spain as an absentee. The general discontent found expression in the revolt of the *comuneros*—a rising chiefly interesting for its constitutional not to say democratic character. Fortunately for Charles a rift soon showed itself between the common people and the nobles, whose support he was able to attract. As a consequence the royalists were able to win the battle of Villalar (April 23, 1521), a small fight that had great results. The *comunero* movement, which had once looked so formidable, collapsed at once, and the king of Castile emerged with fewer constitutional limitations than ever before.

Charles was compelled by the geography of his vast European dominions to be the enemy of France, a state which, though apparently weaker, had the advantage of fighting on interior lines while Charles's government was almost overwhelmed by such serious preoccupations as the Lutheran revolution in the empire and the powerful onslaughts of the Turk. Nonetheless, Charles's invincible Spanish soldiers won resounding victories, such as that at Pavia (1525); and the Spanish people, when they saw Francis I of France brought a prisoner to Madrid, were almost reconciled to their king's international outlook. What lay nearer to their hearts was Charles's successful expedition against Tunis (1535) and his attempt, though unsuccessful, to take Algiers (1541). Before his abdication in 1556 Charles had taught the Spaniards to regard his interests as their own and to pour out their wealth with almost reckless prodigality for his purposes. It has been well said of him that he was born a Fleming and grew up to be a Spaniard.

Philip II (1556–98).—Philip II, having inherited all his father Charles's dominions except the empire, which went to his uncle Ferdinand, was able to be a more truly Spanish king than Charles had ever been. The main purpose to which he devoted his life was the unification of the whole of the peninsula under his rule and the establishment of social justice among all his subjects. To say that he strove for the supremacy of the Catholic faith is only to say that he was a typical 16th-century statesman in his belief that political unity could not exist apart from unity in religion. In pursuance of these purposes he strengthened the Inquisition to enable it to stamp out the first sparks of Protestantism, which appeared in Valladolid and Seville (1559); he managed to get possession of the kingdom of Portugal (1580)—by far the greatest success of his reign, bringing indeed not only the whole Iberian peninsula under a single sceptre for the first time in nine centuries but also the second largest colonial empire in the world into union with the largest of them all, as well as securing to Spain the Atlantic harbours of Portugal and fleets manned by crews of unequalled experience; and he took advantage of a rebellion on behalf of his treacherous secretary, Antonio Pérez (*q.v.*), to reduce the liberties of Aragon (1592) so that that kingdom became hardly less monarchical than the other states of the peninsula.

Many modern writers repeated the views of L. von Ranke (*The Ottoman and the Spanish Empires in the Sixteenth and Seventeenth Centuries*, Eng. trans., London, 1843) that these changes converted Aragon into a virtual despotism. It is necessary, therefore, to point out that the changes made, though numerous, were comparatively small. Philip was too much a Spaniard to harbour the design of levelling out all local independence; he merely desired to give the central government enough power to do away with the most glaring social injustices. Among the chief acts of Philip II's *Cortes* of Tarazona (June 15, 1592) were the change of the office of *justicia* from a life appointment to one terminable at royal pleasure; and the increase in the royal share in the appointment of the *justicia's* lieutenants. In the *Cortes* itself the constitutional theory that all members of every estate (*brazo*) must be unanimous before a measure could become law—a theory that had long been circumvented in practice, since otherwise it would have required a miracle to pass any law whatsoever—was now definitely abrogated. The time that might be spent on statements of grievances (*grueges*), which had previously been theoretically infinite—to the great delay of public business—was now brought within tolerable limits. The right to seats in the *Cortes* was also considered. Henceforth only eight of the higher nobles were allowed to claim as a right their summons to the *Cortes*, and lesser nobles could only sit when the king chose to summon them; and in the same way certain towns were recognized as having the prescriptive right to send representatives, while the king was given the right, after the manner of the Tudors in England, to choose as many others as he thought fit. Such changes as these were obviously moderate and reasonable. They left Aragon with the bulk of its constitution and its ancient rights, liberties and *fueros* only slightly restricted in favour of the crown.

There is little doubt that the mass of the Aragonese benefited greatly from these changes in the customs of their country. They were often hideously oppressed by the aristocratic possessors of liberties and *fueros*. Subsequent experience would lead one to regret that Philip had not been more radical in his demand for their suppression. Thus, for example, Aragonese serfs were frequently put to death at the mere arbitrary orders of their masters; and when the crown, with characteristic zeal for social justice, inquired into these incidents the council of Aragon pointed out that there was nothing to be done, as the killing of peasants by nobles was quite in accordance with the much-prized *fueros* of the kingdom.

But Philip's responsibilities outside Spain were world-wide in

extent and virtually limitless in the expenditure that they demanded. Charles V's unhappy decision to attach the Netherlands to the Spanish monarchy involved his son in a rebellion that he was unable to quell after 30 years' continuous effort. The war with France, which had lasted from his father's reign, was ended at the peace of Cateau-Cambrésis (April 2, 1559), but the wars of religion gradually induced him to come to terms with his former enemies the Guises and to help them with men and money; and finally to attempt to win the crown of France for his daughter Isabella Clara Eugenia. In these plans for France he was foiled by the more able and less scrupulous Henry IV. With England Philip tried to continue under Elizabeth I the friendship that had existed under his wife Mary I Tudor. But the raids of English pirates upon Spanish trade with America and Elizabeth's persecution of Roman Catholic missionaries and her assistance to the rebels in the Netherlands caused strained relations between England and Spain and the consequent sending of the Armada (*q.v.*) of 1588.

Philip's position in the Mediterranean demanded ceaseless war against the Turks. But it was not till 1567 that, recognizing the grave danger present in the pro-Islamic sympathies of the lately converted populations of the south, he issued the severe *pragmática* that forbade all Moorish or Mohammedan practices under pain of the heaviest penalties. The 3½ years of fierce rebellion that followed resulted in a fearful outpouring of blood and treasure and in the transportation of the entire Morisco population of Granada to northern Spain, where the Moriscos were forced to live only in houses sandwiched between those of Old Christians. Then Philip's bastard half-brother Don John of Austria defeated the Turks at Lepanto (Oct. 7, 1571).

Lepanto has been condemned as a barren victory, yet it was regarded throughout Europe as the finest feat of arms of the century. Miguel de Cervantes says of one of the spurious continuators of *Don Quixote* "what I cannot forbear resenting is that he upbraids me with being maimed . . . as if my wound had been got in some drunken quarrel in a tavern and not on the noblest occasion that past or present ages have seen or future ones may hope to see" (introduction to the second part of *Don Quixote*).

The expenses of Spain's great enterprises were, however, beyond anything that even the vast silver imports from America could defray; and the industries of Castile were by now being taxed out of existence. At the same time the great imports of silver from America were causing Spanish prices to rise so steeply as to destroy the export market. Hence the economic decline of Spain was far advanced before the end of Philip's reign, during which there were three virtual declarations of bankruptcy upon debts to foreign bankers: one at the beginning, one in 1576 and one in the last year (1598). Yet in Spanish historical tradition the reign of Philip II passes for an era of great successes and glory. The enemies of Spain—notably Elizabeth I of England—quailed before this strength. The growth, moreover, of literature, art and science in many forms under Philip II gave Spain for a time the cultural and intellectual hegemony of the world. This growth owed at least a little to Philip's patronage and interest and may fairly be regarded as a compensation for his failures.

Philip III (1598–1621).—In territorial extent Philip III's empire was the greatest in recorded history. Philip, born April 14, 1578, was the fourth son of Philip II's fourth marriage. He departed from the traditions of his father by employing a chief minister to carry on the government and by giving much of his confidence to members of the high nobility. The chief minister, Francisco Gómez de Sandoval y Rojas (1552–1623), duke of Lerma, was loudly and probably not without some measure of justice accused by his enemies of piling up a vast fortune for himself and his own favourites at the expense of his overtaxed country. Lerma not only found that Philip II had left an enormous debt and almost all the revenue pledged in advance to moneylenders but also had two major wars upon his hands. He was therefore driven to the last resort of destitute statesmanship, the debasement of the coinage, a measure that drove the silver currency out of the country. In one respect Lerma deserves credit; by his able diplomacy he secured peace with England (1604), the 12 years' truce with the Netherlands (April 9, 1609) and the friendship of the French government after the assassination of Henry IV (1610). An act of more doubtful statesmanship was the expulsion of the Moriscos (1609–14). The Moriscos, however, though treated not without consideration after the resettlement under Philip II, had remained Mohammedan at heart; and it is probable that their hostility to the state was so deep-rooted and their relations with foreign powers so dangerous as to leave no feasible alternative except expulsion. The fall of Lerma (Oct. 4, 1618) was an event of no advantage to Spain, for his son and successor, Cristóbal Gómez de Sandoval y Rojas, duke of Uceda, took the fatal step, through Ambrose Spinola's campaign in the Palatinate, of involving his country in the Thirty Years' War, a step which destroyed all possibility of economic recovery.

Philip IV (1621–65).—As a patron of dramatic art and painting Philip IV was one of the most successful kings in recorded history. The plays and pictures that owe their existence largely to his encouragement are among the world's most precious possessions. Yet from the military and economic point of view his reign is to be regarded as one of the most disastrous in Spanish history.

Having come to the throne at the age of 16, Philip appointed as his chief minister Gaspar de Guzman (b. 1587), count of Olivares (*q.v.*) and first duke of San Lucar, known as the count-duke (*conde duque*). An energetic, patriotic and truculent-looking man, Olivares set to work to restore Spain to its former powerful position in the world by curbing extravagant living and idleness and by changing the country from a bundle of independent states into one centralized Castilian monarchy, a plan not uninfluenced by the example of Richelieu's work in France. "You should not be content," he advised Philip IV, "to be king of Portugal, of Aragon and of Valencia and count of Barcelona, but you should direct all your work and thought . . . to reduce these realms to the same order and legal system as Castile. If your majesty succeeds in this you will be the most powerful prince in the world." Such an aim was unfortunate in coming at a time when Castile, the centre of gravity in the Spanish monarchy, was growing weaker in comparison with the other states, which had centrifugal tendencies. When, therefore, Olivares pushed his centralizing policy on with his habitual tactlessness and violence he was to be met with a crop of revolts of which the French were not slow to take advantage.

The early years of Spain's participation in the Thirty Years' War were marked by military successes that recalled the great days of Spanish victories. Don Fadrique de Toledo destroyed the Dutch fleet off Gibraltar; and more important still was Spinola's capture of Breda after a ten months' siege (1624–25)—an achievement made immortal by the painting of Velazquez. The Dutch were overcome in South America and expelled from Guayaquil and Puerto Rico; and at the same time Moorish pirates were almost swept from the Mediterranean. The fever of glory and conquest had laid hold of Spain, and the name of the king became "Philip the Great"; but the demand for money followed in the footsteps of war. Olivares felt that the eastern states of the peninsula, Aragon, Catalonia and Valencia, used their three close-fisted *Cortes* to evade their fair shares of the burdens of the empire while loyal Castile bore all, including the greater part of the fighting. He therefore went out of his way to show them the scant respect that he considered their due.

The three *Cortes* were convened for Jan. 1626; and the king promised to be present at each and to take the oath as sovereign to respect their rights. The Valencians, however, resented the summoning of their *Cortes* at Monzón, a place outside the kingdom of Valencia, and were ready to take offense before the meeting. When the king asked them at Monzón to grant him 2,000 paid troops to be used as he pleased, as well as a money vote, they refused. It would, they said, be the same as the conscription in Castile, and that was against their liberties. Olivares alternately threatened and bribed and the king himself tried persuasion. At last he formally said that they must vote the supplies without more ado. But the noble estate of the *Cortes* continued to hold out against stronger pressure.

At last the king called them traitors and gave them one day in

which to do his bidding. It was necessary that the vote should be unanimous, and under this pressure all the members gave way except one—Don Francisco Millan. When, however, a hint was given that the garrotte would probably be the recompense for his obstinacy, he acquiesced in the vote. Even so the nobles still cavilled about the amount of their quota and only after much deliberation settled that the vote of money should be reduced to 1,080,000 livres in 15 years, half the sum paid by Aragon, not in coin, but in kind. The king was forced to accept, but the *Cortes* was discontented and would not formally ratify the vote. At length the king sent it a message that he gave it half an hour in which to vote the supply, as he was leaving at once. He then took out his watch and said that the half-hour had already begun. The *Cortes*, thrown into consternation by such a course, deliberated all night. But at last, at 6 A.M., the king sent a message to say that he had determined to abolish the privilege of *némine discrepante* (unanimity). The astonished and indignant Valencian *Cortes* then allowed all its resistance to collapse. The king met the Catalan *Cortes* at Lérida and sought to dazzle it by feasts and splendour, but the Catalans insisted on the repayment of former loans by the king before they would grant a ducat. As they would not yield to threats, Olivares, fearing mischief, suddenly and without warning carried the king back to Castile. The Catalans thereupon voted all that was required of them in the hope of getting the king to come back. The reply was a threat of vengeance for their undutifulness.

In 1639, however, when Richelieu sent French troops into Spanish Roussillon, the Catalans fought loyally and successfully for their province against the invaders. But the Castilians sent to reinforce them behaved so atrociously that in June 1640 the Catalans broke into revolt. Catalonia remained for 16 years a province of France rather than of Spain. It was to return to its allegiance to Philip IV out of dislike of the French rather than as a result of any prowess of Spanish arms. In December of the same year (1640) a well-planned revolution in Lisbon led in turn to the independence of Portugal, which Spain was unable to recover in spite of an exhausting struggle of nearly 28 years (1640–68). As a consequence of these revolts Olivares was overthrown by a court intrigue (Jan. 1643) and succeeded as chief minister by his sister's son Don Luis de Haro.

The earlier actions of the Thirty Years' War had shown that Spanish soldiers had not lost their old quality; but the French victory at Rocroi (May 1643) was not only a decisive defeat of Spanish arms but also the final invalidation of those Spanish tactics that had dominated the battlefields of Europe for nearly 150 years. Though the peace of Westphalia (1648) at last ended the Thirty Years' War, it brought no peace to Spain, with which France remained at war. It was only the Fronde in France that enabled Philip IV to struggle on till the peace of the Pyrenees in 1659, when Cerdaña and Roussillon were ceded to Louis XIV, who next year married his cousin Marie Thérèse (Maria Teresa), Philip's daughter by his first wife, Elizabeth of France. By this time the expense of wars combined with the misgovernment of a corrupt and far too numerous civil service had brought Spain to that lowest point of economic decadence toward which it had been moving with increasing speed for nearly a century.

Charles II (1665–1700).—The fruit of Philip IV's consanguineous second marriage (1649)—with his sister Maria Anna's daughter Mariana (*sc.* Maria Anna also), whose father was yet another Habsburg, the emperor Ferdinand III—the last king of the house of Austria, was so infirm in body and in mind that he was unable to take much significant part in the government of his dominions. Since Charles II succeeded his father in his fourth year his mother ruled in his name as regent, strongly opposed by Philip IV's popular and ambitious bastard son, Don John of Austria. Mariana used as her chief ministers first the German Jesuit Johann Eberhard Nithard and then the Granadine adventurer Fernando de Valenzuela. But these were one after the other driven from power by Don John, with the help of his faction of the grandees.

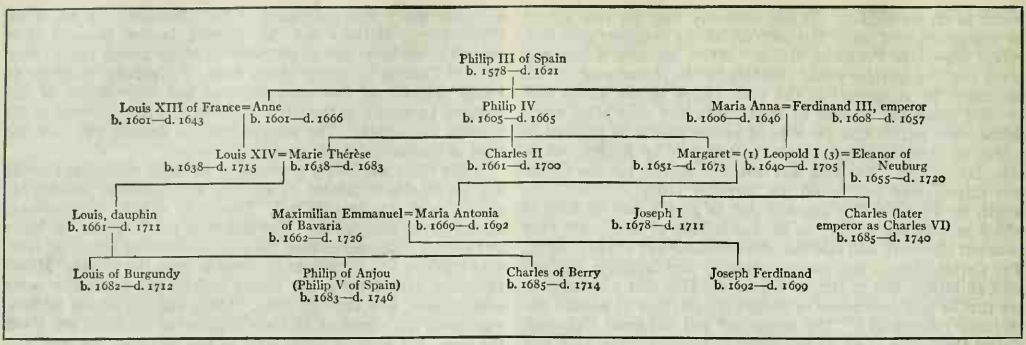
Finally Don John became strong enough to force Mariana into a convent and to persuade the king to marry Marie Louise of Orléans (Aug. 1679) in the interests of France. But Don John in power was a disappointment to his adherents. On his death (Sept. 1679) Mariana and the Austrian faction resumed their control of the king; and on the death of Marie Louise (1689) they induced Charles to marry Maria Anna of Neuburg, daughter of Philip William of Palatinate-Neuburg and sister-in-law of the emperor Leopold I, in the interests of the Austrian claimant to the Spanish succession. The passive Spanish domain was now the bone of contention of the powers of Europe.

While Spain was barely governed at all under the infant Charles and his capricious mother its financial and economic position became almost unbelievably bad. Crowds of reckless civil servants spread havoc everywhere. Corruption of public officers at home and abroad continued to keep the national treasury bankrupt, however crushing was the weight of taxation upon the people. Manufacturing, except for a few coarse articles for immediate home consumption, was virtually dead. Idlers and discharged soldiers, vagabonds and rogues of all sorts swarmed in the cities and about the court, but working hands were so few that even the poor harvests still raised in northern and central Spain were gathered in by French labourers who returned home again with their wages in their pockets. The trade and industry in Madrid, such as they were, were carried on almost exclusively by Frenchmen of whom there were 40,000 in the capital, mostly calling themselves Flemings and Burgundians in order to escape the differential taxation imposed upon Frenchmen. The total revenues of the crown of Castile on the accession of Charles II are given by a contemporary writer, Alonso Núñez de Castro, in *Solo Madrid es Corte* (Madrid, 1669), as amounting to 15,750,000 ducats derived from a great variety of sources. Of all this sum, 8,500,000 ducats are said to be pledged; in any case not more than a third would ever reach the treasury. The average amount received from Aragon, Valencia, etc., appears to have been 2,000,000 ducats, and the average annual revenue from the Indies, derived from a royalty of one-fifth of the precious metals mined and other tributes, is set down in the anonymous *Relation de l'estat et gouvernement* [*d'Espagne*] (Cologne, 1667) at 1,500,000 ducats. It is probably not far wrong to estimate the national income—apart from the royal patrimony, which was now mostly pledged or alienated—at about 9,000,000 ducats (of debased copper) actually encashed. The cost of the royal establishments reached the sum of 1,700,000 ducats a year—about one-fifth of the whole amount received into the national exchequer. The ecclesiastical revenues of the kingdoms of Castile alone amounted to 12,000,000 ducats and the entire revenues of the same kingdoms, clerical and lay, are stated at 113,000,000 ducats. If this really was the case, taxation would not seem to have been so onerous as the complaints would lead us to believe; but since much of the land was now out of cultivation and the country depopulated, it is certain that the income was far inferior to that stated as the nominal value. It will be readily understood from these statements that Spain was in no position to undertake an adventurous foreign policy.

But the restless ambitions of the French king Louis XIV left Charles's dominions small chance of peace, whether in Flanders or in the Mediterranean. Louis fought a series of wars (1667–68, 1672–78, 1683–84; *see* FRANCE: *History*) with the object of annexing portions of them. Complications with the English and with the Dutch led, however, to the frustration of Louis's plans and to the attempts by William III of England to deprive France of the whole of the Spanish inheritance by means of partition treaties, in which the Spanish people were not consulted. Charles II, anxious above all to preserve the unity of the Spanish empire, made his will leaving the whole vast inheritance to Louis XIV's grandson, Philip of Anjou. Having signed it he died on Nov. 1, 1700, in the 40th year of his age and his 80th to judge from his appearance. The last of the house of Austria, he left Spain exhausted and bankrupt. (R. T. D.)

THE SPANISH SUCCESSION

The military and naval history of the War of the Spanish Succession is outlined in the article SPANISH SUCCESSION, WAR OF THE. It may be useful here to examine how the conflict arose. The accompanying genealogy shows the descent of the principal claim-



ants of the succession.

Philip III's two daughters, Anne and Maria Anna, had married, in 1615 and in 1631, Louis XIII of France and the future emperor Ferdinand III respectively; and two sons of these marriages, Louis XIV and Leopold I, married their Spanish cousins, Philip IV's daughters Marie Thérèse and her half-sister Margaret, in 1660 and in 1666 respectively. Marie Thérèse explicitly renounced her claim to the Spanish succession on her marriage to the French king, as her aunt Anne had done; but Margaret made no such renunciation and was moreover named in Philip IV's will as the next heir after his son Charles. Consequently, when Maria Antonia, daughter of Margaret and wife, from 1685, of the Bavarian elector Maximilian Emmanuel, gave birth in 1692 to a son, the electoral prince Joseph Ferdinand, this prince could be regarded as heir presumptive to his great-uncle Charles II of Spain. Leopold I, however, had persuaded his daughter Maria Antonia to bestow her right to her mother's succession on him and on the sons of his third marriage; but the validity of this bestowal, on which the immediate Habsburg claims to the succession were based, was dubious. Also dubious, at the same time, was the Bourbon claim, based on disregard of the French queens' acts of renunciation.

In the precarious lifetime of Charles II, French influences within Spain itself were heavily counterbalanced in the field of international politics by the attitude of the maritime powers. England and the United Provinces, now in personal union under King William III, were strongly opposed to the Bourbon claim, fearing not only the aggrandisement of French power that would result from its success but also the loss of their well-established trade with Spain and the Indies. In Oct. 1698, therefore, Louis XIV and William III agreed, in the first partition treaty (signed at The Hague), to recognize Joseph Ferdinand's rights to Spain, the Netherlands and the Indies. Milan, however, was to go to the archduke Charles, the emperor Leopold's younger surviving son (who would presumably be excluded from the imperial throne by his elder brother); and the rest of Spanish Italy, as well as the Basque province of Guipúzcoa, was to go to the dauphin Louis. Spain was indignant at this partition, and in Nov. 1698 Charles II retorted to it in a will that named Joseph Ferdinand as heir to the whole succession, thereby giving the Bavarian claim a final authority independent of the partition treaty (which moreover the emperor had not approved). But in Feb. 1699 the child Joseph Ferdinand died, and new negotiations were urgently required. The second partition treaty, signed by France, England and the United Provinces in March 1700, offered Spain, the Netherlands and the Indies to the archduke Charles (the dauphin was to get Spanish Italy except Milan, instead of which he would have Lorraine); but the emperor declined to accede to it, hoping to win the whole succession for the archduke in reward of refusing to divide it. To counter this new partition, Charles II made his last will, leaving the Spanish dominions, which were to be kept separate from any other crown and undivided, to Philip, duke of Anjou, second son of the dauphin Louis; failing Anjou, to his younger brother Charles of Berry; failing Berry, to the archduke Charles.

Charles II died on Nov. 1, 1700, and Louis XIV was now confronted with a dilemma. If he accepted the whole succession for his grandson, he would be violating the partition treaty; if he stood by the treaty, the will would prescribe exactly what the treaty had been intended by France to prevent—the passing of the whole succession to the Habsburgs. He chose the former alternative, acknowledging Philip of Anjou as Philip V of Spain. The War of the Spanish Succession ensued. (X.)

THE BOURBON DYNASTY

Under the treaty of Utrecht (*q.v.*), which ended the War of the Spanish Succession in 1713, Philip of Anjou was recognized as king of Spain but lost Gibraltar and Minorca to the English, the Spanish dominions in Italy and Flanders to the Austrians and Sicily to Savoy. To this extent the hopes that had led Charles II to appoint Philip his heir were falsified, though there were compensations. After two centuries Spain in Europe was at last driven back upon itself and given a stimulus, long sorely needed, to put its own house in order. The change in dynasty brought too a radical revision of the whole intellectual background, French taste and French ideas gradually percolating the upper classes of society and changing the course of political thinking and of literature. The Bourbon theory of state was no less absolutist than that of the Habsburgs had been, but it rested on a much more enlightened conception of responsibility and worked efficiently (in the early reigns) through ministers, not corruptly through favourites drawn from an effete nobility.

Philip V (1700–46).—In internal affairs the years of the war were of capital importance in Spanish history in spite of the confusion of a country bitterly divided in its loyalties and overrun by foreign armies, with its capital repeatedly changing hands and the constant levies on manpower and economic resources a running score. The general political and administrative nullity of the Spanish upper class of this generation and the political views of the French monarchy led to the assumption of all real power by the French and, later, the Italian servants and advisers of the king. French economists, Jean Orry and Michel Amelot, marquis de Gournay, became ministers of finance, and Marie Anne de la Trémoille, princesse des Ursins, sent by Louis XIV to counsel the child-queen Maria Louisa of Savoy, was before long the moving spirit behind the throne.

Under such direction important financial and administrative reforms were begun, notably in the collection of the national revenue, which Orry had found in 1701 to amount to little more than one-half of normal expenditure. The *Cortes* of Castile, in which those of Aragon and Valencia (1709) and of Catalonia (1724) were successively merged, was summoned four times only during Philip's long reign, to hear and approve royal decisions. The voting of subsidies, the original basis of their power, continued to be vested in small permanent commissions. In central government, under the chief organ, the *consejo real*, more direct responsibility attached to the secretaries of the subordinate councils (state, grace and justice, war, navy and the Indies, finance, was the normal distribution). From 1714 these were called ministers, foreshadowing the modern cabinet system born in 1787 with their de-

liberation in common as a council of state.

The absolutist and centralizing nature of some of these innovations exacerbated the separatist tendencies of the northeastern regions, which had embraced the Austrian cause against the Bourbon dynasty. Philip was forced to reduce Aragon, Catalonia and Valencia by arms. The Catalans had relied on assurances from England and from the emperor that their *fueros* would be secured to them in the general peace. Deceived in this, they fought on desperately on their own, and Barcelona was only taken in 1714, the year after the signing of the treaty of Utrecht, when the city had been reduced to a shambles. Most of the privileges of these kingdoms now disappeared: those of Aragon and Valencia in 1707; those of Catalonia (exemption from conscription, the right to a separate coinage and some part of its penal law surviving) in 1716. The University of Barcelona was removed to Cervera, and the use of Catalan was forbidden in courts of justice. The Basque provinces, which had sided with Philip, retained their *fueros*. With this exception the writ of Castile now ran throughout Spain as never before; and Philip, dividing the country—formerly a composite of traditional "kingdoms"—into provinces each under a captain general, set about imposing a new unity of administration. Municipal councils, office in which had become purchasable and hereditary, were again made elective and required to submit their accounts annually to and to deposit all surplus moneys with the crown. The promulgation in 1713 of the Salic law, intended to ensure that Spain should never again fall to the Habsburgs through marriage, was to play an unexpected part in the outbreak of the First Carlist War more than a century later.

Maria Louisa of Savoy died in 1714, leaving three sons, Louis, Philip (who died young) and Ferdinand; and Philip V was straightway married to Elizabeth Farnese (*q.v.*), the duke of Parma's daughter, who ruled him thereafter and, exploiting his desire to follow a policy independent of France, dragged him into a series of adventures aimed at securing Italian dominions for her own offspring. The princesse des Ursins she repaid for her part in the match with instant dismissal. Giulio Alberoni, an Italian priest, became head of the government, and for the remainder of the reign the queen's Italian ambitions menaced the peace of Europe and Spain's prospects of recovery. France and England threatened war in defense of the terms of Utrecht. Alberoni, spinning a web of intrigue in half the courts of Europe, was undaunted. He occupied Sardinia in Nov. 1717 and landed 30,000 troops in Sicily in the following July. England acted without declaration of war and destroyed the rising Spanish navy off Cape Passsero in August. General hostilities followed. Spain tried to send an expedition to aid the Jacobites in Scotland (1719), a French army overran the Basque provinces, English warships wrought havoc in Galicia, the emperor regained Sicily. Alberoni, made the scapegoat, was banished in Dec. 1719. By the peace of 1720 the emperor formally renounced his claim to Spain; Philip renounced his to the Austrian territories in Italy (including Sicily); and the succession of Elizabeth Farnese's issue to Tuscany and Parma was acknowledged.

Philip, increasingly a prey to religious melancholia, began in 1721 the construction of the vast La Granja palace, at San Ildefonso near Segovia, planned to remind him of Versailles; and there he withdrew almost entirely from the conduct of affairs. In Jan. 1724 he abdicated in favour of his eldest son Louis (Luis) I, whose death in August caused him reluctantly to resume the crown. An estrangement with France over the succession to the sickly Louis XV brought a momentary *rapprochement* with the emperor Charles VI, but this cooled even before the seizure by imperial troops of the duchy of Parma on the death of the duke, Antonio Farnese, in Jan. 1731. English diplomacy effected the cession of the duchy to Elizabeth Farnese's eldest son Charles. Spain had made great sacrifices to achieve so much of her ambition, but there was more to come. In 1733, the emperor being occupied by the War of the Polish Succession, Spain signed the first "family compact" with France, declared war on Austria and overran Naples and Sicily, of which territories Charles was crowned king in 1733, relinquishing Parma and Tuscany to the emperor.

Gibraltar was a continuing bone of contention with England. Another, even more acute, was English contraband trade with the new world, greatly facilitated by the trading privileges won at Utrecht. War broke out in 1739, and Portobelo on the Caribbean was sacked by the English. The conflict merged soon after in the War of the Austrian Succession, in which Philip appeared in the strange role of claimant to the empire. The second "family compact" (1743) was the prelude to much fighting in Italy, Elizabeth Farnese being set now on winning territories for her second son, Philip. On July 9, 1746, Philip V died, almost insane. For more than 40 of the 46 years of his reign Spain had been at war. His government had been largely French, then largely Italian, with power very rarely in his hands and Spanish interests very rarely in the forefront. More positively, Spain had begun at last, with José de Patiño and José del Campillo, to produce statesmen of its own, who needed only peace to show the nation's powers of recuperation.

Ferdinand VI (1746–59).—Ferdinand, Philip V's third and only surviving son by his first wife, was modest and retiring and much under the influence of his Portuguese queen, Barbara of Bragança, who like him was wedded to a policy of peace. This was the great novelty of the reign. At the end of the War of the Austrian Succession, the treaty of Aix-la-Chapelle (Oct. 18, 1748; Spanish accession Oct. 20) confirmed England in the commercial privileges that it enjoyed in the new world (the South Sea company was two years later, under the treaty of Madrid, Oct. 5, 1750, to receive £100,000 in compensation for the annulment of the *asiento* treaty) and assigned the duchies of Parma, Piacenza and Guastalla to the infante Philip (with reversion to their previous possessors should he succeed his brother Charles in the Two Sicilies or in Spain). Ferdinand henceforth steadfastly rejected the competitive blandishments of France and England alike, unmoved even by offers of Gibraltar or Minorca. Cenón de Somodevilla, marqués de la Ensenada (*q.v.*), and José de Carvajal, his two great ministers, though inclining respectively to France and to England, were Spaniards first and made of the national recovery their major concern. Ensenada, following in Patiño's footsteps, did much to revive agriculture, to stimulate commerce and mining and to strengthen the navy. Fiscal reform and improved communications were other factors in the restoration of a bankrupt country to comparative prosperity. Between the beginning and the middle of the century the population of Spain had risen from 5,700,000 to 7,500,000. In the general spread of enlightenment, exemplified since the change of dynasty in the proliferation of academies and other learned bodies, two details were significant of the new national temper. One was Ferdinand's attracting of foreign scholars and craftsmen to Spain; the other the profound modification in the character of the Inquisition. Under Philip this had continued active, holding 782 *autos-da-fé* and imposing 14,000 sentences during the reign. Under Ferdinand its vigour and rigours were much relaxed and it was no longer an instrument of state policy.

Foreign policy was marked by a series of treaties all aimed at the peaceful settlement of possible motives of discord: frontier issues with Portugal and with England in America (both 1750); the neutrality of Italy, under an agreement with Austria, Tuscany and Parma, which Naples, still reflecting the ambitions of Elizabeth Farnese, refused to sign (1752); rights of ecclesiastical patronage in Spain, settled in Spain's favour by a concordat (1753). Ferdinand's death, childless, in 1759 brought to an untimely end the country's first substantial respite from war for more than a century. The heir to the crown, Ferdinand's half-brother Charles, had already, from his kingdom of Naples, which he now resigned, proved a disturbing influence in Ferdinand's policy of peace and welcomed the wider sphere for action.

Charles III (1759–88).—Charles III was one of the most sincere and one of the most successful of the "enlightened despots" of the 18th century; for Spain he was the last. He came to Spain, aged 43, after a long apprenticeship to kingship and laboured unceasingly to advance the country's material prosperity. Widowed within a few months of his accession, he acted for himself; but his minister of finance was an Italian, Leopoldo de Gregorio, marqués de Squillace, his ideas were those of the French encyclo-

paedists and his attempt to reform the Spanish way of life from the bottom up, as in dress, led to an initial period of friction culminating in the Esquilache revolt of 1766, which caused Charles to fly in alarm from his capital and put an end to the long succession of foreigners at the helm of state.

The count of Aranda (*q.v.*), who now emerged, with the count of Florida Blanca (*q.v.*), as the statesman of the reign, had likewise imbibed advanced ideas, particularly as regards the church, in long residence abroad. In April 1767 Charles decreed the expulsion of the Jesuits, as a body not owning allegiance to the state, from Spain and the American colonies and exerted his whole influence to secure the suppression of their society. One consequence was the nationalizing and secularizing of education. The Holy Office, firmly denied the right of trial in civil offenses, was finally shorn of its importance. In 1780 it burned its last victim, accused of witchcraft, in Seville. A decree of 1773 proclaiming the dignity of labour, even for the nobility, illumined another facet of the national resurgence. Land distribution to peasants, the colonization of deserted areas (6,000 Bavarians were settled in the Sierra Morena), the spread of new industries under foreign instruction, the creation of a network of canals and highways throughout the country, were among the many measures which, combined with continuous administrative reform, improved the lot of the working classes and stimulated all in the common effort. The new national lottery proved a valuable and popular source of revenue. By 1788 the population had reached 10,250,000.

Charles's foreign policy was rooted in a deep distrust of England, the "enemy" both in America and in respect of Gibraltar and Minorca. Hence the third "family compact" (1761–62) and, since England and France were then at grips in the Seven Years' War, the former's declaration of war on Spain in Jan. 1762. The peace of 1763 cost Spain Florida, all its territories east of the Mississippi and more valuable trading concessions. In 1770 the conflict threatened a renewal over Spain's occupation of the Falkland Islands, from which the defection of France caused it to withdraw. England's dispute with its North American colonies confronted Charles with his most acute dilemma. To aid the rebels was tempting; the implications for his own American colonies of such countenancing of revolt only too obvious. In 1779 he opted to aid, following France's example of two years earlier; and in the general settlement of 1783 he recovered Florida and Minorca. Gibraltar, twice besieged, proved impregnable. Substantial administrative advances in the new world too (among the most beneficial was the institution of imperial free trade) were hampered by the overriding demands of the treasury at home. The years 1780–81 in particular were marked by serious uprisings both of Indians and of Creoles. In the Mediterranean an unsuccessful attempt was made in 1775 to seize Algiers, a notorious centre of piracy. The suppression of corsair raids on the Spanish coast was at length achieved by treaty with the regency of Algiers in 1786.

Charles IV (1788–1808).—The fourth Bourbon was at the opposite pole in character and gifts from his father. Sluggish and stupid to the verge of imbecility and much influenced by his wife, Maria Louisa of Parma, a coarse, passionate and narrow-minded woman, he soon became the plaything of events. The French Revolution not only meant the end of the "family compact" on which Spain had relied in its policy of opposition to Great Britain, it involved Spain in a crusade on behalf of Louis XVI and, after his execution in Jan. 1793, in enthusiastic support for the first coalition against France. By then Florida Blanca and Aranda had been replaced by a young guardsman, Manuel de Godoy, who was at once the queen's lover and the favourite of the king. The campaign, undertaken mainly in the Roussillon, was a failure, and French troops advanced into Spain almost to the Ebro. Constrained to make peace by the treaty of Basle (1795)—whence Godoy's title of "prince of the peace"—Spain found itself tied hand and foot to the French republic. The treaty of San Ildefonso (1796) was a virtual renewal of the "family compact," but with terms far more disadvantageous to Spain. A scheme for a joint attack on the English coast was foiled by the battle of St. Vincent, in which John Jervis and Horatio Nelson forced the Spanish fleet to retire to Cádiz. Great Britain further seized Trinidad and set itself to encourage discontent in the Spanish colonies.

The French occupation of Rome and the overthrow of the papal government in favour of a republic were offensive to Spanish opinion, as the dispossession of Ferdinand, duke of Parma, of his territories was to the crown, and Godoy fell from office in March 1798. But Spain was not free to pursue an independent policy, and the return of Bonaparte from Egypt and his emergence as first consul began a period of ever-deepening humiliation. Turning his attention to the peninsula, Napoleon fastened on Godoy as his tool, and a new treaty of San Ildefonso (Oct. 1, 1800), in which Spain ceded Louisiana and six warships against promises concerning the duke of Parma, was followed by Godoy's return to power. In 1801 he led in person an attack on Portugal, at France's behest. The invasion was barely resisted, and Portugal undertook to close its ports to England in return for respect for its territorial integrity.

Napoleon, soon involved in a new war with England, extorted from Spain a new treaty (Oct. 9, 1803) more burdensome still than that of 1796, Spain undertaking to pay a monthly subsidy of 6,000,000 fr. and to enforce Portuguese neutrality. Inevitably Spain too was drawn into the conflict with England: the last remnants of its maritime power were shattered at Cape Finisterre and Trafalgar, and an English force seized Buenos Aires.

Spanish subservience to France was increased and the fate of the Spanish crown deeply compromised by a rivalry for power at court between Godoy and the heir to the throne, Prince Ferdinand, whom Godoy discovered to be intriguing secretly with Napoleon for support. By the convention of Fontainebleau (Oct. 27, 1807) Spain had agreed with Napoleon on plans for the conquest and dismemberment of Portugal, thus giving Napoleon an excuse to send a French army under Andache Junot through Spain to Lisbon, whereupon the Portuguese royal family fled to Brazil. When, shortly after, Godoy induced Charles to order the prince's arrest on charges of plotting to dethrone his father and to murder his mother and Godoy, the resultant scandal and confusion enabled Napoleon to pour more and more troops into Spain, under pretext of reinforcing Junot. Soon even this excuse was dropped, and by Feb. 1808 the French invasion stood clearly revealed as one of conquest.

Godoy now would have fought but, the king refusing, could only counsel instead that the Spanish royal family too should embark for America. It removed to Aranjuez, where a mob instigated by the prince's party provoked a riot and clamoured for Godoy's head. The rioters were only appeased by Charles's announcement (March 17) of his abdication in favour of Ferdinand. Six days later Joachim Murat occupied Madrid. Napoleon meanwhile had advanced to Bayonne, near the frontier, whither at his orders Murat now dispatched Charles, his wife and Godoy. Napoleon had already decided to place one of his brothers on the Spanish throne and by specious promises easily lured Ferdinand likewise to Bayonne. There, confronted by his parents and Godoy in a scene of pitiful degradation, he was forced to restore the throne to his father, who laid it at the feet of Napoleon. The next six years Ferdinand spent under military guard at Valençay; his parents retired to Rome. On May 13 Murat announced to an improvised "junta of regency" at Madrid that Napoleon desired them to accept Joseph Bonaparte as their king.

The War of Independence (1808–14).—Spanish sentiments of loyalty and independence were, however, too profound to be daunted. For the first time Napoleon found himself faced, not by selfish and terrified rulers, but by an infuriated people who rose against the French on May 2, 1808. The rising in Madrid began the movement which ultimately proved fatal to Napoleon's power. At first he treated the novel phenomenon with contempt; but, although Madrid was held by the French, the Spaniards showed great capacity for guerrilla warfare in the provinces. The French were repulsed from Valencia and Pierre Antoine Dupont, who had advanced into the heart of Andalusia, was compelled to retreat and ultimately to capitulate with all his army at Bailén (July 10). The Spaniards now advanced upon the capital, which Joseph Bonaparte had just entered, and expelled him.

To some extent this too was a civil war, for a small but not

unimportant section of Spaniards, who had been much influenced by the doctrines of the French Revolution, continued to see in France the bulwark of liberal principles and looked to Joseph and his constitution of Bayonne as marking the dawn of responsible government in Spain. The struggle for independence, though in itself inevitably productive of new currents of thought in a progressive minority, was to seem in retrospect to have been a struggle no less against the spirit of the age. But the combined defection of the crown, of its senior statesmen, of a large part of the higher nobility and of the intellectual *Afrancesados* made the reaction of the Spanish people, thus left leaderless, the more remarkable. Its first concern was to improvise a provisional organization based on earlier provincial and regional juntas; and *bourgeoisie* and provincial yeomanry together succeeded in creating a system capable of resisting French domination.

The military superiority of the French ruled out hopes of an immediate victory, but Napoleon himself was compelled to come to Spain to direct operations. With this the tide swept once more against the patriots. Napoleon restored his brother in Madrid. Sir John Moore, advancing from Portugal with an English army to the relief of the capital, was forced to retreat; and only after a desperate battle, in which he fell, could his troops re-embark at Corunna. Napoleon, thinking his task completed, had now withdrawn, leaving Marshals Nicolas de Dieu Soult and Claude Victor-Perrin to complete the reduction of the provinces. The two defenses of Saragossa under siege (1808 and 1809) and that of Gerona (1809) are among the most heroic pages in the annals of Spain, but proved alike unavailing. With the fall of Seville in Jan. 1810 the central junta fled to Cádiz. Only the obstinate resistance of the duke of Wellington in Portugal, the continuous activity of the *guerrilleros* and dissensions among the French saved the peninsula from final submission.

The Spanish *Cortes*, if never so entirely suspended as the states-general of France, had become in the course of the 18th century little more than a rubber stamp. Already under Philip V those of the other regions had been merged in the *Cortes* of Castile. Now, under the influence of the ideas of revolutionary France, the reformist patriots of Cádiz planned their re-establishment on a more democratic though still national basis, and in 1810 a new *Cortes* was convened in that city. It included for the first time delegates of the colonies, chosen perforce, like those of enemy-occupied Spain, from among those of their citizens who happened to be then present in Cádiz. The third estate, with 184 deputies, was sufficient to swamp the nobles and the clergy, and the debate on the new constitution soon produced a deep cleavage between *Liberales*—it was then that the term took on its party political connotation—and *Serviles*. The former won the day, and the constitution of 1812 was duly promulgated. Its 384 articles attributed to the Spanish nation the exclusive right to determine its fundamental laws. The king would legislate with the *Cortes*, a single chamber; but royal decrees, to be valid, required the countersignature of a minister, and ministers were responsible to the *Cortes*. Freedom of the press, though not of religious belief, was assured, and legal codes were made uniform for the whole country. The constitution was doubtless unworkable in many respects and never achieved more than a symbolic importance; but by vesting in the people for the first time the responsibility for determining under what laws they should live it closed an epoch in their history and made of constitutionalism the first characteristic of the new age. Its influence was profound also in two other directions. Its radical idealism, offending the mass of the people, threw them into the arms of the liberated Ferdinand and prepared the way for one of the most reactionary regimes in Spain's record, the tendency persisting thereafter for Spanish politics to swing from extreme to extreme. Also it gave warrant for the later identification of liberalism with anticlericalism on the one hand and with centralization on the other, to the great complication and confounding of the political scene in subsequent periods of stress and crisis.

Expulsion of the French.—In the eventual recovery of its territorial integrity and independence, Spain owed a great deal to British assistance. Wellington's campaign of 1812, though it gave him the temporary occupation of Madrid, was again indecisive; but, after Napoleon's disastrous retreat from Moscow and the withdrawal of French troops from Spain to central Europe, he found himself in 1813 fighting at last on equal terms. The defeat of Jean Baptiste Jourdan at Vitoria on June 21 finally decided the issue in the peninsula. Joseph retired altogether, and Wellington fought his way across the Pyrenees into France. Napoleon, after his crushing defeat at Leipzig, recognized the impossibility of retaining his hold on Spain by releasing Ferdinand, who returned to Madrid in March 1814.

Ferdinand VII (1814–33).—Before entering Spain Ferdinand had undertaken to maintain the constitution of 1812; and on reaching Figueras he was met by a formal demand in this sense from the *Cortes*. But he was quick to appraise the true temper of the nation and in rejecting the demand now he was supported alike by the army, the church and the masses. Had he been capable of governing well, he might have ruled despotically for long enough. He was instead base by nature, without real capacity and at once cruel and cowardly. Himself governed by a *camarilla* of low favourites, he changed his ministers incessantly on caprice and was consistent only in his ferocious persecution of the *Liberales*.

Partial revolts were easily crushed. That of Jan. 1820, beginning as a mutiny at Cádiz in protest against the sending of an expeditionary force to crush rebellion in the American colonies, became a revolution. Its leader and hero, Col. Rafael Riego, was one of a number of officers, many of them erstwhile prisoners in France, who now subscribed to liberal doctrines. Six years of tyrannical excesses had turned the army largely against Ferdinand, and for three years, till 1823, he was a prisoner in the hands of a section of his subjects, who proclaimed once more the constitution of 1812. Ferdinand appealed for help to the monarchies of the Holy alliance, who were much concerned to see kingship thus contemned. As early as 1820 the emperor Alexander I of Russia had proposed joint intervention. The project came to nothing because of British opposition and the refusal of Metternich to contemplate the march of a powerful Russian force through the Austrian dominions. In 1822 the question arose again at the congress of Verona, France proposing that it march into Spain, as Austria had into Naples.

Great Britain saw in the demand only a pretext for reviving traditional Bourbon ambitions in the peninsula and protested vigorously; but the mandate was granted by a majority of the powers and on April 7, 1823, Louis Antoine de Bourbon, duke of Angoulême, crossed the Bidassoa at the head of a powerful army. The result was a startling proof of the flimsy hold of liberalism on the Spanish people. What Napoleon had failed to accomplish in years, Angoulême seemed to have achieved in a few weeks. Napoleon, however, had sought to impose upon Spain an alien dynasty; Angoulême came to restore the Spanish king "to his own" and found some Spaniards forthcoming with active support and the majority at least tacitly co-operative. The *Cortes*, carrying the king a prisoner, fled again to Cádiz, where after a brief siege it surrendered. Ferdinand, alleging compulsion, refused to honour the amnesty he had sworn, and the French troops witnessed with helpless indignation an orgy of cruel reaction begun under the protection of their bayonets.

The Spanish Colonies.—The emancipation of Spain's American colonies was an indirect result of Napoleon's detention of Ferdinand and invasion of Spain. Juntas in the various administrative capitals overseas assumed power first in the name of Ferdinand against Joseph Bonaparte, then in the name of the colonies themselves against Spain. Some of these had already become independent *de facto* by 1810, and many more became so in the ten years following; and the recognition of their independence *de jure* was, for Great Britain at least, merely a question of time. When the French invasion of Spain was seen to be inevitable, George Canning informed the French government that Great Britain would not tolerate any subjugation of the Spanish colonies by foreign force. The United States, though declining to act in concert, followed with the famous message of Pres. James Monroe (Dec. 2, 1823) intimating his country's opposition to any interference by Europe in the Americas. The republic of Colombia had

already been recognized by the United States in 1822. Great Britain recognized Mexico and Colombia in 1824 and only delayed recognition of the other new states until they should have given proof of stability. There now remained to Spain, of its far-flung empire, only the islands of Cuba, Puerto Rico and the Philippines.

The Succession Question.—Ferdinand, still childless on the death of his third wife in 1829, married Maria Cristina of Naples in that year and in so doing threatened the mounting hopes of his brother Don Carlos regarding the succession. Don Carlos was likewise an extreme reactionary, and the clash of ambitions threw Maria Cristina on the support of the Liberals. The birth of a daughter, Isabella, in 1830 greatly complicated the issue. By the ancient law of Castile and León women could rule in their own right. This right had, however, been abrogated by an act of 1713 designed to prevent any union of the crowns of Spain and France; and, although Charles IV had restored the former position in 1789, his enactment had never been published and its validity was now hotly disputed. Hence the birth of Carlism, which was for more than half a century to be a disturbing factor in the Spanish scene. When, on Sept. 29, 1833, Ferdinand died, his daughter was proclaimed queen as Isabella II, with Maria Cristina as regent. The First Carlist War broke out almost immediately.

The Regency for Isabella II (1833–43).—Maria Cristina too would have ruled despotically if she could, but the support of the army and control of the machinery of government were not in themselves sufficient to resist the Carlists, in whose ranks were to be found not merely the clergy and other ultraconservative elements but also large numbers of Basques, Catalans and other peoples of the periphery who had been persuaded that their regional *fueros* were in danger and supplied the real fighting strength of the movement in defense of them. The necessities of her position thus drove the regent to accept parliamentary institutions under the constitution of 1837 and the return from England and France of the many exiles of the preceding decade gave much support to the liberal cause. But neither Maria Cristina nor her daughter ever genuinely accepted the principle of representative government and there ensued, within the civil war, a bitter struggle between parliament and the parliamentary cause on the one hand and the crown on the other, to the discredit of both.

The First Carlist War, which was for the crown a dynastic issue and for the nation a conflict between irreconcilable political principles, involved Great Britain in military commitments on the government's side and threatened more than once to disturb the peace of Europe. It ended with the convention of Vergara of Aug. 31, 1839, a recognition of Carlist defeat. Baldomero Espartero, the Liberal general, undertook to recommend to the *Cortes* the substantial confirmation of the Basque *fueros*, and Don Carlos left the country. But the Carlist cause remained alive; and the *Cortes*, which insisted, in the name of national unity and uniformity, on whittling down the *fueros* until little more than administrative autonomy remained, ensured continuing Basque support for Carlism. An earlier measure of kindred intent had been the supersession for administrative purposes, in 1833, of the earlier kingdoms of Spain by 49 provinces on the French model. From that year dates too the Catalan *renaixença*, a resurgence of regional feeling at once cultural and political that was to have far-reaching consequences. Another enactment destined to provoke repercussions for many years was the acceptance in the constitution of 1837 of state responsibility for the upkeep of the church and the consequent disposal by the state of ecclesiastical property. The Jesuits, readmitted by Ferdinand VII, had been re-expelled in 1835.

With the removal of the Carlist threat to the throne the regent was emboldened in her absolutist leanings. The new constitution of 1837, though much strengthening the royal prerogative, was still liberal and bore traces of the English Reform bill of 1832. Maria Cristina now sought to have it overthrown. Failing in this, she directed an attack on the independence of the municipalities in 1840, which provoked a major crisis that culminated in her abdication. Espartero was proclaimed in her place, and Spain had its first experience of the army in political control. After three disturbed years a military rising, with liberal and moderate backing, drove Espartero likewise into exile, and Isabella, then only 13 years old, was declared of age.

Isabella II (1843–68).—The new reign proved but a continuation of the regency, a confused conflict between the attempts of the crown to rule despotically with a mere pretence of a *Cortes* and the nation's growing demand for honest and efficient parliamentary government. New constitutions (1845, 1852, 1855) alternated between the progressive and the reactionary and caused a spurious importance to attach to the letter as distinct from the spirit of constitutionalism. Party leaders were almost always generals, Espartero, Ramón María Narváez (leader of the "Moderates" or extreme conservatives), Leopoldo O'Donnell, Juan Prim. As such they were unable, even where willing, to further the cause of civic responsibility in the nation at large.

The queen's marriage in 1846 provoked international repercussions. Louis Philippe of France, seeking to revive the family alliance, with the aid of the queen mother constrained Isabella to wed her cousin Don Francisco de Asís de Borbón, who was believed to be impotent; on the same day her younger sister Maria Luisa Fernanda was married to Louis Philippe's son Antoine, duc de Montpensier. The affair broke the entente between England and France—Palmerston having only agreed on condition that the second marriage waited on the birth of an heir to the first—and was a contributory cause of the downfall of Louis Philippe (*see* FRANCE: *History*). The church's right to hold and acquire property, restored by a concordat of 1851, was again called in question by the pope's acquiescence, under a new agreement of 1858, in the policy of dispossession. The major external event of the reign was the Moroccan War of 1859–60, in which victory over the Moors in the field was foiled of its full harvest by the opposition of England and France.

Cumulative discontents against the crown, reflected in the emergence of a republican party in 1854, culminated in a mutiny at Cádiz, in Sept. 1868, of a squadron under Adm. Juan Bautista Topete, which became the signal for a general secession. The stand for the queen made at Alcolea by Gen. Manuel Pavía, marqués de Novaliches, was an exception. Even her prime minister, Luis González Bravo, deserted her. Isabella crossed into France an exile, and her reign ended. The revolution of 1868 was the first to be openly directed against the dynasty. The "spurious race of Bourbon," it was boasted, had disappeared; and, as in 1810, sovereignty was again for the moment vested in the people.

AMEDEO AND THE FIRST REPUBLIC, 1870–75

The mass of the nation was, however, not yet ready for a republic, and a constituent *Cortes*, assembled in 1869, voted for a continuance of the monarchy under a different monarch; Gen. Francisco Serrano, a leader of the revolution, was declared regent until one could be found. In the constitution of 1869 may be found, for the first time since that of 1812, the entire program of the liberalism of the period. Monarchist opinion was divided, some favouring Isabella's son Alphonso under a regency, some the widower Ferdinand of Saxe-Coburg (formerly king-consort and then regent of Portugal), others the duc de Montpensier, or a Carlist or even the aged Espartero. The search for a democratic king, said Prim, the prime minister, was like looking for an atheist in heaven. The offer of the crown to Prince Leopold of Hohenzollern-Sigmaringen proved the spark that fired the Franco-German War.

At length, in Aug. 1870, Prince Amedeo (in Spanish, Amadeo) of Savoy, second son of Victor Emmanuel II, accepted the offer of nomination. He was elected on Nov. 3 and landed in Cartagena on Dec. 27. On that day Prim was assassinated. Amedeo's reign lasted nominally till Feb. 1873. It was a period of turmoil and confusion. The monarch, become the tool of intriguing politicians, was looked on as an intruder by men of principle, royalist and republican alike. The Carlists threatened a new rising in the north, there was republican agitation in the towns and when the first honourable excuse presented itself Amedeo resigned.

Republican Interlude.—Disillusionment with monarchy was for the moment complete, and the *Cortes* proclaimed Spain a republic. New elections were held to a constituent *Cortes*, but the republic ran its brief and tempestuous course without a constitu-

tion. Beyond initial acceptance of the federal idea there was little agreement on first principles. Four presidents held office within the year, Estanislao Figueras, Francisco Pi i Margall, Nicolás Salmerón and Emilio Castelar y Ripoll; ministries never knew stability, and many regions disavowed the authority of Madrid and proceeded to act independently. Cantonalists, going beyond federalism to a proposal to split Spain into cantons, seized Cartagena and the Spanish navy in its harbour, which was recovered for the government only by the help of British and German squadrons. The republic had further to contend with a new Carlist revolt, with monarchist opposition and with an insurrection in Cuba.

As it grew ever clearer that only the army could restore authority and order, one major plank in the republican program, the taking of the army out of politics, had perforce to be abandoned, and General Pavía was charged with the pacification of Andalusia. Castelar, reintroducing conscription, provided the means of reducing Cartagena and pursuing the war against the Carlists with vigour. When the *Cortes* reassembled in Jan. 1874 the extreme parties voted against Castelar, whereupon Pavía, then captain general of Madrid, turned the *Cortes* into the streets. The federal republic was dead. A unitary regime followed, under General Serrano, until, in Dec. 1874, another army *pronunciamiento* restored the Bourbons. Alphonso XII, 17 years old and till then at school in England, landed at Barcelona on Jan. 10, 1875.

THE LATER BOURBONS, 1875–1931

Alphonso XII (1875–85).—National exhaustion and disillusion caused a substantial rallying of the nation around the young king. Pacification, military and political, was his first concern. The Carlists were not finally quelled until the beginning of 1876, when a general amnesty was balanced by the forfeiture by the Basque provinces of their differential system of taxation and military service. Part of the large army thus released was then sent to Cuba, where after a further 18 months of fighting the insurgents accepted the convention of El Zanjón (Feb. 12, 1878). Arsenio Martínez de Campos, Spanish commander in chief, was made governor general and held out to the island the prospect of reforms, including a more liberal tariff and eventual self-government.

Alphonso's first premier, the conservative Antonio Cánovas del Castillo, ruled at first dictatorially without the assistance of parliament. Royal decrees set aside most of the legislation and reforms of the previous regime, respecting only—for a term—universal suffrage, on which basis a new *Cortes* was convoked in 1876. The constitution of that year was a compromise between that of 1845 and the more democratic principles of the charter of 1869; it was to remain in force for 47 years. By it the franchise was limited to taxpayers; the senate became in part elective; religious dissidence was tolerated though not recognized; and a considerable freedom of association was granted. Constitutional guarantees might, however, be suspended when the security of the state so demanded, a provision that left a tempting reserve of power in the hands of the crown, while the parallelism of civil and military districts facilitated its invocation.

The Liberals under Práxedes Mateo Sagasta formed the Fusionista party, whose name was indicative of its heterogeneity. But the real opposition came from the Socialist-Republican left. From the founding in 1869 of a Spanish section of the international, labour organization developed rapidly. The split between Marx and Mikhail Bakunin was reflected in Spain on a regional basis, labour being predominantly Marxist in Castile and Anarchist in Catalonia. Regionalism, likewise at its most active in Catalonia, provided another outlet for political dissatisfaction with Madrid.

Cánovas' first administration lasted till Feb. 1881, when, under suspicion of heading the country once again toward absolutism, he gave way to Sagasta and the Liberals. Thereafter, to the end of the century, the two parties alternated in power on an agreed rotatory system based on "made" elections that soon became one factor in the gradual discrediting of the restored monarchy. In Aug. 1883 an attempted republican rising, though soon quelled, led to the discovery of widespread revolutionary activities throughout the country. Fifty Republicans were returned to the *Cortes* at the general election of March 1893. In the country at large the last two decades of the 19th century were scarred by an increasing resort to direct action and crimes of violence, with corresponding waves of harsh official repression.

Alphonso's first wife, his cousin Mercedes, daughter of the duc de Montpensier, had died in May 1878, five months after the wedding. His second marriage (1879) was with Christina (Maria Cristina), a second cousin of the Austrian emperor Francis Joseph. On Nov. 25, 1885, he died, leaving two daughters. Six months later a posthumous son, Alphonso XIII, was born. The queen was proclaimed regent.

Regency for Alphonso XIII.—The regency began with five years of Liberal rule, marked by the civil code of 1889, another major step in the consolidation of the country's still heterogeneous legislation, by the restoration of universal suffrage and of trial by jury and by a liberalizing of the laws affecting the right of association and the liberty of the press. Attempts to curb the political proclivities of the army were less successful and added military to Conservative opposition. The elections of 1890, which returned the Conservatives to office, showed that very few of those lately enfranchised had voted; they showed also that the home office's control of the voting continued as before.

Cánovas now based his economic policy on protection, completely reversing the moderate free-trade policy which had been so beneficial to Spanish foreign trade since 1868. Heavy increases in duties on agricultural imports, particularly on breadstuffs, and the denunciation of all treaties of commerce that contained most-favoured-nation clauses soon made their adverse effects felt and evoked loud protest. Industrial interests alone benefited, and imports of raw materials, notably chemicals, coal and coke, increased. In 1892 Cánovas admitted that never since the restoration had the budget been balanced and that the exchequer was bankrupt. Political bankruptcy could be seen further in the failure of both parties to tackle the problem of the *latifundios* and a land-starved peasantry. Signs of labour unrest multiplied, again chiefly in Catalonia.

Morocco and Cuba.—The Rif tribes bordering the Spanish zone in Morocco had never been good neighbours. In 1894 an incursion into Spanish territory resulted in defeat for the Spanish forces and a severe shock to public opinion at home. An army of 25,000 men, with 30 generals, was sent to Melilla, but hostilities were averted for the time being by a treaty signed in Fez and by the sultan's agreement to pay an indemnity of 20,000,000 pesetas. Earlier, between 1884 and 1886, Spain had established dominion over the vast territory of Río de Oro, south of Morocco, in furtherance of a new African policy of pacific penetration advocated by Joaquín Costa.

The position in Cuba had again grown threatening, and press censorship left Spanish opinion unprepared for its outcome. Cuban autonomist deputies in the *Cortes* had formally tabled the demand for home rule in 1878, but this was rejected, as was a Conservative measure of 1893 for political and administrative reforms tending in that direction. In 1895 a much weaker Cuban Reform bill was passed by all parties, but before it could be promulgated a separatist rising had broken out in the old haunts of Creole dissatisfaction, near Santiago de Cuba, backed by a strong body of sympathy in the United States. Sagasta dispatched a force of 12,000 men to reinforce the 15,000 Spanish troops already in the island, and Cánovas, succeeding once more to the premiership, insisted that submission must precede any concessions. Gen. V. Weyler y Nicolau applied himself to the stamping out of disaffection with a ruthlessness which brought repeated diplomatic representations from the United States, which at the same time was supplying the rebels with arms.

Realizing the danger of more direct U.S. intervention, Cánovas presented to the *Cortes* a bill granting Cuba a substantially enlarged autonomy. His assassination by an anarchist on Aug. 9, 1897, brought Sagasta back into office. Recalling Weyler, Sagasta hastily enacted conciliatory legislation and sought to propitiate the United States, even offering a treaty of commerce which would have allowed the U.S. every advantage in trade with the West Indies possessed by Spain itself. But things had already gone too far, both in Cuba and in the United States, where public opinion

pressed irresistibly on congress and president alike.

Spanish-American War, 1898.—On Feb. 15, 1898, in circumstances never clearly established, the U.S. cruiser "Maine" blew up in the port of Havana, and events moved quickly to a crisis. Spain appealed to courts and governments and to the pope for mediation, but none was disposed to go beyond mild representations to Washington. In April Pres. William McKinley approved a resolution demanding the cessation of hostilities in Cuba with a view to evacuation by the Spanish forces. Sagasta gave the U.S. minister in Madrid his passports and severed relations with the United States; the known inadequacy of Spain's resources for a struggle could not prevent it from yielding to the urge to defend the last remnants of its colonial empire.

The sequel is soon told. The Spanish fleet in the far east was defeated in Manila bay, the Caribbean fleet was destroyed off Santiago de Cuba and all communication between Spain and the colonies was severed. Within a fortnight of the landing of a U.S. force near Santiago the Spanish garrison surrendered (*see* SPANISH-AMERICAN WAR). At the end of July Spain sued for peace. The agreement of Aug. 9 stipulated clearly that Spanish rule in the new world must be considered at an end; and by the peace treaty of Paris (Dec. 12, 1898) Spain renounced unconditionally all sovereign rights over Cuba and Puerto Rico and ceded to the United States, in consideration of a payment of $20.000.000, the Philippine Islands, the Sulus and the largest of the Marianas. The Carolines and a few other Pacific islands Spain sold to Germany in 1899, and a couple of islands omitted in the treaty of Paris were purchased by the United States in 1900. Apart from holdings in North and West Africa, the Spanish empire was at an end.

The Reaction in Spain.—The Liberal goverment was held responsible for the debacle; but the seeds had been sown long before, and the change to a Conservative cabinet was the least significant among its consequences. From the mood of sober analysis and heart-searching that ensued there emerged the "generation of '98", a group of thinkers who through their chosen medium—the essay —effected a renovation not merely of literature but of the country's intellectual life. The debate on the country's role in history led to a cleavage between *europeizantes* and *africanizantes;* it led also to a new interest in Africa as the one remaining field of overseas activity. With the radical reduction of its commitments the army became top-heavy, at once an intolerable drain on the nation's economy and more prone than ever to meddling in politics. The wave of solidarity with Spain that swept through Latin America went far to neutralize remaining antagonisms in this sphere.

Long-due reforms in taxation and the servicing of the nation's debts were carried through by a vigorous Conservative minister of finance, Raimundo Fernández de Villaverde. Bondholders were called on for even heavier sacrifices than taxpayers, and the interest on colonial debts, like that on internal debts and on Spanish holdings of external debts, was made payable in pesetas. But in the main the political scene remained untouched by the new spirit. Sagasta, premier once more in March 1901, was confronted by two major problems. One was the growth of the religious orders and congregations, in defiance of the Law of Associations of 1887 by which all congregations were required to register their members and, when not specifically recognized under the concordat of 1851, to apply for authorization. A proposal now to enforce the law greatly incensed the church. The papal nuncio announced that Rome would only consent to discuss the question on condition that all requests for authorization would be granted. To avoid a crisis just when the young king was about to become of age, Sagasta accepted a *modus vivendi.* A final settlement was not reached with the Vatican until 1912.

Industrial unrest constituted a more pressing peril. Fomented by socialist agitation and by anarchist direct action, it led in Jan. 1902 to serious riots in Barcelona and Saragossa and on Feb. 16 to a general strike in the former city. General Weyler's methods in restoring order all but brought the government down. They provided no solution to anarchism, which under the influence of Georges Sorel was now moving toward syndicalism as a basis for organized opposition to all government as such.

Accession of Alphonso XIII, 1902.—Alphonso XIII was enthroned on his 16th birthday, May 17, 1902. He was still then under clerical and reactionary influences, and his contemptuous treatment of ministers showed from the beginning an intention of intervening actively in politics. His first official act was to countermand a decree closing the military colleges, aimed at reducing the disproportionate officer strength of the army. Sagasta died in 1903. Conservatives and Liberals alike were now suffering from weakened leadership, and there ensued a period of instability and frequent changes of government.

A tendency to disintegration in the political parties was symptomatic of the growing subordination of national to sectional interests. It became an important factor in the eventual collapse of representative government. This political instability contributed to the general unrest; and the old antagonisms, born of the conflict between the industrial north and the agricultural south and of the now traditional cleavages over regional, social and religious issues, were increasingly complicated by labour agitation. The notorious Law of Jurisdictions of 1905, under which attacks on officers and military institutions were to be tried before military tribunals, was a serious capitulation of the civil authority, that had far-reaching consequences.

Alphonso's marriage on May 31, 1906, to Victoria Eugenia (Ena) of Battenberg, princess of Great Britain, marked a stage in his emancipation from clerical-conservative tutelage. Ecclesiastical controversy again loomed large at this period, formal recognition of the validity of civil marriage arousing such agitation among the clergy that the government threatened the bishop of Tuy and the Córdova chapter with prosecution. A new Law of Associations in 1906 brought the Liberal party itself to the verge of disruption, and after three changes of premiers the Conservatives returned to power under Antonio Maura.

Maura (1907–09).—Maura, who had already been premier under the regency, had many qualities of leadership. A sincere Catholic, though no clerical, he pursued first the maintenance of order and the reform of local government. A bill to the latter end was submitted to the *Cortes* in Oct. 1907. Aimed at destroying the power of the *caciques* or local bosses and at educating the electorate in their privileges and responsibilities, it made voting compulsory, abolished official interference with the polls and increased considerably the responsibility of local elected bodies. The bill met with strong Liberal opposition because of its alleged regionalistic tendencies and, though finally approved by the lower house in Feb. 1909, did not reach the statute book.

The elections of 1907 were notable for the return of a strong republican minority and the striking victory of the Solidaridad Catalana, a fusion, resulting directly from the Law of Jurisdictions, of seven Catalan parties in defense of regionalist interests. Of the 44 seats allotted to Catalonia, 41 were gained by its candidates. Maura, ruling rather as an enlightened despot than as the head of a constitutional government, was concerned also to strengthen the navy, to reform and to reconstruct the police force, to institute industrial tribunals and to regulate the conditions of work of women and children.

Trouble long brewing in Morocco broke out in July 1909 with an attack by Rif tribesmen on Spanish railway workers. The need to reinforce the Spanish troops in Africa had for some time been apparent, but Maura had hesitated to lay the necessary estimates before the *Cortes.* The proposal now of the war minister, Gen. Arsenio Linares, to organize a new field force by calling out the reserves, with its vivid memories of the useless miseries of earlier overseas expeditions, provoked another general strike in Barcelona; and a movement at first directed against "conscription" rapidly developed into a revolutionary attack on the established order in church and state that spread to the whole province. Churches and convents were favourite targets for violence. In Barcelona the rising was suppressed after three days of street fighting (July 27–29).

On July 28 martial law was proclaimed throughout Spain, and a military reign of terror continued until the end of September. In the fortress of Montjuich in Barcelona were collected not only rioters caught red-handed but many others, notably journalists,

whose opinions were unpalatable to the government. A sensation was caused by the arrest of Francisco Ferrer, a theoretical anarchist well known in many countries for his anticlerical educational work and especially as the founder of lay schools in Spain. Ferrer was tried by court-martial and shot. The execution aroused widespread indignation throughout Europe and produced a ministerial crisis in Madrid. The government, already weakened by a reverse in Morocco (Sept. 30), was furiously attacked by the Liberals under Segismundo Moret y Prendergast at the opening of the *Cortes* in October, and Maura resigned.

Moret and Canalejas (1909–12).—Moret's Liberal ministry sent General Weyler back to Barcelona to keep order, released most of the Montjuich prisoners, reduced the forces in Morocco and reopened negotiations with Rome for a modification of the concordat. In Feb. 1910 Moret was forced to resign through the king's refusal to grant him a dissolution, and José Canalejas y Méndez, the outstanding Liberal of his day, took office. His government, "inspired by the universal spirit of liberty of conscience," tackled the continuing refusal of the religious houses to comply with the law (in respect of registration and the payment of taxes when engaged in industry and commerce) and announced measures both to restrict the number of religious houses and to implement the religious toleration enshrined in the constitution. A violent Catholic agitation resulted, 72 Spanish bishops and archbishops presenting a formal protest to the government. To attacks from the Vatican Canalejas retorted with the famous "Padlock law" at the end of 1910 forbidding the settling of further congregations in Spain until a revision of the concordat should have been agreed on with Rome.

On the Vatican's demanding that all the obnoxious measures be rescinded as a preliminary to negotiations, the Spanish ambassador to the Vatican was recalled. Hints were dropped in papal circles of a possible *rapprochement* with the Carlist claimant to the throne, and diplomatic relations were not restored until 1912. The overthrow of the monarchy in Portugal on Oct. 5, 1910, and the expulsion thence by the new republican regime of all religious congregations did not ease the government's difficulties.

French activities in Morocco at this time compelled Spain to adopt a more active policy in its zone. To counterbalance the French entry into Fez a Spanish force was landed at Larache in June 1911 and took Alcazarquivir and Arcila. The recruiting of native regiments under Spanish officers did something to lessen popular indignation at home. France was prevented by the Agadir incident from reacting strongly, but a certain tension continued and pervaded the negotiations for a Franco-Spanish agreement following on the Franco-German treaty on Morocco. Concluded on Nov. 27, 1912, that agreement reduced the Spanish zone for the third time in 12 years.

A general railway strike in the middle of 1912 threatened the country with chaos. Canalejas met it by drafting railway workers into military service and compelling them to continue at their posts as soldiers. Such firmness was balanced by a genuine readiness to understand legitimate grievances, which led the Catalanists to make a serious attempt to come to terms. The premier responded by introducing a bill of *mancomunidades* (commonwealths) whereby the four Catalan provincial councils might combine in one joint body (*mancomunitat*) to which the state would grant a substantial measure of devolution. The bill, which was inevitably unacceptable to many Liberals and Conservatives alike, and was regarded in Catalonia as merely a first instalment of Catalan demands, had not passed the *Cortes* when, on Nov. 12, 1912, Canalejas too was assassinated by an anarchist.

The Parties in Decline.—The new premier, Alvaro de Figueroa y Torres, conde de Romanones, was natural heir to the Liberal leadership, but he was unable to maintain the firm party discipline of his predecessor. He pursued meantime Canalejas' policy of seeking to attract and absorb the more moderate wing of the republican left, a policy the desirability of which had been greatly enhanced by the revolution in Portugal. Alphonso himself recognized the need for such an accommodation, and the response of republican leaders like Melquíades Alvarez led to hopes that the disintegration of the regime might yet be checked.

These hopes were dashed, firstly by an internal crisis in the Liberal party over the bill of *mancomunidades*, then by the "implacable hostility" proclaimed by Maura against any policy of collaboration with the antidynastic parties. With the resignation of Romanones in Oct. 1913 the king realized that the Liberal policy which he had consistently followed since 1909 was at an end. Maura's views aroused, however, such intense popular resentment that Alphonso entrusted responsibility instead to his fellow-Conservative Eduardo Dato, a choice that initiated the disintegration of the Conservative party in its turn.

The tangled fate of the bill of *mancomunidades* was resolved, against Dato's judgment, by a royal decree of Dec. 18, 1913, that granted to the four Catalan provinces the right to group together for "exclusively administrative purposes within their competence." This and several similar royal decisions in military matters, which savoured of unconstitutional action, created considerable opposition to the government. But internal difficulties were soon to be merged in the all-absorbing preoccupations and repercussions of World War I.

World War I.—The international situation of Spain at the outbreak of war in 1914 was determined officially by the agreement between Spain, France and England signed at Cartagena in 1907 and confirmed in conversations held there in 1913 between Raymond Poincaré and Romanones, which stipulated that, "should new circumstances arise tending to alter the territorial *status quo*" in the Mediterranean or on the Atlantic coasts of Europe and Africa, the three powers would "enter into communication" in order to take any necessary measures.

The war was obviously such a circumstance. France and England made no sign, but Dato promptly declared for absolute neutrality, Spain having no stake in the conflict. As for the nation, its sympathies were divided. The working classes, most of the intellectuals and the trading communities—the Liberal left—were pro-Ally. The clergy, most of the army and of the bureaucracy and the "idle rich"—the Conservative right—were pro-German. But there was agreement on the one essential of keeping out, and the *Cortes*, resuming its sittings on Oct. 30, 1914, unanimously endorsed the government's decision.

A debate on military reforms brought Dato's government down in Dec. 1915, and Romanones again formed a Liberal ministry. By then the country was profiting from industrial activity on the Allied account, the peseta appreciated and the foundations were laid for the repatriation of the foreign debt. The picture had its reverse. In Morocco military operations were again necessary, this time to defend the mines from hostile tribes; while Spanish shipping was victimized by a ruthless German submarine campaign. By Sept. 1916 Spain had lost more than 30,000 tons by torpedo attack and another 50,000 in obscure circumstances. When, on Jan. 31, 1917, Germany intimated an "absolute blockade" of the Allied coasts, Romanones delivered a spirited answer intended partly to test public opinion.

His resignation on April 19, which cost him the leadership of his party, was not solely the result of the nation's failure to rise to his lead. Toward the middle of 1916 the army had begun to organize secretly its so-called *juntas de defensa* in furtherance first of its professional efficiency, second—and soon chiefly—of its traditional prerogatives, the threat to which came principally from the king's fondness for personal intervention. Before long they were presuming to dictate to the government in a manner which led Romanones to attempt their dissolution. But this new menace to authority from the one bulwark of order loomed larger with each endeavour to exorcize it, and Romanones chose the easier path. Under his successor, Manuel García Prieto, an order to arrest the leaders of the juntas provoked an ultimatum (June 1, 1917) that brought the new government down; and Dato, recalled to office, capitulated before them, conceding to their regulations the force of law. This victory of a movement revolutionary in essence encouraged at first, in a public not alive to its implications, hopes of a renovation of political life. Coupled with early reports of the Russian Revolution, it stimulated the ferment at work in the country, and political manifestoes asking for a "renovation" in government and constitution followed one another (Socialists June 12,

Catalanists June 16, left coalition June 16).

In the teeth of government opposition, an assembly of senators and deputies convoked in Barcelona formally demanded constitutional reform and set up three commissions to report to a second meeting in Madrid. This movement, which might conceivably have provided the renovation so urgently needed, was defeated by the resort to direct action of extreme labour elements. On Aug. 13, three days before the planned second meeting, a revolutionary general strike broke out, led this time by the Socialist U.G.T. (General Workers' union) of Madrid and aimed openly at establishing a socialist democratic republic, which soon threatened to paralyze the whole country. A state of war was again declared. The army now showed its true colours and put down the strike by machine guns, and the juntas remained the only real force left in the country.

Among the victims was the constitution. The new minister for war imposed on the king by the juntas, Juan de la Cierva, dominated a heterogeneous ministry of Maurists, Liberals and Catalanists. Its short life (Nov. 1917–Feb. 1918), in spite of army support, showed that a severe blow had been struck not merely at party government but at cabinet government. Brought down by a telegraph strike, it was followed by a hiatus that only ended when Alphonso, at a midnight meeting of former premiers at the palace on March 21, constrained them under threat of abdication to form a ministry of all the talents, under Maura. This lasted till Nov. 6. The abdication of the kaiser and the toppling of thrones throughout Europe at the end of World War I had a natural accompaniment in the return of the Liberals to power, first under García Prieto, then (Dec. 3) under Romanones, whose first act as premier was to visit Pres. Woodrow Wilson in Paris.

The Postwar Situation.—Spared the war, Spain had passed nonetheless through a very disturbed and anxious four years. The submarine war cost it 65 vessels, in all 140,000 tons, which Germany had agreed to indemnify on a ton-for-ton basis. Against this, the influx of capital went far to revolutionize its economy. Gold reserves in the Bank of Spain, £22,680,000 in 1914, had risen to £88,920,000 in 1918. Not merely was Spain able to resorb in large part its foreign industrial debt and, almost entirely, its national debt and to acquire the ownership of the railways, new industries were founded, not all of which disappeared with the return of peace, and in older industries such as the Basque iron and steel works, plant and conditions were modernized. A mingling of two political currents, one a democratic and constitutional agitation born of the Allied victory, the other a revolutionary agitation akin in the civil sphere to that of the *juntas de defensa* in the military, created an involved situation in Barcelona which issued in the spring of 1919 into open conflict between the military and the civil authorities. When the former expelled the latter from the town, the failure to take up so flagrant a challenge brought the government down, and again there ensued a period of cabinet instability. Its most notable feature was that "made" elections, the mainspring of the old system, were failing now to produce the desired majorities. On March 8, 1921, Dato, who had been engaged in trying to rebuild Conservative unity, was assassinated.

The murder accelerated the final disintegration of the rotatory system which, ever since the death of Alphonso XII, had served the monarchy as substitute for a real constitutional regime on the English model. To its own internal weaknesses was added the triple assault from outside. There were the antidynastic elements, once lumped together under the common label of Republicans, now moving toward socialism on the one hand and toward different varieties of syndicalist or anarchist revolution on the other hand; there were the Catalan home rulers, gradually evolving from autonomy toward separatism under the stimulus of the respect shown by Versailles for the principle of self-determination; and third there was what might be called the military syndicalism embodied in the defense juntas.

With the last-named, governments lacked the moral courage to deal; toward the violence so frequently associated with the first they opposed more violence, failing to recognize legitimate grievances where they existed and tending always to identify the employers' interests with those of authority. To Catalanism opposition in Madrid was instinctive, and the failure of a new approach from Barcelona in 1918–19 did much both to strengthen and to unify support for the movement throughout Catalonia. By 1923, when two strong groups, Estat Català and Acció Catalana, of the right and left respectively, were in agreement on the main objective and had concluded a triple alliance with Basque and Galician autonomists, it was clear that this issue was destined to loom larger and larger in the country's politics.

War in Morocco.—Effective occupation and pacification of the Spanish zone in Morocco were still far from complete, and in guerrilla warfare the tribesmen were often more than a match for Spanish troops, with whom service in Morocco had become notoriously unpopular. In 1921 Manuel Fernández Silvestre, in command at Melilla, embarked impulsively, with Alphonso's personal backing, on an operation which ended in crushing disaster at Anual (July 21). To heavy losses in men and matériel was added a forced withdrawal from the whole eastern zone, the fruit of 12 years of slow penetration. General Silvestre committed suicide.

The repercussions in Spain were profound. The government fell, and Maura was called on to face the emergency. An army of 140,000 men was sent to the rescue of the troops in Morocco. To the growing public demand for an inquiry the new government had at length to give way, and General Picasso was charged with investigating responsibilities, his report being submitted in due course to a committee of the *Cortes*. Meanwhile the *juntas de defensa* had reacted strongly with a campaign of threats and accusations against the *Cortes*, which they sought to hold responsible for the general disorganization underlying the Moroccan crisis. The government retorted with a formal disbandment of the juntas, replacing them by *juntas informativas* subject, in theory at least, to the minister for war. But the juntas were the stronger, and early in 1922 Maura was forced from office. His successor, another Conservative, José Sánchez Guerra, was brought down in December by the debate in the *Cortes* on the Picasso report. The insistence in Socialist sectors of opinion that definite responsibilities attached to specified individuals and should be brought home to them charged the issue with high explosive.

A Liberal coalition followed and was confirmed in office at the general election of April 1923. The sensational feature of this was the return of Socialists for five out of the seven seats in Madrid, a success attributed by their party to the strong line that it had taken on the war in Morocco. The burden which the Moroccan problem imposed on the exchequer (an average of 358,000,000 pesetas annually in 1919–23 as compared with 75,000,000 pesetas for 1909–13 and 146,000,000 pesetas for 1913–19) was aggravated by the gradual wilting of the artificial prosperity of the war years and recurrent industrial troubles at home. For the most part, however, these problems were but facets of the basic inability of government after government to uphold the prestige of authority. In the 21 years since Alphonso's accession to the throne there had been 33 ministries.

Dictatorship of Primo de Rivera (1923–30).—In Sept. 1923 the report of the *Cortes* commission on the Anual disaster was ready. Several politicians were believed to be deeply involved, some as having abetted the king in his fondness for unconstitutional initiative. The reaction of the army, notoriously restive under parliamentary or civilian criticism, was awaited with some concern. Suddenly, on Sept. 13, a dramatic stroke from this quarter brought to an end the system under which the country had lived since 1875. The captain general of Catalonia, Miguel Primo de Rivera, marqués de Estella, rebelled against the government from his headquarters in Barcelona, rejecting its authority and threatening its members with jail. His manifesto declared it imperative to liberate the country from "the professional politicians, the men who for one reason or another are responsible for the period of misfortune and corruption which began in 1898 and threatens to bring Spain to a tragic and dishonourable end." "The wide net of greedy politics," he added, "has caught and imprisoned in its meshes the royal will itself." He announced the setting-up of a military directorate which would find for the Moroccan problem a "quick, dignified and sensible" solution and which would bring home responsibility promptly and justly by the formation

of tribunals "of recognized moral authority."

The premier proposed to the king strong action against the rebels, who included now the military authorities of Bilbao and Saragossa and a council of high officers in Madrid. The king asked for time for reflection, and the government resigned. Two hours later a deputation of generals was received at the palace and the king requested Primo de Rivera, who was still in Barcelona, to take office. Administratively the new directorate struck an original note in that the only minister was Primo de Rivera himself, ministerial departments being left in charge of permanent officials while the "directors" concerned themselves with specific problems as they arose. The one significant exception was the home office, which was entrusted to Gen. Severiano Martínez Anido with the rank of undersecretary of state. Public opinion received the *pronunciamiento* without demonstration, save for a protest from the Socialist party. The directorate declared the country to be in a state of war. A strict press censorship and a ban on political manifestations were imposed. The *Cortes* was dismissed, and the seizure of the files of the Commission on Responsibilities, shelving the further inquiry into the Moroccan disaster which was due to begin on Sept. 15, made clear the reason for the timing of the coup. When, two months later, the king refused to acknowledge his obligation under the constitution to convoke a new *Cortes* within three months of a dissolution, he made equally clear his personal responsibility in the collapse of the parliamentary regime and rendered it inevitable that the eventual fall of the dictatorship should involve that of the throne. In November king and dictator paid an official visit to Italy and expressed open admiration of Fascism.

The compelling problem was still Morocco, and it was there that the dictator won his greatest success and the nation's gratitude. The first phase, initiated in the summer of 1924, was one of cutting losses and retreating to a base line near the coast sufficient to cover Ceuta and Melilla. Primo de Rivera claimed later that in this he was interpreting the popular will, "which was weary and sceptical of any other possible solution." He himself led the operation in the field. It evoked opposition from the army and was not effected without heavy loss, but by the end of the year it was completed.

Abdel Krim's sudden attack on the French protectorate in 1925, which led to combined action by France and Spain against the rebellious tribes, made possible a complete reversal of policy. The brilliant landing (Sept. 8) in Alhucemas bay, again led by Primo de Rivera in person—the other generals having refused—was a thrust at the heart of the enemy's power, and within a month Abdel Krim was in flight. The pacification of the entire Spanish zone was not completed until 1927, but already Spain's position and prestige in Africa had been retrieved.

Primo de Rivera was now tempted to consider his mission at an end and withdraw, but there was substantial popular support for the view that the dictatorship could contribute no less to the solving of the nation's internal difficulties. He decided, however, that the military directorate had by now served its purpose and, replacing this in Dec. 1925 by a predominantly civilian cabinet, sought to base his regime on more normal political foundations. The Unión Patriótica (1926) was an attempt to form a nonparty organization of all such good citizens as would accept his own motto of "Country, Religion, Monarchy"; the "supreme national assembly" (Sept. 1927) represented a pseudo *Cortes* largely nominated by the dictator and given a merely advisory function.

Primo de Rivera remained the fount of power and, in his zeal for reform, was driven inevitably to abuse that power, both by attempting to force the pace and by riding roughshod over dissident susceptibilities. His successes lay accordingly in such spheres as public works, finance and public order. In the conciliation of opposition elements he failed to make headway. The Confederación del Ebro, associating the many interests concerned in the utilization of that river, was a notable innovation and the forerunner of others of the same nature. Ambitious schemes for the electrification of the railways and for road development were set on foot. Income tax was introduced in Dec. 1926 and gradually worked its way into the national finances. Morocco had ceased to be an intolerable drain on the exchequer, and in 1927, for the first time in 20 years, no loan was raised. Economic policy was directed to the correction of the heavy adverse balance of trade by stimulating home industries, canalizing investment and creating state monopolies. In the labour field an elaborate system of corporations was initiated; deriving from the Italian model, it won the approval of many Socialist leaders. An agrarian credit bank gave promise of new life to the small farmer; the other great problem in agriculture, the breaking-up of the *latifundios,* the regime left untouched.

In 1925 the government undertook the reorganization of provincial and municipal administration. The right to levy certain rates was delegated to municipalities, with a corresponding widening of their range of action; and female and proportional suffrage were instituted in local elections. One important provision in the decree of provincial reform was the abolition of the *mancomunitat,* the central body in Barcelona responsible for co-ordinating the work of Catalan local councils. In its Catalan policy the directorate had soon reversed the sympathetic attitude to autonomy voiced in Primo de Rivera's first utterances, which were evidently directed to winning support in Catalanist circles for his coup d'état. Army opposition to any measures of Catalan devolution combined with the inherent logic of dictatorship to create an attitude and a policy described in Catalonia as one of systematic persecution.

In external relations interest centred chiefly on the League of Nations. In 1926 Spain claimed a permanent seat on the council of the League in fulfilment of repeated promises by the French and British governments. Similar claims by Brazil, Poland, China and Persia and opposition from Germany, Sweden and others to any increase in the number of permanent seats led to a confused situation and to much resentment in Spain. At the June meeting of the council Spain's representative was ominously absent, and on Aug. 9 Spain signed a treaty with Italy stipulating mutual neutrality in the event of unprovoked aggression. Further to stress this independent mood, Primo de Rivera announced in the latter month his desire for such a drastic change in the Tangier convention as would have amounted to the setting up in Tangier of a Spanish protectorate. The assembly and the council of the League decided in December to institute a special semipermanent seat on the council for Spain. Spain refused this and gave notice of withdrawal from the League.

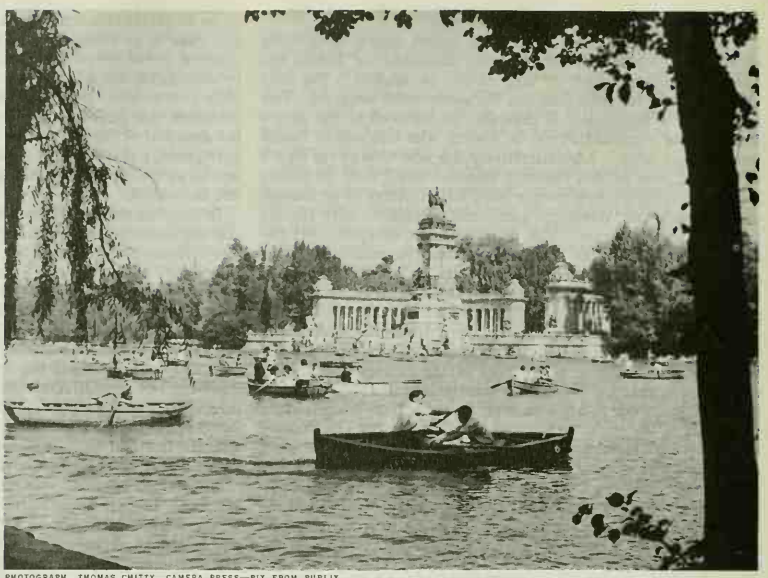
Negotiations with France over Tangier led finally to a number of French concessions which improved the security position in the Spanish zone of Morocco. This was a factor when, on the invitation of the president of the League council in March 1928 to reconsider its attitude, Spain announced its readiness to remain a member of the League and to accept such status on the council as the League might determine.

The year 1926 witnessed three attempts to overthrow the dictatorship by force. A military rebellion at Valencia in June, led by old General Weyler and Gen. Joaquín Aguilera, was nipped in the bud. Trouble with the artillery corps arose from a measure in July abolishing its cherished tradition of accepting promotion by seniority only as a guarantee against favouritism. A compromise settlement was, the corps claimed, broken by Primo de Rivera, and in September the artillery officers too prepared a revolt. Its failure was followed by the abolition of the corps. An amnesty later restored all but a few to their posts, but much ill-feeling remained. In November Col. Francesc Macià, exiled leader of the separatist Estat Català party, launched a futile attempt to invade Catalonia from France and set up a provisional government there for the restoration of "the national sovereignty."

By the end of 1928 civil discontent with the dictatorship had combined with army unrest to render the dictator's position more difficult. Continual dissensions between dictator and king, riots in the universities and protests from intellectuals were symptoms of the general unease. The world economic crisis of 1929 was soon reflected in a serious fall in the peseta. In Jan. 1930 the finance minister, José Calvo Sotelo, resigned; and when, shortly after, the captains general of the army intimated that they no longer stood behind the dictator his fall became inevitable. On Jan. 28, 1930, Primo de Rivera admitted defeat and, laying down the reins,

crossed the Pyrenees. He was already broken in health and died in Paris on March 16.

Fall of the Monarchy.—A temporary government under Gen. Dámaso Berenguer now sought to prepare the way for a return to constitutional rule. Unrest, however, continued. Revolutionary strikes multiplied all over the country. The ranks of republicanism were swollen by all those who refused to absolve the king from his responsibility in the abrogating of the constitution, and a revolutionary committee, pledged to bring about a republic, concluded with Catalan autonomists the pact of San Sebastián (Aug. 17) whereby it undertook, in return for their support, to introduce a bill for regional autonomy as soon as it was in power. The peseta, which had dropped from 29 to the pound sterling to 36.50 during 1929, fell to 51 by the autumn of 1930. An abortive military rising took place at Jaca (Dec. 15). On Feb. 14, 1931, the government resigned. A new administration under Adm. Juan Bautista Aznar was charged with convoking, first, municipal elections, and then parliamentary elections for a constituent *Cortes*, but it was understood that the former would be taken as a test on the one great issue and that, if the outcome were held to warrant it, the revolutionary committee would attempt to seize power.



PHOTOGRAPH, THOMAS CHITTY, CAMERA PRESS—PIX FROM PUBLIX

BOATING ON THE LAKE AT RETIRO PARK, MADRID. IN THE BACKGROUND IS A MONUMENT TO ALPHONSO XII

Full results of the municipal elections of April 12, 1931, which went off quietly enough, were never published. The only figures issued showed that 22,150 seats had gone to Monarchists and 5,875 to Republicans. But in Barcelona the Republicans had a three-to-one majority of votes; in Madrid, Niceto Alcalá Zamora y Torres, leader of the revolutionary committee, headed the poll; and most of the large cities showed republican majorities. Alcalá Zamora took the bold course of demanding Alphonso's abdication and the immediate transfer of powers to his committee. The king, after some parleying, agreed to leave Spain, but without formal abdication, and on the night of April 14 he embarked at Cartagena for France. The revolutionary committee became the provisional government of the second republic.

THE SECOND REPUBLIC, 1931–36

King Alphonso's action, and the fact that the army was no longer behind the monarch, allowed the establishment of the new regime without bloodshed. But the initial optimism did not allow for long the illusion that the nation as a whole had turned to fervent belief in responsible democracy or achieved the competence to exercise it. Right-wing elements—army, church, landed aristocracy—were prepared to tolerate the republic so long as it respected what they believed to be their prescriptive rights. Left-wing elements—Anarcho-Syndicalists and the more extreme Socialists (Communism was as yet an insignificant force in Spain)—regarded its advent as but a stage to the overthrow of the established order of society. The regional autonomists, fortified by the pact of San Sebastián, appeared to Madrid to be intent on disrupting the national unity. Moreover, the new government, attempting to steer a middle course through all these dangers, was forced by the quantity of social problems facing it to action that of necessity invaded sphere after sphere of vested interests. The dictatorship had been merely the culmination of a long denial to the mass of the people of education in political responsibility. The republic found itself heir to the consequences, and the challenge to the political stature of the nation was as searching as that to the statesmanship and administrative ability of those now come to authority.

The whole course of the republic was marked by revolutionary strikes, anticlerical outrages and incipient revolts against the regime, now of the right, now of the left, now of the regions of the periphery. In May 1931 churches were burned by the mobs in Madrid, Valencia and throughout southern Spain. The monarchist newspaper *ABC* was suspended and the primate, Pedro Cardinal Segura y Sáenz, expelled from Spain for alleged monarchical indiscretions. A Communist rebellion was crushed in Jan. 1932, a right-wing rising under Gen. José Sanjurjo in the following August. The economic life of the country was seriously jeopardized by its constant subordination to political agitation.

Parliamentary elections held in June 1931 gave the left-wing parties 315 out of the 466 seats in the new constituent *Cortes*. The Monarchists had 1 deputy. The *Cortes* met in July, and on Dec. 9 the new constitution was promulgated. This defined Spain as "a democratic republic of workers of all classes," with authority "emanating from the people" and no official religion. It bound Spain to respect for the rules of international law and to the renunciation of war as an instrument of national policy. Provision was made for the granting of autonomy to regions which could justify the demand for it, but any federation of autonomous regions was forbidden. Titles of nobility were abolished. Free expression of opinion was guaranteed, and both sexes were given the vote from the age of 23. All property and wealth were declared subordinated to the interests of the national economy; forcible expropriation of property, against adequate indemnity, might be authorized by the *Cortes*. Marriage was to be founded on sex equality, and unions were made dissoluble "as a result of mutual disagreement or on the petition of either party." Primary education would be, not merely compulsory and free, as nominally under the monarchy, but exclusively lay.

Legislative authority was assigned to a unicameral *Cortes*, with four years as the maximum length of each parliament. The president, to be chosen by an electoral college comprising all deputies and an equal number of *ad hoc* members, was to hold office for six years. His powers included the nomination of the prime minister and considerable authority over the *Cortes*. Amendment of the constitution required an absolute majority in the *Cortes* and the

lapse of at least four years from the date of promulgation.

The most hotly debated clauses related to religion. New measures in this field included the disestablishment of the church, the cessation from Dec. 1933 of state payment of the clergy (a form of compensation for past expropriations agreed upon in the concordat of 1851) and the submission to a special law of all religious confessions and congregations. Freedom of conscience and of worship was recognized, subject to government sanction for all public manifestations of religion. Religious orders requiring a vow of obedience to an authority foreign to the state were to be dissolved and their property nationalized for educative and charitable purposes. Other orders would likewise be dissolved if considered a peril to the safety of the state. None might hold more property than was necessary to its sustenance, and none could engage in industry, commerce or education. Finally, it would be "permissible" for the property of all the orders to be nationalized.

These provisions occasioned a first serious political crisis. Manuel Azaña y Díaz, who as war minister had been active in reorganizing the army and reducing its officer strength, violently attacked the religious orders in the *Cortes*, announcing as the government's intention the dissolution in Spain, under the "vow of obedience" clause, of the Society of Jesus (this took effect on Jan. 23, 1932) and the application of the other clauses with the utmost rigour. When, after an all-night sitting, the clauses were approved by a half-empty house (178 votes to 59), the prime minister, Niceto Alcalá Zamora, and the home secretary, Miguel Maura, resigned (Oct. 14, 1931). Azaña thereupon became premier, and in December Alcalá Zamora was elected first president of the republic.

A week after Azaña's accession came the notorious Law for the Defense of the Republic, under which certain constitutional guarantees might be suspended at the discretion of a single minister. By it the home secretary was empowered to cancel meetings, dissolve associations, close clubs and suppress newspapers. Eleven "acts of aggression" against the republic were named, some in the vaguest terms, and made punishable with fines or exile. The frequent and severe invocation of this law gave Spain in many ways less freedom than it had enjoyed under Primo de Rivera.

Left-Wing Government (1931–33).—Once the constitution was promulgated, Azaña recast his government, which became predominantly Socialist. The chief measures of 1932 were the Agrarian law and the Statute of Catalan Autonomy. The agrarian problem, one of the most pressing inherited by the republic, hinged on the prevalence throughout central and southern Spain of huge estates held by absentee landlords and worked by peasants at starvation wages. The new law (Sept. 15) provided for expropriation at discretion, against payment in government bonds to the value of the estate as assessed for taxation. Its administration was entrusted to a state-subsidized Institute of Agrarian Reform, representative of both owners and landworkers.

The movement for regional autonomy was strong in the Basque provinces (although not successful there until Oct. 1, 1936, after the outbreak of the Civil War); it was weaker among Galicians and Valencians, who were far less united in their claims. In Catalonia its urgency was extreme. The fall of the monarchy had been followed there by the immediate proclamation of a "Catalan state," in a manifesto envisaging a "confederation of Iberian peoples." Appeals from Madrid secured the substitution of the historic and less provocative term *generalitat* of Catalonia and resignation to the status of an autonomous region within a unitary Spain. The necessary plebiscite of the Catalan people (Aug. 2, 1931) gave a vote of 592,961 for the draft Autonomy statute and of 3,276 against. Despite considerable opposition within the *Cortes* and from various interests without, the statute, much revised, was approved on Sept. 9 by 314 votes to 24. By it the *generalitat* was empowered, among much else, to organize its own police services and be responsible for public order, to execute much state legislation and to found and maintain its own educational institutions. Catalan and Castilian were to be "co-official" languages. Though bitterly disillusioned by the mutilations suffered by the statute, Catalonia, under its presidents Francesc Macià and—on his death in 1933—Lluis Companys, became on the whole the most contented part of Spain and the least affected by strikes and revolts. The Law of Religious Confessions and Congregations, implementing the clauses of the constitution dealing with religious activities, also had a stormy passage, lasting seven months (Oct. 1932–May 1933). The proposal that teaching by the orders was to cease after 1933 made necessary alternative provision for the education of at least 350,000 children—an impossible undertaking. Already in June 1931, in Madrid alone, 50,000 children were without schools. The *Cortes* had authorized the creation of 27,000 new schools, 7,000 to be finished within a year; but efficient teachers could not be thus improvised. Misgivings were widespread and were shared by the president himself, who deferred signing the law until the expiry of the period allowed by the constitution.

During these two years, despite the government's efforts to cope with its many major problems, opinion had been running strongly against the left. Opposition was much exacerbated in Jan. 1933 by the ruthless handling of disturbances in the village of Casas Viejas. In April partial municipal elections, held in 2,500 rural districts, gave impressive Conservative majorities. In the autumn Azaña resigned; and, after the Radical leader Alejandro Lerroux had failed to form a government, the *Cortes* was dissolved and elections fixed for Nov. 19. In them the new female vote, totalling 6,500,000, could be foreseen as likely to strengthen any swing to the right.

Centre-Right Government (1933–35).—The election results showed how deep was the disillusion with the republican experiment to date. In the new *Cortes* the right held 207 seats, the centre 167 and the left 99. The centre thus held the balance of power, and—with occasional infiltrations from the right—governed for the next two years, generally under Lerroux. Changes of government were numerous (seven in 1935 alone). The most important, in Oct. 1934, admitted three members of the right and precipitated revolutions in Catalonia and Asturias. In Barcelona, on Oct. 6, the *generalitat* again proclaimed the Catalan state of the federal Spanish republic. The central government reacted strongly. After some street fighting, Companys and his fellow-ministers were arrested, the Autonomy statute was annulled in fact if not formally and Catalonia was again ruled from Madrid. The extremist revolt in Asturias (Oct. 6–13), which resulted in 1,335 killed and 2,951 wounded, was liquidated with extreme severity, reprisals being taken all over Spain. It was during this period that the Falangist (Spanish Fascist) party was founded by José Antonio Primo de Rivera, son of the former dictator. The right-wing leader in the *Cortes* was Gil Robles.

Government policy during this biennium was directed broadly to obstructing previous legislation of the left until, in Dec. 1935, revision of the constitution should become legally possible. Thus the substitution of lay for religious schools was indefinitely postponed, the clergy had part of their state stipends restored to them, the transfer of powers to Catalonia was delayed and a new Agrarian law treated the large landowners with respectful benevolence. But as the date in question drew near it became evident that this negative policy and the severity of the government's repressive measures were in their turn producing a widespread reaction. Azaña, arrested on suspicion after the Catalan revolt, tried, set free and later rearrested, became a popular hero and the centre of progressive hopes, the "strong man" who would soon return to power and carry to completion the reforms already begun. Some looked on him as a possible dictator of the left. By the end of 1935 the centre-right could hold on to power no longer. The president acceded to a dissolution, and new elections were announced for Feb. 16, 1936.

The Popular Front (Feb.–July 1936).—From the political manoeuvring preparatory to the elections Azaña emerged as the leader not only of his own party but also of the group of parties and trade unions (Republicans, Left Republican union, Socialists, Communists, Trotskyists, Syndicalists and Anarchists) that now banded together, in spite of the previous abstention of some from political action, to ensure electoral victory for the left. This popular front, so-called after a similar French combination, won by a substantial majority (left 256, centre 52, right 165), and no sooner were the results announced than tumult broke loose. Officials dis-

possessed during the preceding two years clamoured for reinstatement; prisons were burst open; bands of peasants seized land and defied the law to eject them; churches, seminaries, monasteries and the offices and houses of capitalists were set on fire; there was a wave of murders by gunmen. During the first four months of popular front rule 113 general and 218 partial strikes took place, while 170 churches, 69 clubs and the offices of 10 newspapers were fired by incendiaries, who attempted to burn a further 284 buildings, 251 of them churches.

Politically, the event of this period was the deposition, on the charge of unconstitutional dissolution of the *Cortes*, of President Alcalá Zamora. Having tried for so long to steer a middle course, the president found himself under attack by both right and left. Azaña had gone so far as to break off relations with him. Only 5 deputies voted for the president, 238 against. On May 10 Azaña was elected, almost unanimously, as his successor.

The new prime minister, Santiago Casares Quiroga, proved wholly unable to cope with the indiscipline of the country. While his government worked out fresh measures of reform, Spanish fascism grew, much as republicanism had grown in 1930–31, as a protest against the virtual usurpation of power by extremist violence. The premier flung down a challenge: "In its relations with fascism, the government is a belligerent." The challenge was taken up, and soon violence and counterviolence had relegated the activities of the *Cortes* to insignificance.

On July 13, in revenge for the murder of a Communist lieutenant in the shock troops, the former dictator's one-time finance minister, José Calvo Sotelo, was arrested and murdered by men in officers' uniform. Four days later an army mutiny broke out in Spanish Morocco, led by Gen. Francisco Franco Bahamonde, which spread next day to the mainland and within 48 hours had involved the whole country in civil war. The second republic had lasted just three times as long as the first, but it ended in much the same way and for much the same reasons.

THE CIVIL WAR, 1936–39

Everywhere garrisons revolted. In the south Seville (under Gen. Gonzalo Queipo de Llano), Córdova, Granada and Cádiz supported the rebels; in the north, all Galicia, most of León and part of Asturias, with Burgos (for three years the nationalist headquarters), Salamanca, Valladolid, Segovia, Pamplona and Saragossa. Both in Madrid and in Barcelona the mutinies failed (otherwise there might have been no war), and the government began the struggle with the whole of eastern and most of southern Spain under its control. The Basque provinces illustrated the tug of conflicting loyalties. Vizcaya and Guipúzcoa, setting autonomy above all else, threw in their lot with the regime pledged to secure this for them; Alava sided, as also did Navarre, with the insurgents, who claimed to have risen in defense of religion and traditional values.

The nationalists, as these called themselves, representing that theirs was no mere *pronunciamiento* but a *movimiento nacional*, lost several outstanding men early, notably General Sanjurjo and Gen. Emilio Mola, killed in aeroplane crashes, and Gen. Manuel Goded Llopis, shot after the unsuccessful rising in Barcelona. They had, however, the bulk of the army, of the armed police and of the arsenals. The popular front had to rely chiefly on untrained volunteers and on such arms, outside the arsenals of Madrid and Barcelona, as the unions had been storing for their own purposes. Much time was necessary to turn improvised militias into an efficient army, and by then the nationalists had greatly improved their positions.

Casares Quiroga had resigned the premiership on the outbreak of war. President Azaña, choosing a moderate, Diego Martínez Barrio, to succeed him, found his choice vetoed by the leaders of the extremist groups. The "strong man" was compelled to give way and, from then until his own resignation in Feb. 1939, was but a figurehead. The new government, thus chosen in contravention of the constitution, lost its claim to legality, and the constitution was no longer taken seriously.

By the end of July 1936 foreign intervention had begun. From the U.S.S.R., the All-Union Central Council of Trade Unions made the republicans a first payment of 12,145,000 roubles; both the U.S.S.R. and Mexico began to send war matériel; extremist volunteers (mainly Communist) from many countries entered Spain in large numbers and formed the international brigade. The nationalists received combatants, technicians and large supplies of war matériel from Italy and Germany, the fruit seemingly of prior consultation with Rome and Berlin. Realizing the danger of international complications, France approached the leading powers, and as a result formal embargoes were placed on the export of war matériel and a nonintervention committee, representing 27 nations, was set up in London. Intervention continued nonetheless throughout the war.

With the capture of Badajoz in the middle of August the nationalists were enabled to unite their northern with their southern forces; and while the former took Irún and San Sebastián early in September, the latter advanced rapidly upon Madrid. The capital was saved by the detour made to relieve the nationalist garrison imprisoned in the Alcázar of Toledo (Sept. 27). Thanks to this delay the international brigade was moved up in strength before the arrival of the nationalists on Nov. 6; and instead of submitting to a triumphal entry the Madrileños held out through a 28 months' siege.

As the threat to the capital developed, the government, now under Francisco Largo Caballero and enlarged by the inclusion of both Communists and Anarcho-Syndicalists, moved to Valencia (Nov. 7), leaving Madrid in charge of a defense council under Gen. José Miaja. President Azaña had already gone to Barcelona. There the *generalitat* was allowed to retain nominal power and even to take over services reserved by the Autonomy statute to the state; but real power had passed to the unions, which collectivized industry and commerce (Oct. 28) and were responsible for the ban on public worship that only ended with final defeat.

Apart from the taking of Málaga, little was achieved by the insurgents in the winter of 1936–37, two attempts to cut the roads to Madrid conspicuously failing. In April 1937 a new and successful offensive was begun in the north, with Bilbao as its first objective. The city's "iron ring" of defenses held out for nearly two months, but on June 19 the Basque capital fell and one of Spain's two autonomous regions was again centralized. Santander was taken on Aug. 25, and Gijón, the last town to hold out in the north, on Oct. 21.

On both sides the war was marked, especially in its early stages, by a ruthlessness which astounded the civilized world. Not only were churches burned or desecrated and public religious observances forbidden throughout republican Spain; ten bishops and many thousands of priests, religious and devout members of the laity were murdered in cold blood, for no political activity or crime. The nationalists for their part organized mass executions (*e.g.*, in Badajoz), and the Basque town of Guernica was subjected to terroristic air bombardment.

On Oct. 1, 1936, General Franco assumed the leadership of nationalist Spain and outlined the features of his proposed "broadly totalitarian" rule. These included abolition of popular suffrage and of regional autonomy; equitable and graduated taxation; complete religious tolerance and a new concordat with Rome; in foreign affairs, suppression of all "Sovietic contacts" and preferential treatment for "nations of related race, language or ideology." A technical council of state was set up, to be succeeded in Jan. 1938 by a cabinet on the orthodox pattern. On April 19, 1937, the two main nationalist groups—Falangists and Traditionalists—were fused into a single party, the Falange Española Tradicionalista, and all other parties were dissolved. The organization of the F.E.T. was developed four months later by the creation of a national council of about 50 members and a *junta política* or party cabinet invested with considerable powers.

Much constructive legislation was promulgated even during the war, such as the Labour charter of March 9, 1938, which, under captions like "Work for All," "Remuneration and Security," "Protection and Production," gave the worker an undertaking, somewhat vaguely worded, that the "new state" had not come simply to restore the *ancien régime*. In agrarian policy the objective was to give every labourer a living wage and every peasant family a

small holding and to improve sanitary and housing conditions in the villages. Little idea was given as yet of how such aims were to be accomplished. Another feature of the Labour charter was the creation of vertical syndicates combining into one organism, under state direction, all the elements functioning within a single service or branch of production. This departure earned for the new regime the adjective "national-syndicalist."

The republican cause suffered from internal differences much more fundamental than were to be found on the other side, and serious clashes occurred. A miniature civil war within the Civil War was caused by an Anarcho-Syndicalist rising that raged for a week in Barcelona (May 3–10, 1937). On this followed a major crisis in the Valencian government, leading to the formation of a new cabinet, under Juan Negrín, from which the Anarcho-Syndicalists were eliminated; they were not again represented in the government until April 1938. The *generalitat* underwent numerous crises on a smaller scale (six between Sept. 1936 and June 1937 alone), which were due also, principally, to the clash between centralizing and decentralizing groups, though, as the combat became fiercer, these drew more closely together. Large-scale plot and spy trials were frequent, some of them implicating the anomalously placed Trotskyists, whose group, known as the P.O.U.M. (Partido Obrero de Unificación Marxista) was the Cinderella of the republican family.

By the end of the first year of war, General Franco held 35 of Spain's 50 provincial capitals and 119,690 sq.mi. of territory. But the republicans had as yet scarcely organized their "new army," and the struggle was far from over. The loss of the north and the consequent unification of front simplified their practical problems; and when, on Dec. 4, 1937, they launched an unexpected offensive in Aragon, it won them the city of Teruel (Jan. 9, 1938). This was to be their last, as it was almost their first, important gain and was lost once more on Feb. 22. With the spring the nationalists regained the initiative, relieved Huesca, penetrated into Catalonia, captured Lérida (April 3) and pursued the enemy up the Segre valley to Tremp. An almost simultaneous push farther south carried them to the Mediterranean (April 15), and soon they had carved themselves a corridor 40 mi. wide, threatening Castellón to the south and Tortosa to the north. Republican Spain was thus cut in two, and its government, which had removed from Valencia to Barcelona on Oct. 31, 1937, deputed General Miaja, the defender of Madrid, to act as civil and military governor of its central and southern territory.

International Activity.—The efforts made by the London committee to prevent, or at least to lessen, foreign intervention were many and complicated. Though the entry of foreign combatants was made illegal (Feb. 21, 1937), they continued to enter, despite a coastal control system inaugurated on March 7. All the powers save Great Britain and France soon withdrew from this, and Great Britain proposed its supersession by port control and by the granting to both parties of belligerent rights as soon as the withdrawal of foreign combatants already in Spain should have made substantial progress. In fact General Franco never withdrew more than 10,000 of these during the war, and the republicans, though more amenable, had not completed withdrawal at its close. Belligerent rights were, therefore, never granted.

In the summer of 1937 intensified pronationalist submarine warfare in the Mediterranean led to a nine-power conference's being held at Nyon, Switz. This organized (Sept. 14) a new system of maritime zone patrols and agreed that any submarine attacking a non-Spanish merchant ship should itself be attacked. As the fighting moved more and more toward eastern Spain, attention became increasingly centred on the bombing of Mediterranean ports, which often resulted in great damage to neutral shipping and much loss of life among civilians. Protests were made by several powers and by the Vatican against the repeated bombing of open towns (Granollers suffered 700 casualties in one air raid, and Falset, near Gandesa, was almost completely destroyed).

The idea of peace negotiations, which the powers would have supported, was rejected throughout by General Franco, who insisted on the unconditional surrender which he eventually obtained. Signs that the republicans were envisaging defeat first came in April 1938, when Negrín formulated 13 points as his irreducible minimum. At the time of the Munich conversations (Sept. 1938) these were given great publicity, and there was talk of a more moderate republican government which might canvass for an armistice. That this was not formed may have been because of a checking of the nationalists on the Ebro (July 26–Nov. 18, 1938), which gave the republicans new hopes of an indefinite prolongation of the struggle. These proved vain. Their army was exhausted, their people half-starved, their territory cumbered by about 3,000,000 refugees from the west. When at length the nationalists were able to recross the Ebro, even the severe winter could not hold back their final sweep to victory.

The End of the War.—The war had been one of lightning offensives and wearisome delays, and it conformed to type to the end. The last great offensive began on Dec. 23. A general advance in Catalonia soon developed into an attack on Barcelona which, menaced from both northwest and southwest, fell on

BY COURTESY OF THE SPANISH EMBASSY, LONDON

VIEW OF THE OLD CITY OF BARCELONA. THE TWO LARGE TOWERS AND SPIRE ARE PART OF THE CATHEDRAL BEGUN IN 1298, COMPLETED IN THE LATE 19TH CENTURY

Jan. 26, 1939, after a campaign of only 34 days. The government fled north to Figueras, where, on Feb. 1, 62 members of the republican *Cortes* held a meeting in the castle vault and Negrín hurled defiance at the enemy, while announcing that his 13 points were now reduced to 3. Pursued northward along the three main roads to France, the republican forces soon lost the rest of Catalonia. Hundreds of thousands of refugees poured into France, able-bodied men of military age being at first sent back, though eventually admitted and interned. When Gerona fell on Feb. 4, the government moved to a village near the French frontier. Across this, a day later, passed a succession of presidents: Azaña, Companys and the Basque president José Antonio de Aguirre.

On Feb. 28 Azaña resigned, Negrín and his cabinet returning to continue the struggle from Madrid while the nationalists prepared to move south to attack the capital. With no heating in the city, only two ounces of food daily and hundreds dying of starvation, attack proved unnecessary. On March 5 all parties

except the Communists formed a new council of defense, again under General Miaja, ousting the Negrín government. A revolt against the council by Communist and other Negrín supporters broke out two days later but was suppressed after a week of heavy fighting, and the council addressed itself to preparing for surrender. On March 28, without resistance, 200,000 nationalist troops marched into the capital. By the following day the rest of republican Spain had surrendered, and Burgos radio announced that the civil war was over. It had lasted two years and 254 days and was estimated to have cost 1,000,000 lives.

GENERAL FRANCO'S GOVERNMENT

Great Britain and France had recognized General Franco's government on Feb. 27; the United States followed suit on April 1. A victory parade in Madrid on May 19 was followed within a few days by the departure from Spain of the German and Italian combatants; and during the summer the various ministries returned to the capital. These few months saw the first stages of reconstruction, to speed which General Franco "adopted" the most severely damaged towns and granted extraordinary credits for house building, the repair of public buildings and the rehabilitation of the transport system. The Falange, reorganized, outnumbered the Traditionalists in the government, and Ramón Serrano Súñer, appointed minister of the interior and later foreign minister, emerged as the second strongest figure in the administration. He was brother-in-law to the *caudillo* (leader), as General Franco was now called in his capacity as head of the Falangist party.

World War II.—With the outbreak of hostilities between Germany and the Allies in Sept. 1939, General Franco was quick to reaffirm Spain's neutrality. Six months earlier, in March, Spain had joined the anti-Comintern bloc; but a much more significant action, as it proved, was the conclusion only four days after this of a treaty of friendship and nonaggression with Portugal. This agreement, reinforced by a protocol of July 1940, seemed calculated to rule out any step capable of involving Spain in war with Great Britain's oldest ally. The Soviet-German pact came as a great shock to Spanish idealism, and the overrunning of Catholic Poland brought consternation to "the most Catholic country in Europe." Germany's attack on the U.S.S.R. in June 1941 helped to restore the perspective, and Spain dispatched its blue division of volunteers to fight by Germany's side in the Ukraine. In a speech of July 18, 1942, General Franco reiterated his stand: "Yesterday, today and tomorrow, for the countries of Europe there exists only one danger—Communism." Earlier he had promised 1,000,000 volunteers for the defense of Berlin should this ever be threatened by the Russians.

Meanwhile, Spain had used the occasion of Italy's entry into the war to alter the status of Tangier, and on Nov. 24, 1940, the administration of the international zone was formally combined with that of Spanish Morocco. England refused to recognize this unilateral action and protested against attacks on British property in Tangier; but a *modus vivendi* was reached on Feb. 26, 1941, when Great Britain, while reserving all rights, accepted the new situation meantime. The redefinition, likewise in June 1940, of Spain's own status as one of nonbelligerency and the emphasis laid on its new-found "will to empire" reflected confidence in an impending Axis victory and the desire to share in any redistribution of colonial territories. The presence of German troops at the Pyrenees after the French collapse weighed heavily thereafter on Spanish policy; but although General Franco journeyed to Hendaye, Fr., on Oct. 23, 1940, to meet Hitler and to Bordighera, It., to meet Mussolini four months later (Feb. 12, 1941), he succeeded in avoiding irrevocable commitments. The creation in Nov. 1940 of the Council of Hispanity, its concern being "all activities tending to the unification of culture, economic interests and power throughout the Hispanic world," was aimed at the recovery of prestige and influence in Spanish America. The atmosphere across the Atlantic, as country after country identified itself with the Allied cause, was, however, unsympathetic in the main toward Spain and its totalitarian leanings, Argentina being the notable exception.

By late 1942 the course of the war was giving General Franco cause for second thoughts. The dismissal of the violently germanophile Serrano Súñer from the foreign ministry in September was one sign that Spain was cautiously reverting from nonbelligerency to neutrality. The Allied landings in North Africa in November, followed by the German occupation of the whole of France, threatened Spain for a moment with involuntary involvement in hostilities, and partial mobilization was decreed. But the danger passed, and in December an Iberian bloc was formally cemented with Portugal. This, described by Gen. Francisco Gómez Jordana y Souza, now foreign minister (it was he who had signed the original nonaggression pact with Portugal in 1939), as "an instrument of peace capable of playing its part in international policy," confirmed the new direction; and with the collapse of Italy in July 1943 Spain began to urge the desirability of a compromise peace, in which it aspired to play the role of mediator. The Allies however were still not convinced that Spanish neutrality was genuine and in Feb. 1944 the United States, to encourage its stricter observance, imposed an embargo on oil shipments to Spain. This bore fruit in an agreement of May 2, Spain giving satisfaction on various outstanding Allied grievances concerning exports of wolfram to Germany, German agents in Tangier, Italian ships in Spanish ports and the blue division.

Ostracism by the United Nations.—The ending of the war in Europe in May 1945 and the political landslide to the left that swept the continent, from Great Britain to Greece, as its first consequence confronted the Spanish regime with a position of acute difficulty. Overtures concerning Spain's claim to a voice in the peace settlement and urging the need for a western entente against Soviet expansion were rudely brushed aside. The Potsdam declaration of July branded Franco's regime as unfit to associate with the United Nations, one-third of whom were Spain's daughter nations in Latin America. In August, at an international conference in Paris on Tangier, the U.S.S.R. but not Spain being represented, Spain was given a month in which to remove its troops from the city. In that same month Mexico, which had never recognized the new Spanish regime, granted hospitality and recognition to a Spanish republican government in exile. The culmination of this general ostracism was the decision of the United Nations in Dec. 1946, by 34 votes to 6, to demand the replacement of the regime by one "deriving its authority from the governed," coupled with the recommendation that all member nations should recall their ambassadors or ministers from Madrid. Apart from Portugal and Argentina, Spain had few friends left. This danger had been foreseen. Once the tide of war had clearly turned, the "totalitarian state" gave way to a new term, "organic democracy." A pseudo *Cortes* was inaugurated (March 1943), and Franco began to drop hints of a restoration of the monarchy. In Oct. 1944 a first approach to elections took place, for the renewal of the syndical organization. This last was the residual sphere of influence of the Falangist party, whose program of national syndicalism had earlier supplied the doctrinal basis for the regime. But the party's entanglement in government was coming to be felt as an embarrassment. Party control of press, radio and propaganda was transferred in August to the ministry of education. The Falangist salute was abolished by decree in the following September, and the title of *caudillo* was supposed to recede with the party into the background.

Relations With the Church.—The appointment as foreign minister, in the far-reaching cabinet reshuffle (July 1945), of Alberto Martín Artajo, former president of Catholic Action and not a party member, was likewise aimed at foreign opinion and particularly at closer relations with the Vatican. Repeal of the divorce law, restoration of religious education and of budget provision for the church, and the return to the Jesuits of their property and prerogatives had followed close on victory in the Civil War, but there had been no progress toward the new concordat promised in the program of the "new state." A *modus vivendi* reached in June 1941, to allow the filling of vacant sees, then numbering about 20 out of 55, merely revalidated the first four articles of the concordat of 1851. The church distrusted certain of the totalitarian implications of the regime and by 1945 was demanding (as in a pastoral letter issued on Aug. 28 by the archbishop of Toledo and primate

of Spain, Enrique Pla y Deniel, to mark the end of the world war) an end to the "constituent period" and a return to a constitution "in conformity with Spain's historic tradition and the degree of political education of the Spanish people."

The new concordat between the Holy See and Spain, confirming Spanish loyalty to the Church of Rome, was signed only on Aug. 27, 1953. There were ecclesiastical objections to some reforms of the examination system and curricula in Spanish schools before the Secondary Education act passed the *Cortes* at the end of Feb. 1953.

Franco and the Monarchy.—As during the war, however, a restoration of the kingship was still to Franco only the last resort. In March 1945 the pretender Don Juan, third son of Alphonso XIII (who had designated him as his successor before his death in Rome on Feb. 28, 1941), had called on Franco in a manifesto to resign, offering Spain a new constitution to be determined by popular vote, a legislative assembly, guarantees of political liberty and individual rights, recognition of regional characteristics and an amnesty. Only the mention of an amnesty provoked any response. On April 29 it was announced that all political charges against republicans were being dropped and exiles were invited to return. In October came a proclamation of full pardon for all prisoners convicted of political offenses committed during the Civil War. Coming six and a half years after the end of hostilities, the detail was eloquent of the animosity with which the regime had persecuted its political opponents. The execution on Oct. 17, 1940, of Lluis Companys, former president of Catalonia, who had been taken in France and handed over by the Vichy government, was perhaps the blackest and most impolitic stain on the record.

Political evolution did not signify any abandonment of "the fundamental principles of the state," which were declared to be sacrosanct in a charter of rights, the *Fuero de los Españoles*, submitted to the *Cortes* on July 19, 1945; and it gradually became clear that by a restoration Franco understood, not a return to the traditional monarchy, but only such a solution for the problem of the succession as would guarantee the continuity of those principles under a guise more acceptable to public opinion. This was made explicit in a Law of Succession published on April 1, 1947, by which Spain constituted itself a kingdom and the successor to Franco was declared to be that person of royal blood with the best right who fulfilled certain conditions, swore to observe the fundamental laws of the regime and was approved by a two-thirds vote of the *Cortes*. Failing such a king, a council of the kingdom, acting jointly with the government, would nominate a regent.

In Dec. 1954 Franco and Don Juan agreed on the education in Spain of the 16-year-old Infante Don Juan Carlos, eldest son of the latter, as presumptive heir to the former (then aged 62). The step provoked renewed Falangist opposition to the idea of an eventual restoration.

In Feb. 1955, Franco, who in July 1945 had announced to the world that the Falange no longer wielded any political power, made known his decision to strengthen that power. The Falangist national council was increased from 100 to 150 members, thus raising automatically the membership of the *Cortes* from 460 to 510. A major task, he added, would be to convince Spanish youth of the merits of the monarchical system.

International Relations.—The exclusion of Spain from Marshall aid and the failure of repeated approaches to the United States for loans were aggravations of the acute economic difficulties that had weighed on the regime ever since the end of the Civil War. A succession of droughts and bad harvests, together with the refusal of foreign credits, caused shortages of food, of water, gas and electricity and of raw materials and machinery for industry, all gravely impeding the nation's recovery and contributing to black-market abuses on a scale that threatened alike public morale and administrative efficiency. They did not suffice seriously to threaten the regime. Apathy, born of the devastation of the Civil War and the desire at all costs to avoid another, had become a major political factor. The army was still behind the regime, and the psychological effect of the ostracizing of Spain by the United Nations had been to bind Spaniards together in defiance of outside interference. By late 1950 the withdrawal of ambassadors was seen to have had the opposite effect to that intended, and on Nov. 4, by 38 votes to 10, with 12 abstentions, the United Nations rescinded its resolution of Dec. 1946, thus paving the way for Spain's return to the family of nations.

A week later Spain was admitted to full membership of FAO (to UNESCO not until Nov. 1952, and to the United Nations on Dec. 14, 1955). Diplomatic relations were resumed with the United States (Dec. 1950), Great Britain (Jan. 1951) and other powers; and in July 1951 the United States initiated discussions with Spain on the latter's contribution to western defense. These fructified in the bilateral agreements of Sept. 26, 1953, providing for (1) the construction and joint use of naval bases at Cádiz and Cartagena and air bases near Madrid, Saragossa and Seville; (2) economic aid to Spain, and (3) aid in the organization and equipping of Spain's defense forces. The term for (1) was of 10 years, renewable for two periods of 5 years, each government retaining the right of annulment. By March 1958 the total U.S. aid to Spain reached $811,800,000 as follows: defense aid $341,000,000; Export-Import bank loans and credits $106,600,000; Commodity Credit corporation wheat sale (1954) $20,000,000; sales of agricultural surpluses $254,800,000; grants of food surpluses to charities $89,400,000. Wheat and agricultural surplus sales were for pesetas. Rationing ended in Spain, after 16 years, on June 1, 1952.

BIBLIOGRAPHY.—B. Sánchez Alonso, *Fuentes de la historia española e hispanoamericana*, 3 vol. (Madrid, 1927–46), and *Historia de la historiografía española*, 3 vol. (Madrid, 1944–50); A. Ballesteros y Beretta, *Historia de España y su influencia en la historia universal*, 9 vol. (Barcelona, 1918–41); R. Altamira y Crevea, *Historia de España y de la civilización española*, 4 vol. to 1808, 4th ed. (Madrid, 1928–29), with continuation to 1923 by P. Zabala y Lera, 2 vol. (Barcelona, 1930), *A History of Spanish Civilization*, Eng. trans. by P. Volkov (London, 1930), and *A History of Spain*, Eng. trans. by M. Lee (New York, 1949); R. Menéndez Pidal (ed.), *Historia de España* (Madrid, 1935 *ff.*), and *The Spaniards in Their History*, Eng. trans. by W. Starkie (London, 1950); J. P. de Oliveira Martins, *A History of Iberian Civilization*, Eng. trans. by A. F. G. Bell (London, 1930); M. A. S. Hume, *The Spanish People* (London, 1901); W. C. Atkinson, *Spain, a Brief History* (London, 1934); A. Mousset, *Histoire d'Espagne* (Paris, 1947).

For pre-Roman and Roman Spain: P. Bosch Gimpera, *El Poblamiento antiguo y la formación de los pueblos de España* (Mexico, 1944), and *Etnología de la península ibérica* (Barcelona, 1932); A. Schulten and P. Bosch Gimpera, *Fontes Hispaniae antiquae* (Barcelona, 1922–40); A. Schulten, *Hispania: geografía, etnología, historia*, (Barcelona, 1920), *Tartessos* (Hamburg, 1922), and *Numantia* (Munich, 1914–31); L. Pericot, *La Prehistoria de la península ibérica* (Barcelona, 1923); H. Obermaier, *Fossil Man in Spain*, Eng. trans. by C. D. Matthews (New Haven, 1924), and, with H. Breuil, *The Cave of Altamira at Santillana del Mar*, Eng. ed. (Madrid, 1935); J. R. Mélida, *Arqueología española* (Barcelona, 1929); P. Dixon, *The Iberians of Spain* (New York, 1940); Rhys Carpenter, *The Greeks in Spain* (Bryn Mawr, 1925); C. H. V. Sutherland, *The Romans in Spain* (London, 1939).

For mediaeval Spain: A. Castro, *España en su historia: Cristianos, Moros y Judíos* (Buenos Aires, 1948), Eng. trans., *The Structure of Spanish History*, by E. L. King (Princeton, 1954); R. Menéndez Pidal, *The Cid and His Spain*, Eng. trans. by Sunderland (London, 1934); H. J. Chaytor, *A History of Aragon and Catalonia* (London, 1933); E. Lévi-Provençal, *Histoire de l'Espagne musulmane*, vol. i–ii, 2nd ed. (Paris-Leyden, 1950), and *L'Espagne musulmane au Xe siècle* (Paris, 1932); H. Mones, *La Chute du califat umayyad de Cordoue* (Cairo, 1948).

For the Spanish colonial empire: R. B. Merriman, *The Rise of the Spanish Empire in the Old World and in the New*, 4 vol. (Cambridge, Mass., 1918–34); C. Haring, *The Spanish Empire in America*, 2nd ed. (New York, 1952), and *Trade and Navigation Between Spain and the Indies in the Time of the Hapsburgs*, vol. xix of "Harvard Economic Studies" (1918); S. de Madariaga, *The Rise of the Spanish American Empire* and *The Fall of the Spanish American Empire*, 2 vol. (New York, London, 1947); Earl J. Hamilton, *American Trade and the Price Revolution in Spain*, 1501–1660, vol. xliii of "Harvard Economic Studies" (1934).

For the 16th, 17th and 18th centuries: M. A. S. Hume, *Spain, Its Greatness and Decay, 1479–1788*, 3rd ed. (Cambridge, 1925); R. Merton, *Cardinal Ximenes and the Making of Spain* (London, 1934); Conde de Cedillo, *El Cardenal Cisneros* (Madrid, 1921); R. Trevor Davies, *The Golden Century of Spain* (London, 1954); H. J. Seaver, *The Great Revolt in Castile* (Boston, 1928); F. Brandel, *La Méditerranée et le monde méditerranéen à l'époque de Philippe II* (Paris, 1949); C. J. Bratli, *Philippe II, roi d'Espagne* (Paris, 1912); Cristóbal de Castro, *Felipe III* (Madrid, 1944); G. Marañón, *El Conde-Duque de Olivares* (Madrid, 1953); A. Cánovas del Castillo, *La Decadencia de España*, 2nd ed. (Madrid, 1910); H. C. Lea, *History of the Inquisition of Spain*, 4 vol. (New York, 1906–07), and *The Moriscos of*

Spain (Philadelphia, 1901); José Amador de los Ríos, *Historia de los Judíos de España y Portugal*, 3 vol. (Madrid, 1876); M. Menéndez y Pelayo, *Historia de los heterodoxos españoles*, vol. ii (Madrid, 1880); J. Klein, *The Mesta; a Study in Spanish Economic History, 1273–1836*, vol. xxi of "Harvard Economic Studies" (1920).

For the modern period: M. A. S. Hume, *Modern Spain, 1788–1898*, 3rd ed. (London, 1923); H. B. Clarke, *Modern Spain, 1815–1898* (Cambridge, 1906); S. de Madariaga, *Spain*, 2nd ed. (London, 1942); A. Ramos Oliveira, *Politics, Economics and Men of Modern Spain, 1808–1946*, Eng. trans. by T. Hall (London, 1946); G. Brenan, *The Spanish Labyrinth*, 2nd ed. (1950); G. Maura Gamazo, *Bosquejo histórico de la Dictadura*, 2 vol. (Madrid, 1930); L. Jiménez de Asúa, *Proceso histórico de la constitución de la República española* (Madrid, 1932); H. Buckley, *Life and Death of the Spanish Republic* (London, 1940); F. Borkenau, *The Spanish Cockpit* (London, 1937); C. A. Gómez, *La Guerra de España, 1936–39*, 2 vol. (Madrid, 1939); J. Beneyto Pérez, *El Nuevo Estado español* (Madrid, 1939); J. Pemartín, *¿Qué es "lo nuevo"?* 3rd ed. (Madrid, 1940); J. M. de Areilza and F. M. Castiella, *Reivindicaciones de España*, 2nd ed. (Madrid, 1941); José Plá, *Gibraltar* (London, 1955). (W. C. An.)

TABLE I.—*Population of the Provinces of Spain*

Provinces	Area in sq.mi.	Pop. in 1930	Pop. in 1950	Density per sq.mi. in 1950
New Castile:				
Madrid	3,090	1,383,951	1,823,410	590.1
Guadalajara	4,709	203,908	208,652	44.3
Toledo	5,925	489,396	533,654	90.1
Cuenca	6,588	309,526	344,033	52.2
Ciudad Real	7,622	491,657	572,589	75.1
Old Castile:				
Burgos	5,481	355,299	390,058	71.2
Logroño	1,946	203,789	231,010	118.7
Santander	2,108	364,147	405,420	192.3
Ávila	3,144	221,386	259,534	82.5
Segovia	2,682	174,158	203,488	75.9
Soria	3,977	156,207	164,575	41.4
Palencia	3,096	207,546	236,008	76.2
Valladolid	3,155	301,571	348,185	110.4
Asturias:				
Oviedo	4,207	791,855	895,804	212.9
León:				
Salamanca	4,758	339,101	415,127	87.2
Zamora	4,082	280,148	316,493	77.5
León	5,937	441,908	551,072	92.8
Estremadura:				
Badajoz	8,358	702,418	817,703	97.8
Cáceres	7,705	449,756	548,256	71.2
Galicia:				
Corunna	3,051	767,608	971,641	318.5
Lugo	3,815	468,619	521,213	136.6
Orense	2,695	426,043	494,283	183.4
Pontevedra	1,695	568,011	714,666	421.6
Andalusia:				
Almería	3,388	341,550	361,769	106.8
Granada	4,838	643,705	793,338	164.0
Málaga	2,813	613,160	756,083	268.8
Córdova	5,300	668,862	790,242	149.1
Jaén	5,209	674,415	781,228	150.0
Cádiz	2,827	507,972	693,267	245.2
Seville	5,429	805,252	1,101,595	202.9
Huelva	3,907	354,963	369,722	94.6
Valencia:				
Castellón de la Plana	2,579	308,746	330,257	128.1
Valencia	4,239	1,042,154	1,344,365	317.1
Alicante	2,264	545,838	634,632	280.3
Murcia:				
Albacete	5,739	332,619	400,731	69.8
Murcia	4,369	645,449	755,850	173.0
Catalonia:				
Lérida	4,691	314,435	323,460	69.0
Gerona	2,264	325,551	322,371	142.4
Barcelona	3,033	1,800,638	2,215,901	730.6
Tarragona	2,426	350,668	356,864	147.1
Aragon:				
Huesca	5,849	242,958	237,681	40.6
Zaragoza (Saragossa)	6,608	535,816	609,393	92.2
Teruel	5,721	252,785	243,269	42.5
Navarre:				
Navarre	4,056	345,883	383,354	94.5
Basque Provinces:				
Vizcaya (Biscay)	836	485,205	554,302	663.0
Guipúzcoa	728	302,329	371,024	509.6
Álava	1,176	104,176	114,139	97.1
Canary Islands:				
Las Palmas	1,565	250,991	379,977	242.8
Santa Cruz de Tenerife	1,329	304,137	427,796	321.9
Balearic Islands	1,936	365,512	419,628	216.8
Total	194,945	23,563,867	28,039,112	143.8

POPULATION

The important remains found throughout the Iberian peninsula of most phases of civilization, from the later Palaeolithic Age onward, argue a population of considerable racial complexity. The peninsula, generally, is inhabited by long-headed peoples..

The Basque language (*q.v.*) is a mysterious survival in the western Pyrenees (Navarre) and in the Basque provinces to the west. Catalan, widely spoken in northeastern Spain, in Valencia and in the French Roussillon, is allied to Balearic speech and Provençal and differs from Castilian Spanish, though both are derived from Latin. Galician is akin to Portuguese.

It was usually the eastern, southern and western coasts of the peninsula that maintained contacts with ancient Mediterranean civilizations. This is one factor in the contrast in social and cultural traditions observable between Galicia and Asturias, on the one hand, and the rest of Spain. But in the main, physical considerations, such as the natural differences between pluviose and arid Spain, have contributed no less than historical phenomena to characterize and differentiate the various regions.

The population of Roman Spain, an estimated 6,000,000 to 7,000,000 in the 1st century A.D., had risen possibly to 9,000,000 by the end of the empire. On the basis of separate regional enumerations of households, the population at the end of the 16th century has been placed at some 8,000,000. Economic decay, war and pestilence probably brought about a decline in numbers during the next century. The effects of emigration to the new world were probably not so severe as contemporary writers supposed.

Modern demographic data begin with the 1787 census, which showed a population, for Spain and the "adjacent islands" (Baleares and Canaries), of 10,268,150. By 1857 this had grown to 15,454,514 and the 1897 census enumerated 18,065,635. Relatively, Spain fell behind other European countries in the 19th century, while these countries were experiencing rapid industrialization.

Population increased 26.7% from 1900 to 1930 and the ravages of the Civil War of 1936–39 did not prevent an increase (9.8%) in the decade 1930–40. The 1950 census showed a population of 28,039,112, an increase of 6.8% over 1940. By Dec. 1956 the figure was officially estimated at 29,202,785.

TABLE II.—*Towns of More Than 150,000 Inhabitants**

Town	1900	1930	1950	1957
Madrid	539,835	952,832	1,527,894	1,848,901
Barcelona	533,000	1,005,565	1,276,075	1,428,777
Valencia	213,550	320,195	503,886	550,969
Seville	148,315	228,720	374,138	425,541
Málaga	130,109	188,010	274,847	304,232
Zaragoza	90,118	173,987	244,015	282,394
Bilbao	83,306	161,987	216,417	254,072
Murcia	111,539	158,724	217,934	236,049
Córdova	58,275	103,106	160,347	181,570
Las Palmas	44,547	79,444†	153,856	180,074
Corunna	43,971	75,508†	127,618	157,477
Granada	75,900	118,179	154,589	153,715

*There were 12 towns with populations from 100,000 to 150,000.
†1931 estimate.

Religion.—Roman Catholicism is the established religion of Spain. There are 10 metropolitan and 64 suffragan sees. The primate of all Spain is the archbishop of Toledo.

With important exceptions, Spanish liberalism has been anticlerical. The influence of the church was curbed during the revolutionary period of 1868–74, but the restoration again favoured the growth of monastic and ecclesiastical institutions. Disestablishment was one of the first major objectives of the second republic in 1931. The clergy were removed from the state pay roll, religious bodies were prohibited from teaching and trading and the Jesuit

order was expelled. After the civil war of 1936–39, the new regime restored the church, substantially, to its prerepublican position; confiscated property was returned to the orders, which were authorized to acquire by expropriation additional real property; religious education was reintroduced, divorce was suppressed and the cemeteries were returned to ecclesiastical jurisdiction. On June 7, 1941, the government signed an interim agreement with the Holy See regulating appointments to the Spanish hierarchy. In consultation with the papal nuncio the executive was to nominate six persons for each episcopal vacancy. The pope would then designate three of these as a panel from which the chief of state was to fill the post. This procedure was confirmed in the new Concordat of Aug. 27, 1953. The dignitaries of the church subscribed to an oath of fealty to the state. (R. S. S.; X.)

Education.—"Every Spaniard has the right to education, whether at home or in private or public centres of instruction, at choice" (*Fuero de los Españoles*). A Primary Education bill of July 1945 envisaged the building of 100,000 primary schools additional to the 54,000 then in existence, of which 3,000 had been established since the Civil War. Church and other private interests may found schools; the church has the right to inspect the religious instruction which is compulsory in all. All instruction is free and compulsory in theory from 6 to 12 years of age (the full primary course runs to 15) and must be given in Castilian, with separation of the sexes save in sparsely populated areas. The same official textbooks are used throughout the country. Figures for 1951–52 showed 60,613 public primary "school units" (*i.e.* classes or one-teacher schools) with 1,083,057 boys and 1,035,703 girls, being 49.4% of the population of school age. Incomplete figures for pupils in private schools totalled 698,669. There were 53 teachers' training colleges. The officially admitted figure for illiteracy over 10 years of age was 14.24% in 1950. Attendance at primary schools was again declared compulsory by a decree of May 1955.

Secondary education follows lines laid down by a law of Sept. 1938. Of 234,627 pupils (girls 83,230) in 119 schools in 1951–52, more than one-half were at schools run by the religious orders, one-quarter at private lay schools (both fee-paying) and the remainder at free state schools. The course (*bachillerato*) runs from 10 to 17 years of age, terminating with a state examination conducted by the universities, to which this serves as entrance qualification. Choice of textbooks is free, subject to official approval; French is the first foreign language. Teachers in state schools must be college or university graduates, and there is again segregation of the sexes. There were 23 *institutos laborales* or technical secondary schools for working-class pupils, and 36 commercial colleges.

University education is regulated by a law of July 1943. There are 12 state universities: Barcelona, Granada, La Laguna (Canaries), Madrid, Murcia, Oviedo, Salamanca, Santiago, Seville, Valencia, Valladolid and Saragossa (these serving as centres of the country's 12 educational districts), with a total in 1951–52 of 53,434 students (women 7,955) who pay fees. Madrid (17,125 students) has seven faculties (philosophy and letters, law, political economy, science, medicine, pharmacy, veterinary science), and can alone confer the doctorate. Murcia and La Laguna, the most recent foundations, have a faculty of law, with one or two other isolated departments. Entrants must pass, in addition to the state examination mentioned above, a qualifying faculty examination. Hostels (*colegios mayores*) house a proportion of the students. The university city of Madrid, largely destroyed in the Civil War, was rebuilt and enlarged. Religious instruction is compulsory for all students, but theological training is reserved to church seminaries. There is no university autonomy, rectors and teaching staff being state-appointed.

GOVERNMENT, ADMINISTRATION AND DEFENSE

The "new state" promulgated no formal constitution but five fundamental laws of the nation: the Labour charter (March 1938), the Law of the *Cortes* (July 1942), the *Fuero de los Españoles* (a charter of rights, July 1945), the National Referendum law (Oct. 1945) and the Law of Succession (June 1947). With the last-named the formative period of the regime was held to have ended. Its article 1 defined Spain politically as "a Catholic, social and representative state hereby constituted, in accordance with tradition, a kingdom." General Franco, head of the state, would be succeeded eventually by a king, being a person of the blood royal acceptable to the council of the realm, the government and the *Cortes* or, failing such, by a regent equally acceptable.

The council of the realm, instituted by the Law of Succession, is the supreme consultative body, advising the head of the state on all matters of his exclusive competence, such as declaring war or making peace and the choice of a successor. Seven of its 14 members are ex officio, 4 elected by the *Cortes* and 3 nominated by the head of the state. The *Cortes*, established by a law of July 1942 as the "superior organism for the participation of the Spanish people in the work of the state," discusses and may propose laws, the head of the state, who with his ministers normally initiates legislation, having the power of veto. Ministers, members of the Falangist national council, civil governors, mayors of provincial capitals and heads of universities and of other learned bodies are members (*procuradores*) ex officio. The head of the state nominates as many as 50 more; a further 50 are elected to represent provincial municipalities; and syndicalist associations elect representatives of employers, technicians and workers, in equal numbers, up to a total of one-third of the whole. Elected members serve for three years. The *Cortes* functions both as a whole, to consider all major legislation, and by commissions. It has no control over the government, though members may ask questions of ministers. Laws of outstanding importance are submitted further to national referendum. The head of the state presides over the cabinet, on which the secretary-general of the Falangist party sits without portfolio.

Local Government.—The statutes of regional autonomy passed by the republic were abrogated after the Civil War (Navarre and Álava, however, having sided with the nationalists, were in 1941–42 confirmed in possession of their traditional *concierto económico*, whereby in return for a block contribution to the exchequer they enjoy rights of financial autonomy); and legislation dealing with provincial and municipal government was modified. The 50 provinces—named, except for Navarre, the 3 Basque provinces and the Balearics, after their capitals—are officially regarded as intermediary administrative units between state and municipality. The provincial council (*diputación*) is elected partly from representatives of municipalities, partly from economic, cultural and professional bodies; councillors serve for six years, half their number changing every three years. The displacement of politically disaffected *diputaciones* by administrative commissions (*comisiones gestoras*) nominated by the civil governor, a measure introduced under the dictatorship of Primo de Rivera, was invoked on occasion under Franco. Even without this, the civil governor, appointed by the government and responsible to the minister of the interior, effectively controls the life of the province. He is answerable for public order, through the civil guard and in the last resort the army, exercises delegated state authority and may suspend decisions of the council. The *mancomunidad* or interprovincial association for the integration of common interests—best exemplified in that of the four constituent provinces of Catalonia instituted in 1914 and suppressed in 1939—continues in the Canary Islands (two provinces since 1928).

A municipal law of July 1945 renewed the bases of municipal administration. In towns of 10,000 inhabitants and more the mayor (*alcalde*) is appointed by the government, in smaller communities by the civil governor of the province. (Municipalities may be either urban or rural and number in all 9,212.) The first elections since 1936 took place in Nov.–Dec. 1949, one-third of each council (*ayuntamiento*) being elected from among representatives of heads of families, the other two-thirds, in equal numbers, from those of syndical organizations and of economic, cultural and professional bodies, all Spaniards over 21 years of age having the vote. The mayor is regarded as a delegate of the central government and under its orders, whether intimated directly or through the civil governor, the council tending in consequence to fill a merely advisory function. The Institute of Local Administration founded in Madrid in 1940 had among its objects the training of local gov-

ernment officials and employees.

Judicial System and Law.—The principle of the independence of the judiciary, proclaimed by the republican constitution of 1931, remained for some time in abeyance after the Civil War, the new regime availing itself freely of legal process for political ends, notably through the tribunal of political responsibilities (1939) and the tribunal for the repression of Freemasonry and Communism (1940). Penal legislation at this stage was often retrospective, and the oath required of all judges, magistrates and fiscals bound them first to unconditional adhesion to the *caudillo*, only secondly to "the upright and impartial administration of justice." By the *Fuero de los Españoles* (1945) no Spaniard may be held for more than 72 hours without being handed over to a judicial authority or be sentenced unheard or for a crime antedating the law in question. Judges and magistrates are now irremovable save as is provided for by law.

"Pending the definitive organization of the administration of justice in the New State," the supreme court was reconstituted in Aug. 1938. The president and 15 of its 20 members became direct, the remaining 5 indirect, government appointees. Appeals are not admitted to its administrative division (*contencioso administrativo*) against government decisions. There are 16 regional high courts (*audiencias territoriales*); in each province an *audiencia provincial* serves as court of first instance for criminal offenses. Below these are the 568 district courts, which hear civil cases in first instance, and the 9,338 municipal and related courts which deal with petty offenses and minor civil issues in each municipality. The jury system exists except for military trials.

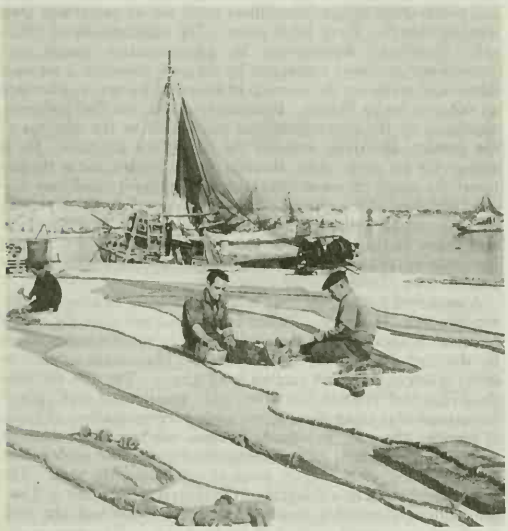
Spanish law is founded on Roman law, Gothic common law and surviving elements of mediaeval customary and local law. Modern substantive law includes the civil code of 1889, much indebted to that of France; the penal code (1870; revised 1877 and 1932); the commercial code (1885), the labour code (1926) and the codes of civil and criminal procedure (1881–82). Numerous decrees and laws of the "new state" have modified or abrogated important sections of the statute law in force in 1936.

Defense.—On the conclusion of the Civil War the army was reorganized (July 1939), the standing army consisting of eight corps in Spain, corresponding to the eight military regions (*capitanías generales*) of Madrid, Seville, Valencia, Barcelona, Saragossa, Burgos, Valladolid and Corunna (numbered in that order and known by number), two in Morocco, two general commands in the Balearics and the Canaries, an independent cavalry division and the general artillery reserve, making in all 25 divisions. A ninth military region (Granada) was constituted in Jan. 1944, of one division only. General Franco, as head of the armed forces, became chief of the supreme general staff and of the national defense council. Compulsory military service (two years) was instituted in Aug. 1940. The Falangist militia was dissolved in Dec. 1943.

The navy possesses 6 cruisers built between 1923 and 1936 (from 4,680 to 10,670 tons), 18 destroyers, 19 frigates, 6 mine layers, 6 submarines, 14 mine sweepers, 16 motor torpedo boats and various smaller craft, with, in 1954, 4 submarines and 3 anti-submarine frigates under construction. The naval dockyards are at Ferrol, Cartagena and, on a smaller scale, Cádiz. Officers and ratings, including marines, total 39,000.

An "army of the air" was created by decree of Oct. 1939, which divided Spain into five air regions, with the Balearics, the Canaries and West Africa, and Morocco as three more. There are an air branch (with special sections including parachutists) and air infantry (for nonflying duties, including aerodrome defense). Personnel are trained at an air staff college, an air force academy and various technical schools.

Of the total ordinary budget for 1954, 7,505,200,000 pesetas or 28.85% was allotted to the armed services. (W. C. An.)



BY COURTESY OF HOMER MC K. REES

FISHERMEN OF THE VILLAGE OF SAN CARLOS DE LA RÁPITA MENDING THEIR NETS. THE VILLAGE IS ON THE MEDITERRANEAN, NEAR ALICANTE

ECONOMICS

Agriculture.—Spain is essentially an agricultural country. More than 55% of the population work on the land, about 47% as casual labourers, 29% as landowners and 16% as tenant farmers. In the mid-1950s more than 50% of the national income was derived from agricultural products, which supplied the bulk of the nation's food and were the principal source of foreign currency. In general the land is arid and poor. About two-thirds of it suffers from aridity; in certain small irrigated regions around the edges of the peninsula there does exist a very productive soil, but the large central tableland (*meseta*) consists chiefly of moorland, steppe pasture or desert. The agricultural yield per hectare was thus less than that of other western European countries. In 1955, of a total land area of 50,349,000 ha., arable land covered 19,986,000 ha., permanent meadows and pastures 9,563,000 ha. and forested land 13,813,000 ha.

The progress of agriculture was not only handicapped by climatic conditions, it was also long retarded by the resistance of the landed classes to legislation for its development. The best land was for centuries in possession of the church or nobles. Cultivation, particularly in the central provinces, was sacrificed to large-scale sheep farming. During the first half of the 19th century extensive areas of common land were purchased by the middle classes. Estimates made in 1930 showed that in Estremadura, Andalusia and La Mancha about 7,000 landed proprietors (most of them absentees) owned about 6,000,000 ha. Living conditions for the tenant farmers and the landless labourer (often unemployed) were precarious. In attempting to redress this, the republican government met such opposition that little improvement was made before the outbreak of the Civil War.

Agricultural technique remained on the whole out-of-date, and mechanization (even in the 1950s) was rare. Irrigation, which is really vital to Spanish agriculture, had long been planned. Lack of capital and later the Civil War held up its development. Until the Civil War Spain did, however, manage to be self-supporting in many agricultural products and even to produce a surplus for export in some. In the immediate years following the conflict, agricultural production fell by 20% and cereal by 30%. The decrease in cereal acreage was the result of the imposition of control prices, the lack of fertilizers and machinery. Producers found more profit in industrial crops such as cotton, tobacco, esparto, vegetables and fruits (whose prices were not controlled). Some slight recovery in cereal production took place however after 1945 under the encouragement of the government. Nevertheless, whereas before 1935 Spain was almost self-sufficient in essential foodstuffs, in later years considerable quantities (particularly of

wheat) had to be imported.

Under Franco attempts to improve agriculture were begun by the establishment in Oct. 1939 of the National Institute of Land Settlement, a successor to the republic's Institute of Agrarian Reform. Subsequent legislation included steps to buy, develop and break up among small holders various large and unproductive estates; to select certain areas for irrigation and general development; to reorganize and combine existing agricultural organizations; to introduce a more helpful system of loans; and to promote some technical agricultural training.

In the 1950s, however, land was still divided very unequally (about 45% of farming land being occupied by farms of more than 100 ha. and 30% by those of less than 5 ha.), and much depended on casual labour. Most small farms were uneconomical, and the rate of agricultural unemployment was high (permanent workers were about 8%). Spain remained one of the least fertilized countries in Europe because of a lack of nitrogenous fertilizers.

Crop Distribution.—Wheat was by far the most important cereal of the country, being chiefly grown in parts of León, of Old and New Castile and of Aragon and, to the southwest, in Estremadura and Andalusia. Except in good years, yield before 1939 was less than 1 metric ton per hectare. Barley and to a lesser extent oats and rye were the next important cereals. Barley (the chief grain fodder) was the principal product of the southeast but was subordinated to other cereals elsewhere; oats and rye were grown on the poorer soils, with rye (and maize) predominating in the northwest. Rice, a staple food in Spain (the second producer in Europe), was chiefly grown in Valencia and Murcia. Other widely grown products were pulses (again a staple diet) and potatoes, whose output had after 1945 become inadequate for home consumption. (*See* Table III.) Onions occupied an area of 29,000 ha. in 1956, which produced 556,000 tons.

TABLE III.—*Agricultural Production*
(In 000 metric tons)

Crop	1931–35	1948–52	1954	1955	1956	1957
Wheat*	4,392	3,622	4,708	4,003	4,215	3,510
Barley	2,391	1,909	2,295	1,718	1,551	1,670
Oats	670	519	526	506	452	580
Rye	551	482	526	493	511	550
Maize	709	520	751	616	714	700
Rice, paddy	293	272	401	389	384	480
Potatoes	5,010	3,348	3,939	4,081	4,307	...
Pulses†	453	336	379	395	369	...
Olives	1,824	1,760‡	1,496	1,350	1,858	...
Tobacco	7.1	19.0	34.3	33.3	33.5	...

*Includes spelt. †Dry beans, dry peas, broad beans, chick peas and lentils together ‡1947–52.

Olives were one of the greatest sources of agricultural wealth; the chief producing district was Andalusia with less important areas in Catalonia, Castile and Estremadura. Around 98% of the crop was used to extract olive oil and the latter's annual average production in 1954–56 was 319,000 tons. The cultivation of vines (1,444,000 ha. or 2.9% of the total area in 1950) was widespread; table wines were produced in La Rioja district and Catalonia, dessert wines in Valencia and Málaga, sherries in the southwest and brandy in Jerez. Must wine was exported in large quantities to France for blending. Annual average wine production was 19,760,000 hl. in 1931–35 and 14,990,000 hl. in 1948–52.

Fruit growing and market gardening (13% of the national production) were second only to cereals in Spain, and an increasing acreage was devoted to this type of agriculture. Citrus fruits (particularly oranges), stone and soft fruits and vegetables formed the basis for a large dried or tinned export trade.

Esparto, hemp, flax and cotton were produced in appreciable quantities as increasing attention was paid to industrial crops. Tobacco was principally grown in Granada.

The most important industrial food crops were sugar cane and sugar beet; annual average sugar production in 1954–56 was 370,000 tons, including 35,000 tons of cane sugar.

Livestock.—Draft and working animals in the 1950s still played an important part in Spanish agricultural life, as methods of tillage were primitive, roads bad and tractors insufficient. Plowing was done by cows or oxen; mules and asses were used for transport. Horses, mules and asses were bred chiefly throughout Catalonia and the basin of the Ebro, in the Cantabrians and in Andalusia. Sheep were bred in all parts of the country and especially in the provinces of Estremadura and Old Castile. The three basic breeds were churro (long-wooled), merino (short-wooled) and manchegan. The wool of the churro (found mostly in Old Castile) is a coarse variety; the fleece of the merino (of which Spain long had a monopoly) is curly and silky and yields a fine wool.

Pigs were raised chiefly in the western provinces and in Galicia. Cattle were bred mostly in the northwest and north. Dairy farming was practically confined to this area and even there was on a small scale. (*See* Table IV.)

TABLE IV.—*Livestock*
(In 000 head)

Animal	1935	1939	1948	1950	1953	1955
Cattle	4,215	3,739	3,300	3,112	4,396	3,011
Sheep	17,526	21,779	15,921	16,344	17,233	16,312
Goats	4,692	6,692	4,222	4,135	3,754	3,428
Pigs	5,134	6,942	2,668	2,688	4,332	...
Horses	808	555	607	642	650	2,560
Mules	1,176	1,117	1,070	1,089	1,156	563
Asses	1,479	805	747	732	790	908

Fishing.—Spain has a coast line of nearly 2,000 mi.; but good natural harbours are found only in Galicia, which province, with its sheltered inlets, was the main source of the fishing industry. Principal fish were sardines, anchovies, hake, tunny and bream and shellfish and crustaceans in large quantities. From the second decade of the 20th century the catch showed a steady increase: 135,600 tons in 1917; 408,500 tons in 1934 and 749,100 tons in 1956.

While the value of the total catch amounted to 307,321,000 pesetas in 1934, that of 1956 ran to 4,615,587,000 pesetas. By 1956 the fishing fleet consisted of 46,959 vessels, totaling 390,863 tons.

Forestry.—The area occupied by woodlands in 1945 totaled 7,292,000 ha., of which 2,610,000 were under pines, 873,000 under oak, 418,000 under beech, 125,000 under poplar and 3,266,000 under miscellaneous trees. Cork (of which the Iberian peninsula has almost a monopoly) was produced from the cork oak principally in Andalusia and Catalonia, and the centre of the industry was Catalonia (production 75,000 tons in 1950).

Minerals.—Spain has, since the time of the Phoenicians, been known as a rich storehouse of mineral wealth and was, before the discovery of deposits outside Europe, a leading world producer of copper, mercury, lead and iron ore. The exploitation of other regions, together with Spain's political instability, lack of capital, etc., contributed during the 20th century toward a serious decline in the relative importance of the country's output. The number of workers in the mines was 156,850 in 1953.

Coal.—Spain being geologically composed largely of igneous rocks, carboniferous strata are relatively rare. The largest occur in the Cantabrian mountains of the north, where the Asturian coal field is the most important. It produces about 75% of Spain's coal output. In the south there are two main areas and several minor ones in the lower hills of the Sierra Morena. Coal production was greatly stimulated by World War I, reaching an annual production of more than 7,250,000 tons in 1918. After a temporary decline in the immediate postwar years, it again reached this mark in 1935, but the Civil War brought a setback in production. After the outbreak of World War II, production again increased considerably. The greater proportion of Spanish coal was of an inferior quality, and coking coal had to be imported. Spanish production of metallurgical coke, which was 768,000 tons in 1929, increased in 1956 to 1,596,000 tons.

Iron and Copper Ore and Pyrites.—Iron ore and pyrites were chiefly found in the Cantabrian provinces, particularly Vizcaya, where ores with an average iron content of 50% occur. Iron ore was also found in central and southern Spain, but its quality was not so uniformly good nor was it as easily worked as in the north. Deposits of iron pyrites with a large sulfur content were found in Huelva and Seville. Spain for many years was a leading producer of copper ore and pyrites (in which a small proportion of

copper was usually found with iron). The principal zone was in the Sierra Morena, particularly western Seville (El Castillo), and Huelva (the region between the Odiel and the Guadiana and along the Río Tinto). Extraction of iron ore (metal content) rose between 1946 and 1956 from 750,000 tons to 2,068,000 tons; that of pyrites from 791,000 tons to 2,281,000 tons. Extraction of copper ore (metal content) was in 1950–56 about 7,500 tons yearly.

Lead Ore.—In the earlier years of the 20th century Spain was second to the United States in the production of lead. The chief area was to the east of the Sierra Morena, particularly the districts of Linares, La Carolina (in the province of Jaén), with a less prolific one to the west in Córdova. A second area lay in the southwestern provinces of Murcia and Almería. Deposits also existed in the north, especially in Santander. Extraction of lead ore (metal content) rose between 1946 and 1955 from 36,800 tons to 63,400 tons (61,200 tons in 1956); that of zinc ore (metal content) from 39,400 tons to 92,300 tons (87,200 tons in 1956).

Mercury.—Spain remained one of the chief producers of mercury. Most important were the mines of Almadén in the central province of Ciudad Real (exploited since Roman times), which in the 1950s still possessed immense reserves. Production of mercury in 1950–57 averaged 1,531 tons yearly.

Other Minerals.—Other mineral ores found in Spain were manganese (in the provinces of the Asturias, Huesca, Ciudad Real and Huelva), tin ore together with tungsten (in the northwest). Extraction of manganese ore (metal content) rose between 1950 and 1955 from 7,400 tons to 16,600 tons; that of tin concentrates (metal content) remained stationary at about 972 tons yearly. Extraction of other ores reached in 1950–55 the following yearly averages: bauxite 7,500 tons; tungsten (WO_3 content) 1,217 tons; antimony (metal content) 497 tons. Extraction of potash (K_2O content) rose between 1946 and 1956 from 135,000 tons to 225,000 tons. Production of salt in 1955 was 1,172,000 tons (including 720,000 tons evaporated).

Industries.—The industrial areas of Spain comprised a few small regions within the predominantly agricultural country. Industrial development was seriously retarded by long periods of internal strife and by such handicaps as difficult communications, lack of fuel and of capital and a certain social and religious traditionalism. After World War II there had been an attempt, not too successful, to decentralize industry. Catalonia remained, however, by far the most important industrial area of the country.

Textiles.—The principal textile area was also Catalonia, where modern technique was first applied to the industry. There 90% of the cotton industry, 75% of the wool and most of the rayon became concentrated. Less important areas were Valencia, the Basque provinces and Andalusia. In 1954 there were more than 4,000 firms (50% cotton, 25% wool, 12.5% silk, etc.); they employed around 300,000 workers (two-thirds in Catalonia). Cotton represented in the 1950s more than 18% of Spanish economic activity. The cotton industry had about 90,500 looms (including 11,254 automatic) with 2,190,000 spindles in 1955. The woolen industry had in the same year 8,998 looms (including 396 automatic) with 589,000 spindles. The manufacturing of natural textiles was stationary, the average yearly production in 1951–55 being 56,500 tons of cotton yarn and 13,800 tons of wool yarn. However, production of rayon yarn rose between 1950 and 1955 from 9,940 tons to 14,080 tons and that of rayon staple from 13,530 tons to 32,530 tons.

Metallurgy.—Spain's output of minerals was only devoted in a minor degree to the domestic metallurgical industry. The principal area of this was centred on Bilbao (which produced the greater part of Spain's pig iron and steel); a second area was around La Felguera, Albana, Moreda-Gijón in the Asturias and Santander (iron and steel); a third area was in Valencia (steel); and a fourth area was in the producing regions of the Sierra Morena. Production of steel in 1954 surpassed that of the peak year 1929 to reach 1,100,000 tons, but it was estimated that when the new steelworks at Avilés, where the first blast furnace was lighted in Sept. 1957, would be in full operation the total annual output would be 2,500,000 tons.

The main centres of shipbuilding were at Bilbao, Barcelona, Gijón, Valencia, Cartagena and Cádiz; locomotives, rolling stock and machinery were manufactured at Bilbao, Barcelona, Valencia and Saragossa; armament centres were at Trubia, Oviedo, Seville and Eibar. The number of workers in the metallurgical industries in 1950 was about 100,000.

For production *see* Table V.

Chemical Industry.—Domestic mineral resources formed the basis of a diversified and growing chemical industry, though sometimes their exploitation was found uneconomical because of transport costs. Phosphates (largely imported from Morocco) were converted into superphosphates with sulfuric acid produced from pyrites. Such production was found chiefly in the provinces of Santander, Valencia, Murcia and Barcelona. Spain was also a large producer and exporter of potassic salts and by the 1950s was trying to develop the domestic exploitation of nitrogenous fertilizers so as to lessen dependence on imports from Chile. Refinery of crude petroleum (near Cartagena) reached in 1954 over 3,000,000 tons.

TABLE V.—*Industrial Production*
(In 000 metric tons; electricity in 000,000 kw.hr.)

Product	1935	1946	1954	1955	1956
Coal*	7,028	10,684	12,398	12,425	12,851
Lignite	312	1,336	1,755	1,828	1,935
Electricity	3,272	5,411	10,480	11,922	13,923
Pig iron	349	491	905	995	944
Steel	595	641	1,100	1,213	1,243
Sulfuric acid	261	288	721	799	859
Superphosphates†	973‡	287	1,414	1,436	1,519
Nitrogenous fertilizers§	...	3.7	32.6	43.5	44.0
Cement	...	1,835	3,323	3,752	3,999

*Including anthracite, extraction of which rose between 1935 and 1956 from 696,000 tons to 2,275,000 tons. †About 18% P_2O_5. ‡1929. §In terms of nitrogen.

Miscellaneous Industries.—The food industries were mainly based on the milling of cereals, the production of wines and oils and the curing of fish. The refining of olive oil and processing of wine were widespread; fish products were manufactured chiefly in the northern and northwestern coastal towns; sugar refining was found especially in Saragossa and Granada. Good timber resources were rare; but two valuable forestry products were cork and resin. The old established papermaking industry, chiefly dependent on raw imports, was centred in the Basque provinces and Catalonia. Other industries included the production of cement, china, pottery and glass, leather and tobacco. In 1956 Spain launched 93,900 gross registered tons of shipping, and produced 6,900 passenger cars and 4,700 trucks. The assembly of passenger cars reached 12,000 in 1955.

Electric Power.—The installed capacity of Spanish power stations rose between 1929 and 1956 from 877,000 kw. to 4,306,000 kw. (including 3,345,000 kw. in hydroelectric plants). The output of electric energy in 1929 was 2,433,000,000 kw.hr. (including 1,982,000,000 kw. generated in hydroelectric plants); in 1956 the output was 13,923,000,000 kw.hr. (including 11,463,000,000 from hydroelectric plants). The north of the country was the chief source of hydroelectricity, of which one-half was derived from the Pyrenean rivers and from the Ebro and its tributaries and one-quarter from the Cantabrian rivers. Plants of a much lesser capacity supplied power drawn from the rivers of the Sierra Morena and the Sierra Nevada.

Foreign Trade.—Spain's foreign trade in relation to its size and geographical position was relatively small. Trade was mostly with northwestern Europe and with the United States. Mineral ores and foodstuffs were exchanged for manufactured goods. A secondary trade was maintained with Latin America and with the Spanish colonies, in which small quantities of manufactured articles were re-exchanged for raw materials. Trade policy was closely controlled. After 1918 a high customs barrier was erected, particularly for cotton.

Before 1930 Spain's imports (worth 2,000,000,000 pesetas annually) were roughly matched by its exports. From that year until after 1946 Spanish trade suffered from abnormal conditions. The international crisis of the early 1930s was followed within three years by the Civil War, and trade sank to a quarter of what it had been before 1930. During World War II there was a pronounced

recovery, but after 1945, with the exception of 1951, the balance of trade was constantly unfavourable. (*See* Table VI.)

Main sources of Spanish imports in 1956 were: continental European Payments union countries 35%; United States and Canada 27%; United Kingdom 9%; Latin America 9%. In the same year main destinations of exports were: continental E.P.U. 40%; United Kingdom 15%; United States and Canada 15%; Latin America 13%.

From 1946 foodstuffs, mainly fruit and vegetables, accounted for about two-thirds of Spain's exports, the remaining third being equally divided between raw materials (chiefly minerals) and manufactured goods. Raw materials (mainly cotton, petroleum and

TABLE VI.—*Foreign Trade*
(In 000,000 of gold pesetas*)

Item	1935	1940	1946	1953	1954	1955	1956
Imports	875	621	900	1,825	1,879	1,890	2,347
Exports	586	394	734	1,476	1,422	1,366	1,351
Balance	−289	−227	−166	−349	−457	−524	−996

*1 gold peseta = 32.67 U.S. cents.

products and fertilizers) represented two-fifths of Spain's imports, manufactured goods accounted for the other two-fifths and foodstuffs (mainly wheat) for the remaining fifth.

Communications.—The well-developed Roman road system had been allowed to fall into decay, and in medieval times the only good paved roads were the pilgrimage ways. In the latter part of the 18th century the main (royal) routes were improved, but after the Peninsular War little repair or development took place until the 1920s. The total length of roads in 1956 was about 221,000 km., of which 120,764 km. were macadamized. Because the railway system was insufficient, roads played a constantly increasing part in public transport. Between 1913 and 1956 the number of passenger motorcars rose from 105,700 to 152,610, and that of commercial motor vehicles from 94,700 to 115,452 (including 9,599 buses). There were 2,803 regular motorcoach services which in 1956 carried over 146,152,000 passengers, as compared with 123,283,000 passengers carried by the railways.

The first Spanish railway was opened in 1848, but later construction proceeded very slowly. By 1956 the total length of lines reached 18,066 km., including 4,938 km. of small gauge; only 1,782 km. of the Spanish standard gauge (1.676 m.) had double track and only 1,006 km. were electrified. Railways of standard gauge were nationalized in 1949: they are known as RENFE (Red Nacional de Ferrocarriles Españoles).

The rivers of the country are unsuitable for inland navigation because of variations in flow and in gradient and on account of their use for irrigation. Very few navigation canals were constructed. The Guadalquivir is the most important waterway, being navigable by ocean-going vessels for 54 mi., to Seville.

Spain's merchant marine increased from 884,000 tons in 1914 to 1,286,000 in 1924; in 1956 Spain had 1,246 sea-going vessels totaling 1,438,000 gross registered tons.

Finance.—From 1808 expenditure always exceeded revenue. Rarely were estimates of revenue fulfilled, and the balance of unpaid debts carried on from year to year was considerable. National accounts worsened considerably during the Civil War 1936–39 and the cost of reconstruction afterward absorbed any increase in revenue. The gap between revenue and expenditure was narrowed after 1939 by increased taxation on income, profits and real property and in the excise duties. (*See* Table VII.)

TABLE VII.—*Budgets*
(In 000,000 current pesetas, actual)

Item	1935	1940	1946	1954	1955	1956
Revenue	4,555.1	5,378.4	10,221.3	28,837.1	31,616.0	37,165.8
Expenditure	4,557.7	5,593.7	11,257.5	27,190.0	34,879.1	44,441.0
Balance	−102.6	−215.3	−1,036.2	+1,647.1	−3,263.1	−7,275.2

The note issue of the Bank of Spain, which reached 4,800,000,000 pesetas in 1935, rose to 9,300,000,000 pesetas in July 1940. Between Dec. 1948 and Dec. 1956 it rose from 18,300,000,000 pesetas to 56,100,000,000 pesetas. Deposit money increased during the same period from 29,900,000,000 pesetas to 87,200,000,000 pesetas. Spain's national income was estimated at 107,500,000,000 pesetas in 1947 and at 310,800,000,000 pesetas in 1956.

The export exchange rates in 1957 were 42 pesetas to the U.S. dollar and 117.60 pesetas per £ sterling.

General index of the wholesale prices (1948 = 100) advanced from 22 in 1936 to 38 in 1940, 80 in 1946 and 158 in 1956. The cost of living rose over the same period (1948 = 100) from 22 in 1936 to 39 in 1940, 80 in 1946 and 140 in 1956.

Social Welfare and Labour.—In 1908 a national institute was established for granting old-age pensions and administering a scheme of social insurance; its funds came from endowments, state subsidies and donations.

Extensive schemes for social and labour reform were envisaged by Franco's administration in 1938. Compulsory social insurance was introduced in 1939. Based on contributions from employers and workers, it provided for family subsidies and national health benefits. This was extended in 1948 to old-age pensions and sickness benefits. Cost-of-living bonuses were from time to time granted to various categories of workers, and subsidies were allotted to provide employment in various agricultural areas. Unemployment in 1957 was insignificant (91,100) but pay was low. Often double employment assisted the standard of living.

Spain has a long history of trade corporations and guilds, which in the middle ages were often of a religious origin. Guilds were abolished by a royal decree of 1836, and Spain's later domestic politics were scarcely favourable to the growth of more modern unions. Under the second republic (1931–36) the unions were unfettered by supervision and became active political as well as social and economic organizations.

Under Franco a system of national or vertical syndicates (from the raw material to the marketing stage in every industry) was created for each industry and for agriculture. These syndicates came to control the whole economic activity of the country. In each district national syndicalist centres or horizontal syndicates backed by labour courts settle such problems as working hours, conditions of labour and industrial disputes. (J. D. L.; X.)

BIBLIOGRAPHY.—*Anuario Estadistico de España;* G. Brenan, *The Face of Spain* (1950); Ministerio de Información y Turismo, *España 1958.*

SPAIN, CODES OF. The codes in force in modern times cover a considerable part of the legislation of Spain. (For medieval codes, *see* FUERO; GERMANIC LAWS, EARLY.)

The chief codes include the following: the *Código civil*, comprising the law on status, rights and duties of persons (primarily Spaniards) considered as active members of society from birth to death; it went into force on May 1, 1889, and in the same year was applied to Cuba, Puerto Rico and the Philippine Islands (as in the Code Napoléon, the Roman concepts were followed, and the code in fact carried out the directions of the *Ley de Bases* of May 11, 1888); the *Código de comercio* of 1885, which went into force on Jan. 1, 1886, following the code of 1829, framed on the model of the French code of commerce with improvements taken from French jurisprudence and the previous local laws of Spain, such as the *Novisima Recopilacion* and the Ordinances of Bilbao; the *Código de enjuiciamiento civil* (civil procedure) of Feb. 3, 1881; the *Código de enjuiciamiento criminal* of Dec. 22, 1872; the *Código penal* of 1870; the *Ley Hipotecaria* (mortgage law) of 1870; and the labour code of 1926. The codes were modified by subsequent legislation.

(*See* SPAIN: *Government, Administration and Defense.*)

SPALATIN, GEORGE, the name taken by George Burkhardt (1484–1545), an important figure in the history of the Reformation. He was born on Jan. 17, 1484, at Spalt (whence he assumed the name Spalatinus), near Nuremberg, where his father was a tanner. He went to school at Nuremberg, and afterward to the University of Erfurt, where he took his bachelor's degree in 1499. There Nikolaus Marschalk, the most influential professor, made Spalatin his amanuensis and took him to the new University of Wittenberg in 1502. In 1505 Spalatin returned to Erfurt to study jurisprudence, and was welcomed by the little band of German humanists of whom Mutianus was chief. He became a teacher in the monastery at Georgenthal, and in 1508 he was ordained priest by Bishop Johann von Laasphe, who had

ordained Luther. In 1509 Mutianus recommended him to Frederick III the Wise, the elector of Saxony, who made him tutor to his nephew, the future elector, John Frederick. The elector sent him to Wittenberg in 1511 and procured for him a canon's stall in Altenburg. He managed all the elector's correspondence.

Spalatin's letters to Luther have been lost, but Luther's answers remain and are extremely interesting. Spalatin read Luther's writings to the elector, and translated for his benefit those in Latin into German. He would have dissuaded Luther again and again from publishing books or engaging in overt acts against the papacy, but when the thing was done none was so ready to translate the book or to justify the act.

On the death of Frederick the Wise in 1525 Spalatin no longer lived at the Saxon court. But he attended the imperial diets, and was the constant and valued adviser of the electors, John and John Frederick. During the later portion of his life, from 1526 onward, he was chiefly engaged in the visitation of churches and schools in electoral Saxony, reporting on the confiscation and application of ecclesiastical revenues and he was asked to undertake the same work for Albertine Saxony. He was also permanent visitor of Wittenberg university. He died on Jan. 26, 1545, at Altenburg.

His writings include *Annales reformationis*, ed. by E. S. Cyprian (1718), and "Das Leben und die Zeitgeschichte Friedrichs des Weisen," published in *Georg Spalatins Historischer Nachlass und Briefe*, ed. by C. G. Neudecker and L. Preller (1851). A complete list of them may be found in A. Seelheim's *George Spalatin als sächs, Historiographer* (1876).

See the article on Spalatin, by T. Kolde, in Herzog-Hauck, *Real-encyklopädie*, Bd. xviii (1906).

SPALDING, ALBERT (1888–1953), U.S. violinist who won an international reputation, was born at Chicago, Ill., on Aug. 15, 1888. At seven he commenced his musical studies under Ulpiano Chiti in Florence, continuing them with Juan Buitrago in New York city and Narcisse Lefort in Paris. Spalding's debut was made in 1905 in Paris, where in 1906 he appeared with Adelina Patti. His first American appearance was with the New York Symphony orchestra in 1908. During World War I he served with the U.S. aviation corps in Italy. He played at La Scala, Milan, in 1919.

His compositions include many works for the violin. He wrote an autobiography, *Rise to Follow* (1943), and a novel, *A Fiddle, a Sword and a Lady* (1953).

Spalding retired in 1950 and died on May 26, 1953, in New York city.

SPALDING, ALBERT GOODWILL (1850–1915), U.S. baseball player and sporting goods manufacturer, was born in Byron, Ill., on Sept. 2, 1850. At 17 years of age he was a leading pitcher for a Rockford baseball team. He was a pitcher and captain of the Boston baseball club (1871–75) and twice won more than 50 games a season. Spalding joined the Chicago National League team in 1876 and served as manager (1876–77), pitcher, captain, secretary and president (1882–91).

Spalding and his brother organized in the meantime a firm that eventually became the A. G. Spalding & Bros. company, and with the addition of a manufacturing branch, one of the world's largest sporting goods companies.

He served as editor of *Spalding's Official Baseball Guide* (1878–80) and in 1907 appointed a commission to establish the exact date of the origin of the modern game (*see* BASEBALL: *History of Baseball*). He died at Point Loma, Calif., on Sept. 9, 1915.

SPALDING, a market town and urban district in the Parts of Holland, Lincolnshire, Eng. Pop. (1961) 14,821. Area, 12.2 sq.mi. It is the centre of an extremely fertile fenland area, visited in spring for its flowering bulb fields; it cans locally grown fruit and vegetables and has a large beet-sugar factory and a livestock market. The Welland river passes through the town by a deep channel. The parish church (once the priory church) of SS. Mary and Nicholas dates from 1284. Ayscoughfee hall (15th century, now belonging to the town) was the house of Maurice Johnson, a founder of the Spalding Gentlemen's society (1710), probably the earliest antiquarian society in England; its museum includes birds and antiquities.

SPALLANZANI, LAZZARO (1729–1799), Italian priest and biologist, a skilled experimenter in physiology, was born at Scandiano in the province of Modena on Jan. 12, 1729. At the age of 15 he was sent to the Jesuit college at Reggio di Modena, and soon went to study law at the university of Bologna, where his kinswoman, Laura Bassi, was professor of physics; she, it is said, first interested him in science. After taking orders, he was professor at Reggio, Modena (1756) and later at Pavia, where he enriched the museum by his collections from journeys along the shores of the Mediterranean. In 1785 he visited Turkey, where he made many observations. He visited Vesuvius and Sicily in 1788 and embodied his researches in *Viaggi alle due Sicilie* (1792–97). Spallanzani studied the circulation of the blood, respiration, gastric digestion, the senses of bats, the electricity of the torpedo, the breeding of eels and the regeneration of different appendages of Amphibia, following the muscles, nerves and bones with a microscope. By filtering the semen he proved that spermatozoa were necessary for fertilization in different animals, and succeeded in artificially inseminating a bitch. Spallanzani supported the preformation theory of embryology (*see* EMBRYOLOGY, EXPERIMENTAL: *Preformation and Epigenesis*). In a controversy with J. T. Needham, he disproved (1765) the spontaneous generation of Infusoria and other microscopic organisms which did not develop in vegetable infusions that had been boiled long enough and kept in properly closed vessels. Some of his experiments on spontaneous generation were very similar to those later performed by Pasteur.

Spallanzani's great works are the *Prodromo di un'opera da imprimersi sopra le reproduzioni animali* (1768) and *Dissertazioni di fisica animale e vegetabile*, 2 vol. (1780). He died at Pavia on Feb. 11, 1799. (A. C. CE.)

SPANDAU, formerly a suburb of Berlin, Ger., now a district of Berlin. Pop. (1961) 172,663. One of the oldest places in the Altmark, it received civic rights in 1232. It became a favourite residence of the Hohenzollern electors of Brandenburg, and was fortified in 1577–83. In 1635 it surrendered to the Swedes, in 1806 to the French and in 1813 to Prussia. Spandau was the chief garrison town of Prussia and contained the imperial war treasure. It was united with greater Berlin in 1920.

SPANGENBERG, AUGUST GOTTLIEB (1704–1792), German Moravian bishop, Count Zinzendorf's successor, was born on July 15, 1704, at Klettenberg, on the south of the Harz mountains, where his father, Georg Spangenberg, was court preacher and ecclesiastical inspector of the countship of Hohenstein. Left an orphan at the age of 13, he was sent to the *Gymnasium* at Ilefeld, and from there in 1722 he passed to the University of Jena to study law. Johann Franz Buddeus received him into his family, and a stipend was procured for him.

He soon abandoned law for theology; he took his degree in 1726 and began to give free lectures on theology. He also took an active part in a religious union of students, in the support of the free schools for poor children established in the suburbs of Jena and in the training of teachers. In 1728 Count Zinzendorf visited Jena, and Spangenberg made his acquaintance; in 1730 he visited the Moravian colony at Herrnhut. A *collegium pastorale practicum* for the care of the sick and poor was in consequence founded by him at Jena, which the authorities at once broke up as a Zinzendorfian institution. But Spangenberg's relations with the Moravians were confirmed by several visits to the colony, and only the accident of an unfavourable appeal prevented his appointment as chief elder of the community. His free lectures in Jena led to an invitation from Gotthilf Francke to the post of assistant professor of theology and superintendent of schools connected with his orphanage at Halle. He started work there in Sept. 1732, but differences between the Pietists of Halle and himself soon became apparent. He found their religious life too formal, external and worldly; and they could not sanction his comparative indifference to doctrinal correctness and his incurable tendency to separatism in church life. Spangenberg's participation in private observances of the Lord's Supper and his intimate connection with Count Zinzendorf brought matters to a crisis. He was offered by the senate of the theological faculty of Halle the alternative of doing penance before God, submitting to his superiors and separating himself from Zinzendorf, or leaving the matter to the decision of the king, unless he preferred to "leave Halle quietly." The case came before the king and, on April 8, 1733, Spangenberg was con-

ducted by the military outside the gates of Halle.

At first he went to Jena, but Zinzendorf at once sought to secure him as a fellow labourer, though the count wished to obtain from him a declaration which would remove from the Pietists of Halle all blame with regard to the disruption. Spangenberg went to Herrnhut and found among the Moravians his lifework, having joined them at a moment when the stability of the society was threatened. He became its theologian, its apologist, its statesman and corrector, through 60 long years of incessant labour.

For the first 30 years (1733–62) his work was mainly devoted to the superintendence and organization of the extensive missionary enterprises of the body in Germany, England, Denmark, Holland, Surinam, Georgia and elsewhere. It was on an island off Savannah that Spangenberg startled John Wesley with his questions and profoundly influenced his future career. One special endeavour of Spangenberg in Pennsylvania was to bring over the scattered Schwenkfeldians to his faith. In 1741–42 he was in England collecting for his mission and obtaining the sanction of the archbishop of Canterbury. During the second half of this missionary period of his life he superintended as bishop the churches of Pennsylvania, defended the Moravian colonies against the Indians at the time of war between France and England, became the apologist of his body against the attacks of the Lutherans and the Pietists and did much to moderate the mystical extravagances of Zinzendorf, with which his simple, practical and healthy nature was out of sympathy. The second 30 years of his work (1762–92) were devoted to the consolidation of the German Moravian Church. Zinzendorf's death (1760) had left room and need for his labours at home. At Herrnhut there were conflicting tendencies, doctrinal and practical extravagances, and the organization of the brethren was very defective. In 1777 Spangenberg was commissioned to draw up an *idea fidei fratrum*, or compendium of the Christian faith of the United Brethren, which became the accepted declaration of the Moravian belief. As compared with Zinzendorf's own writings, this book exhibits the finer balance and greater moderation of Spangenberg's nature. In his last years Spangenberg devoted special attention to education.

Spangenberg died at Berthelsdorf, on Sept. 18, 1792. In addition to the *Idea fidei fratrum*, Spangenberg wrote, besides other apologetic books, a *Declaration über die seither gegen uns ausgegangenen Beschuldigungen sonderlich die Person unseres Ordinarius (Zinzendorf) betreffend* (1751), an *Apologetische Schlussschrift* (1752), *Leben des Grafen Zinzendorf* (1772–75), and his hymns are very well known beyond the Moravian circle. (*See* also Moravian Church.)

SPANISH AMERICAN ARCHITECTURE: *see* Ibero-American Architecture.

SPANISH AMERICAN LITERATURE: *see* Ibero-American Literature.

SPANISH-AMERICAN WAR. This war originated in the Cuban struggle for independence from Spain, which began in Feb. 1895. (*See* Cuba: *History.*) The destructive Cuba conflict was injurious to U.S. investments in the island which were estimated at $50,000,000, and almost ended U.S. trade with Cuban ports, normally valued at $100,000,000 annually. On the insurgent side the war was waged largely against property, taking the form of the destruction of growing sugarcane, sugar mills and other economic establishments. Of more importance than its effect on U.S. property interests was the appeal to American humanitarian sentiment. Under the Spanish commander, Capt. Gen. Valeriano Weyler y Nicolau (nicknamed the "Butcher"), supposedly loyal Cubans were herded into so-called "reconcentration" areas in and around the larger cities; those who remained at large were treated as enemies. In the reconcentration areas the Spanish authorities made no adequate provision for shelter, food, sanitation or medical care for the *reconcentrados*, thousands of whom died from exposure, hunger and disease. These conditions and other alleged atrocities perpetrated by the Spanish were graphically portrayed to the U.S. public by the more sensational newspapers, notably Joseph Pulitzer's *New York World* and William Randolph Hearst's recently founded *New York Journal*. Humanitarian concern for the suffering Cubans was added to the traditional American sympathy for a colonial people struggling for independence. While these aspects of the war created a widespread popular demand for action to halt it, the U.S. was continually inconvenienced by the necessity of patrolling the coastal waters to prevent gunrunning to the insurgents, and by demands for aid from Cubans who had acquired U.S. citizenship and then were arrested by the Spanish authorities for participating in the rebellion.

The popular demand for intervention to stop the war and assure Cuban independence gained support in congress. In the spring of 1896 both the senate and the house of representatives declared by concurrent resolution that belligerent rights should be accorded the insurgents. This expression of congressional opinion was ignored by Pres. Grover Cleveland, who firmly opposed intervention, as did his successor, William McKinley, during his first year in office. In the fall of 1897, indeed, prospects for a peaceful settlement in Cuba brightened. In Spain the Sagasta ministry, taking office upon the assassination of Premier Cánovas, announced a new policy for Cuba: Captain General Weyler was to be recalled; his reconcentration camps were to be abandoned; and a considerable degree of autonomy was to be granted the Cubans.

These concessions came too late. The insurgent leaders would now settle for nothing short of complete independence. The war went on in Cuba, and a series of incidents brought the United States to the brink of intervention. Riots in Havana in December led to the sending of the battleship "Maine" to that port, as a precaution for the safety of U.S. citizens and property. On Feb. 9, 1898, the *New York Journal* printed a private letter from the Spanish minister in Washington, Dupuy de Lôme (filched from his correspondent in Cuba), describing President McKinley as "weak and a popularity-hunter." De Lôme immediately resigned and the Spanish government tendered an apology. The sensation caused by this incident paled beside that of six days later. On the night of Feb. 15 a mighty explosion sank the "Maine" at her Havana anchorage, carrying more than 260 seamen to their deaths. Responsibility for the disaster was never determined. A U.S. naval board found convincing evidence that an initial explosion outside the hull (presumably from a mine or torpedo) had in turn touched off the battleship's forward magazine. The Spanish government offered to submit the question of its responsibility to arbitration, but the American public, prompted by the *New York Journal* and other sensational papers, held Spain unquestionably responsible. "Remember the 'Maine,' to hell with Spain!" became a popular watchword.

President McKinley, still desirous of peace, now found himself under intense pressure for intervention; the pressure came from members of his own party, the Republican majority in congress, as well as from the Democratic minority. U.S. business interests, in general, opposed intervention and war. But such opposition diminished after a speech in the senate, March 17, by Sen. Redfield Proctor of Vermont, who had just returned from a tour of Cuba. In matter-of-fact and unsensational language, Proctor described his observations in the war-torn island, the suffering and death in the reconcentration areas, the devastation elsewhere, the evident inability of the Spanish to crush the rebellion. His speech, as the *Wall Street Journal* remarked on March 19, "converted a great many people in Wall Street."

Yielding to the pressure, the president on March 27, 1898, sent what amounted to an ultimatum to Madrid. Let Spain, he asked, abandon reconcentration, declare an armistice and accept U.S. mediation in peace negotiations with the insurgents.

The Spanish government was caught upon the horns of a cruel dilemma. It had not readied its army or navy for war with the United States, nor had it warned the Spanish public of the necessity of relinquishing Cuba. War meant certain disaster. The surrender of Cuba might mean the overthrow of the government, perhaps of the dynasty. The ministry clutched at the only straws in sight. On the one hand, it sought support from the chief European governments. Aside from the British, these governments were sympathetic to Spain but were unwilling to give it more than weak verbal support. Their six ambassadors in Washington (including the British ambassador) called upon McKinley on April 6 and begged him in the name of humanity to refrain from

armed intervention in Cuba. McKinley replied urbanely that if armed intervention should come, it would be in the interest of humanity. Meanwhile, Spain was going far in the acceptance of McKinley's terms of March 27—so far that the U.S. minister in Madrid, Stewart L. Woodford, assured his government that, granted a little time and patience, Spain could work out a solution acceptable to both the United States and the Cuban insurgents. Spain would end the reconcentration policy in fact as well as in name. Instead of accepting U.S. mediation, Spain would seek the pacification of the island through the Cuban Cortes, about to be elected under the autonomy program. Spain at first stated that an armistice would be granted only on application from the insurgents, but on April 9 announced one on its own initiative.

By this time, however, President McKinley had yielded to the war party in congress. Brushing aside the Spanish concessions as unsatisfactory, he declared, in a special message to congress on April 11, that "the war in Cuba must stop." From congress he asked authority to use the armed forces of the United States "to secure a full and final termination of hostilities between the Government of Spain and the people of Cuba." Congress responded emphatically. In resolutions of April 20 it declared that "the people of Cuba are, and of right ought to be, free and independent," demanded that Spain at once relinquish authority over Cuba and withdraw its armed forces from the island, and authorized the president to use the army and navy of the United States to enforce that demand. A fourth resolution renounced for the United States any idea of annexing Cuba. As it was proposed by Sen. Henry M. Teller of Colorado, it came to be known as the Teller amendment. Upon being informed of the signing of the resolutions, the Spanish government at once severed diplomatic relations and on April 24 declared war upon the United States. Congress declared war April 25 and made the declaration retroactive to April 21.

The war thus begun was pathetically one-sided. Spain, as noted above, was in no sense prepared for war with a formidable power. Neither was the U.S. army but the outcome of the war was dependent on sea power, and in this element the United States completely outclassed its opponent. Spain had nothing to match the four new battleships, "Indiana," "Iowa," "Massachusetts" and "Oregon," which formed the backbone of the North Atlantic squadron (the "Oregon" after her notable voyage from the Pacific coast). Even more superior to their antiquated antagonists at Manila were the protected cruisers of Commodore George Dewey's Asiatic squadron. Thanks largely to the energy and enthusiasm of the young assistant secretary of the Navy, Theodore Roosevelt, the U.S. ships had engaged in battle maneuvers and target-practice and were well supplied with fuel and ammunition. Officers and men were confident and aggressive, whereas their Spanish opponents knew that they were doomed to defeat.

The first blow fell in Manila bay, May 1, 1898. Commodore Dewey, picked by Roosevelt for the command, led his squadron into the bay before dawn and in a leisurely morning engagement destroyed the anchored Spanish ships by gunfire. American casualties amounted only to seven men slightly wounded. Dewey remained in control of the bay while a military force was sent out to assist him in taking possession of the city of Manila. By the end of July some 11,000 U.S. troops under Maj. Gen. Wesley Merritt had arrived in the Philippines, and on Aug. 13 they occupied Manila.

Meanwhile, attention had centred on Cuba. Upon the declaration of war a Spanish fleet of four armoured cruisers and three destroyers, commanded by Adm. Pascual Cervera, had steamed westward from the Cape Verde Islands. Its whereabouts remained unknown until late in May, when it was located in Santiago harbour on the south coast of Cuba. The North Atlantic squadron under Rear Adm. William T. Sampson and the so-called Flying squadron under Commodore Winfield Scott Schley thereupon blockaded the harbour entrance. An army of regulars and volunteers (including Theodore Roosevelt's regiment of "Rough Riders," minus their horses) embarked at Tampa and landed on the Cuban coast east of Santiago. The American objective now was to catch Cervera between army and navy, thus forcing him either to surrender or to come out and fight. On July 1, in the hard fought battles of El Caney and San Juan hill, U.S. troops penetrated the outer defenses of Santiago. Their hold here was so precarious, and the incidence of malaria and other diseases was so widespread, that their commander, Maj. Gen. William R. Shafter, considered withdrawing to await reinforcements. This idea was abandoned when, on July 3, Admiral Cervera, under orders from Havana, led his squadron out of Santiago harbour and tried to escape westward along the coast. In the ensuing battle all of Cervera's ships, under heavy fire from the U.S. fleet, were beached in a burning or sinking condition. American losses were insignificant. Two weeks later the city of Santiago surrendered to General Shafter.

The war was now over for all practical purposes, and on July 18 the Spanish government requested the good offices of France in arranging a termination of hostilities. Before the fighting ended,

PHOTOGRAPH, CULVER SERVICE

THE ROUGH RIDERS ON SAN JUAN HILL AFTER ITS CAPTURE. THEODORE ROOSEVELT, LIEUTENANT-COLONEL OF THE REGIMENT, IS SHOWN IN THE CENTRE

however, another American expeditionary force, commanded by Gen. Nelson A. Miles, occupied Puerto Rico. Armistice negotiations conducted in Washington ended with the signing of a protocol, Aug. 12, 1898. Besides ending hostilities, this agreement pledged Spain to surrender all authority over Cuba and to cede Puerto Rico and an unnamed island in the Ladrones (Marianas) to the United States. In the Philippines, Spain consented that the United States should occupy the city and harbour of Manila until the conclusion of a peace treaty that would determine the final disposition of the islands. Peace commissioners were to meet in Paris not later than Oct. 1.

The great question now confronting McKinley and his advisers was what, if anything, to demand of Spain in the Philippines. It seems certain that McKinley, in proposing intervention in Cuba, had had no thought of acquiring an empire on the other side of the globe; nor is there any reason to suppose that many members of congress or of the public at large contemplated such an outcome of the war. Dewey's dramatic victory at Manila, however, had called sudden attention to a spot of great potential strategic importance. Theodore Roosevelt and his friend, Sen. Henry Cabot Lodge, devotees of the sea power doctrines of Capt. A. T. Mahan,

saw in Manila bay a base that might greatly enhance the sea power of the United States. Recent European aggressions in China seemed to many businessmen to threaten the American market. Manila appealed to them as a base from which U.S. interests in China could be defended. The Protestant churches saw the easy victory at Manila as a divine summons to missionary work in the Philippines. The British and Japanese governments, furthermore, let it be known that they would be pleased to see the United States keep the islands. There were difficulties about any other course. Restoration of Spanish rule promised only chaos similar to that from which Cuba had just been rescued. Nor did the Filipino people have the education, training or experience needed for successful self-government, though an active group headed by Emilio Aguinaldo had actually proclaimed independence.

Swayed by these varied considerations and by his appraisal of popular sentiment, McKinley decided, after long deliberation, that the United States must take possession of the entire archipelago of about 7,000 islands and 7,000,000 inhabitants. This demand was reluctantly agreed to by Spain, with the stipulation that the United States should pay Spain $20,000,000, nominally for public buildings and public works in the Philippines. The treaty, signed Dec. 10, 1898, conformed to these terms. Spain relinquished Cuba and ceded to the United States the Philippines, Puerto Rico and Guam in the Ladrones. The treaty was vigorously opposed in the U.S. senate as inaugurating a policy of "imperialism" in the Philippines and was approved on Feb. 6, 1899, with a margin of a single vote. Two days earlier hostilities had begun at Manila between U.S. troops and Aguinaldo's insurgents. For over three years the Filipinos carried on guerrilla warfare against U.S. rule. (*See* PHILIPPINES, REPUBLIC OF THE: *History.*)

The Spanish-American War, short as it was, and relatively inexpensive in both resources and human life, was an important turning point in the history of both antagonists. Though disastrous for Spain in immediate results, it was followed by a remarkable renaissance in Spanish life, both intellectual and material. As Salvador de Madariaga has written: "Spain felt then that the era of overseas adventures had gone, and that henceforth her future was at home. Her eyes, which for centuries had wandered to the ends of the world, were at last turned on her own home estate." (Salvador de Madariaga, *Spain: a Modern History,* Frederick A. Praeger, Inc., New York, 1958.) There ensued two decades of significant progress in agriculture, development of mineral resources, industry and transportation. At the same time there came to the front a brilliant group of thinkers and writers, the "Generation of 1898," who gave Spain a prominence in the intellectual and literary life of Europe that she had not enjoyed for centuries.

For the victorious United States the consequences were entirely different. It is a truism that the United States became a world power as a result of the war. It now had insular possessions in the Caribbean and stretching across the Pacific, including Hawaii, whose peaceful annexation had been hastened by the war. The war made certain that an American canal would be cut through the isthmus of Panama. It stimulated enthusiasm for the navy, which soon grew from fifth or sixth to second place among the world's war fleets. It prompted drastic reform in the U.S. army (*see* STAFF, MILITARY) and it promoted the career of the nation's first world-minded president, Theodore Roosevelt. Within a few years the U.S. had made the Caribbean an American lake, was taking a leading part in the politics of the far east and was preparing in spite of itself to play a determining role in the affairs of Europe.

See also references under "Spanish-American War" in the Index.

BIBLIOGRAPHY.—Orestes Ferrara, *The Last Spanish War: Revelations in Diplomacy* (1937); F. B. Freidel, *The Splendid Little War* (1958); Margaret Leech, *In the Days of McKinley* (1959); Salvador de Madariaga, *Spain: a Modern History* (1958); Walter Millis, *The Martial Spirit* (1931); J. W. Pratt, *Expansionists of 1898* (1936); H. F. Pringle, *Theodore Roosevelt: a Biography* (1931). (J. W. PR.)

SPANISH BROOM, a leguminous shrub. *Spartium junceum,* sometimes treelike, with leafless-or nearly leafless green, rushlike branches and showy yellow, fragrant flowers. It is a native of the Mediterranean region and naturalized elsewhere.

Spanish broom is cultivated as an ornamental and for the fibre in its stems. This fibre has been used since ancient times for cordage, textiles, nets, etc.

All parts of the plant contain the alkaloid spartiin and have tonic, diuretic and emetic properties. (J. W. TR.)

SPANISH GUINEA, a former overseas Spanish territory of West Africa, created into two metropolitan provinces of Spain on April 1, 1960, and comprising: (1) Río Muni (continental Guinea) with an area of 10,045 sq.mi., including the estuary islets of Corisco (6 sq.mi.), Elobey Chico (0.1 sq.mi.) and Elobey Grande (0.9 sq.mi.); and (2) the insular province of Fernando Póo (779 sq.mi.) and Annobón (6.6 sq.mi.), comprising two lofty and volcanic islands. The respective provincial capitals are Bata and Santa Isabel. They are Spain's only wet-tropical provinces.

Río Muni.—This mainland province, with a population (1960) of 183,377 including 3,496 white, lies on the west-central coast of Africa between the republics of Gabon and Cameroun. The present territory is but a fragment remaining from the 18th-century Spanish claims to the coast between the Ogooué and Niger rivers. Apart from sedimentary rocks of the coastal plain, almost all of the territory consists of the basic rocks of Africa, ancient metamorphic rocks such as granito-gneiss, diorites and gabbros. These have undergone very long leaching, decomposition and erosion, so that the soils to which they give rise, are mediocre. Most of the country lies within the Río Benito basin, with the mouths of the Río Campo and Río Muni forming the seaward ends of the northern and southern boundaries. Other boundaries are straight lines east-west and north-south. The highest parts are in the centre and east, where several plateaus reach about 4,000 ft. The coastal hills are the edge of these plateaus. The climate is typically equatorial on the coast, with greater daily variation on the plateaus. The wet seasons are from February to June and from September to December. Forests are mainly on the coastal plain and hills.

The Economy.—There are about 30 timber concessions near the Campo, Benito and Muni, which rivers are used for floating and export. The main species cut are Okume (*Okumea klaineana*), mahoganies (especially *Khaya kleinei*), African walnut (*Lovoa trichilioides*) and Abura or Uwen (*Mitragyna ciliata*). The cut areas south and east of Bata are sometimes replanted with oil palms or with sisal. Cocoa and coffee, the most remunerative crops, are mostly grown by Africans in the northeast, although there are some large European-owned coffee plantations in the coastal hills. Cocoa exports amount to about 2,000 metric tons and coffee to 3,800 tons annually.

Though there is a considerable tidal range, surf is absent from the coast. Consequently, offshore handling of ships at Bata (which has a small pier), Río Benito and Puerto Iradier (in the Río Muni estuary) is easy. Fishing is also important, and is largely in the hands of Annobonese and Canary islanders. Roads join the coastal towns with those of the interior, such as Evinayong (south-centre), Niefang (north-centre), Ebebiyin (northeast) and Nsoc (southeast). The three rivers are navigable for some distance from the sea. There are regular air services from Bata to Fernando Póo, Douala (in Cameroun) and Spain, and shipping services from Bata and Puerto Iradier. Real development began only after the Spanish civil war and World War II, at a time when Spain had little money to invest here and little or no international aid. Available resources had previously been largely concentrated on the far more fertile and healthier Fernando Póo.

Fernando Póo (or Po).—Southwest trending faults extend from near Lake Chad to the Bight of Biafra and far out into the Atlantic ocean. Along these faults, Tertiary and Quaternary volcanism threw up the islands of Fernando Póo, Príncipe, São Tomé and Annobón. All have an equatorial type of climate, locally modified by relief which lowers temperatures and causes greater daily variation at higher altitudes, as well as higher rainfall and relative humidity. The lowlands are hot and humid, the uplands cooler though humid. Average annual rainfall is about 100 in., mostly between May and October.

Fernando Póo is the largest island and about 20 mi. from the mainland in the Bight of Biafra. It has a typical northeast to

southwest orientation, and is about 25 mi. wide by 34 mi. long. It rises sharply and impressively from the sea, the highest point being Santa Isabel peak (9,842 ft.). This and other peaks are extinct volcanic cones and there are beautiful crater lakes, the port of Santa Isabel being in a crater breached by the sea.

There are indications of a west-northwest to east-southeast fault across the narrowest part of the island, which may explain the bays of San Carlos in the west and Concepción in the east, as well as the rugged mountain range of the south. This ends abruptly on the southern coast and, helped there by a monsoonal rainfall of some 400 in. per annum, these mountains are deeply trenched by torrents. This sector is inaccessible and virtually unused, although the hydroelectric power potential is high.

The Economy.—The east, north and west coast lands are narrow and almost completely occupied by European-owned plantations producing cocoa up to 2,000 ft., although some estates grow coffee up to 3,000 ft. African (Bubi) farms raise subsistence crops. The volcanic soils are rich and highly productive, but absentee-landlordism and immigrant labour give rise to social and economic problems. Cocoa exports average 20,000 metric tons annually, depending upon the availability of labour, and coffee exports are about 1,600 tons. Costs of production are higher than in Ghana, but there are preferential duties for Spanish Guinea cocoa and coffee in Spain.

Above the plantations is tall and canopied rain forest with many mosses, tree-ferns, epiphytes and lichens because of the high relative humidity and mist. Between about 4,000 ft. and 5,300 ft. pasture lands were won by forest clearance, though giant tree ferns and lichens survive in valleys of torrential streams. A Spanish ranching company keeps Galicia, Swiss and Canary Island cattle around Moka in the south-centre, and the island is self-sufficient in dairy produce although some meat must be imported. Temperate vegetables are grown at Moka in an African peasant cooperative enterprise. On steep slopes above the Moka pastures is a narrow belt of montane woodland, and from 5,500 ft. upward bracken and the beautiful *Lobelia columnaris*, reaching 7 ft., are found. Above this zone lie poorer pastures and moorland. The island has sea links with Calabar, Douala and Bata and air services to the latter two towns.

Population.—The truly indigenous population numbered 20,975 in 1950. Most are more or less pure Bubi of Bantu stock, the descendants of immigrants from the mainland. Originally, they were finely developed people and their ancient centre was Moka. After contact with the Portuguese and Spanish they degenerated and diminished rapidly because of alcoholism, the wider spread of disease, and despair at the loss of their land. They have recovered somewhat and are increasing slowly, but are unwilling and too few to work on the European plantations. Their houses are rectangular, with fireplaces and thatching down exterior walls, both devices to counter the high humidity.

The other indigenous people are the Porto (or "mixed" people) who live on the coast and are engaged in professional, clerical and commercial pursuits. They are descended from slaves freed from slave ships captured by the British, from former slaves from Sierra Leone and Jamaica (these speak Creole English), from Cubans who remained from a Cuban penal settlement established in 1879 and from intermarriage of these with Portuguese, Spanish and Bubi. Porto is a Bubi corruption of "Portuguese." Plantation labour is mostly provided by immigrants (17,066 in 1950), chiefly Ibo from eastern Nigeria, of whom 90% are men, and this sexual unbalance leads to local problems. There were also 2,434 non-Africans in 1950, largely officials, planters or plantation managers, and Catholic missionaries.

The provincial capital, Santa Isabel, stands on a plateau about 200 ft. above the deep crater harbour, and is a spacious town with a dignified cathedral. In the graveyard are buried Richard Lander and other explorers. Its population in 1950 was 11,098, out of 40,475 for the whole island; non-Africans were 1,572. The other towns are San Carlos on the west coast and Concepción on the east.

Annobón.—This is the smallest of the volcanic isles and the farthest to the southwest, 400 mi. from Fernando Póo. It rises sharply to an extinct volcano of 2,438 ft. There is also a large crater lake and much of the island is rugged and useless. The population (1950) of 1,403 presses on the land, and about 400 Annobonese work on Fernando Póo or in Río Muni, mostly as fishermen. Fishing is likewise important off Annobón but the main income is from the export of palm kernels, copra, cocoa and coffee.

History.—The island of Fernando Póo was discovered by Fernão do Pó, probably in 1472. At first it was called Formosa ("beautiful") but received its present name within the next 80 years. Annobón was probably discovered by Ruy de Sequeira on a New Year's day (hence the name) between 1472 and 1475, most likely that of 1474. By the treaty of Tordesillas, the Portuguese had exclusive rights in Africa, and it was not until 1778 that they agreed to cede to Spain the islands of Annobón and Fernando Póo as well as rights on the mainland coast between the Ogooué and Niger rivers. These cessions were designed to give Spain her own source of slaves in Africa for transport to Spanish America, where, in exchange, the Spanish confirmed the rights of the Portuguese west of the 50° W. meridian in what is now Brazil. The Spanish were soon decimated by yellow fever on Fernando Póo and they withdrew in 1781. No occupation was made on the mainland.

After the British abolition of the slave trade in 1807, bases were required by the Royal Navy for the effective suppression of the trade. Fernando Póo was unoccupied and lay in a strategic situation from which the Niger mouths and the Slave Coast could be watched for slavers. In 1827 the Spanish leased bases for this purpose to the British at Port Clarence (now Santa Isabel), a fine deep-water harbour on the north coast, and in San Carlos bay on the west coast. In the absence of the Spanish, the British also became responsible for administering the island. Thereafter, the British landed many freed slaves, in default of knowing their origin or of being able to repatriate them. Freed slaves also came to the island from Sierra Leone and Jamaica, and the descendants of these several groups continue to speak a form of English. Because of the existence of these freed slaves and the absence of any Spanish administration, the United Kingdom made several unsuccessful offers to Spain for the purchase of Fernando Póo; *e.g.*, from 1839 to 1841. In 1843 the Royal Navy was able to concentrate its antislavery patrol at Freetown, Sierra Leone, and its buildings on Fernando Póo were sold to a Baptist mission.

In 1844 the Spanish made a second effort at effective occupation of Fernando Póo and their first exploration of the mainland. The latter was carried out in the two decades ending in 1877. Meanwhile, the Spanish expelled the British Baptists from Fernando Póo in 1858, whilst by the use of it as a penal settlement for Cubans from 1879 onward they added to the diverse origins of the present peoples. Following the Spanish-American War of 1898, Spanish Guinea remained as Spain's last significant tropical colony. Profiting from the weakness of Spain, France was able to confine mainland Spanish Guinea to its present limited extent. Economic development started only at that time, and was concentrated on the richer and healthier Fernando Póo. The mainland has received significant attention only since the Spanish civil war of 1936–39.

In 1959 the status of Spanish Guinea was changed and the region was reorganized into two provinces of overseas Spain, each under a civil governor. (*See* introduction above.) The citizens, including Africans, were granted the same rights as the citizens of Spain.

BIBLIOGRAPHY.—R. J. Harrison Church, *West Africa*, pp. 519–525 (1960); *Atlas Histórico y Geográfico de Africa Española* (1955); J. Nosti, *Notas Geográficas, Físicas y Económicas sobre los Territorios Españoles del Golfo de Guinea*, Instituto de Estudios Africanos (1947); L. B. Corella, *Guinea*, Instituto de Estudios Africanos (1957).

(R. J. H. C.)

SPANISH LANGUAGE. Among the modern languages descended from Latin, Spanish shows the greatest vitality and promise of future importance over a vast territory. In the Christian north of the Iberian peninsula, the earliest Hispanic vernacular texts (glosses) go back to the 10th century, the first works of fine literature, displaying strong dialectal colouring, to about 1150.

By 1270 Alfonso the Learned created standard Old Spanish. The late 15th century witnessed the introduction of preclassical Spanish. Classical Spanish, the vehicle of the Golden Age literature (*c.* 1530–1680), was followed by Modern Spanish.

Spanish is also known (paradoxically, more so in Latin America than in Spain proper) under the name of Castilian. Historically, this is a reminder of the fact that at the outset it merely represented a northern variety of Hispano-Latin folk speech. It emerged after A.D. 800 in Old Castile (around Burgos), then by the 11th century spread fanwise southward to New Castile (Toledo, Madrid) in the wake of Christian "reconquest" and, especially after the merger of Castile with León and in the late 15th century with Aragon, became the official language of all Spain. In the peninsula, Castilian has gradually crowded out, except in secluded rural districts, its erstwhile rivals, the Hispano-Roman dialects of León, Asturias, Santander, Navarre, and Aragon.

In the south, where the Moorish occupation lasted generally until the early 13th century (much longer around Málaga and Granada), Western Arabic dialects were used among Moslems, while a strikingly archaic Hispano-Roman dialect ("Mozarabic") survived until 1100 amid the dwindling Christian minority. Modern southern dialects (Extremeño, Andalusian, Murcian) represent prongs of Castilian overlying dimly recognizable earlier speech forms.

The language of the 500,000 Spanish Jews (Balkan peninsula, north Africa, Israel, New York city, Buenos Aires) is based squarely on 15th-century Spanish, containing heterogeneous admixtures. Portuguese, in the extreme west, and Galician, confined to the northwest, jointly form a language apart, with close grammatical and lexical, but not phonic, ties to Spanish. In the northeast, Catalan, including Balearic and in strongly diluted form Valencian, constitutes a separate Romance language, influenced by Provençal. Basque, a relic of a pre-Indo-European family, straddles the western Pyrenees.

Outside the peninsula, Spanish is spoken or understood in virtually all South America except Brazil, in most of continental and insular Central America, in the Canary Islands, in the Tangier zone and adjoining sections of Morocco, along stretches of equatorial Africa's Atlantic coast and, on a decreasing scale, in the Philippine Islands. Its hold on the new world is secure: if it has yielded ground to Portuguese in southern Brazil and is retreating before English in California, Arizona, New Mexico, Colorado and Texas, it is steadily gaining ground, by compensation, in areas long dominated by indigenous Indian languages (Paraguay and Mexico's Yucatán peninsula are characteristic bilingual zones) and is eagerly accepted by immigrants' families in Buenos Aires and Montevideo. The influx of Cubans and, later, Puerto Ricans has made Spanish the second-ranking language in the New York metropolitan area.

Phonology and Orthography.—The sound system is characterized by five sharply delineated, short to medium-length vowels, lacking relevant variants (*i, e, a, o, u*). Two important rising diphthongs (*ie, ue;* also, only after *c: ua*), typically confined to the stressed syllable, outbalance several infrequent falling diphthongs (*ai, au, ei, eu, oi, ui*). Consonants, except for *n* in special cases (*ennoblecer, perenne*), are not doubled: *abad, Meca, ilustrar; rr* intervocalically contrasts with *r* (*perro* ~ *pero*), forming with *r* word-initially (*rey*) and after *n, s* (*Enrique, Israel*) a separate sound normally produced by five vibrations of the tongue. Voiceless stops (*p, t, k;* the last spelled mostly *c* or *qu*, according to environment) are uttered virtually alike in all positions. The corresponding voiced stops *b, d, g* are reduced to spirants between vowels, as in *abogado*. The typical syllable consists of consonant plus vowel; hence few consonants occur word-finally (*-d, -l, -n, -r, -s, -z*, anciently also *-ch, -ll, -ñ*, etc.); of these, *-d, -n, -r*, and especially *-s* are inflectionally important. Dialect speech goes further, allowing *-d, -r, -s* to disappear, or *-l* and *-r* to merge.

Until *c.* 1600, Spanish contained a greater variety of consonants: *x* in *exe* (mod. *eje*) equaled English *sh*, *j* in *ojo* was like *z* in English "azure" (these two merged, and their product became [χ], like *ch* in German *Bach*). The affricates *ç* [=ts] and *z* [=dz] collided; their amalgam was pronounced either [θ], like *th* in "thin" (most of Spain) or like voiceless *s* in "mass" (part of Andalusia, Spanish America); nowadays, *c* before *e, i* and *z* otherwise are merely variant spellings of the same sound, as are also *g* before front vowels and *j*. The lone remaining affricate *ch*, pronounced as in English, represents a unit phoneme and rates as a separate letter. Voiced *s*, as in English "as," became voiceless between vowels (*casa*), surviving where adjacent to another voiced consonant (*desde*). The [v] sound, as in Old Spanish *bever, Córdova*, has disappeared; today *v*, tolerated for historical reasons, duplicates the orthographic role of *b:* both letters represent the same sound, released as a stop or as a spirant according to its position. Colloquially, palatal *l* (conventional spelling, *ll*, rated as a single letter) tends to become the semiconsonant [*j*], more rapidly so in the new world than in Spain.

Noteworthy are the traditional frequency of *a* and *s* and the infrequency of stops and of *f* (compare *hijo* with its counterparts in cognate languages). The sound [h], absent from standard Spanish, occurs dialectally, in words like *holgar* and *ojo*. Hence the letter *h* is predominantly ornamental, being used in deference to Latin tradition (*hora*) or as a holdover from earlier native practice (*hierro*, from older *fierro*); it is functional only in the segment *hue-*, where its role is comparable with that of mute *u* in *gue-, gui-*, and as an ingredient of *ch*.

Grammar.—Inflection, derivation, composition and syntax show the typical features of a conservative Romance language, pointing up its resemblance to Portuguese, Old (but not Modern) French and Italian as regards, for instance, the sparing use of definite and indefinite articles and of subject personal pronouns, the scope of imperfect and preterit, the formation of future and conditional, the partial dependence of the gender on the final vowel. Pronouns alone show traces of declension (*yo, me, mí*) and originally of the "neuter" (*lo mío, lo otro;* secondarily, *lo bueno*). Plural formation depends on interplay of termination and accentual pattern (*hombre-s, mujer-es, cortés* ~ *corteses, lunes* invariable). Nouns and pronouns in *-o* produce feminines in *-a;* otherwise the feminine is more sharply characterized in substantives (*vendedor-a*) and in adjective-substantives (*trabajador-a*) than in pure adjectives (*cortés* invariable).

The conjugation is based on a closely integrated system of tenses (simple or compounded with *haber*) and moods (indicative, subjunctive; the latter suggesting uncertainty or purpose); the reflexive construction has overlaid the passive voice. Peculiarly modern are the almost interchangeable use of two past subjunctives (*-se, -ra*) and the obsolescence of the future subjunctive. The only productive conjugation class is *-ar;* for verbs derived from substantives: *-ear* (*telefonear*); the *-er* and *-ir* verbs form closed groups. Stray relics of the Latin paradigm are found among "strong" (=radical-stressed) preterits (*dije, hice*) and past participles (*dicho, hecho*).

Vocabulary.—Viewed in the historical (stratigraphic) perspective, only a handful of important Spanish words, including some animal names, may be safely ascribed to the chronologically oldest layer, having been absorbed by Hispano-Romans from the language of ancestral Iberians and Celtiberians or of neighbouring Basques; *e.g., ard-illa* "squirrel," *bruja* "witch," *izquierdo* "left," *perro* "dog," *pizarra* "slate," *zorro* "fox." The core of the lexicon consists of a relatively archaic layer of Latin words, reflecting (frequently in contrast to French and to Italian) the preference of colloquial Latin throughout the last two centuries of the Republic; characteristic are *hermoso* "beautiful" as against French *beau*, Italian *bello; arena* "sand" as against French *sable*, Italian *sabbia; día* "day" as against French *jour*, Italian *giorno;* and *comer* "to eat" as against French *manger*, Italian *mangiare*. Moreover, Latin, as the medieval language of Church, administration and scholarship, contributed numerous learned and semilearned words, which show "retarded" sound development, having skipped certain evolutionary stages; *e.g., iglesia*, Old Spanish *eglesia* "church." Since the early days of humanism, there has been an incessant flow of Latinisms and Hellenisms into Spanish, less numerous, however, by a considerable margin than those absorbed into English (thus, Spanish lacks counterparts of "obstreperous" and "preposterous" and infrequently uses those of "acrimonious" and

"exacerbate").

Additional lexical strains that have contributed to the present wealth of the Spanish vocabulary are traceable to Gothic (*hato* "herd," *ropa* "clothing," *brotar* "to sprout," and numerous proper names, including *Alfonso*, *Fernando* and *Elvira*); to Byzantine Greek (*botica* "drug store"); to Arabic (frequently recognizable by the initial segments *al-*, *a-* or the termination *-í: al-calde* "mayor," *a-zucena* "lily," *balad-í* "frivolous," "paltry"); and to Old French and Old Provençal, especially words relating to the monastic and the chivalrous or courtly spheres (*monje* "monk," *fraile*, *Fray*, *Frey* "friar," *ruiseñor* "nightingale," *vergel* "flower and fruit garden" and others ending in *-el*).

Italian was very influential in the period 1450–1625, supplying hundreds of specific terms which, in many instances, continue to be used in numerous arts, crafts and trades, such as painting, navigation and banking.

French reached the peak of its influence in the 18th century (when the controversial Afrancesados introduced short-lived *remarcable*, *retornar*, etc.).

English, in particular U.S., civilization began to colour the lexicon of sports, science, commerce and politics after 1900. Over the centuries, the indigenous new world languages have supplied only terms in restricted domains (mostly fauna and flora).

BIBLIOGRAPHY.—*Frequency data:* (sounds) T. Navarro, *Estudios de fonología española* (1946); (phraseology) H. Keniston, *Spanish Idiom List* (1929), (constructions) *Spanish Syntax List* (1937) and *The Syntax of Castilian Prose: the Sixteenth Century* (1937); (vocabulary) V. García Hoz, *Vocabulario usual, común y fundamental* (1953). *Phonetics:* T. Navarro Tomás, *Manual de pronunciación española*, 6th ed. (1950). *Historical phonology:* A. Alonso, *De la pronunciación medieval a la moderna en español*, vol. i (1955 *et seq.*). *Phonemics:* E. Alarcos Llorach, *Fonología española*, 2nd ed. (1954). *Grammar:* (inventory of difficulties) M. M. Ramsey, *A Textbook of Modern Spanish*, rev. by R. K. Spaulding (1956); (historical) R. Menéndez Pidal, *Manual de gramática histórica española*, 9th ed. (1952). *Language history:* (introductory) R. Lapesa, *Historia de la lengua española*, 4th ed. (1959); (advanced) R. Menéndez Pidal, *Orígenes del español*, 4th ed. (1956). *Syntax:* (descriptive) S. Fernández Ramírez, *Gramática española: los sonidos, el nombre y el pronombre* (1951). *Lexicon:* (standard dictionary) R. Academia Española, *Diccionario de la lengua española*, 18th ed. (1956); (bilingual) E. B. Williams, *Holt's Spanish and English Dictionary* (1955); (overseas) A. Malaret, *Diccionario de americanismos*, 3rd ed. (1946); F. J. Santamaría, *Diccionario general de americanismos*, 3 vol. (1942); (etymological) J. Corominas, *Diccionario crítico etimológico de la lengua castellana*, 4 vol. (1954–57). *Historical and regional varieties:* (substratum) R. Menéndez Pidal, *Toponimia prerrománica hispana* (1953). *Foundations:* (Romance) A. Kuhn, *Die romanischen Sprachen* (1951); C. Tagliavini, *Le origini delle lingue neolatine*, 3rd ed. (1959); (Hispano-Arabic) E. K. Neuvonen, *Los arabismos del español en el siglo XIII* (1941); (medieval) Menéndez Pidal, *Cantar de Mio Cid*, rev. ed. (1944–46); (new world) R. J. Cuervo, *El castellano en América* (1946) and *Apuntaciones críticas sobre el lenguaje bogotano*, 9th ed. (1955); A. Alonso (ed.), *Biblioteca de dialectología hispanoamericana*, with emphasis on Argentina, Santo Domingo, old and new Mexico, 7 vol. (1930–49); C. E. Kany, *American-Spanish Syntax*, draws on written sources, 2nd ed. (1951); (Gypsy-Spanish) C. Clavería, *Estudios sobre los gitanismos del español* (1951); (Judeo-Spanish) C. M. Crews, *Recherches sur le judéo-espagnol dans les pays balkaniques* (1935). *Bibliography:* (general) H. C. Woodbridge and P. R. Olson, *A Tentative Bibliography of Hispanic Linguistics* (1952); (specialized) M. W. Nichols, *A Bibliographical Guide to Materials on American Spanish* (1941); (current) in the two leading journals, *Revista de Filología Española* (Madrid) and *Nueva Revista de Filología Hispánica* (Mexico City). (Y. M.)

SPANISH LITERATURE. Spanish literature falls into three divisions: Castilian, Catalan and Galician. Galician, for linguistic and historical reasons, is more conveniently dealt with under Portuguese. In the two Senecas, Lucan, Martial and Quintilian—as also in Prudentius—Spain had contributed much to Latin literature before the separation of the Romance languages; while the writings of Spanish Moslems and Jews form an important branch of the literature of Arabic.

CASTILIAN LITERATURE

By the time of the Moslem invasion (711) the Latin spoken in the Iberian peninsula was already in process of transformation into Romance. By the 10th century, the *Glosas emilianenses* and the *Glosas silenses* reveal traces of a vernacular already substantially developed. Mocádem, a blind poet who flourished near Córdova at the beginning of the century, was recorded 200 years later as having taken "expressions in vulgar Arabic or Romance, called them *markaz* and built on them his *zéjels* or *muwassahas*." After 1948 numerous examples of the *markaz*, or theme stanza, written in a Romance idiom transliterated into Hebrew or Arabic, were recovered from Hebrew and from Arabic *muwassahas* dating from the 11th century onward. These provided not only the earliest texts in Mozarabic (the Romance dialect of Spaniards living under the Moslems) but also evidence of a popular poetry that may have begun in the time of Mocádem or earlier and can be seen now to explain much in the traditional lyric types (*e.g.*, the *villancico*) of the later middle ages and the Renaissance. The poem as a whole was generally a love song sung by the woman (comparable to the later Galician *cantiga de amigo*), and the motif in Romance was a cry of passion on which the whole was an elaboration or to which it led up.

Heroic Poetry (11th and 12th Centuries).—The earliest surviving monument of Spanish literature and one of its most distinctive masterpieces is the *Poema de mio Cid*, an epic poem (*cantar de gesta*) of the middle of the 12th century (the existing manuscript is an imperfect copy of 1307). This work of 3,730 irregular lines (from 11 to 18 syllables, with marked caesura and traces of a four-beat accentual rhythm) grouped in assonating *laisses* tells of the fall from and restoration to royal favour of the Castilian noble Rodrigo Díaz de Vivar, known by the Arabic title of *sidi* ("lord"), who lived from 1040 to 1099. The historical setting and personages, the topographical detail, the realism of tone and treatment and the proximity in time of the anonymous poet to his hero have allowed the advancing of "historicity" as a first characteristic of the poem and of the Castilian epic in general. The latter two of its three *cantares*, however, while equally convincing in detail, are untrue to recorded history; the conception of the whole is essentially poetic and is handled with much imaginative artistry; and the according of a mere seven lines to the Cid's greatest achievement, the taking of Valencia from the Moslems after a nine-months' siege in 1094, makes clear the subjective scale of values. However, the poem caught the popular imagination and lived on in later epic, chronicle, ballad and drama, helping to fix the popular conception of the Castilian character. The only other surviving epic text (apart from a 100-line fragment of uncertain filiation on *Roncesvalles*, of the early part of the 13th century) is the wholly fanciful and decadent *Cantar de Rodrigo* of the 14th century, telling of the Cid's early manhood; it was from this that the later legend of the Cid took shape.

Frequent allusions in the vernacular chronicles to the heroic narratives of minstrels (*juglares*) make it clear that this *mester de juglaría* was much richer than the scarcity of surviving manuscripts would suggest. Thanks to the same vernacular chronicles, which from their beginning late in the 13th century accepted these narratives as being of historical validity and prosified them at length in their pages, it has been possible to reconstruct the themes and even, on occasion, fragments of the poetic text of epics on *Los siete Infantes de Lara*, *Fernán González* (both set in the 10th century), *El Cerco de Zamora* (again involving the Cid and perhaps artistically the best) and others. *Bernardo del Carpio* provides a hero of the days of Roncesvalles in defense of the national prestige against the French Carolingian epic; the remaining themes are an integral part of the feudal history of 10th- and 11th-century Castile, close to the time of telling and less probably related in their origins to the very different French epic than to a remote Visigothic past.

Learned Narrative Poetry (13th and 14th Centuries).—A new school of erudite poetry, much indebted to France and linked not with the itinerant *juglar* reciting to popular audiences but with the monastery and a literate public, became known as the *mester de clerecía*. Regularly adopting the French alexandrine in the set mould of the *cuaderna vía* or "fourfold way" (four-line single-rhyme stanza) and dealing with religious, didactic or pseudo-historical matter, it was a product of the study. The school is best and probably first exemplified in Gonzalo de Berceo (d. after 1246), the earliest Spanish poet known to us by name, a cleric of the same San Millán de la Cogolla that had produced the

Glosas emilianenses, in La Rioja. Apologizing for his inadequate Latinity, he versified unwearyingly in the vernacular the lives of Spanish saints, the miracles of the Virgin and other devotional themes with an ingenuous candour and an accumulation of picturesque and affectionately observed detail which largely atone for his prolixity (about 13,000 lines) and his fidelity to his sources. Also of the 13th century and in the same form are the *Libro de Alexandre*, much more erudite and pretentious, the *Libro de Apolonio*, a romantic tale of Greek extraction, and the *Poema de Fernán Gonçález*, a clerical reworking of a lost epic on the 10th-century count of Castile who won his county's independence from León.

The Beginnings of Prose.—A major influence on prose was exercised by Arabic. The gates of oriental learning and story were opened to Spain by the capture from the Moslems of Toledo (1085), which became a centre of translation from oriental languages. In 1120 Pero Alfonso, a Jew converted to Christianity, introduced Indian fable to the non-Arabic reader with his *Disciplina clericalis*, a collection of tales translated into a barbarous Latin. The translation from the Arabic of the "beast fable" *Kalila e Dimna* (1251) is the first essay in storytelling in the Spanish language. The romance of the Seven Sages (or *Sendebar*) was translated under the title of *Libro de los engannos e assayamientos de las mugeres* (1253), and other collections of eastern stories followed.

Alphonso the Wise.—The middle of the 12th century saw the recovery of Córdova, Valencia and Seville by the Christians, the kingdom of Granada alone remaining in Moslem hands. An indication of the more propitious intellectual atmosphere resulting was seen in the founding of the first universities (Palencia *c.* 1212, Salamanca *c.* 1243); and under the personal stimulus of Alphonso X of Castile and León (Alfonso el Sabio, "the Wise;" reigned 1252–84) literature achieved official prestige. Alphonso, in whose chancery Castilian replaced Latin as the language for public documents, may be called the father of Castilian prose. He organized a vast enterprise of translation and compilation aimed at fusing in the vernacular all available knowledge, classical, oriental, Hebrew and Christian. The works undertaken, often under his personal editorship, included the great legal code *Las siete Partidas* (a mine of curious information on the life and customs of the time) and compilations from Arabic sources such as the *Libro de saber de astronomía*, the *Tablas alfonsíes*, the *Lapidario* and the *Libro de los juegos*. With the *Primera Crónica general*, covering the history of Spain from its first settlers, and the *Grande e general Estoria*, an attempt at universal history from the creation, Alphonso was the founder of Spanish historiography in the vulgar tongue: the former work, carried by Alphonso to A.D. 711 and completed in the following reign, gave rise to a whole series of chronicles over the following two centuries and was the most influential single work of the Spanish middle ages; the latter work did not reach later than the parents of the Virgin Mary. Himself a poet, Alphonso was responsible too for one of the greatest collections of mediaeval poetry and music, the *Cantigas de Santa María* in Galician (then the accepted vehicle for lyrics).

The 14th Century.—The period of translation and compilation was succeeded by one of brilliant original creation. It is best represented in the prose of Alphonso's nephew the infante Don Juan Manuel (1282–*c.* 1349) and in the poetry of Juan Ruiz, archpriest of Hita (*c.* 1283–*c.* 1350). Juan Manuel's *Libro de Patronio*, known also as *El Conde Lucanor* (1335), a collection of 50 tales each pointing a moral, still drew heavily on Arabic sources; but by its individuality of style and treatment, the connecting thread of plot and the relating of fable to the day-to-day problems of life in high position, it ranks high in the beginnings of Spanish fiction. The same author's *Libro del caballero e del escudero* and *Libro de los estados*, both betraying indebtedness to the Catalan Raimon Lull (Ramón Llull), likewise pursue in the guise of fiction a doctrinal purpose, the education of a prince.

Earlier, under the influence in part of the Arthurian Legend (*q.v.*) (*see* also TRISTAN), which from the turn of the 13th century had been circulating in Spain in translation, there had appeared (*c.* 1305) the first Spanish novel of chivalry, *El Caballero Cifar*, the work possibly of the archdeacon Fernán Martínez. Cifar's squire Ribaldo was a distant precursor of Sancho Panza and thus an early token of the constant impinging of the real on the ideal so characteristic of Spanish literature. Contemporary with this was the creation, also related to the Breton cycle, of the *Amadís de Gaula* (*c.* 1304–12), destined to father a progeny that held the imagination of kings, courtiers and scholars through the 16th century. Sentimental idealism, lyrical atmosphere and supernatural adventure notwithstanding, the weight of evidence is against the traditional ascription to Portugal, where the work is first found mentioned *c.* 1450; the author may have been a Galician troubadour, João de Lobeira, living in exile at the Castilian court.

Juan Ruiz, archpriest of Hita, a man of the people, was the most intensely alert and individualized of early writers. His *Libro de buen amor* (1330, expanded in 1343) represents a collection of the most disparate elements—Ovid, Aesop, the *Pamphilus de amore*, the *Carmina Burana* and the liturgy—with a strong suggestion of Moslem ways of thinking in the mingling of eroticism with devotion and the invitation to the reader to interpret for himself the author's very equivocal teachings. But the diversity of adventure was presented as a satiric autobiography and stamped with the author's exuberant personality; and his Trotaconventos, ancestress of Celestina, is the first great character creation in Spanish literature. The dominant metre was still the *cuaderna vía*, handled with a narrative vigour and a plasticity that it had never known before; but the interspersing of the text with lyrics, religious, pastoral-farcical, amorous or satirical, of great metrical variety pointed to the approaching supersession of the "fourfold way."

Continuing evidence of the exotic strands that were being woven into the texture of Spanish letters may be found in the *Poema de Yuçuf* (assigned to the same 14th century) and in the rabbi Sem Tob's *Proverbios morales* (*c.* 1355). The former, an Aragonese version in *cuaderna vía* of the story of Joseph, is based not on the Bible but on the Koran and is written in Arabic characters; it is the chief representative of the so-called *aljamiado* ("barbarian") literature, Spanish in language, Arabic (or Hebrew) in tone and script. The *Proverbios* of Sem Tob (Shem Tobh), comprising 725 alexandrine couplets with internal as well as end rhyme, introduced to Spanish the grave sententiousness of Hebrew poetry with extreme aphoristic concision; his chief sources were the Old Testament, the Talmud and Ibn Gabirol (Avicebron). The principal surviving manuscript, much later than the date of composition, is in Hebrew characters. Sem Tob was a poet too in the Hebrew tongue.

The chancellor Pero López de Ayala (1332–1407) dominates the later part of the 14th century both in poetry, with his *Rimado de palacio*, the last major relic of the *cuaderna vía*, whose 8,200 lines distilled the essence of a lifetime of experience and disillusion in high places, and in prose as a historian of things seen. His chronicles of the four reigns (the last incomplete) of Peter I, Henry II, John I and Henry III of Castile infused new life in the vernacular into the tradition of the Latin *cronicón* and greatly stimulated the writing of personal, contemporary history. An early humanist, he translated and imitated Livy and Boccaccio together with Boethius, St. Gregory and St. Isidore.

The 15th Century.—The early part of the 15th century witnessed a first tentative renewal of Spanish poetry under Italian influence, which was to extend and to produce ever deeper effects over the next 150 years. The contrast is strong, during the long reign of John II (1406–54), between the social anarchy of a feudalism in its death throes and the cultivation of polite letters, which was now becoming an expected mark of birth and breeding. Collections such as the *Cancionero* made for the king *c.* 1445 by Juan Alfonso de Baena, containing 583 poems by 55 poets ranging from highest nobles to humblest versifiers on the fringes of their service, show not merely the now decadent tradition of the Galician-Portuguese troubadour school but the new stirrings of a much more intellectual poetry which used symbol, allegory and classical allusion in the treatment of themes of high moral, philosophical or political intent. Francisco Imperial, son of a Genoese

settled in Seville, drew on Dante in the first years of the century to present a challenge which Íñigo López de Mendoza, first marqués de Santillana (1398–1458), Juan de Mena (1411–56) and others duly took up. Santillana, like Juan Manuel and Pero López de Ayala, was poet and scholar as well as soldier and statesman, a collector of the masterpieces of foreign literatures and a stimulator of translations. His *Prohemio e carta al condestable de Portugal* (*c.* 1449), the earliest work of literary history and criticism in Spanish, draws upon his reading in Italian, French, Catalan, Provençal, Galician and Portuguese as well as in the classics in translation. Imperial had already sought to acclimatize the Italian hendecasyllable. Santillana's 42 sonnets "fechos al itálico modo" marked the beginnings of the formal enrichment of Spanish poetry at the dawn of a new age; though these, like his long poems *La Comedieta de Ponza*, the *Diálogo de Bías contra Fortuna*, or *La Defunsión de don Enrique de Villena*, which in thought and structure likewise show the poet at school in Italy, are today neglected in favour of his charming *canciones, decires* and *serranillas* of native inspiration, his role as precursor is still outstanding.

Juan de Mena's vast allegorical poem of the drama of history past, present and to come (*El Laberinto de Fortuna*, known also as *Las Trezientas*, *c.* 1444) was a more conscious attempt to rival Dante; felicitous often in isolated narrative passages out of Spain's heroic past, it is weighed down by pedantry and by over-Latinization of syntax and vocabulary. His metre, the common vehicle of learned poetry throughout the 15th century, is the *verso de arte mayor*, characterized less by its variable length (9–14 syllables, with 12 as the norm) than by its two hemistichs and strong ternary rhythm; accentual verse will not be found again in Spanish until the 19th century. The other outstanding name in the century's poetry is Jorge Manrique (*c.* 1440–79), who achieved immortality with one poem, the *Coplas a la muerte de su padre*.

In prose the age saw the first appearance of books of travel, notably the accounts of a journey to Samarkand (1403–06) of Henry III's ambassador Ruy González de Clavijo (*Historia del gran Tamorlán*) and of the travels in Europe and the near east (1435–39) of Pero Tafur (*Andanças e viajes por diversas partes del mundo*). Contemporary history took on a biographical dress in two admirable collections of portraits of eminent Castilians, Fernán Pérez de Guzmán's *Generaciones e semblanzas* (part of a larger *Mar de historias*, *c.* 1450) and Hernando del Pulgar's *Claros Varones de Castilla* (1486).

The Renaissance and the Siglo de Oro.—The unification of Spain under the "Catholic kings" and Columbus' discovery of the New World, following the introduction of printing (1474) and concurrent with the full play of cultural traffic with Italy (where Naples had been since 1443 a dependency of Aragon), may be taken as opening the era of the Renaissance in Spain. With the exception of the more popular veins associated with oral tradition, the mediaeval poetic achievement lapsed into oblivion: the *Poema de mio Cid* and the corpus of the *mestcr de clerecía* were not to be recovered and printed until 1779, the *Auto de los reyes magos* (sole surviving fragment, from the early 12th century, of mediaeval dramatic activity) not until 1863. National heroic themes survived through ballad and chronicle, early beneficiaries of the printing press.

What conferred on Spanish literature its distinctive mark in this period was, however, the wealth of new experience born of the overseas adventure, reflected not merely in historical and scientific writing but also in the detached and questioning attitude, so strikingly exemplified in the theatre, to the rediscovered norms of classical authority. The first Spanish humanists were also the first grammarians and lexicographers of any Romance tongue: the *Gramática castellana* (1492) of Antonio de Nebrija (Lebrija) was written under the sense of impending imperial responsibilities. Juan Luis Vives (1492–1540), a figure of European eminence, the brothers Juan and Alfonso de Valdés and others among the best brains and noblest minds of contemporary Spain were friends or followers of Erasmus, whose writings circulated in translation from 1520 until their suppression after his death in 1536, and whose influence has been seen in St. Ignatius of Loyola, founder of the Society of Jesus, and even, a generation later, in Fray Luis de León.

The masterpiece of the early Renaissance is the *Tragicomedia de Calixto y Melibea* (*c.* 1499), a novel in dialogue form by a converted Jew, Fernando de Rojas. The dominant character, a go-between elaborated from Juan Ruiz's Trotaconventos and depicted with a realism unsurpassed in Spanish letters before or since, soon caused the work to be popularly rebaptized *La Celestina*, thereby obscuring its significance as a tragedy of idealism in the new world of the senses liberated by the Renaissance. The analysis of passion and the dramatic conflict with nemesis that its pursuit involves were worked out with such psychological intensity on the intersecting planes of the noble and the ignoble, of mediaeval and Renaissance, as to make of the work the first indubitable and universal masterpiece of Spanish prose. It gave rise to innumerable sequels and imitations and was in due course translated into most European languages.

Romancero and Early Drama.—Spanish ballads (*romances*) form the strongest link in a chain of tradition that runs unbroken from mediaeval heroic epic to 20th-century poetry and drama; they lie at the heart of the national consciousness; and their expansion and capacity for survival, from Salonica to Chile, from the Low Countries to North Africa, accurately reflect the far-flung boundaries of Spain's prestige in its age of greatness. The ballad form is clear: the 16-syllable line with equal hemistichs (or two octosyllables) and a single assonance, suggesting a regularizing of the old epic versification, though it is doubtful whether we may infer from this that the *romancero* originated in the fragmentation of the decadent epic. The earliest datable ballads treat of frontier incidents (late 14th century) or lyrical themes (early 15th century), and it has been shown that only 5 out of about 35 "old" ballads on the Cid bear relation to known epic sources, while two-thirds of the whole vast corpus are lyrical in character and betray French or other exotic influence. It is on the ballads on mediaeval heroic themes, nevertheless, that the peculiar importance of the kind in Spain hinges: they formed everyman's source book on national history and on the national character. Circulating first as *pliegos sueltos*, the traditional ballads were collected in the *Cancionero de romances* (Antwerp, *c.* 1550) and in the *Silva de varios romances* (1550) and thereafter repeatedly. Soon the form was taken up and exploited for lyrical purposes in the *romances artísticos* of the most famous poets of the age, to remain ever afterward the chosen medium for popular narrative verse.

The origins of drama in Spain may be presumed, in default of evidence, to have been like those of drama in France and England. A tradition of religious performances in the vernacular is supported by references in legal and other documents and by the fragmentary *Auto de los reyes magos*, already mentioned; and one may infer, as also from a passage in *Las siete Partidas*, the existence of a secular drama interpreted by *juglares*. Juan del Encina (1468–*c.* 1529) marked the first stage in the emancipation of the drama as we know it from ecclesiastical leading strings, by way of performances for a noble patron. The *representaciones* and *églogas* in his *Cancionero* (1496) began as pastoral-religious, cast in a rustic dialect, *sayagués*, which was later accepted as a convention; but he soon turned to the debating of purely secular themes or to the vividly depicted farce of the *Auto del repelón*. Three plays dating from a long stay in Italy (*Églogas de Fileno y Zambardo, de Cristino y Febea* and *de Plácida y Vitoriano*) show Encina's conception of drama transformed by the experience from an ingenuous mediaevalism to one of bold Renaissance experimenting in the realm of sensuous indulgence, fortified by classical myth and allegory.

Encina's Portuguese disciple Gil Vicente (*c.* 1465–*c.* 1536), a court poet at Lisbon who wrote 11 of his 44 plays entirely and 17 of them partly in Spanish, showed a great advance in naturalness of dialogue, acuteness of observation and humour of situation. His weakness lay in plot, his peculiar distinction in a lyricism that expressed itself in the finest Spanish poetry of the period and in a vivid, plastic representation of allegory (as in the trilogy of the *Barcas*) that looks forward to Calderón. Bartolomé de Torres

Naharro (d. after 1530), who settled at Naples, published there in his *Propalladia* (1517) six of his eight comedies, all again performed before a noble audience. These show a sense of plot and a first awareness of dramatic theory with the division in form into five acts (soon to be reduced to three) and in kind into *comedias a noticia* and *comedias a fantasía* (the former realistic, as the *Comedia soldadesca* and the *Tinellaria*, satirizing respectively an army of occupation and the servants' hall in a cardinal's palace; the latter romantic and novelesque, the *Comedia serafina* and the *Himenea* being of particular interest as foreshadowing the cloak-and-sword comedy with the *pundonor* or "point of honour" as theme).

The emergence of the drama to the market place and the creation of a public were largely the work of Lope de Rueda (*c.* 1510–65), who from 1551 toured Spain with his modest troupe, his still more modest properties and a repertoire of his own composing. His four prose comedies all owned Italian models and were clumsy in construction: it is in the ten comic *pasos* (interludes performed between the acts of longer plays) that his dramatic merit resides, and with it the distinction of fathering the one-act play which, constantly renewing itself, has today some claim to be held the most living and popular of all dramatic forms in Spain.

The first dramatist to realize what might be made of the *romancero* in the theatre, in the sense of stirring the public's most responsive chord, was Juan de la Cueva (*c.* 1550–*c.* 1610): most of his 14 comedies and tragedies (1583) were taken, it is true, from classical antiquity and written in the manner of Seneca, but in three of them (*Los siete Infantes de Lara, El Reto de Zamora, La Libertad de España por Bernardo del Carpio*) he turned to themes of Spain's early heroic story, already familiar in ballad, and thereby set an example that accounts in large part for the immense fertility of the *comedia* and for the close accord thenceforth established between dramatist and public. Juan de la Cueva was not an accomplished dramatist, but he helped to found a drama "national" in subject-matter and free from pseudoclassical norms.

The 16th-Century Lyric.—The earlier attempt to Italianize Spanish poetry had failed through an overambitious intellectualizing of theme, on a basis of pedantic imitation, at a time when language and metrics were still incapable of sustaining the burden. A century later Juan Boscán (d. 1542), by reintroducing Italian metres (sonnet, *ottava rima, canzone,* tercet, blank verse), prepared the way for a much greater poet, his friend Garcilaso de la Vega (1503–36), with whom the lyric was reborn. The importance of the new metres went much beyond the formal: in Garcilaso the most significant novelty is the verse paragraph of the *silva* (a free combination of hepta- and hendecasyllables), allowing a new concern with the analytical expression of thought and emotion. To a mastery of the poetic process derived from close study of Petrarch, Boccaccio and Sannazaro and also of Theocritus, Virgil, Tibullus and Horace he added the gift of infusing an intense personal note into the most characteristic Renaissance themes. His meagre output of eclogues, elegies and sonnets, published posthumously in 1543, won him recognition for centuries as "the prince of Castilian poets" and largely determined the course of lyric poetry throughout the *siglo de oro.*

Fray Luis de León (1527–91), adopting chiefly Garcilaso's *lira* stanza (7, 11, 7, 7, 11) for poems charged with a deep emotional sincerity, typified a so-called "Salamanca school," with its emphasis on content over form. Fernando de Herrera (*c.* 1534–97) headed a contrasting school of Seville which, deriving equally from Garcilaso, was concerned rather with Petrarchan subtleties of refined sentiment; in a quartet of remarkable odes he gave vibrant expression to heroic themes of the day. A movement of protest in defense of the short native metres, associated chiefly with the name of Cristóbal de Castillejo, was greatly reinforced after 1550 by the printing of *romanceros* and by the evolving drama which, while employing much metrical variety for the procuring of specific effects, found in the octosyllable the accepted vehicle for narrative and description. Throughout the 17th century poets were to write in both the old and the new, their practice still allowing a notable degree of equation between the old and "verse" and between the new and "poetry."

For epic poetry the metre was the *octava real* (*ottava rima*) and the models Ariosto and Tasso, with the reserve that the themes and the heroes to be sung were those of overseas conquest and expansion or of defense of the empire and of the faith at home. Alonso de Ercilla y Zúñiga came nearest to achievement with his *Araucana* (1569–90), telling at first-hand of native resistance to the conquest in Chile; its defects sprang from a wavering concept of epic theory, now martial-heroic unrelieved, now amorous-fantastical with admixtures of prophecy and pedantry. Bernardo de Balbuena's *Bernardo,* on a familiar mediaeval theme, and Lope de Vega Carpio's *Dragontea* (Drake) and *Corona trágica* (Mary Stuart) may stand for the multitudinous activity in a kind that laid its compulsion on every poet of ambition.

History.—Prose before the Counter-Reformation produced some notable Lucianesque dialogues, from the pen especially of Alfonso de Valdés (*Diálogos de Lactancio y un arcediano* and *de Mercurio y Carón,* both 1528). His brother Juan wrote in Naples a *Diálogo de la lengua* (*c.* 1535) which, though not published till 1737, is still a document of the highest critical value. In style these stand at the other extreme from the *Relox de príncipes* and the *Cartas familiares* of Antonio de Guevara (*c.* 1480–1545), which enjoyed a European notoriety in their day as the Spanish counterpart of euphuism. History in the tradition of Alphonso the Wise continued to be cultivated, patriotism waxing higher as Spain's imperial greatness cast its shadow over Europe; its last great flowering was seen in the *Historia de España* of Juan de Mariana (*c.* 1536–1624), the translation of which from its original Latin (1592) into Spanish (1601) by the author himself marked the final triumph of the vernacular for all literary purposes. Diego Hurtado de Mendoza gave in his *Guerra de Granada* (*c.* 1572) an extreme example of the concern to write history upon the classical model; the *Anales de la corona de Aragón* (1562–79) of Jerónimo de Zurita exalted by contrast serious documentation and impartiality over style.

But the real landmarks in historical writing came from the New World and showed the transmuting of vital experience into literature with a vividness till then unknown in Spain. The letters of Columbus, the *Cartas de relación* of Hernán Cortés and many another narrative by humble *conquistadores* void of literary artifice or pretension opened up new horizons to the reader and, as in the attempt to capture exotic landscapes in words, enlarged the resources of the language. Most engaging of them all was the *Historia verdadera de la conquista de la Nueva España,* written in old age by Bernal Díaz del Castillo (1492–*c.* 1581), who had a phenomenal memory, an infallible eye for the telling detail and a passion for distributive justice. Bartolomé de las Casas (1474–1566), "father of the Indians," wrote history for noble yet partisan purposes: on his *Brevísima Relación de la destruyción de las Indias* rests a large measure of responsibility for the *leyenda negra* which Spain's enemies were for so long after to use against it.

Mysticism.—The great period of Spanish mysticism coincided closely with the Counter-Reformation, though it had its antecedents. Raimon Lull had been a notable precursor in 13th-century Catalonia; and the *Dialoghi di amore* (1535) of the expatriate Spanish Jew Judas Abrabanel ("León Hebreo"), which both in Italian and in Spanish translation exercised a profound influence on 16th-century Spanish thought, allowed the infiltration of many a Platonic image and doctrine. The literary importance of the mystics derives not only from their subject matter but also from the consideration that in the constant striving to transcend the limitations of language there were liberated resources of expression, especially in the sphere of psychological probing, till then unknown. In the *Libro de su vida,* in the *Libro de las moradas* and in the *Cartas* of the untutored St. Theresa (1515–82) many of the novelist's greatest gifts can be appreciated in embryo. Fray Luis de León, humanist and theologian, did not attain to the highest flights of the mystic ascent, but in his prose *De los Nombres de Cristo,* as in his poems, he moves by his passionate devotion, his sincerity and his profound feeling for nature, the whole finding expression in a style of singular purity. St. John of the Cross

(1542–91) rests on a pre-eminence unparalleled in Spanish literature by virtue of three mystical poems (*Noche oscura*, *Cántico espiritual*, *Llama de amor viva*) of which his prose writings were long and involved expositions. The poems throb with religious ecstasy and represent the most exalted expression in Spanish of the most exalted of human experiences, mystic union.

The Novel.—Popular taste in the novel was dominated for a century by the monstrous progeny of the *Amadís de Gaula*, which had achieved print in 1508. These interminable romances kept alive certain ideals of mediaeval chivalry, but, having lost touch with life at every point, represented pure escapism and in due course evoked, along with moral censure, various literary reactions. The pastoral novel was one. Akin to the eclogue naturalized from Italy by Garcilaso, it was modelled on Sannazaro's *Arcadia* and inspired by nostalgic echoes of the golden age. Its shepherds were courtiers and poets who, no less than the knights-errant, turned their backs on material reality, with the purpose, however, of exploring the more freely, against an idealized and carefree setting, the realm of the emotions. In Spain the first and best example was the *Diana* (*c.* 1559) of Jorge de Montemayor; to the vogue it started, Cervantes with his *Galatea* and Lope de Vega with his *Arcadia* alike subscribed. In Gaspar Gil Polo's *Diana enamorada* (1564), which was outstanding among the many sequels and imitations, the story became little more than framework for a lyrical anthology.

A more positive reaction against the inflated tales of chivalric prowess was seen in the picaresque novel (*q.v.*) initiated in 1554 with the brief, anonymous *Lazarillo de Tormes*. The word *pícaro*, of doubtful origin, does not appear in the text; the kind, native to Spain, was to exert influence in many foreign literatures. The *pícaro* was the antihero, innocent of ideals and concerned only to keep alive, who lived on his wits and, as he passed from master to master, saw and depicted life from underneath. Important for its part in guiding fiction back to direct observation of life, the picaresque formula had nothing to contribute to the development of the novel as an art form. The seed sown by the *Lazarillo* did not in fact bear fruit for nearly 50 years; and with the *Guzmán de Alfarache* (1599) of Mateo Alemán, whose interminable moralizings showed the Counter-Reformation's influence on literature at its heaviest, and Quevedo's *Buscón* (1626), where realism has become a cynically grotesque caricature of all human values, the kind, though it continued abundant throughout the 17th century, may be held to have given what it had in it to give.

Miguel de Cervantes Saavedra (*q.v.*; 1547–1616), the culminating figure in his country's literature, produced in *Don Quixote* (1605–15) the prototype of the modern novel and still its greatest exemplar. Nominally a satire on the already moribund romances of chivalry, the conception allowed of the presentation of reality on two levels, the "poetic truth" of Don Quixote and the "historic truth" of Sancho Panza. In pursuing the constant interaction of these equally valid but rarely compatible attitudes to life Cervantes revealed the true scope of the novel as a philosophical commentary on existence, and in tracing the constant interplay of character between master and squire—in contrast with the static characterization of previous fiction—he established psychology as one of the touchstones of excellence. In the *Novelas ejemplares* (1613), 12 tales of which the 5 belonging to the romantic-adventure type of the Italian *novellieri* may be discounted, Cervantes particularized his claim to be the first to write *novelas* (short stories) in Spanish by differentiating between those which interest by the action and those whose merit lies in the mode of telling, "making something out of nothing."

Later Drama.—The *comedia* achieved its formula and its splendour in the genius of Lope de Vega (1562–1635) toward the close of the 16th century, to fill thereafter with an incredible fertility of invention the 100 years prior to the death of Calderón in 1681. Its manual was Lope's own *Arte nuevo de hacer comedias en este tiempo* (1609), which firmly rejected the "rules," opted for a mixed kind (*comedia* is in Spain an all-embracing term) and for metrical variety and made of the *vulgo* the ultimate arbiter of taste. This was essentially, therefore, a "social" drama, ringing a thousand changes on the accepted foundations of society, respect for crown, for church and for the human personality, the latter symbolized in the "point of honour" that Lope commended as the best theme of all "since there are none but are strongly moved thereby." The *pundonor* was in truth a matter largely of convention, "honour" being equivalent, in a very limited and brittle sense, to reputation. Here too the drama was social; characters, to merit esteem, had perforce to abide by all the prescriptions of their fellows. It followed that this was a drama less of character than of action or of intrigue, which rarely if ever grasped the true essence of tragedy or knew the cathartic function. What the great Spanish playwrights did possess was a remarkable sense of stagecraft and the ability to make the most intricate plot grip to the end, however arbitrary the final tying or untying of the knots.

Lope, the "prodigy of nature," as Cervantes called him, who claimed authorship of 1,800 plays, some composed within 24 hours, and of whom we still possess around 500, swept public and fellow-dramatists alike into vassalage by his powers of creation. Little that he wrote was perfect, but he had an unerring sense for the theme and the detail that could move an audience, conscious of being on the crest of its country's greatness, to respond to a mirroring on the stage of some of the basic ingredients thereof. Because of him the *comedia* became a vast sounding board for every chord in the Spaniard's consciousness, a "national" drama in the truest sense. His plays have been classified into religious, mythological, pastoral, novelesque and the like. In reality only two groupings matter: the native historical, and the *capa y espada* ("cloak-and-sword") plays of contemporary manners. Lope ransacked mediaeval chronicle, the *romancero*, popular legend and song for heroic themes, chosen for the most part as throwing into relief some aspect of the national character or of that social solidarity on which the Spain of his own day rested. *Fuente Ovejuna*, *Peribáñez y el comendador de Ocaña*, *El Caballero de Olmedo*, *La Estrella de Sevilla* (concerning the authorship of which there is now some doubt) count among the best.

The cloak-and-sword play, a type which weighed more and more in the *comedia* after Lope, was pure diversion, written to a number of hollow conventions with much disguising, mistaking of identities, duelling, falling in and out of love and false alarms about "honour" and a duplicating of the plot below stairs between *gracioso* (the gallant's humorous or would-be humorous lackey) and lady's maid. Though contemporary, it could in no sense be held to give a true picture of society, but it gave pleasure by its infinite resource and dexterity, its sparkling dialogue, its tripping convolutions in the relations between the sexes and its constantly implied paean to the *Carpe diem* in an irresponsible world. The essentially lyrical character of much of the *comedia* further explains its charm.

The greatest of Lope's immediate successors was Tirso de Molina, pen name of Fray Gabriel Téllez (*c.* 1584–1648), whose *El Burlador de Sevilla*, though far from his best play, achieved fame as presenting the Don Juan legend on the stage for the first time. *La Prudencia en la mujer* figures among the greatest of Spanish historical dramas, as *El Condenado por desconfiado* does among theological, while his cloak-and-sword comedies include some of the liveliest of the kind. In the delineation of character, especially feminine character, Tirso had few equals. Juan Ruiz de Alarcón (*c.* 1581–1639) struck a distinctive note which some have sought to relate to his Mexican birth and upbringing, others to his physical deformity. His 20 plays were sober, studied and imbued with serious moral purpose; *La Verdad sospechosa* inspired Corneille's *Le Menteur*. The dramatic conflict between love and honour posed in *Las Mocedades del Cid* by Guillén de Castro (1569–1631) was the source similarly of Corneille's *Le Cid*. But moral purpose and dramatic conflict are infrequently to be found among the plays of Antonio Mira de Amescua, Luis Vélez de Guevara, Juan Pérez de Montalbán and the many others who with more or less distinction, if with dwindling originality, bridged the transition from Lope to Calderón.

Culteranismo and Conceptismo.—In poetry and prose the early 17th century was marked by the rise and spread of two interrelated movements. The one, known as *culteranismo*, of which Luis de Góngora (1561–1627) was the archpriest, resumed, as one of its facets, earlier attempts to ennoble the language by re-Latinizing

it. In Góngora's *Soledades* and *Polifemo y Galatea* this process was not only carried to extremes in vocabulary, syntax and word order, so that the reader too must needs be something of a Latinist, but expression was garbed in the full panoply of classical myth and allusion and further complicated by every subtlety of which metaphor is susceptible. Yet this poetry was something other than a deliberate cult of obscurity. Theme and treatment make clear the attempt to achieve a poetry not merely "pure" but timeless, to do for Spanish what the lyric poets of antiquity had done for Greek and Latin, to write poems which, rediscovered after the lapse of hundreds or thousands of years, could be acclaimed still fresh and immortal. Indeed after 300 years Góngora was so rediscovered; but for his age his example invited many who lacked his genius to imitate slavishly his procedures. Gongorism disfigures many a good play by Calderón.

The twin affectation was *conceptismo*, which played on ideas as *culteranismo* did on language. The *conceptista* purveyed thought and sought to shock the reader into attention by the subtlety and also by the violence of his conceits. Aiming always at the semblance of profundity, the quintessential, his style was concise and epigrammatic and demanded for comprehension not the well-stocked mind but one infinitely nimble and alert. *Conceptismo* thus belonged properly to prose, inclined to satire, since it was much concerned with stripping off appearances, and found its most adequate outlet in the essay. Francisco de Quevedo (1580–1645), the master satirist of his age in the literary, the political and the social sphere and one of Spain's greatest wielders of language, was, in *Los Sueños,* its outstanding exponent. Baltasar Gracián (1601–58), its preceptist, reduced all its refinements to an exact code in his *Agudeza y arte de ingenio*, as in a series of treatises (*El Héroe, El Discreto, El Oráculo manual y arte de prudencia*) he sought to codify also the art of worldly living. Gracián, if doubtfully a philosopher, was a thinker, and his allegorical novel *El Criticón,* a pessimistic interpretation of life as a "daily dying," was to influence La Rochefoucauld, Schopenhauer and Nietzsche.

Calderón.—Pedro Calderón de la Barca (1600–81) spanned the century of Spain's decline, and his drama constitutes a studied expounding of patterns of behaviour (the *pundonor*) or of belief. He wrote 80 *autos sacramentales* (symbolical plays performed on Corpus Christi day), and these, together with *La Vida es sueño, La Devoción de la cruz* and *El Mágico prodigioso,* which gave what dramatic tension was possible to a highly intellectualized exposition of Catholic dogma, make of him the perfect literary representative of the Counter-Reformation. In the so-called "honour" plays—*El Médico de su honra, A secreto Agravio secreta venganza, El Pintor de su deshonra* and *El mayor Monstruo los celos*—the cold logic with which an unreal code is pursued to the extreme becomes repugnant. *El Alcalde de Zalamea,* a reworking of a play by Lope, is perhaps his most successful play on the human level. *La Vida es sueño,* to the extent that it raises profound philosophical issues on a plane of universality, has given him his great fame abroad.

Francisco de Rojas Zorrilla (*Del Rey abajo, ninguno*) and Agustín de Moreto (*El Desdén con el desdén*) stand out among those who revolved in the orbit of Calderón. After his death Spanish drama lay dead for the next hundred years. *Culteranismo* and *conceptismo,* though symptoms rather than causes, had played their part too in the stifling of imaginative literature, and by the close of the century nothing but the shell remained of the greatness that had been the *siglo de oro.*

The 18th Century.—With the advent of the Bourbon dynasty in 1700 and its establishment after the War of the Spanish Succession a profound renewal of the country's intellectual life took place. Numerous academies were created, the most influential being the Real Academia Española founded in 1713 to maintain the purity of the language (*Diccionario de autoridades,* 1726–39; *Gramática de la lengua castellana,* 1771). Men of letters once again went to school abroad and discovered how far Spain had deviated from the intellectual courses of western Europe.

Moreover, a new spirit of inquiry into the national heritage was to carry back to Luis de León, to Herrera and to Garcilaso an awareness that in the early decades of the century had seemed limited to Góngora, Quevedo and Calderón (worst of models for an imitative generation) and was indeed to go back earlier still, to unearth the forgotten mediaeval literature. Gregorio Mayáns y Siscar (1699–1781), sometime royal librarian, edited Juan Luis Vives (in Latin), Juan de Valdés and Luis de León and, in a study of Cervantes, initiated Cervantine criticism. The Augustinian Enrique Flórez, embarking on the vast *Teatro geográfico-histórico de la iglesia de España* better known as *España sagrada* (29 vol., 1747–73; increased to 52 vol. by 1918), helped resurrect the whole cultural background of mediaeval Christian Spain. A landmark of even greater importance was the publication for the first time in Tomás Antonio Sánchez's *Poesías castellanas anteriores al siglo XV* (1779–90) of the *Poema de mio Cid,* the works of Gonzalo de Berceo, the *Libro de Alexandre* and Juan Ruiz's *Libro de buen amor.* From all this critical activity there resulted a debate between old and new which, waged throughout the middle decades of the century, compelled either side to reason its position and marked the birth of a new critical approach to literature.

Two names stand out: Ignacio de Luzán, educated in Italy and well-grounded in foreign literatures, whose *Poética* (1737) introduced to Spanish readers the aesthetic of Muratori, Boileau and René Le Bossu and launched the great polemic; and Benito Feijóo (1676–1764), a Benedictine whose *Teatro crítico* and *Cartas eruditas* (8 vol. and 5 vol., 1726–60) comprised more than 280 essays ranging over the entire universe of human knowledge. Assailing error, prejudice and superstition wherever he found them and so bitterly assailed in his turn that the king intervened to protect him, Feijóo made a monumental contribution to the intellectual emancipation of Spain. Imaginative prose produced the *Noches lúgubres* of José Cadalso (published posthumously in 1792), looking forward to romanticism; and the *Fray Gerundio* (1758–68) of Francisco José de Isla, a satire in the guise of biographical fiction on affectation in the pulpit, which looked back to the picaresque novel.

Poetry, moribund for nearly 100 years, raised a timid head at length in a small group at Salamanca, led by the Augustinian Diego González, which toward 1775 turned for inspiration to Luis de León, just as two decades later a group at Seville sought to revive the glories of Herrera; and again, though the lines were now more blurred, Salamanca stood for content, Seville for form. Juan Meléndez Valdés (1754–1817), who learned to think from John Locke and to feel from Edward Young, best exemplified the combination of new influences at work. A conscious artificer rather than a great poet—though the nearest to such among the Salamancans—he helped poetry through the painful apprenticeship necessary to its rehabilitation.

For drama, the second half of the century saw the great battle of the "rules," meaning chiefly the unities that the *siglo de oro* had spurned. *La Raquel* (1778) of V. A. García de la Huerta, combining a traditional Spanish theme with neoclassic construction, showed of what the reformist school was capable. The same author's vast compilation of earlier drama, the 16-vol. *Theatro hespañol* (1785–86), made apparent in its total omission of Lope, Tirso and Alarcón how unsubstantial were the grounds on which condemnation rested.

It fell to Ramón de La Cruz (1731–94) to bridge the gap, unwittingly but very effectively, by his fecund resurrection on a somewhat ampler scale of the earlier *paso* or *entremés* of Lope de Rueda, Cervantes and Luis Quiñones de Benavente (1589–1651). Based on acute satirical, though in the main good-humoured, observation of the contemporary Madrid scene, his one-act *sainetes* could not, in their brevity, transgress the unities or offend the purist; at the same time they delighted the *vulgo* and brought the drama back from the realms of pseudoantiquity to the function of commenting on life and society. Leandro Fernández de Moratín (1760–1828) applied the lesson to the full-length play and produced in *El Viejo y la niña, La Comedia nueva* and *El Sí de las niñas* comedies imbued with deep social seriousness which, written to a meticulous observance of the unities, were still good theatre. The *Comedia nueva,* a remorseless flaying of such writers as L. F. Comella (1751–1812) who had reduced to absurdity what they fondly believed to be the Calderón tradition, absolves one from any

need to read their handiwork. Moratín's dialogue here ranks among the best prose of the 18th century.

Romanticism.—The first third of the 19th century proved but a prolongation of the 18th the Napoleonic wars and their long aftermath casting a blight on letters. The inspiriting odes of a Nicasio Gallego or a Manuel José Quintana or the impassioned prose of a Gaspar Melchor de Jovellanos, one of the noblest figures of the age (1744–1811), show neoclassic objectivity and proportion set aflame with the emotion of outraged patriotism. War against France did not mean the repudiation of French literary influences. The thousands of liberals driven into exile by Ferdinand VII after 1823 sought refuge for the most part in France (some too in England). Having drunk there the heady wine of romanticism, they flocked back to Spain on Ferdinand's death in 1833.

Such fortuitous contacts were far from being the sole source of Spanish romanticism. The ground for this had been prepared in Cadiz from 1814 in a notable debate on literary values initiated by the German J. N. Böhl de Faber; in Barcelona with the founding in 1823 of the review *El Europeo*, its three editors a Catalan, an Englishman and an Italian; and in Madrid with Agustín Durán's critical essays on the drama of the *siglo de oro* and his vast *Colección de romances antiguos* (1828–31). Romanticism in France, moreover, was largely a revolt against native literary tradition and a turning abroad, notably to Spain, for new sources of inspiration; whereas in Spain it implied the rejection of an alien neoclassicism and to that extent a return to the spirit of its own earlier classics. All the formal characteristics of Spanish romantic drama —the mingling of kinds, the rejection of the unities, metrical variety (this now extended to the inclusion of prose)—had characterized the drama of Lope and his contemporaries, who had moreover treated many of its themes (taken preferably from the Spanish middle ages). It is from a study of these themes that the novelty and significance of the new school becomes apparent. Whereas the drama of the *siglo de oro* had rested directly on social solidarity, romantic drama was in revolt against society: its protagonist staked everything on his right to be himself, *especie única*, and, deliberately flouting every convention, defied society to stand in his way. In France, tragedy was by romanticism deflected into drama; in Spain, drama into tragedy.

The movement in Spain, coming to fruition a generation later than elsewhere in Europe, knew a correspondingly short and chequered life. The six plays that epitomize its dramatic achievement fell within the decade 1834–44: Francisco Martínez de la Rosa (1787–1862), *La Conjuración de Venecia;* Mariano José de Larra (1809–37), *Macías;* Ángel Saavedra, duke de Rivas (1791–1865), *Don Álvaro o la fuerza del sino;* Antonio García Gutiérrez (1813–84), *El Trovador;* Eugenio Hartzenbusch (1806–80), *Los Amantes de Teruel;* José Zorrilla y Moral (1817–93) *Don Juan Tenorio;* so too the poetry of José de Espronceda (1808–42). Rivas' *Don Álvaro* came nearest, in the theatre, to expressing a "philosophy" of romanticism; and the preface, from the pen of Antonio Alcalá Galiano, to his *Moro expósito* to supplying a foundation of precept. But the movement never possessed either the consistency of a school, or a leader, or a corpus of doctrine. Espronceda was the one romantic who lived his romanticism, if with an element of Byronic posturing; his *Estudiante de Salamanca, Canciones* and unfinished *Diablo Mundo* were the only subjective lyricism of value that the period produced and marked too a milestone in the enlargement of poetic form. In comparison, the *Romances históricos* of Rivas and the *Cantos del trovador* of Zorrilla were little more than exercises, brilliant within the limitations of their kind, in the retelling of picturesque legend. In prose Larra, Rivas, Espronceda and others tried and failed to acclimatize the historical novel after Sir Walter Scott, a form that never struck roots in Spain. Larra, the outstanding prose writer and most critical mind of his generation, dissected society pitilessly in his *Artículos de costumbres*, which, with the *Escenas matritenses* (1836–42) of Ramón de Mesonero Romanos and the *Escenas andaluzas* (1847) of Serafín Estébanez Calderón, helped to prepare the ground for the resuscitation of the novel.

Post-Romantic Poetry.—Three poets bestride the third quarter of the century and, in their diversity, reveal how one of romanticism's major services was its liberation of the individual personality, whereby literature was set free from compulsory subscription to any school or doctrine. Gustavo Adolfo Bécquer (1836–70), accounted a romantic, was concerned only to give expression to his own tortured emotions. His 76 brief *Rimas*, which owe something to Heinrich Heine, bear the stamp of a deeply moving sincerity. In his *Leyendas* Bécquer proved himself too a prose stylist in a richly poetic and imaginative vein. Ramón de Campoamor (1819–1901) affected the epigrammatic in his *Doloras, Pequeños poemas* and *Humoradas*, names indicative of a sustained attempt at originality and at bringing poetry back into the realm of ideas. The contemporary fame as a thinker of this bourgeois, utilitarian mind has not withstood the lapse of time; but he had a gift of mildly ironic observation, and later dramatists on occasion found in a quatrain by him matter for a play. Gaspar Núñez de Arce (1834–1903) likewise endowed poetry with a purpose, the social purpose of stimulating the civic virtues and defending the body politic against dangerous heresies. His *Gritos del combate* represented a noble if despairing attempt, at times overdeclamatory in tone, to stem the tide. Among somewhat later poets were Salvador Rueda (1857–1933; *Cuadros de Andalucía, Aires españoles*), an exuberant local colourist who has been taken as a precursor of modernism; José María Gabriel y Galán (1870–1905; *Castellanas, Extremeñas, Campesinas*), a *laudator temporis acti* singing the traditional virtues of the Castilian countryside; and Rosalía Castro (1837–85), a true poetess of Galicia, whose one volume of poems in Castilian, *En las orillas del Sar* (1884), invited comparison with Bécquer.

The 19th-Century Novel and Drama.—The last-named poets, each identified with his *patria chica*, shared with the novelists a quality that was to characterize Spanish fiction on its rebirth. For more than two centuries, from long before the period of *afrancesamiento*, the kind in which Spain had made its most striking contribution to world literature had been virtually extinct. The revival during the romantic period of the *cuadro de costumbres* was a pointer to where the native bent lay, and some *novelas de costumbres* timidly reidentified the novel with realistic social observation. It was Fernán Caballero (pen name of Cecilia, daughter of Böhl de Faber) who, with *La Gaviota* (1849), determined the particular technique of observation that was to rehabilitate the novel. This simple tale of romantic disillusion captivated less by its plot than by its Andalusian setting and by vividness of colour and detail. Twenty-five years later, coinciding with the restoration of the Bourbon monarchy (1874), came the beginning of the great period of the regional novel.

Pedro Antonio de Alarcón (1833–91), a waverer between the esoteric-romantic and the pedagogic, achieved his masterpiece in *El Sombrero de tres picos*, a sparkling tale of Andalusian peasant malice. A fellow-Andalusian, Juan Valera (1824–1905), brought much psychological insight and a classical note of cultured detachment to four novels (*Pepita Jiménez, Los Ilusiones del doctor Faustino, El Comendador Mendoza* and *Doña Luz*) which depended less on local realism, acutely observed though this was, than on the author's knowledge of men and women. José María de Pereda (1833–1906), graduating through a long series of *Escenas montañesas*, emerged as the greatest of the school. His realism in depicting his native Santander was no longer merely picturesque but a powerful re-creation of nature as the abiding reality, more important than the individual and therefore than the plot. *Sotileza* and *Peñas arriba*, perhaps his best novels, revealed a champion too of traditional values—religion, the family, country life—in a dissolving society typified by the big city.

Emilia Pardo Bazán (1851–1921) attempted, in two novels of Galicia, *Los Pazos de Ulloa* and *La Madre naturaleza*, to combine the Spanish Christian outlook with French naturalism, which she had already examined curiously and with many reserves in *La Cuestión palpitante*. The attempt made clear the incompatibility, and in later novels (*La Quimera, La Sirena negra*) the return to a genuinely Spanish realism, in which spiritual values also found their place, was complete. Armando Palacio Valdés (1853–1938) was the novelist of Asturias even when, as in the perennially popular *La Hermana San Sulpicio*, the stage was transferred to

Andalusia or, in *La Alegría del capitán Ribot*, to Valencia. He likewise experimented, in mid-career and somewhat half-heartedly, with naturalism (*La Espuma, La Fe*) but was saved by his most engaging quality, humour.

Besides this cult of regionalism, the fecundity in the novel of Benito Pérez Galdós (1843–1920) created a world in itself. Coming to Madrid from the Canary Islands in 1862, Galdós was better placed to see Spain from outside, in the round; and, after living through one of the most turbulent chapters in Spanish history, he addressed himself to the task of imaginatively re-creating its antecedents, beginning with the war against Napoleon. The 46 vol. of his *Episodios nacionales*, written from 1873 to 1879 and from 1898 to 1912, embraced 70 years of the country's record, from Trafalgar to the early years of Alphonso XII, and taught generations of Spaniards much of whatever history they knew. Other series (*Novelas de la primera época, Novelas españolas contemporáneas*) completed the revelation of Spain to Spaniards; embodying the author's liberal and challenging views, they constituted a concerted campaign to regenerate and Europeanize Spanish society and thought. *Gloria, La Familia de León Roch, Fortunata y Jacinta, Ángel Guerra, Misericordia* are individual excellences in the rich endowment of this giant among Spanish novelists.

The drama in the later 19th century, disorientated after the brief romantic efflorescence, produced few outstanding works and no clear conception of goal or function. Manuel Tamayo y Baus (1829–98), who began by oscillating between, and even attempting to fuse, classical tragedy and romantic drama, gradually evolved from verse to prose and from the historical-heroic to the contemporary-social; he achieved fame by the isolated *Un Drama nuevo* (1867), a powerful portrayal of a Yorick compelled to act on the stage his own personal tragedy. Adelardo López de Ayala (1829–79) used considerable technical gifts to pillory bourgeois vices (*El Tejado de vidrio* and *Consuelo*). The Nobel prize for literature (1904) set a seal on José Echegaray (1832–1916), whose 63 plays fall into two main groups: in the first, imbued with a violent neoromanticism, drama became melodrama with a constant striving after effect and a basic falsity of character, passion and situation alike; the second comprised a serious, often tragic drama of social problems in which Ibsen could be seen as one of several foreign influences (*El gran Galeoto, O Locura o santidad, De mala Raza, El Hijo de don Juan*), though here too passions were apt to be torn to shreds and dénouements to rely on the violent dispatch. Joaquín Dicenta, still close to Echegaray in his search after effect, brought the drama nearer to earth with an unpleasantly realistic study of working-class conditions in *Juan José* (1895). Failure to master the problems involved in the change of medium robbed Galdós of full success in the theatre, to which he turned increasingly in his concern to stimulate the social conscience (*Realidad, La Loca de la casa, Electra, El Abuelo*).

The Generation of 1898.—Once again, as in 1700, the turn of the century brought about a radical change in literary values and directions. Over the previous two decades and more a mood of scathing analysis had been developing with the growing conviction that behind the peace and order of a nominally representative regime much rottenness lay concealed. Galdós sensed and gave voice to the disquiet; and it evoked, in Ángel Ganivet's *Idearium español* (1897), one of the most searching analyses of the Spanish character ever written. With 1898 came the brief and disastrous war with the United States, which left Spain stripped of the last vestiges of its trans-Atlantic and Pacific empire. The imperial cycle begun in 1492 had ended in ignominy, and thinking Spaniards embarked on an analysis of their country's ills, a scrutiny of its possible destiny and an attempt to shock the national mentality out of its *aboulia*. Among literary consequences were the informing of the novel with a new seriousness of purpose and the emergence of the essay—critical, psychological, philosophical—to a new importance. These novelists and essayists constitute the "generation of '98." Rarely in agreement in their diagnosis or on the treatment proposed for the patient, they reinvigorated Spanish letters to a point where these again commanded respect abroad and, as part of the process, laid Spain open to English, French, German, Russian and other influences on a scale never known before.

Miguel de Unamuno (1864–1936) dominated the literary scene for a generation. In him the national problem was intertwined with his obsession with the personal problem of immortality. The former he studied acutely in his essays *En Torno al casticismo* and in the *Vida de don Quijote y Sancho*, the latter in his most important work, *Del Sentimiento trágico de la vida*. A deliberately provocative rather than systematic thinker, he aimed at sowing spiritual disquiet. The novel was to him an extension of the essay, as the theatre was of the novel, a medium for discussion of the fundamentals of personality, hence stripped of externals and as timeless as the *Soledades* of Góngora (*Niebla, Abel Sánchez, Tres Novelas ejemplares y un prólogo*).

Azorín (pen name of José Martínez Ruiz, 1873–) concerned himself with the re-interpretation of earlier literary values and of the Spanish countryside (*El Alma castellana, La Ruta de don Quijote, Clásicos y modernos*). An artist in criticism and a miniaturist of fine sensitivity, he contributed powerfully to the deflation of the hollow rhetoric that had vitiated much of 19th-century writing. José Ortega y Gasset (1883–1955) brought to the analysis a mind formed in Germany in the study of neo-Kantian thought. From criticism and psychology (*Meditaciones del Quijote, El Espectador*) he passed to national problems (*España invertebrada*), then to international (*El Tema de nuestro tiempo, La Rebelión de las masas*). His *Revista de Occidente*, founded in 1923 and directed to keeping Spain abreast of contemporary foreign thought, proved a major intellectual force.

Pío Baroja (1872–1956) repudiated tradition, religion and the cult of the individual and sought to be an arch-European, fixing his gaze on the future and making a cult of action for social purposes. The *Memorias de un hombre de acción* (a series of novels set in the early part of the 19th century) are thus less characteristic than his numerous trilogies on the contemporary scene (*La Raza, La Lucha por la vida, Agonías de nuestro tiempo*). Insensitive to art for art's sake, Baroja often depresses the reader through his concern to arouse discontent with material conditions; but his novels, full of rude vigour and a fierce independence of spirit, are seedbeds of ideas. There was rude vigour too, together with richness of colour and greater narrative skill, in Vicente Blasco Ibáñez (*q.v.*; 1867–1927), a survivor into the 20th century of the regional novelist (*Flor de mayo, La Barraca*, on his native Valencia) who wrote also on contemporary social problems (*La Bodega, La Horda*) from the standpoint less of the generation of '98 than of the anarchist and who won international renown with novels of World War I (*Los cuatro Jinetes del Apocalipsis, Mare nostrum*). Ricardo León (1877–1943) was a conscious stylist who sought to recapture the traditional Christian values of a society threatened with materialism (*Casta de hidalgos, Alcalá de los Zegríes, El Amor de los amores*).

The 20th Century.—In Ramón Pérez de Ayala (1880–1962) the novel is at once a satisfying art form and a forum for philosophical discussion tinged with Asturian humour. *Troteras y danzaderas* recaptures something of the vein of Juan Ruiz. *Belarmino y Apolonio*, a projection of the old debate between faith and reason, makes its characters almost symbolic, as do *Tigre Juan* and its continuation *El Curandero de su honra* on the traditional theme of honour. Gabriel Miró (1879–1930) was a lyric poet in prose, with gifts of description and evocation (*Figuras de la pasión del Señor, Años y leguas*) which retard the action of his novels (*La Novela de mi amigo, El Libro de Sigüenza, Nuestro Padre San Daniel*) but have given him the reputation of a supreme artist in words. In the work of nearly all these writers can be seen qualities and tendencies incompatible to some extent with the basic art of the novel, which as a kind was already in decline in 1936. The civil war of 1936–39 dealt a heavy blow to imaginative writing in general. Novelists of promise who made their name later were: Camilo José Cela, *La Familia de Pascual Duarte* (1942); Carmen Laforet, *Nada* (1944); and Juan Antonio de Zunzunegui, *La Quiebra* (1947).

The drama also achieved new vigour with the new century under the stimulus of Jacinto Benavente (1866–1954), another Nobel prize winner (1922) and Spain's greatest playwright since Calde-

rón. With him the drama returned to reality, after the extravagances of Echegaray, by way of social criticism, declamatory verse giving way to prose, melodrama to comedy, formula to experience, impulsive action to dialogue and the play of minds. After a first period of effervescence and irony, Benavente showed a preoccupation with aesthetics as he mastered his medium, and later with ethics. *Los Intereses creados, Los Malhechores del bien, La Noche del sábado, La Malquerida* are evidence of the extent to which he broadened the scope of the theatre. By contrast, Gregorio Martínez Sierra (1881–1947) remained true to a more traditional Spanish atmosphere, in which idyllic sentiment, though deep-rooted always in Spanish realities, breathed a sensuous joy in nature and in life (*Teatro de ensueño, Canción de cuna, El Reino de Dios, Nuestra Esperanza*). The brothers Serafín (1871–1938) and Joaquín (1873–1944) Álvarez Quintero, writing as one, were pure *costumbristas*. If they contributed nothing new to dramatic art, the infinite *joie de vivre* of their long succession of Andalusian comedies never failed to entertain (*La Puebla de las mujeres, El Centenario, La Señora de Alfaqueque*). Among other dramatists of note are Manuel Linares Rivas (*El Abolengo*, 1904, *La Garra*, 1914) and José López Pinillos ("Pármeno"), author of *Esclavitud* (1918) and *Embrujamiento* (1923).

The poetic drama was cultivated with success by Eduardo Marquina (1879–1946), who found inspiration in the heroic past (*Las Hijas del Cid, En Flandes se ha puesto el sol*); by Jacinto Grau (1877–1958; *El Conde Alarcos, El Hijo pródigo*); by Francisco Villaespesa (*El Alcázar de las perlas*, 1911); by Ramón del Valle-Inclán (*Cuento de abril*, 1910; *Voces de gesta*, 1912); and by Luis Fernández Ardavín (*La Dama del ermiño*, 1921; *La Maja*, 1928). Federico García Lorca (*c.* 1898–1936) stands far above any of these. His drama (*Bodas de sangre, Yerma, La Casa de Bernarda Alba*) was poetic in more than the formal sense, presenting the elemental passions with an intensity that made the characters poetic symbols of man's tragic impotence to arrest his fate.

One other vein of dramatic activity, commonly held to fall below criticism, adds persisting vitality to a long and honourable descent. It is the *género chico* ("small kind"), heir to earlier *paso, entremés* and *sainete*, which within the limitations of the short sketch, frequently with music, achieves a great variety of dramatic effect and for almost a century has charmed the *vulgo* whose way of life supplies it with setting and theme. Ricardo de la Vega (1839–1910) and Carlos Arniches (1866–1943) span that century between them and are outstanding in the kind, which the brothers Quintero likewise notably enriched.

Rubén Darío, Latin America's greatest poet, brought with him to Spain in 1892 the movement known as *modernismo*, of which he was the leading exponent. In its ideas modernism was a reaction against 19th-century bourgeois materialism and a search for other and, more specifically, for aesthetic values in life. In literature it was a reaction against naturalism drawing heavily in its poetic norms on French symbolism and parnassianism alike, with elements of a revived romanticism. Endowed with great technical virtuosity, Darío enriched beyond measure the musical resources of Spanish verse, notably in the sphere of rhythm; and under his influence poetry in Spain experienced a renewal comparable only with that effected by Garcilaso almost four centuries earlier.

Though so nearly coincident in time with the movement represented by the "generation of '98," modernism had scarcely any contact with it. Initially it stood for the cult of the ivory tower, wherein the poet took refuge from his fellows and from life and sought self-realization in the pursuit of esoteric beauty. It produced, however, one prose writer, the novelist Ramón del Valle-Inclán (1869–1936), whose four *Sonatas* typify the four seasons of his hero's life. Valle-Inclán later wrote historical novels (the trilogy *La Guerra carlista*, the two books of *La Rueda ibérica*), but here too his vision was purely aesthetic. For the rest, the modernists were poets; and predominantly they were Andalusians, the old contrast still holding between the attraction of form, which interested them, and that of content, with which the northerners of the "generation of '98" were concerned.

Juan Ramón Jiménez (1881–1958), beginning to write under the aegis of Darío, soon discovered poetry to be for him not a withdrawal but the whole of life, reality itself, and his style, initially ornate, became more and more bared, with a turning to blank verse, as he strove to capture the poetic essence of each passing moment. Using and gradually winnowing all the resources of modernism, he became the poets' poet and a powerful influence on the younger generation. The brothers Machado exemplify well the Andalusian-Castilian dichotomy: Manuel (1874–1947), always at heart of his native Seville, was the modernist, with a detached Bohemian attitude to life and the attitude to poetry of the craftsman; Antonio (1875–1939) early lost his heart to Soria, in Old Castile, and his best poetry, written with an infinite simplicity, almost poverty, of artifice, dealt with the recesses of the poet's soul or with the unchanging soul of Castile. Unamuno, Valle-Inclán and Pérez de Ayala all figure in the tale of modern poetry, though their poems were but the prolongation in another medium of their prose works. Of the three Unamuno, the least accomplished technically, is the most rewarding as he seeks to distil into verse the intensity of his spiritual striving.

The poetic stature of García Lorca grew steadily after his death in 1936. As in his plays, he is a poet of fundamentals and has shown of what starkness of feeling and effect the *copla* and *romance* of secular tradition are still capable. His *Romancero gitano* (1928) and *Poema del cante jondo* (1931) reveal the potency of the inspiration that he drew from Andalusian ways and folklore; the former won a wider public throughout the Spanish-speaking world than any other work of poetry of the 20th century. The other major poets of the middle of that century were still those of the pre-civil war generation, contemporaries of García Lorca and heirs, as he was, to Darío and Juan Ramón Jiménez: Pedro Salinas (1892–1951), Jorge Guillén (1893–), Gerardo Diego (1896–); Dámaso Alonso (1898–), Vicente Aleixandre (1900–), Rafael Alberti (1902–), Luis Cernuda (1904–1963) and Manuel Altolaguirre (1906–). The civil war dealt hardly with them too, driving some into exile, stifling the inspiration of others; in the main, their best work was already done. The immediate postwar generation was characterized by its minority groups, fugitive reviews and dispersive tendencies.

BIBLIOGRAPHY.—J. Simón Díaz, *Bibliografía de la literatura hispánica* (Madrid, 1950 ff.); J. Fitzmaurice-Kelly, *A New History of Spanish Literature* (Oxford, 1926); J. Cejador y Frauca, *Historia de la lengua y literatura castellana*, 14 vol. (Madrid, 1915–22); G. T. Northup, *An Introduction to Spanish Literature* (Chicago, 1925); L. Pfandl, *Geschichte der spanischen Nationalliteratur in ihrer Blütezeit* (Freiburg, 1929; Spanish trans., Barcelona, 1933); A. F. G. Bell, *Contemporary Spanish Literature* (New York, 1925) and *Castilian Literature* (Oxford, 1938); E. Mérimée, *A History of Spanish Literature*, Eng. ed. by S. G. Morley (New York, 1930, London, 1931); A. Valbuena Prat, *Historia de la literatura española*, 3rd ed., 3 vol. (Barcelona, 1940); J. Hurtado and A. González Palencia, *Historia de la literatura española*, 6th ed. (Madrid, 1949); G. Diaz-Plaja, ed., *Historia general de las literaturas hispánicas*, 3 vol. (Barcelona, 1949 ff.); G. Brenan, *The Literature of the Spanish People* (Cambridge, 1951); J. Marías and G. Bleiberg, ed., *Diccionario de literatura española* (Madrid, 1949); C. Barja, *Libros y autores clásicos* (Brattleboro, 1929), *modernos* (Los Angeles, 1933), *contemporáneos* (New York, 1935); S. de Madariaga, *The Genius of Spain and Other Essays on Contemporary Spanish Literature* (Oxford, 1923); J. B. Trend, *Alfonso the Sage and other Spanish Essays* (Boston and London, 1926); M. Menéndez y Pelayo, *Estudios y discursos de crítica histórica y literaria*, 7 vol. (1941–42), *Orígenes de la novela*, 4 vol. (Madrid, 1905–15; reprinted 1943), and *Historia de las ideas estéticas en España*, 5 vol. (Madrid, 1883–91; reprinted 1946–47).

CATALAN LITERATURE

Early Poetry.—The Catalan language is properly to be regarded as a branch of Peninsular rather than of southern Gallo-Romance. It shows nonetheless many traces of kinship with Provençal, and the literature in its origins used the *langue d'oc* and all the poetic forms cultivated by troubadours north of the Pyrenees: Guillem de Bergadá, Hug de Mataplana, Ramon Vidal de Besalú, Guillem de Cervera, Cerverí de Girona were genuine Provençal poets in the same sense as were those of Limousin, Quercy or Auvergne. Ramon Vidal (late 12th-century) was preceptist as well as poet; his *Dreita Manera de trobar* became the code for Catalan poets writing in a tongue based supposedly on Limoges and known as *lemosí*. This conventional language of poetry was not proof against the contagion of the spoken tongue,

and there resulted a composite idiom well illustrated in Raimon Lull (Ramón Llull, *c.* 1232–1315) who, for all the earnestness of his later religious verse (*Lo Desconort*, of 69 12-line stanzas in monorhyme; *Lo Consili*), never wholly divested himself in his poetry of his troubadour beginnings. Ramon Muntaner (1265–1336), writing plain Catalan in prose, is still a troubadour in verse. In the later part of the 14th century, there was apparent a gradual waning of Provençal influence and a turning instead for inspiration to northern France, whence derived a new cult of long narrative poems on novelesque themes, including the *matière de Bretagne*. With the themes came the metre, known as *noves rimades*, a sequence of octosyllabic rhymed couplets. There exist poems of this class by Pere March, Guillem de Torroella, Bernat Metge; other poems, anonymous, include the *Llibre dels set savis de Roma*, of over 3,000 lines, the *Facet o Libre de cortesia* of half that length, and several which carry the new interest in the *langue d'oïl* even to the incorporation on occasion of passages of French poetry. Another form, the *codolada*, was native in origin; used also for narrative but more especially for satire, it combined, again in rhyming couplets, an eight-syllable line with a four-syllable one.

The 15th Century.—The 15th century is the great period of Catalan poetry. It dates from the institution at Barcelona in 1395, under the auspices of John I of Aragon (1387–95), of a "Consistory of the *Gay Saber*" on the model of that of Toulouse, with "jocs florals" which produced in themselves little poetry of note but typified a new and propitious atmosphere of royal encouragement. This continued under Martin I (1395–1410) and Ferdinand I (1412–16) and contributed much to the emancipation of style from alien influences. As the century advanced, Valencia emerged as a new focus of literary activity; many of the great names of the age, in prose and verse, were Valencians.

With Joan Berenguer de Masdovelles, Francesc Ferrer, Pere Torroella, Pau de Bellviure, Antoni Vallmanya and, above all, the Petrarchan Jordi de Sant-Jordi and Ausiàs March there developed a school which, as regards versification, is noted for its almost exclusive use of eight-line decasyllabic *cobles* with crossed or "chained" rhymes and a final four-line *tornada* and, in inspiration, shows a turning away from France to Italy, where from 1443 an Aragonese court held sway in Naples. The *Cants d'amor* and *Cants de mort* of Ausiàs March, (1379–1459), the most inspired and profound, if also the most difficult, poet of his age in the peninsula, contain the finest verses ever written in Catalan; they exerted great influence in 16th-century Castile and still do on modern Catalan poets. Very different in class and spirit is the *Spill de les dones* of Jaume Roig, a bitter and caustic satire upon women of more than 16,000 four-syllable lines which derives a very piquant interest from its really authentic element, the vivid picture of contemporary Valencian life and manners. Joan Roiç de Corella is the third great Valencian lyricist, heir to Sant-Jordi and March and perhaps the best representative of the Renaissance spirit, which flowered more fully there than in Barcelona. After this bright period of efflorescence, Catalan poetry faded rapidly, the victim largely of political circumstance. From the union of Aragon with Castile there resulted a predominance of Castilian throughout Spain which inflicted a death blow on Catalan letters. The fact that it was a Catalan, Joan Boscà i Almogàver (in Castilian Juan Boscán), who inaugurated in the Castilian tongue a new school of poetry, and that Castilians regard him as a landmark in the history of their Renaissance muse, is important and revealing; by the year of publication of Boscán's works (1543), Catalan poetry had already been dead for half a century.

Mediaeval Prose.—The earliest prose works in Catalan are later than the poems of the first troubadours of the Provençal school. Though the oldest document dates from *c.* 1100 (the text of an oath imposed on certain barons by a bishop of Urgel), and the *Homilies d'Organyà* are attributed to the 11th century, literary prose goes back no further than the close of the 13th. Beginning as history, written close to, and often by prime movers in, the events that it describes, it has the advantage of being entirely original, the language being the true everyday speech found in charters from about the time of James I's accession to the Aragonese throne in 1213. Its four great 14th-century chronicles are the crowning ornament of mediaeval Catalan prose. Two of them—that of James I, apparently compiled shortly after his death in 1276 with the help of memoirs dictated by himself, and that of Ramon Muntaner (1265–1336), relating at length the expedition of the Catalan company to the Morea and the conquest of Sardinia by James II—are distinguished alike by the artistic skill of their narration and by the quality of their language, inviting comparison with Villehardouin, Joinville and Froissart. Bernat Desclot's chronicle deals particularly with the reign of Peter the Great (1276–85). The account of the reign of Peter the Ceremonious (1336–87) is ascribed to Bernat Descoll but was planned and revised by the king himself. The four works collectively present the perfect expression of the spirit and the greatness of mediaeval Catalonia.

Raimon Lull, the *Doctor Illuminatus*, born in Majorca, stands on an eminence apart through his vast encyclopaedic production, totalling about 180 works in Catalan, Arabic and Latin and covering every branch of mediaeval knowledge and thought. His *Libre de contemplació en Déu* (1272) marks the beginning of Catalonia's golden age of literature; an exhaustive theological treatise, it is also a mine of information on contemporary society. *Blanquerna*, his literary masterpiece, is a social-biographical romance which lays the foundations of Catalan fiction. *Fèlix de les meravelles del món* and the *Libre del orde de cauayleria* are other doctrinal works with a narrative framework. Inserted in the *Blanquerna* is the *Libre d'amic e amat*, a brief masterpiece of mysticism. A century later, Francesc Eximenis (1340–1403) represents something of the same intellectual amplitude, if without the originality or driving force, in *Lo Crestià*, planned as a comprehensive work in 13 books on theology, morals and politics for the layman but surviving in 4 books which were perhaps the only ones written. From its pages emerges a vivid picture of the mediaeval scene. The *Libre de les dones*, at once a book of devotion and a manual of domestic economy, is similarly informative on women's way of life at the time.

At the close of the 14th century Bernat Metge inaugurates the so-called "classical age." Well versed in Italian literature, he translated in *Valter e Griselda* Boccaccio's story of Griselda from the Latin version of Petrarch; in *Lo Somni* (1398), a work which clothes scholastic learning with much poetic imagination, he achieved the stylistic masterpiece of Catalan prose. Joanot Martorell's chivalric romance *Tirant lo Blanc* (1460) was notable in its kind for the theme, drawn from Muntaner, of the real adventures of the Catalans in the near east (hence the sparing of it in the famous scrutiny of Don Quixote's library). The anonymous *Curial e Güelfa*, of the end of the century, draws similarly on Desclot and is the only other Catalan novel in this vein. The beginnings of the drama are represented by the *Misteri d'Elch*, a 15th-century Assumption play based on an earlier *Representació;* sung throughout to traditional music, it is still performed on Aug. 14 and 15 every year at Elche.

The 16th-18th Centuries.—With the loss of political independence, literary and linguistic independence was lost, and Catalan fell to the level of a patois, kept alive over the next 200 years in the countryside and in the pulpit. The 16th century furnishes a single poet worthy of the name: Pere Serafí, some of whose *Cants d'amor* (1565), written in imitation of Ausiàs March but less obscure, are graceful enough to merit remembrance. In prose only scholars, chiefly antiquaries and historians, still wrote in Catalan; such were Miquel Carbonell, compiler of the *Chròniques d'Espanya* (1513), Francesc Tarafa (*Crònica de cavallers catalans*) and Anton Beuter, of whose *Crònica general de toda España* the first part (dealing with Valencia) appeared in Catalan in 1538. Forty years of research and abundant documentation give interest to the *Crònica universal del Principat de Cathalunya* of Jeroni Pujadas, of which likewise only the first part (1609) was written in Catalan. Thereafter the eclipse was almost complete. Catalan remained only as the language of folk songs and ballads; in these (collected in the *Romancerillo catalán* [1853] of M. Milà i Fontanals) it lived on until the reawakening at the beginning of the 19th century.

The Renaixença.—In 1814 appeared the *Gramàtica y apologia de la llengua cathalana* of Josep Pau Ballot i Torres, a forerunner

of the literary and linguistic renaissance which marked the age of romanticism in Catalonia. The pioneers of the rehabilitation of Catalan for literary purposes soon discovered, however, the inadequacies for the expression of their spiritual and intellectual needs of the ancient tongue that they sought to resurrect. The very popularity of the satirical poems and farces of a Josep Robreño (1780–1838) was a serious obstacle, because of the corruption of the language in which they were written, as was the later advocacy by Frederic Soler (1839–95) of "Catalan as now spoken" in opposition to literary refinement. The development of modern Catalan literature has been directly dependent on the studied enrichment and purifying of the language as a vehicle for contemporary thought; in this the Institut d'Estudis Catalans, founded in Barcelona in 1907, played a notable part. B. C. Aribau's patriotic *Oda* (1833) and the poems of Joaquim Rubió i Ors (1818–99) and Victor Balaguer (1823–1901) prepared the way, as did the reinstitution in 1859 of the *jocs florals*, for the imagination and mysticism of Jacint Verdaguer (1845–1902), a great epic poet (*L'Atlàntida, Canigó*) whose gift is still essentially lyrical. Miquel Costa i Llobera (1854–1922) cultivated with much distinction a classical perfection of form; formal beauty is also the mark of his fellow-Majorcan Joan Alcover (1854–1926). In Joan Maragall (1860–1911) Catalonia found its first truly great modern poet, one who looked forward rather than back and whose poetry is suffused with deep spiritual intent. In this spiritual quality, as also, if less happily, in his theory of inspiration as absolving from the need for perfection of form, he exerted a powerful influence on later poets. Josep Carner (1884–), Josep-Maria López-Picó (1886–) and Josep-Maria de Sagarra (1894–1961) are outstanding names, among many of distinction, in the poetry of the 20th century.

The foundations of modern Catalan prose were laid by the critical writings of Rubió i Ors, Francesc Pi i Margall, president of the Spanish republic of 1873, and Josep Torras i Bages, archbishop of Vich (*La Tradició catalana*, 1892). Fiction reached a modest level of achievement late in the century with Josep Pin i Soler (1842–1927), Narcís Oller (1846–1930) and Emili Vilanova (1840–1905). One of the best and most influential writers in prose was the essayist Eugeni d'Ors ("Xenius," 1882–), part of whose *Glossari* (1906–17) had already been translated into Castilian before the author abandoned Barcelona for Madrid and began himself to write in that tongue. His philosophical novel *La ben Plantada* is one of the most notable works in modern Catalan literature.

In the theatre Catalan dramatists have produced plays of considerable originality. Angel Guimerà (1847–1924) achieved a European reputation with *Terra baixa*, which inspired a German and a French opera (*Tiefland, La Catalane*) and was translated into 12 languages. Beginning with historical tragedy (*Gala Placidia, Judith de Welp*), Guimerà turned to rural drama in *La Boja*, to comedy in *La Baldirona* and *La Sala d'espera* and to modern tragedy in *Martha Rosa* and *La Festa del blat*. The many social dramas of Ignasi Iglésias, inspired originally by the early works of Gerhardt Hauptmann, included one nearly a masterpiece, *Els Vells;* passing later under the influence of Maeterlinck (*Focfollet, Cendres d'amor*), he eventually cultivated pure comedy (*Girassol*). Adrià Gual, author of several works of fantasy, did his best work as director of the *Teatre Intim*, founded in Barcelona in 1898, which familiarized the public with the great drama of all countries and ages. The painter Santiago Rusiñol achieved repute in the theatre too with a number of vivid and imaginative plays (*L'Alegria que passa, La bona Gent, El Místic*).

Hardly had Catalan literature thus won itself once more a place in the sun, when there supervened two calamities, the dictatorship (1923–29) of Primo de Rivera, who banned the use of any language other than Castilian, and the civil war of 1936–39. The expatriation of many intellectuals led to the development of a considerable publishing activity in Catalan in Mexico City and Santiago de Chile.

BIBLIOGRAPHY.—J. Massó i Torrents, *Revista de bibliografia catalana* (Barcelona, 1900–07) and "Bibliografia dels antics poetes catalans," *Anuari de l'Institut d'Estudis Catalans* (Barcelona, 1913); J. Amade, *Bibliographie critique pour l'étude des origines et des premières manifestations de la renaissance littéraire en Catalogne au xixe siècle* (1925); M. Garcia Silvestre, *Història sumària de la literatura catalana* (1932); M. de Montoliu, *Manual d'història crítica de la literatura catalana moderna* (to 1860; 1922). (W. C. An.)

SPANISH MOSS, the name given to an epiphytic, herbaceous plant (not a true moss), *Tillandsia usneoides*, of the bromelia family (Bromeliaceae), that hangs in long festoons from the branches of trees in tropical and subtropical America forests, being especially conspicuous on the live oak in the southern United States. The whole plant is covered with scaly hairs for the absorption of water, giving it a gray colour. It is used, after processing, as an excellent substitute for horsehair.

Ramalina reticulata, a lichen called Californian Spanish moss, grows supported on trees in humid valleys of the western coastal range of the United States.

SPANISH SUCCESSION, WAR OF THE, the name given to the general European war which began in 1701 and ended with the treaties of Utrecht and Rastatt in 1713–14. Its monotonous round of marches and sieges is often quickened by the genius of three great soldiers, Marlborough, Eugene and Villars, while Peterborough and Galway, Catinat and Vendôme, though less highly gifted, were men of unusual and conspicuous ability. As usual in these wars, manoeuvres, threats and feints played the principal part in field warfare. The tangible assets, at the conclusion of peace, the effective seizure of fortresses and provinces was in most cases the principal object with which kings and princes made war. Nevertheless, the object of this war, at least in the case of England and Holland, was less to add a few cities and districts to their own domains than to cripple the power of Louis XIV. The ambition of the *Grand Monarque*, by placing on the throne of Spain his grandson Philip, had brought into politics the fear not merely of a disturbance but of an entire overthrow of the balance of power. Thus the instrument of Louis's ambition, his magnificent army, was (above all for England) an object in itself and not merely an obstacle to the attainment of other objects. Many of the allies, however, had good reason to fear for their own possessions, and others entered the alliance with at least the hope of acquiring a few material gains at small expense. On the side of the allies therefore, throughout the war, there was a perpetual struggle, between offensive activity and defensive passivity. The action of Louis XIV in the matter of the Spanish succession had been foreseen, and William III of England devoted his last years to providing against the emergency by the formation of a coalition to deal with it, and the production of a claimant for the Spanish throne, the archduke Charles. The coalition naturally grew out of the Grand Alliance (*see* GRAND ALLIANCE, WAR OF THE) and consisted of Austria, some of the German states, Great Britain, Holland, Denmark and Portugal. On the other side Louis XIV was supported by Spain—where Philip, recognized as heir by the dying Charles II, had been promptly installed—Bavaria and Cologne. A doubtful ally was the duke of Sâvoy, whose policy was to secure and aggrandize himself by adhering at each moment to the stronger party.

Italy, Chiari 1701.—The war began, to all intents and purposes, with the handing over of the fortresses in the Spanish Netherlands to the French in March 1701. England and Holland at once began their preparations, but neither state was able to put an army in the field in the year—England because its peacetime army was insignificant, and Holland because it dared not act alone. In Italy, however, the emperor took the initiative, and an Austrian army under Prince Eugene, intended to overrun the Spanish possessions in the peninsula, assembled in Tirol in the early summer. The French army under Catinat blocked the defile of Rivoli, but Eugene reconnoitred passages over the mountains between Roveredo and the Vicenza district in neutral Venetian territory. On May 27, taking infinite precautions as to secrecy and requesting the Venetian authorities to offer no opposition so long as his troops behaved well, Eugene began his march by paths that no army had used since Charles V's time, and on the 28th his army was on the plains. Catinat was surprised, for he had counted upon Venetian neutrality, and when, in the search for a passage over the lower Adige, Eugene's army spread to Legnago

and beyond, he made the mistake of supposing that the Austrians intended to invade the Spanish possessions south of the Po. He thinned out his line to cover all the front as far as that river. But Eugene found an unguarded spot. With the usual precautions of secrecy (deceiving even his own army), he crossed the lower Adige in the night of July 8–9. Catinat at once concentrated his scattered army backward on the Mincio, while Eugene turned northward and regained touch with his old line of supply, Roveredo-Rivoli. For some time Eugene was in great difficulties for supplies, as the Venetians would not allow his barges to descend the Adige. At last, however, he made his preparations to cross the Mincio close to Peschiera and well beyond Catinat's left. This was executed on July 28 and Catinat fell back quietly to the Oglio. But his army resented his retreat before the much smaller force of the Austrians and, early in August, his rival Tessé reported this to Paris, whereupon Marshal Villeroy, a favourite of Louis, was sent to take command. The new commander was the least competent of all the French senior officers. Before long he attacked Eugene in a well entrenched position at Chiari (Sept. 1) and was thoroughly defeated. In the winter a fresh mishap befell the French. Eugene surprised Cremona on the night of Feb. 1, 1702, and, after a confused fight, drew off, taking with him Villeroy as a prisoner. The rest of the French army retreated, while Eugene quietly resumed his winter quarters and his blockade of Mantua.

DRAWN BY BARTHOLOMEW & SONS, LTD., EDINBURGH

Marlborough's First Campaigns.—With the year 1702 the real struggle began. Villars and one or two others of Louis's counsellors urged the king to concentrate his attention on the Rhine and the Danube, where, they pointed out, was the centre of gravity of the coalition. This advice was disregarded, and with political aims, the largest French army was employed on the side of the Meuse, while the Rhine front was entrusted to smaller forces acting on the defensive. In Italy the balance of power remained unchanged, except that one of Louis's best generals, Vendôme, was sent to replace the captured Villeroy. In the Low Countries, Ginkel, earl of Athlone, the interim commander of the allies (English, Dutch and minor German states), was at the outset outmanoeuvred by the French (Boufflers), and the momentary threat of a French invasion had a lasting effect on the Dutch authorities, whose timidity thereafter repeatedly ruined the best-laid schemes of Marlborough, who was obliged to submit to their obstruction and their veto. This handicap, moreover, was not the only one under which Marlborough suffered. Unless it is realized and borne in mind that the great captain was struggling against factiousness and intrigue in England and against jealousies, faintheartedness and disagreements among the states who lent their contingents to his miscellaneous army, the measure of his achievements in ten years seems small. But in fact it was marvellous. Under 18th-century conditions of warfare, and with an army so composed that probably no other man in Europe could have held it together at all, obstructed and thwarted at every turn, he yet brought Louis XIV and France to the very edge of ruin.

In this theatre of war the French, in concert with the garrisons of the Spanish Netherlands, had fortified a line of defense more than 70 mi. long from Antwerp to Huy, as well as another line, longer but of only potential importance, from Antwerp along the Scheldt-Lys to Aire in France. Beside the "lines of Brabant" Boufflers held all the Meuse fortresses below Huy except Maestricht. Marlborough concentrated 60,000 men (of whom 12,000 only were British) about Nijmwegen in June, and early in July, having made his preparations, he advanced directly by Hamont on Diest. Boufflers hastily fell back, in order to regain the Brabant lines. Marlborough, with the positive object of bringing his opponent to battle at a disadvantage, won the race and awaited the arrival of Boufflers's tired army to strike it a paralyzing blow. But at the critical moment the Dutch deputies forbade the battle, content to see the army that had threatened Holland with invasion driven off to a safe distance without bloodshed (July 22). After this experience Marlborough thought it prudent to pacify the Dutch by besieging the Meuse fortresses, several of which fell in rapid succession (Sept.–Oct.). Boufflers fell back within the defended area of the lines of Brabant, and the campaign closed with the capture of Liège by the allies (Oct. 12). Marlborough was created a duke on his return to England in November.

In the next two years Bavaria was the centre of gravity of the French operations, and only campaigns of the methodical and non-committal kind were planned for Italy and the Low Countries. In the Netherlands the French field army was behind the lines of Brabant, the Spanish troops in the lines of Flanders (Antwerp-Ghent-Aire). Together the two considerably outnumbered Marlborough (90,000 against 50,000), but the duke managed to be first in the field. After capturing Bonn in May, Marlborough's plan was to break the immensely long line of defence of the French and Spaniards by the capture of Antwerp. One Dutch corps under Coehoorn was to assemble in the Sluys-Hulst region, and another under Opdam at Bergen-op-Zoom while Marlborough, after manoeuvring Villeroy's field army out of the way, was to join them before the fortress. Marlborough executed his own share of the movement with his usual skill, but the Dutch generals enabled the French to emerge from the manoeuvre with a handsome victory. Great projects were now entertained by the French, nothing less than the capture of Vienna by a Franco-Bavarian-Hungarian army being the intention. The elector of Bavaria insisted that Villars should cross the Black Forest and join him, which Villars was unwilling to do thus early in the year, as two-thirds of his officers were as usual on leave or detached on recruiting duties. Courtier though he was, the marshal would not stir even in spite of the king's orders until he was ready. At the end of April, leaving Tallard alone to defend Alsace against the margrave of Baden, Villars plunged into the defiles of the Black Forest and on May 8 joined the elector at Ebingen. All seemed favourable for the advance on Vienna, but at the last moment the elector half repented of his alliance with

the enemies of Germany and proposed instead a junction with Vendôme in Italy by way of Tirol. This proposal came to nothing, the Tirolese were soon roused to revolt by the misconduct of the ill-disciplined Bavarians, and Vendôme, who, like Luxembourg, was a giant in battle and a sluggard in camp, would not stir.

Hochstett, 1703.—Villars had posted a protective force at Ulm to contain the Margrave's army should it turn back upon him, and this, after an engagement at Munderkingen (July 31) induced the cautious Louis to return to the Rhine. Five weeks later, however, the Margrave returned in full force, and moving by the right bank of the Danube reached Augsburg on Sept. 6. The elector, returning from his futile Tirol expedition, had already rejoined Villars at Dillingen, and the marshal persuaded him to attack the Austrian army that covered Vienna before the two imperial generals could join forces. The result was the battle of Hochstett (Sept. 20) in which the elector and Villars won a great victory, at a loss of only 1,000 men to the enemy's 11,000. Rarely indeed had an 18th-century general so great an opportunity of finishing a war at one blow. But even Villars saw no better use for the victory than the unimpeded junction of his own army and Tallard's and winter quarters in Württemberg. Though the grand advance on Vienna came to nothing, the French won important successes and established an army in Bavaria. In Italy, on the other hand, Vendôme, although no longer opposed by Eugene, achieved nothing.

Rhine and Danube Campaign, 1704.—The campaign of 1704, though in the Low Countries and in Italy practically nothing was done, is memorable for what was perhaps the greatest strategical operation in the 18th century, Marlborough's march to the Danube. At the outset the elector and Marsin (Villars' successor) were on the Iller, between Ulm and Memmingen, Tallard between Strasbourg and Landau, Villeroy as usual between the Brabant lines and the Meuse. On the other side the Margrave Louis was in the Stockach-Engen region. Responsible for guarding the whole of the Middle Rhine as well as for opposing the elector he was weak everywhere, and his defence of the Rhine was practically limited to holding the "lines of Stollhofen," a defensive position near Bühl in Baden. With Breisach and Kehl in their own hands, the French were more or less closely in touch with their comrades in Bavaria, and Tallard convoyed a large body of recruits for Marsin's army through the Black Forest defiles. But in doing so he lost most of them by desertion, the Margrave's army dogged his march, and in fact no regular line of communication was established. Marlborough's purpose at any rate was quite definite—to transfer a large corps from the Low Countries to Bavaria and there in concert with the allies in that quarter to crush the elector decisively, but rather than be burdened with Dutch counsellors he chose to forgo the assistance of the Dutch troops. These were left under Overkirk to defend the Meuse, and English and English-paid troops alone took part in the venture.

Marlborough's March to the Danube.—Marlborough calculated that as he progressed up the Rhine the French would collect to prevent his crossing, instead of themselves passing over to join the elector and Marsin. Thus the expedition would reach the Neckar mouth, without its true purpose being suspected, and once there Marlborough would vanish from the ken of the defenders of the Rhine, to reappear on the Danube where he was least expected. On May 12 the army crossed the Meuse at Ruremond, on the 23rd it reached Bonn, on the 29th Mainz. On June 1 the puzzled French noted preparations for bridging the Rhine at Philipsburg. But two days later the English had turned to their left into the valley of the Neckar. The endurance of Marlborough's corps, as displayed in the long march from Ruremond, was not the least extraordinary feature of the operation. For 18th-century troops such performances were generally provocative of desertion, and involved the ruin of the army that attempted it. But Prince Eugene, we are told, was astonished at the fine condition of the army. On the French side meantime all was perplexity, and it was not until a week after the Margrave and Marlborough had joined, that Villeroy's main body from the Meuse started for Alsace to watch Eugene's corps, or rather the Stollhofen-Buhl position. This meant conceding both the initiative and the superiority in numbers to Marlborough.

Campaign on the Danube, 1704.—The duke had now manoeuvred himself with brilliant success from one theatre of war to another, and had secured every advantage to himself. From before Ulm he sidled gradually along the north side of the Danube in the hope of finding an unguarded passage. He and the Margrave exercised the general command on alternate days, and when on his own day he arrived opposite Donauwörth, knowing Louis's caution, he thought that direct attack was better than another two days' extension to the east. Moreover he needed a walled town to serve as a magazine. In the late afternoon of July 2 the army was flung, regardless of losses, against the entrenched hill of the Schellenberg at Donauwörth, where the elector had posted a strong detachment. The attack cost 6,000 men, but it was successful, and of the 12,000 Bavarians on the hill only 3,000 returned to their main body, which had now moved from Ulm to Lauingen. Passing the river, the allies besieged and took the small fortress of Rain, and thence moved to the neighbourhood of Augsburg, thoroughly and deliberately devastating the country-side so as to force the elector to make terms.

There were now five armies in the field, two allied and three French. The centre of gravity was therefore in Villeroy's camp. If that marshal followed Tallard, even Eugene's junction with Marlborough would not give the latter enough force. If Tallard alone joined the elector and Eugene Marlborough, the game was in the hands of the allies. But none of the possible combinations of two armies against one was attempted by either side. Eugene when informed that Tallard was on the move slipped away from Villeroy to join Marlborough. Tallard and the elector, aware of Eugene's march, were content to join forces peaceably at Augsburg. Villeroy, in whose hands was the key of the situation, hesitated and finally tried in vain to detain Eugene (who was already far away) in the Stollhofen lines. The last stage of the campaign was brief. Marlborough and Eugene had in mind a battle, Tallard and Marsin a war of manoeuvre to occupy the few weeks now to be spun out before winter quarters were due. The two allied armies met in the Danube valley on Aug. 6. If the enemy remained on the south side Eugene was to cross, if they recrossed to the north bank Marlborough was to follow suit. The margrave Louis of Baden had been sent off to besiege Ingolstadt as soon as Eugene had come within a safe distance. When therefore the French and Bavarians were reported opposite Eugene on the north side, Marlborough crossed at once, and without waiting for the margrave the two great soldiers went forward. On Aug. 2 (old style; 13, new style) they attacked and practically destroyed the armies of Tallard, Marsin and the elector (*see* BLENHEIM, BATTLE OF).

Campaign of 1705.—The campaign of 1705 was uneventful and of little profit to either side. Marlborough's army had returned to the Low Countries. Villeroy had also returned to Brabant and retaken Huy. With him was the now exiled elector of Bavaria. On July 18, after a series of skilful manoeuvres, Marlborough forced the lines of Brabant at Elissem near Tirlemont, but not even the glory of Blenheim could induce the Dutch deputies to give him a free hand, or the Dutch generals to fall in with his schemes. King Louis was thus able to reinforce Villeroy betimes from Villars's Lorraine army, and the campaign closed with no better work than the razing of the captured French entrenchments. On the Rhine Villars, with a force reduced to impotence, carried on a spiritless campaign against the Margrave Louis. In Italy there was serious fighting. Here Vendôme's army engaged in the attempt to subdue Victor Amadeus of Savoy and was so far successful that the duke implored the emperor for aid. Eugene was sent with new reinforcements, opposed to which was a force under Vendôme's brother Philippe, called the Grand Prior. This man, a lazy dilettante, let himself be surprised by Eugene's fierce attack on the line of the Adda. The day was restored however, and the Austrians beaten off, thanks to Vendôme's opportune arrival and dauntless courage (battle of Cassano, August 16). Nevertheless, the subjugation of Piedmont was put off until the next year.

Ramillies, 1706.—1706 was a bad year for the French. At the very outset of the campaign in the Netherlands, Villeroy, hearing that some of the allied contingents that composed Marlborough's

army had refused to join, ventured out from his new defensive lines along the Dyle and moved towards Namur. Marlborough hurried forward to intercept him before he could reach the shelter of the Meuse fortresses and was able (May 12) not only to win but also to profit by the glorious victory of Ramillies on the 12th of May.

In Italy the campaign had, as before, two branches, the contest for Piedmont and the contest between the French forces in Lombardy and the Austrian second army that sought to join Victor Amadeus and Starhemberg. The latter, repulsed by Vendôme at Cassano, had retired to Brescia and Lake Garda; Vendôme followed, and in April 1706, profiting by Eugene's temporary absence, attacked the Imperialists' camps. His onset broke up the defence completely (battle of Calcinato, April 19), and he hustled the fragments of the Imperialist army back into the mountains, where Eugene had the greatest difficulty in rallying them. Until the middle of June Vendôme completely baffled all attempts of Eugene to slip past him into Piedmont. He was then, however, recalled to supersede Villeroy in Belgium, and his feeble successor entirely failed to rise to the occasion. Philip of Orleans, with Marsin, was besieging Turin. As soon as he knew of Vendôme's departure Prince Eugene emerged afresh from the mountains, and, out-manoeuvring the French in Lombardy, hurried towards Turin. Victor Amadeus, leaving the defence to the Austrian and Piedmontese infantry, escaped through the besiegers' lines and joined his cousin with a large force of cavalry. On Sept. 7 they attacked the French lines round Turin. Owing to the disagreements of their generals, the various corps of the defenders, though superior in total numbers, were beaten in detail. Marsin was killed and Orleans retreated ignominiously to Pinerolo. The battle of Turin practically ended the war in Italy.

Both in the north and in the south the tide had now receded to the frontiers of France itself. But it is from this very point that the French operations cease (though only gradually it is true) to be the ill-defined and badly-joined patchwork of forays and cordons that they had hitherto been. In the place of Tallards, Marsins and Villeroys Louis made up his mind to put his Villars, Vendômes and Berwicks, and above all the approach of the allied armies roused in the French nation itself a spirit of national defence which bears at least a faint resemblance to the great uprisings of 1792. For the gathering of this unexpected moral force 1707 afforded a year of respite.

Campaign of 1708.—In 1708 Eugene foresaw a shift of the centre of gravity and arranged with Marlborough to transfer his army which was ostensibly destined for the Rhine campaign to Brabant, since the French was markedly superior in numbers to Marlborough's and hardly inferior to Marlborough's and Eugene's combined. King Louis had put his young grandson and heir, the duke of Burgundy, at the head of the great army which assembled at Valenciennes, and gave him Vendôme as mentor. But the prince was pious, mild-mannered, unambitious of military glory and also obstinate, and to unite him with the fiery, loose-living and daring Vendôme, was, as Saint-Simon says, "mixing fire and water." At the end of May Vendôme advanced to engage Marlborough before Eugene should join him. As the French came on towards Brussels, Marlborough, who had concentrated at Hal, fell back by a forced march to Louvain. Vendôme having thus won the first move, there was a pause and then the French suddenly swung round to the west, and began to overrun Flanders, where their agents had already won over many of the officials who had been installed by the allies since 1706. Ghent and Bruges surrendered at once, and to regain for King Philip all the country west of the Scheldt it only remained to take Oudenaarde.

Oudenaarde.—The duke discussed the situation with Eugene. Marlborough was half inclined to wait for Eugene's troops, for he knew that Vendôme was no ordinary opponent, but Eugene counselled immediate action lest the French should escape, and relying on his own skill and on the well-known disunion in the French headquarters, Marlborough went forward. As he approached, the enemy gave up the siege of Oudenaarde and took up a position at Gavre, 7 mi. lower down the Scheldt. Here the dissensions in the French headquarters became flagrant. Vendôme began to place part of the army in position along the river while the duke of Burgundy was posting the rest much farther back as another line of defence. The allied main body, marching with all speed, crossed the Scheldt at all hazards. In the encounter-battle which followed (*see* OUDENAARDE) Marlborough separated, cut off and destroyed the French right wing. The French retreated in disorder on Ghent (July 11) with a loss of 15,000 men. Marlborough was prevented from carrying out his desire for a prompt move on Paris after Oudenaarde. For the moment Berwick, recalled from Alsace, manoeuvred about Douai, while Vendôme remained near Ghent, and between them Marlborough's and Eugene's armies devoted themselves to the siege of Lille. In this town, one of Vauban's masterpieces of fortification, the old Marshal Boufflers had undertaken the defence, and it offered a long and unusually gallant resistance to Eugene's army. Marlborough covered the siege.

On Dec. 8 the brave old marshal surrendered, Eugene complimenting him by allowing him to dictate the terms of capitulation. Ghent and Bruges were retaken by the allies without difficulty, and, to add to the disasters of Oudenaarde and Lille, a terrible winter almost completed the ruin of France. In despair Louis negotiated for peace, but the coalition offered such hard terms that not only the king, but his people also, resolved to fight to the end.

1709, Malplaquet.—When spring came round Marlborough proposed to make a daring thrust toward Paris, masking the fortresses, but this scheme was too bold even for Eugene, who preferred to reduce the strong places before going on. Lille having been successfully besieged, Tournai was the next objective, and the allies suddenly and secretly left their camps before Lille as if for an attack on the Douai lines (June 26–27). But before noon on the 27th they had invested Tournai. A few days afterwards their siege guns came up from Menin by water (down the Lys and up the Scheldt) and the siege was pressed with intense vigour. But it was Sept. 3 before the citadel capitulated. Then Marlborough, free to move again, transferred his army secretly and by degrees to the river Haine, and burst through the French lines almost unopposed. Mons was weakly held, and Marlborough hoped by the rapidity of his operations to take it before Villars could interrupt him. But Villars also moved quickly, and his eager army was roused to enthusiasm by the arrival of Boufflers, who, senior as he was to Villars, had come forward again at the moment of danger to serve as his second in command. Thinking that the allies were somewhat farther to the east than they were in fact, the French marshal marched secretly, screened by the broken and wooded ground, to the south of the fortress, and occupied the gap of Aulnois-Malplaquet (Sept. 9) to work feverishly to entrench himself. Marlborough at once giving up the siege of Mons brought his army forward to attack as rapidly as his brigades came on the scene. At this crisis the duke submitted the question of battle—unwillingly, as one may imagine—to a council of war, and Eugene himself was opposed to fighting an improvised battle when so much was at stake, and it was not until the 11th that the duke delivered his attack on the now thoroughly entrenched position of the French. The battle of Malplaquet (*q.v.*) was by far the most desperately contested of the war. In the end Boufflers, who took command when Villars was wounded, acknowledged defeat and drew off in good order. Eugene was wounded, and Marlborough, after the most terrible experience in any soldier's lifetime, had only enough energy remaining to take Mons before he retired into winter quarters. The loss of the French is given variously as 7,000 and 12,000. The allies sacrificed no less, probably more, than 20,000 men.

Campaign of 1710.—In 1710 Villars lay entrenched behind a new series of lines, which he called *Ne plus ultra* and which extended from Valenciennes to the sea. Marlborough made no attempt to invade France from the side of Mons, for Villars at the head of the army which had been through the ordeal of Malplaquet was too terrible an opponent to pass by with impunity. In England, too, the anti-Marlborough party was gaining the upper hand in the queen's council. So Marlborough took no risks, and returning to the Lille side, captured Douai (June 26) and Béthune (Aug. 26). No attack was attempted upon the lines. In Dauphiné,

Berwick repulsed the Austrians and Piedmontese.

The year 1711 was Marlborough's last campaign, and it was remarkable for the capture of the *Ne plus ultra* lines by manoeuvres that must be recorded as being the *ne plus ultra* of the 18th century way of making war by stratagem. In May the sudden death of the emperor completely altered the political outlook, for his successor Charles was the coalition's claimant to the throne of Spain, and those who were fighting for the "Balance of Power" could no more tolerate a new Charles V. than they could see Louis XIV. become a Charlemagne.

In accordance with a strategic policy of passive endurance the marshal Villars remained on the defensive behind his lines, and Marlborough determined to dislodge him. What force could not achieve, the duke trusted to obtain by ruse. The lines extended from the sea along the Canche, thence to Arras, and along the Sensée to Bouchain on the Scheldt. As the western part of the lines, besides being strong, were worthless from the invaders' point of view because their capture could not lead to anything, Marlborough determined to pass the barrier between Arras and Bouchain. On July 6 Marlborough marched away to the west, as if to attack the lines between Arras and the headwaters of the Canche. Villars followed suit. The plot of the comedy now thickened. Marlborough lost his usual serenity, and behaved in so eccentric a manner that his own army thought him mad. He sent off one part of his forces to Béthune, another back to Douai, and ordered the small remainder to attack the lines between the Canche and Arras, where, as every one knew, Villars's whole army was massed. In the night of Aug. 4–5 the main army slipped off westward, at the highest possible speed. The Scarpe was crossed and then the pace was increased, though thousands of the infantry fell out and scores died from exhaustion. Five hours ahead of the French army the allies crossed the great lines unresisted. The troops concentrated at Cambrai when Marlborough, declining Villars's offer of a battle, manoeuvred farther to the east and invested Bouchain. The siege, covered by a strong "line of circumvallation" which Villars did not attempt to attack, ended with the surrender of the place on Sept. 13, and so terminated a series of manoeuvres so extraordinary as to be almost incredible. In December Marlborough was dismissed from the service in disgrace.

But Holland and Austria determined to make one last effort to impose their own terms on Louis. Eugene's army, which had been used in 1711 to influence the imperial election instead of to beat Villars, was brought back to the Low Countries. Reading the meaning of Marlborough's fall, he quietly made preparations to take over the various allied contingents into Imperial or Dutch pay. So when England seceded, Ormonde only marched away with 12,000 sullen men, while 100,000 remained with the prince.

Misfortunes at Versailles helped Eugene in his first operations, for two successive heirs apparent to the crown died within a month and all was in confusion, not to speak of the terrible misery that prevailed in the country. But the old king's courage rose with the danger and he told Villars that if the army were beaten he would himself join it and share in its fate. Villars, though suffering still from his Malplaquet wound, took command on April 20, and spun out time on the defensive until the end of May, when Ormonde's contingent withdrew. Eugene, as the defection of England had made further operations near the sea unprofitable, took Le Quesnoy (July 4) and moved thence on to Landrecies, which was closely invested. Then followed the last serious fight of the war, the battle of Denain, which saved the French monarchy and completed the disintegration of the coalition.

Denain.—In order to protect his camps around Landrecies, Prince Eugene constructed the usual lines of circumvallation with such speed that Villars, on coming up, found that they were too formidable to attack. Villars anxiously looked out for an opportunity of breaking through. At Denain, the besiegers' route crossed the Scheldt and he resolved to attack them there. On the night of July 23 the French army moved off silently, and by 9 A.M. on the 24th had completely deployed on the north bank of the Scheldt. Eugene galloped away to bring up his army from Landrecies. But, long before it arrived, Villars's troops successfully stormed the lines. Villars followed up his victory at once, capturing Marchiennes and St. Amand, and in these places all Eugene's reserve stores, pontoons and guns. On Aug. 2 Eugene broke up the siege of Landrecies and retreated by a roundabout route to Mons, while Villars's lieutenants retook Douai and Bouchain (September–October). Before the next campaign opened the treaty of Utrecht had been signed, and although the emperor continued the struggle alone for another year, the enfeebled combatants accepted Villars's captures of Landau (July 22, 1713) and Freiburg (Nov. 21) as decisive. The treaty of Rastatt, between Austria and France, was signed on March 7, 1714. (X.)

NAVAL AND MILITARY OPERATIONS IN SPAIN

Allied Victory at Cádiz.—Admiral Sir George Rooke was instructed to seize Cádiz as a base for future operations in the Mediterranean. It was judged inexpedient to compel the Spaniards to acknowledge the Archduke Charles as their rightful king, by means of a bombardment, and they, being not unnaturally somewhat backward in receiving the candidature of an Austrian Prince, lent no support to the Anglo-Dutch arms. Consequently the attempt on Cádiz could not be pushed too vigorously and may be said to have failed. Rooke was already steering north again when he learnt that the Spanish treasure fleet had entered Vigo Bay. The Allies at once determined to prevent the French obtaining the treasure by capturing it themselves, and, entering the bay, Ormonde's troops were landed to attack the forts, while Rooke pushed forward a special advance squadron and forced the boom which the French had laid across the entrance to the inner harbour (Oct. 12, 1702). In the furious engagement which followed, Rooke sank and captured the 24 French warships and the 17 Spanish galleons, and recovered bullion worth two million pounds.

Portugal Joins Allies.—In 1703 Portugal joined the Allies so that Lisbon could be used as a naval base, but Admiral Clowdisley Shovell was sent to the Mediterranean too late to achieve any object, and many of his ships on their return were destroyed in the Channel during the great storm of Nov. 27. Great efforts were made in 1704 to bring about something decisive. Rooke sailed in February and landed the Archduke Charles at Lisbon, together with 2,000 English and Dutch troops to co-operate with the Portuguese in an invasion of Spain.

Rooke carried his fleet through the Straits, hoping to be able to make an attack on Toulon in concert with the duke of Savoy, who had recently joined the alliance. This, in conjunction with Marlborough's march across Germany to Blenheim, would have been a fine counterstroke to the French thrust at Vienna, but as often happened afterwards, Savoy failed to co-operate. An attempt was made to win over Barcelona to the archduke's cause, but the governor refused to admit the allied forces. Meanwhile, the Comte de Toulouse succeeded in bringing the French Brest fleet into the Mediterranean and though Rooke sighted him, his ships were too foul to catch the French, who entered Toulon unopposed, and so united their principal Atlantic and Mediterranean forces. Rooke therefore passed the Straits and met Shovell and the Dutch between Cape St. Vincent and Cádiz, and then reentered the Mediterranean to watch for the French.

Gibraltar Taken by Allies.—At this juncture the allied leaders took the memorable decision to capture Gibraltar and so advance their sphere of activity 300 miles beyond Lisbon. After lengthy preparations the attack was begun by Admiral George Byng, who bombarded the fortress with a special squadron on July 23 (1704), while Prince George of Hesse-Darmstadt, the Imperial representative, landed 1,800 English marines and cut off the fortress on the land side. Gibraltar being only weakly held, an attacking force was soon landed from the ships, and the place surrendered next day. Rooke left Prince George and 900 marines in Gibraltar and went in search of Toulouse. On Aug. 9 the fleets were in sight of each other, but the French retreated and then attempted to double back again close inshore. Rooke, however, sighted them slipping past and brought them to action off Málaga on Aug. 13. The French were decisively beaten.

The allies returned to Lisbon to refit, leaving Prince George with all the marines and some ships, guns and stores to defend Gibraltar. He was soon attacked by land and sea in overwhelm-

ing force, but was relieved by Admiral Leake, who arrived from Lisbon where he had been left in command by Rooke, who had gone home. The French now sent Marshal Tessé with a siege train to command the attack by land, and Prince George was again hard pressed, but Leake again reinforced and revictualled him in December. At last, on March 10, 1705, Leake destroyed de Pointis' force at Marbella, and Tessé raised the siege.

Allies Invade Spain.—The Allies now took the offensive, Shovell and Lord Peterborough coming out from England as joint commanders. Picking up Leake at Lisbon and Prince George at Gibraltar, they sailed to Barcelona and besieged it. It was finally captured and occupied on Oct. 3, 1705, but Prince George was killed during the assault. Philip, the Bourbon candidate, hurried to Catalonia with Marshal Tessé to recapture it, and a squadron put out from Toulon. Leake, however, was able to foil all the French efforts and forced them to raise the siege on April 30, 1706. Leake, who was again left in command of the fleet, made good use of his time by capturing Cartagena (June 1, 1706) and Alicante (Aug. 24, 1706), and then sailing to the Balearic Islands and capturing Mallorca (Majorca) and Iviza in September.

Meanwhile, the allied troops advanced from Portugal and captured Madrid (June 26, 1706) and proclaimed the archduke Charles. The Castilians, however, being rallied by Berwick, compelled the allies to retreat southeast to Valencia, on the coast.

Closing Operations.—In July and August combined operations were again attempted against Toulon, but again Savoy defaulted, and Shovell, returning with 12 of the line, was wrecked on the Scilly Isles on Oct. 22, 1707. Leake now operated in the Mediterranean, reducing Sardinia for the Imperialists in August 1708. In September General Stanhope captured Minorca.

The war in Spain languished but Stanhope in 1710 persuaded the Imperialist troops to join him in another invasion, and they actually succeeded in bringing the archduke Charles to Madrid. Previous conditions, however, returned, and the allies were again driven east to the coast by the duc de Vendôme.

Stanhope was defeated and captured on Dec. 9, 1710, at Brihuego, and the Imperialists were only just able to make good their retreat to Barcelona, which they held till the Peace of Utrecht. At sea, Admiral Wager captured a Spanish treasure fleet off Cartagena in May 1708, and Duguay Trouin captured an English convoy off the Lizard on Oct. 20. In 1710 Commodore Martin's squadron, assisted by American Colonists, captured Nova Scotia. *See* also references under "Spanish Succession, War of the" in the Index. (G. A. R. C.; W. C. B. T.)

SPARK (SPARKING) **PLUG:** *see* AUTOMOBILE; INTERNAL-COMBUSTION ENGINE.

SPARKS, JARED (1789–1866), U.S. historian, who won great distinction as biographer, editor and educator, was born at Willington, Conn., on May 10, 1789. He attended Phillips Exeter academy and graduated from Harvard college in 1815. Sparks was installed as minister of the First Independent church (Unitarian) of Baltimore, Md., in 1819, resigning the post in 1823. From then until 1830, under his ownership and editorship, the *North American Review* became the arbiter of literature in New England. He fostered the growing interest in the American Revolution with his multivolumed "lives and letters" of Franklin, Washington and Gouverneur Morris, 12 volumes of that war's diplomatic correspondence and 25 volumes entitled *The Library of American Biography*. Occupying the McLean chair at Harvard, 1839–49, he was the first U.S. professor of secular history. On Feb. 1, 1849, Sparks was chosen president of Harvard. Harried by administrative difficulties and ill-health, his presidency, which lasted four years, closed his most fruitful years of scholarship. He died at Cambridge on March 14, 1866.

Sparks believed that patriotism obliged him, when editing source materials, to omit passages likely to cause international ill will and sometimes to "embellish" what "the Fathers" had actually written. Therefore the exacting scholarly standards of a later age rendered obsolete much of what he had edited. To his admiring contemporaries, however, he was the "patriarch of American history," the "American Plutarch" and the peer of his friend George Bancroft (*q.v.*).

Sparks's own manuscripts and papers are at Harvard university, while his private library and maps are at Cornell university, Ithaca, N.Y.

BIBLIOGRAPHY.—H. B. Adams, *The Life and Writings of Jared Sparks*, 2 vol. (1893); J. S. Bassett, *The Middle Group of American Historians* (1917); Samuel Eliot Morison, "Jared Sparks," in D. Malone (ed.), *Dictionary of American Biography*, vol. xvii (1935). (W. T. HN.)

SPARROW, a word which originally meant any small bird, sparrow now is restricted to those of the types enumerated below.

1. The house sparrow (*Passer domesticus*) is "the sparrow" *par excellence*. Thriving best in the neighbourhood of man, this bird is common throughout the northern parts of Europe and Asia and has been introduced into North America (in 1851) and the British colonies.

2. The tree sparrow, or mountain sparrow (*P. montanus*), both sexes of which resemble the male house sparrow, but are distinguished by the rufous crown and double barred wings, is a local species in England.

In China, it is abundant and replaces the house sparrow near houses and in towns.

3. The hedge sparrow or dunnock (*Prunella modularis*), belonging to the thrushes (*q.v.*), is a little brown-backed bird with an iron grey head and neck. It is widely distributed throughout Europe.

4. The reed sparrow or reed bunting is, as the second name implies, a bunting (*q.v.*).

5. The numerous American sparrows are also buntings (*q.v.*). Among them may be mentioned the song sparrow (*Melospiza melodia*) of which about 20 subspecies are recognized. It has a fine song.

The chipping sparrow (*Spizella passerina*) is a confiding, humble little bird, with no great attractions but his manners. The savannah sparrow (*Passerculus sandwichensis*) is abundant in the maritime provinces of Canada, but breeds south to Iowa.

JOHN N. GERARD

WHITE-THROATED SPARROW (ZONOTRICHIA ALBICOLLIS)

The grasshopper sparrow (*Ammodramus savannarum*) is also a common form, but its skulking habits and insectlike notes cause it to be overlooked.

The lark sparrow (*Chondestes grammacus*) is common in the Mississippi valley east of the great plains; it has a pleasant if hurried song.

The fox sparrow (*Passerella iliaca*) has also a fine song; it breeds in north Canada and Alaska, wintering from Ohio southward.

SPARTA, the capital of Laconia (*q.v.*), was for long the most powerful city of ancient Greece. The name refers in Greek to the city alone, the alternative name LACEDAEMON to both city and territory; "Spartiate" designated a member of the ruling aristocracy, but the people and state were normally "the Lacedaemonians," this term including the subject perioeci (*q.v.*).

Sparta lay at the northern end of the plain of the Eurotas (Evrotas) river on its western bank, where the routes into Laconia from north and northwest join. The central territory was this plain, between the high peaks of Taÿgetus (Taïyetos) to the west and Parnon to the east, reaching the sea at the port Gythium (Yithion) to the south. Sparta also controlled the mountainous district eastward, with the coast to Cape Malea; some Arcadian border districts to north and northwest; Messenia (*q.v.*) to the west: altogether about two-fifths of the Peloponnese. The Eurotas valley and the Messenian plain of Stenyclarus are, for Greece, conspicuously fertile, and their climate kindly.

Prehistory.—Sparta itself has produced no Mycenaean remains: Amyclae, a few miles south, and Therapne, above the east bank of the Eurotas, were then the important centres, the latter perhaps the capital of the kingdom assigned in the *Iliad* to Menelaus, though no palace comparable to those of Mycenae or Pylos had been found by 1960.

Tradition ascribed the foundation of Sparta to Dorian invaders claiming descent from Heracles (*see* DORIANS). The Dorian dialect, spoken at Sparta, belongs to that division of the Greek-speaking peoples which entered Greece last; and at the Amyclaeum the latest Mycenaean pottery is followed by a variety of Proto-geometric which marks a fresh start and does not, like some varieties elsewhere, show stylistic continuity with Mycenaean. The earliest finds from Sparta show some, but not much, contact with the outside world.

Detailed accounts by later Greek writers of Sparta's gradual conquest of Laconia have an improbable ring. Spartan tradition made much of the capture of Amyclae, in the later version the work of the eighth generation of kings, but more likely an incident of the original conquest: importance was also attached to the capture of Helos in the south, in the belief that the serfs called helots (*q.v.*) took their name from it.

Conquest of Messenia.—The First Messenian War (*c.* 735–715) under King Theopompus marks the beginning of authentic Spartan history, though it represents the completion of the conquest rather than Sparta's first venture west of Taÿgetus. Samos helped Sparta, and the war may be connected with the contemporary Lelantine War (*q.v.*). The best land was divided among the Spartiates, and the Messenians joined the helots in servitude. After the war Tarentum in south Italy, the only Spartan colony, was founded. To the east, Spartan expansion was checked by a defeat at Hysiae in the pass leading down from Arcadia to the coast south of Argos: Pausanias dates this to 669, and the victor might be King Pheidon. Two generations after Theopompus' conquest the Messenians broke out in a long revolt, later associated with the Messenian national hero Aristomenes. Argos and the Arcadians helped Messenia, while Elis supported Sparta, and after a hard struggle the rebellion was crushed.

Archaic Sparta.—The British excavations begun in 1906 revealed a civilization more open and less austere than classical literature had suggested. The main sites were those of Artemis Orthia (earliest temple perhaps of the 9th century) and Athena "of the bronze house."

Laconian artists show their best in bronze, mainly known from small human and animal figures, but there is some evidence that the enormous, sparely decorated mixing-bowl from Vix near Châtillon-sur-Seine may be Spartan, and this may give an idea of, *e.g.*, the bowl which Herodotus says was sent as a present to Croesus. Stone sculpture, to judge by the reliefs extant, remained ungainly and provincial, but literary tradition knew some famous Spartan artists. The Artemis temple produced innumerable small lead figurines, cheerful and attractive trifles; and small objects carved in bone and ivory whose workmanship shows close connections with Ephesus. Laconian pottery (once ascribed to Cyrene in Africa) shows the influence of metalwork in its shapes and its precise decoration, painted on a white ground. It developed unevenly, influenced by Corinthian art but generally conservative, and seldom achieved satisfactory composition: the well-known cup depicting Arcesilaus of Cyrene watching the weighing and storage of wool is at least fresh and lively. After the middle of the 6th century it declined like all fabrics except Attic.

Of native craftsmen, Gitiadas who made Athena's Bronze House in the middle of the 6th century was best remembered. From abroad, Theodorus of Samos built a famous portico, and Bathycles of Magnesia late in the 6th century the elaborate throne for Apollo at Amyclae. Many foreign poets and musicians came, among the earliest Terpander and Polymnastus from the eastern Greek world, Thaletas from Crete, but these are little more than names. Late in the 7th century Alcman, probably of Lydian origin, is known from the long fragment of a *partheneion* ("maiden song") and many shorter quotations: he had much to say of Spartan girls, horses and food, and was more sensitive to landscape than most Greeks. But Tyrtaeus' military exhortations, composed during the Messenian revolt, are native work, though later he was inappropriately claimed as an Athenian.

Within its fairly narrow limits, this society shows a gay and lively mind, open to the outside world. But the Messenian revolt underlined the dangerous position of a small aristocracy which excluded the free perioeci from government and held down a helot population many times its own number, and this must have hampered Spartan development. The evidence of archaeology and Alcman, however, provides no decisive evidence about the date at which Sparta was bound by its other peculiar institutions.

Political and Military Institutions.—In the 5th century and later, the Spartans believed that their whole system of life was the work of a single early lawgiver, Lycurgus (*q.v.*). Some modern historians substantially accept the Greek theory, as it was eventually fixed, that he carried out a large reorganization in the 9th century; but the readiness of the Greeks to ascribe everything Spartan to Lycurgus, and the character of the institutions themselves, make this doubtful. The institutions may not all be of the same date, and it is possible that later reformers sheltered behind Lycurgus' name, fostering the ascription of everything to him, their own reforms included.

Two kings held equal power, from the separate families of the Agiadae and Eurypontidae. The origin of this unique system is lost in the dark ages. Later the kings held certain priestly and other privileges, and a Spartan army abroad was almost always commanded by one of them: Aristotle characterizes them as "hereditary generals." The two families to the end produced many gifted and influential leaders.

Greek aristocracies regularly curtailed the king's power by setting up other executive magistrates, and there is evidence that the Spartan ephors (*see* EPHOR) once represented "the city" as against the kings.

The *gerousia* (*q.v.*) no doubt descends from a council of nobles summoned in Homeric fashion by the kings. The assembly included all Spartiates over 30 who had undergone the prescribed training and paid their dues to *phiditia* (*see* below). It could pronounce only on issues set before it by the *gerousia*, but in practice it discussed and decided all major questions, such as war and peace. It often tried kings whose conduct had caused dissatisfaction; and through the election of ephors it indirectly controlled the administration.

The relation between kings, *gerousia* and assembly in the process of reaching decisions of state is the main subject of the Great Rhetra which Plutarch quotes as an oracle given to Lycurgus; it also refers to the numbers of the *gerousia* and to tribal organization. It has been accepted as a 9th-century document, but its provisions seem more appropriate to the 7th century, and it is here taken as an enactment of that time, one of the pillars of the classical constitution, designed to secure Sparta from the discontents which in many cities produced tyranny by guaranteeing, under certain limitations, the rights of the assembly.

The army, based originally on the three kinship tribes common to all Dorian communities, was regrouped when the citizen body was reorganized on a local basis, the five villages (*obai*) of Sparta. This system probably dates from the 7th century (*obai* are mentioned in the Rhetra) and was in force at the battle of Plataea (479), where the perioeci fought in separate units but in equal numbers with the Spartiates. Before the end of the 5th century there were six regiments called *morai*, in which Spartans and perioeci served together, and at Leuctra (371) the Spartan element was little more than a cadre.

The navy was commanded by an admiral who held office for one year only.

Social Institutions.—Spartiates lived under state supervision from birth, when weakly children were exposed on Taÿgetus. From 14 to 20, graded in year-groups, they underwent a severe training in discipline and endurance. Thereafter they might be admitted to the messes called *phiditia* in which male Spartans ate and passed much of their lives, and on admission they became full citizens, called "equals" (*homoioi*): it is not clear whether in the classical period there was any large residue of those who failed for admission or had not land enough to pay their dues, in kind, but a class of "inferiors" is mentioned once in the 4th century. Foreigners were sometimes allowed to undergo this training, even some selected helots (*mothakes*), but these could not become citizens.

Similar but less strict institutions existed among the Dorians

of Crete, and there are many parallels outside Greece, so that the system may be a general inheritance which elsewhere lapsed so early that it was forgotten. But at Sparta it served military needs, and it may have been deliberately preserved and strengthened there. The assertion of equality must also have been a deliberate assertion at some particular date: in practice it proved increasingly difficult to maintain.

The evidence for Lycurgus' division of land into equal lots is suspect, coming mainly from the time of the 3rd-century reformers (*see* below). But land was the basis of a Spartan's citizenship, and for long there were conventional or legal restrictions on its alienation. The accumulation of other forms of wealth was hindered by Sparta's retention of a clumsy currency of iron spits when other Greeks, from the 7th century onward, coined silver; and by the prohibition of private ownership of precious metal. Spartiates might not engage in trade, and there were sumptuary laws, *e.g.*, against elaborate building.

Over-all, this was a hardy society, freed to an unusual extent from the Greek preoccupation with agriculture, using its leisure for hunting and the perfection of its rigorous military training. Plato complained that Lycurgus educated his citizens in the warlike virtues to the neglect of the others.

Peloponnesian League and Persian Wars.—When Croesus of Lydia in the 540s sought Greek help against the rise of Persia, he found that Sparta was the strongest city and made his alliance with it. By the end of the 6th century Sparta headed an organized league covering most of the Peloponnese and reaching out beyond it. The Spartans' rise was due primarily to their military power. They suffered one heavy defeat in the middle of the 6th century when they tried to reduce Arcadian Tegea to a status like that of Messenia, but Herodotus says that their other wars of this period were successful, and they had mastered Tegea before Croesus' appeal. The basis of the settlement with Tegea was alliance, not subjection, an important stage in the development of the Peloponnesian league. They also defeated Argos in the Thyreatis, a border district on the coast which the Spartans now firmly held: this involvement, they said, prevented them from helping Croesus when Cyrus of Persia overthrew him.

A policy of deposing tyrants is ascribed to the powerful Chilon, who was ephor *c.* 556 (he has also been credited, but by a misunderstanding of the evidence, with increasing the power of the ephorate). Sicyon was liberated soon after 556; an unsuccessful attempt was made, with Corinthian help, to overthrow Polycrates of Samos in 525; Hippias was driven from Athens in 510. It is not a very full record, but Sparta gained the reputation of supporting constitutional government against tyranny, and this must have helped the growth of the league.

King Anaxandridas, who died soon after 520, played some part in all this, but his attitude is not clear and it may not have been consistent. His energetic son Cleomenes I (*q.v.*) quarreled with Athens after its liberation and made several attempts to coerce the Athenians, thwarted mainly by his colleague Demaratus and the Corinthians. These differences were perhaps the occasion for giving a formal constitution to the Peloponnesian league, for the first recorded league conference occurred about 500. If these developments damaged Sparta's prestige, it was restored by Cleomenes' victory over Argos at Sepeia without allied help (*c.* 494), and at the end of his reign he was on good terms with Athens, at whose request he took hostages from Aegina when the Aeginetans gave tokens of submission to King Darius shortly before Marathon (490). But to achieve this he had to get his opponent Demaratus deposed, and overstrained the great personal influence he had in Sparta: the opposition triumphed, and after a brief exile and restoration he committed suicide in prison. The troubles in which this reign ended probably included also a Messenian revolt in 490, but ten years of quiet enabled Sparta to recover, and in 480 the Greeks accepted without question Spartan leadership during Xerxes' invasion.

Since Cyrus' first conquest of Asia Minor frequent appeals had been made to Sparta for help against the Persians. This is the more striking, since after the failure of the expedition against Samos in 525 Sparta consistently refused to send help overseas: this was presumably from military calculation rather than indifference, and several Spartan actions of this period can be interpreted as intended to build up resistance to Persia inside Greece. Above all, Sparta had never shown any sign of compromising with the Persians.

For the operations of 480 and 479 *see* GRECO-PERSIAN WARS. Sparta's performance has been criticized, but under Spartan leadership the war was won: Spartan heroism at Thermopylae was never forgotten, and the last land battle at Plataea was very much Sparta's victory, under the regent Pausanias. But Salamis had demonstrated Athenian naval power. Athens came forward as the leader of the eastern Greeks, and in 478, though Pausanias' expedition to Cyprus and Byzantium was a military success, his conduct so much alienated the non-Peloponnesian allies that they formed the Delian league (*q.v.*) under Athens.

Struggle With Athens.—The Spartans acquiesced in the transfer of the naval hegemony to Athens, and this withdrawal must have weakened the great influence they had acquired. Pausanias (*q.v.*) probably wished to fight Athens instead, but the peace party prevailed, and his eventual disgrace and death did further harm to Sparta's prestige. The Spartans had now to fight to retain their position even inside the Peloponnese.

The Spartan army, trained for a set battle in the Greek heavy infantry style, was ill-adapted for distant and scattered operations, and the Spartans failed to improvise the machinery necessary for the running of an overseas alliance. The peace party saw further expansion as a danger to Sparta's traditions, and trusted the Athens of Cimon (*q.v.*); and the helots were an ever-present menace. Democracy was another danger: in the 6th century the Spartans could appear as the leaders of a crusade for constitutional government, but in the 5th they were left on the defensive as a bulwark of oligarchy, for they could not accept democracy without disrupting their whole system.

Democracy was adopted in Elis (471) and Mantinea: Themistocles, in exile at Argos, had perhaps some hand in this. Tegea also revolted, with some help from Argos, and Arcadia had to be subdued in two hard-fought battles. In 464 a disastrous earthquake caused heavy casualties at Sparta and precipitated a Messenian revolt. Defeated in the field, the rebels were hard to dislodge from their stronghold on Mt. Ithome, and the Spartans called for help from their allies, Athens included. But they then dismissed the Athenian contingent, distrusting its loyalty, and this insult wrecked Cimon's policy of co-operation with Sparta. The Athenians allied themselves with Sparta's enemy Argos (462/461).

The Spartans, having weathered the crisis but still dangerously weakened, took no direct part in the ensuing war except in 457, when they tried to revive the Boeotian league as a counterweight to Athens: but they had a hard fight at Tanagra to secure the army's return, and Athens soon after overran Boeotia. Disaster in Egypt (454) checked Athens' progress, and a truce was made in 451 for five years. In 446 Sparta invaded Attica in support of a Megarian revolt from Athens, but retired without a battle and concluded the Thirty Years' peace, which implicitly recognized the Athenians' right to their empire.

The breakdown of this peace in 431 and the outbreak of the Peloponnesian War (*q.v.*) have been variously explained. Thucydides thought the basic cause was Spartan distrust of Athens, and this is probably correct: incidents just before the war suggested that Athens had not really renounced expansion, while Sparta had recovered and was ready to proclaim a crusade to liberate Greece from Athenian tyranny. But Spartan invasion and devastation were inadequate weapons against Athens, supplied with corn from overseas, and the cities of the empire were not eager enough for liberation. Fighting in the northwest, which could never be decisive, went in Athens' favour. Pylos on the Messenian coast, taken in 425, served as a refuge for fugitive helots, and Spartiate prisoners captured there caused Sparta great anxiety. But in 424 Athens was defeated near Delium by the Boeotians, and the success of the enterprising Spartan general Brasidas in the north gave the Spartans something to bargain with, so in 421 they gave up their crusade and signed the compromise peace of Nicias.

Several Spartan allies refused to sign this peace, which was never fully effective. In 418 Sparta actually fought Athens and allies at Mantinea, where the Spartan victory restored their position inside their own league, and in 414 Athens formally broke the treaty by raiding Laconian territory. The Spartans, resuming open war, fortified a permanent post at Decelea in Attica. The loss of Athens' expedition to Sicily (413) gave a fresh turn: Sparta, hesitantly at first, crossed the Aegean to provoke revolt in Asia Minor, where the remainder of the war was mainly fought to sever or protect Athens' vital access through the Hellespont to the grain of south Russia. The problem of maintaining campaigns at a distance had been partly solved by the use of detachments of allied troops and freed helots under officers called harmosts. In spite of Athenian victories at Cyzicus (410) and Arginusae (406), Lysander finally gained control of the Hellespont at Aegospotami (405) and Athens was starved into surrender (404). But the fleet which won the war was maintained only with Persian subsidies, notably those of Cyrus the younger to Lysander from 407 onward, and Persia's price was the surrender of the Greek cities of Asia Minor.

Spartan Empire.—Lysander (*q.v.*) aimed at a tightly controlled empire, and had secured political control of many cities through committees of ten (decarchies) of his personal adherents; and Athens was governed by "the Thirty" with a Spartan garrison. But jealousy of Lysander's power and distrust of his methods grew at Sparta. The turning point came with Thrasybulus' democratic rebellion at Athens, and King Pausanias was allowed to mediate a settlement by which democracy was restored (403): on his return to Sparta he was put on trial, but Lysander lost this trial of strength, and his decarchies were abolished. Sparta's prestige remained enormous, but Spartan troops and harmosts could not be everywhere.

In 401 Cyrus revolted against his brother Artaxerxes II with some Spartan help. When he was killed and the satrap Tissaphernes began to reconquer the coast of Asia Minor for Persia, the Greek cities there obtained Spartan help in a series of inconclusive campaigns, including those of Agesilaus II (*q.v.*) in 396 and 395. Meanwhile Sparta's continental ambitions in central Greece provoked Thebes to form a coalition with Athens, Argos and Corinth against Sparta. They were defeated in 394 near Corinth, and again by Agesilaus' returning army at Coronea in Boeotia, but the destruction of the Spartan navy at Cnidus, by the Athenian Conon with a largely Persian fleet, largely compensated this. Sporadic land fighting went on around Corinth, and Thrasybulus in 389 recovered much of Athens' naval power. Thereafter Sparta and Persia rediscovered their common interest: Antalcidas with Persian and Syracusan help gained control of the Hellespont, thus paralyzing Athens, and imposed the king's peace or peace of Antalcidas (spring 386) which assigned all Asia Minor to Persia and proclaimed the autonomy of all other Greek states.

Sparta, led by Agesilaus, made ruthless use of this peace to dissolve the Boeotian league and any other combination thought dangerous. The seizure of the citadel of Thebes in 382 marked the height of Spartan power; but Thebes recovered its independence in winter 379, and an unsuccessful Spartan attempt to surprise the Piraeus alienated Athens, whose Second Confederacy spread widely, defeating the Spartan navy at Naxos. But the war exhausted both Sparta and Athens; peace was signed in 375/374 but immediately broken; at the next conference, in 371, Thebes refused to recognize the independence of the other cities of Boeotia, and when King Cleombrotus tried to impose this by force the Spartans were disastrously defeated at Leuctra (July 371).

Leuctra ended Sparta's domination. Impressive as it looked, it was never fully effective without Persian help, especially against Athens, and its effects were wholly negative. Spartan institutions, however much admired by theorists, were not a model that could be copied in an increasingly democratic Greece; and Sparta's great prestige could not suppress the resentment caused by Spartan interference. Greece as a whole was appreciably weaker since the 5th century.

Decline.—Epaminondas' liberation of Messenia (winter 370/369) cut off nearly half Sparta's resources, but Spartan power was not quite broken and Thebes could not dominate the Peloponnese: the Theban victory at Mantinea (362), marred by Epaminondas' death, only increased the confusion. Sparta held aloof from peace conferences, refusing to recognize the independence of Messenia, while some of the kings capitalized their military skill as mercenary leaders abroad. The Spartans took no part in resistance to Philip of Macedonia: Agis III led an anti-Macedonian movement while Alexander was conquering Persia, but was defeated and killed by Antipater at Megalopolis (331). Toward the end of the 4th century the city, hitherto proudly unwalled, was partly fortified; Areus I, who showed some military energy but was killed before Corinth (*c.* 265) during the Chremonidean War, struck coins in his own name; land had for some time been fully alienable and was falling steadily into fewer hands, but there was less of it since the loss of Messenia, and other territory was transferred to Sparta's neighbours by Philip (338) and later by Antigonus Gonatas. The pattern of Spartan life was changing.

Agis IV by peaceful agitation in 244, and Cleomenes III by force in 227, attempted to restore the old order. For details *see* AGIS; CLEOMENES: the basic plan was to cancel debts, redistribute land on the system ascribed to Lycurgus, bring up the numbers of the citizen population by admitting perioeci and others, and re-establish the ancient discipline. To the Hellenistic world, where the propertied classes had a firm grip on city politics, this was dangerous social revolution. Agis was soon suppressed. Cleomenes, more effective in his seizure of power, so alarmed the Achaean league by his military achievements and ambitions that Aratus (*q.v.*) of Sicyon reversed the policy of years and called in the Macedonian king Antigonus Doson; Cleomenes was defeated at Sellasia in 222, and died soon after in Egypt. The reforms were cancelled, but trouble continued. The Achaean leader Philopoemen defeated and killed a "tyrant," Machanidas, but the still more radical Nabis (*q.v.*) achieved some power and was partly tolerated by Rome till his assassination. Philopoemen then captured Sparta (188) and abolished the remains of the Lycurgan constitution: but many problems remained, especially the numerous exiles left over from this period of trouble, and Spartan discontents served as some excuse for Rome's war on the Achaeans (146).

Roman Sparta.—After Achaea became a Roman province, Sparta, though technically a free city, had no further individual history. Much territory had been lost, many of the perioecic cities being now detached in an independent league, and Sparta was wholly cut off from the sea: but under the wealthy citizen Gaius Julius Eurycles it took Augustus' side in the Roman civil wars and was rewarded with an outlet at Cardamyle (Kardhamili) on the Messenian gulf. Some picturesque features of the Lycurgan training were retained for the young—the flogging ordeal at the temple of Artemis Orthia, suitably brutalized, proved a tourist attraction—but their elders led the ordinary life of Roman provincials.

Medieval Sparta.—Sparta suffered severely during the invasion of Greece by the Visigoth Alaric in A.D. 395-396, and from subsequent invaders. Nevertheless, it remained a prosperous town and bishopric, often known as Lacedaemonia under the Byzantine empire. On the establishment of the Latin empire in Constantinople in 1204, the bulk of the Peloponnese became the Frankish principality of Achaea under the Villehardouin family, who built a new fortress and city named Mistra on a spur of Taÿgetus near Sparta. Michael Palaeologus recovered Mistra for the Byzantine empire as part of the ransom of William Villehardouin, captured at the battle of Pelagonia in 1259; thereafter it was the capital of the despotate of the Morea (*i.e.*, the Peloponnese), with a brilliant life of about two centuries (*see* MISTRA). From 1460 it was Turkish, with a Venetian interval from 1687 to 1715.

Modern Sparta.—The existing town (modern Gr. SPARTI) was built on the ancient site in 1834, after the establishment of Greek independence, and is the capital of the *nomos* (prefecture) of Lakonia. Pop. (1961) 10,412.

See also references under "Sparta" in the Index.

BIBLIOGRAPHY.—General histories of Greece; H. Michell, *Sparta* (1952); K. M. T. Chrimes, *Ancient Sparta* (1949; very technical and somewhat adventurous). *Archaeology:* principally *Artemis Orthia,* ed.

by R. M. Dawkins (1929). *Pottery:* E. A. Lane, "Lakonian Vase-Painting," *Annual of the British School at Athens*, xxxiv, pp. 99 ff. (1933–34). *Sculpture:* E. Langlotz, *Frühgriechische Bildhauerschulen* (1927; some doubtful attributions). *Other Finds From Sparta:* M. N. Tod and A. J. B. Wace, *Catalogue of the Sparta Museum* (1906). (A. As.)

SPARTACUS, leader in the Gladiatorial War against Rome (73–71 B.C.), a Thracian by birth. He served in the Roman army, but seems to have deserted and been sold as a slave. Destined for the arena, he, with some fellow gladiators, broke out of a training school at Capua and took refuge on Mt. Vesuvius (73 B.C.), where they were joined by other runaway slaves. They defeated two Roman forces in succession, overran most of southern Italy and ultimately were about 90,000 strong. Spartacus defeated both consuls of 72 and fought his way successfully toward the Alps. As the slaves refused to leave Italy, he returned to Lucania, hoping to cross to Sicily; but he was thwarted by the new Roman commander, M. Licinius Crassus, who had eight legions. His large army divided; the Gauls and Germans were defeated first, and Spartacus ultimately fell fighting in a pitched battle (71). Many slaves who escaped northward were cut to pieces by Pompey on his way home from Spain. Spartacus was both competent and, apparently, humane. His reputation has suffered from the terror that his name inspired throughout Italy.

See T. Rice Holmes, *The Roman Republic,* (1923); H. Last, in *Cambridge Ancient History,* ix (1932).

SPARTANBURG, a city in the piedmont section of South Carolina, U.S., the seat of Spartanburg county, is in the foothills of the Blue Ridge mountains, 32 mi. N.E. of Greenville. The name of the city and county commemorate the Spartan regiment of South Carolina militia which fought in the American Revolution. Established first as a courthouse village in 1785, the town was in an area noted prior to 1860 for iron works, cotton planting and a few cotton mills. Nearby resorts at mineral springs drew a clientele from the entire state. After 1865 growth was accelerated by the intersection of three major rail lines at Spartanburg. By 1900 a number of cotton mills had been established on nearby streams. The trend to manufacturing continued in the 20th century; less than 12% of the county labour force is employed in agriculture and about 45% in manufacturing. Textile mills dominate county industry, which has been diversified with producers of bathroom fixtures, electrical equipment, food products, textile machinery, wearing apparel and chemical products. With the decline of cotton farming, farmers in the Spartanburg area turned to vegetable and peach production. The city was incorporated in 1831 and has a council-manager form of government in effect since 1947.

Spartanburg has a symphony orchestra, a civic band and a public-school symphony. It is the site of Converse college, a private college for women, founded in 1889; Wofford college, Methodist, for men, founded in 1854; Spartanburg Junior college, a coeducational Methodist institution founded in 1911; and the South Carolina School for the Deaf and Blind (1849). Located near the city is Croft State park. The nearby mountains offer many recreational opportunities.

The population of Spartanburg is about 67% native-born white and 32% Negro. For comparative population figures *see* table in SOUTH CAROLINA: *Population.* (L. P. J.)

SPARTINA. A genus of grasses (family Gramineae) belonging to the tribe Chlorideae. Of the seven or eight species, most are North American, and as a group are commonly called cord, salt or marsh grass; salt meadow grass (*S. patens*) and smooth cord grass (*S. alterniflora*) are common on the salt marshes of New England. In Europe *S. stricta* occurs on the salt flats of southeast England, France, Portugal and the Adriatic. *S. alterniflora* has been known at Southampton from 1829, where it was probably introduced by shipping from America.

The genus is remarkable for the sudden appearance of a new distinct form at Southampton in 1870, *Spartina townsendi.* This tall-growing and robust form spread widely on the soft tidal muds of both sides of the English channel. *Spartina townsendi* is exceptional in colonizing soft mud previously bare, for the vigour of its spread and for its power of holding silt and raising the level of the mud. It is probably a hybrid between *S. stricta* and *S. alterniflora. Spartina spartinae,* gulf cord grass, native to the southern coastal plain of the U.S., produces wiry culms which are used, in combination with broomcorn, in the manufacture of household brooms and brushes. (F. W. O.)

SPASM: *see* CRAMP; MUSCLE AND MUSCULAR SYSTEM.

SPATHE, the large sheathing bract, or pair of bracts, enclosing an inflorescence on the same axis, called a spadix (*q.v.*). Spathes assume various forms and are often highly coloured, as in canna. (J. M. BL.)

SPEAKER. Originally the name given to the presiding officer of the English house of commons, "speaker" has been adopted by many legislatures which have based their procedure on that of the British parliament. The office of speaker exists throughout the Commonwealth, but there are usually modifications of the British pattern: for example, in Australia, Canada and New Zealand the office is regarded as the privilege of the party in power and the occupant of the chair changes with the government. The office of speaker in the United States house of representatives is described below.

GREAT BRITAIN

Functions and Powers.—The speaker is the mouthpiece or representative of the house of commons in its relations with the crown, the house of lords and bodies or persons outside parliament. This function is largely confined to formal or ceremonial occasions. The speaker's main tasks are to preside over debates (except when the house is in committee), to enforce the observance of all rules for preserving order and to advise the house in its proceedings (for the presiding officer of the house of lords, *see* LORD HIGH CHANCELLOR). Conscious endeavour over the last 150 years has succeeded in making the office the objective embodiment of the rules and laws of the house, freeing it from any suspicion of partisanship. The speaker rules on points of order as they arise in debate and such rulings constitute precedents by which subsequent speakers are guided. He decides whether a proposal for an emergency debate should have precedence over the business already set down, and whether a prima facie case has been made out in complaints of breaches of privilege. He has the power to select amendments and to accept or reject a motion for the closure of debate. He can check irrelevance or repetition. Disorderly conduct can usually be restrained by a simple call to order; the speaker can direct the withdrawal of offensive expressions, or, in serious cases, "name" a member, which leads to immediate suspension from the house. If general disorder arises the speaker may either suspend the sitting for a period or adjourn the house at once without putting the question. The calling of members to speak in debate is entirely in the hands of the speaker, and his main concern is to ensure that a variety of points of view is heard. In addition to his duties in the chair the speaker controls much of the machinery of the house. His department includes the library of the house, and he is responsible, through an editor whom he appoints, for the production of the official report of debates (*Hansard*). He is also the chairman of the boundary commissions which review parliamentary constituencies, and from time to time he is called on to preside over conferences on electoral reform.

Rules, custom, ceremonial and deliberate policy combine to give the speaker, once elected, great authority. He is elected for a whole parliament (normally of about four or five years duration) and by custom dating back to 1835 is re-elected for as long as he wishes, no matter what changes of government may occur. The usual tenure of office is about 10 or 15 years. The house of commons' mace is carried before him on his daily procession to the house at the beginning of its proceedings. He is robed in a gown and full-bottomed wig; members bow to the chair on entering and leaving the chamber; all speeches are addressed to him; and when he rises to his feet all members must resume their seats and hear him in silence. The speaker's actions or rulings may not be criticized incidentally in the course of debate but only on a substantive motion, and such motions are rare. The speaker receives an annual salary of £5,000, which is charged on the con-

solidated fund and not voted annually, and in addition the annual allowance for parliamentary expenses paid to all members of parliament; he has an imposing official residence within the palace of Westminster. Formerly the first commoner in the land, the speaker since 1919 takes precedence immediately after the prime minister and lord president of the council. On retirement the speaker is offered a viscountcy and is voted a pension of £4,000 a year for life.

As important as the speaker's authority is his impartiality. On election a speaker severs all his ties with political parties and clubs. He takes no part in debate and votes only in the event of a tie, and then only in accordance with rules which preclude an expression of opinion upon the merits of the question. A. W. Peel (speaker, 1884–95) summed up the position when he said: "I know how necessary it is for any man who aspires to fill that great office to lay aside all that is personal, all that is of party, all that savours of political predilection, and to subordinate everything to the great interests of the house at large." J. W. Lowther (speaker, 1905–21) quoted with approval the remark that the office of speaker "does not demand rare qualities. It demands common qualities in a rare degree." The qualities chiefly required are good temper, patience, tact, judgment, and dignity in voice and bearing. A spice of wit, which can dissolve a strained situation in a burst of laughter, is invaluable.

Election of the Speaker.—At the beginning of a new parliament the sovereign commands the house to elect a speaker. When a new speaker has to be chosen the government of the day makes every effort, by informal discussion with the opposition and all shades of back-bench opinion, to discover a candidate who will command general support. Contested elections for the speakership are rare—the last two contests being held in 1895 and 1951. On the latter occasion W. S. Morrison (speaker, 1951–59), a former minister from the Conservative side of the house, was opposed by members of the Labour party, who put forward Major James Milner, for many years chairman of ways and means and deputy speaker, as their candidate. The Labour party had not, by the early 1960s, provided the house with a speaker, and the feeling that it should be enabled to do so was widespread within the party in 1951, and again in 1959 when Morrison retired. Although the election of Sir Harry Hylton-Foster in 1959 was not contested by the Labour party, strong dissatisfaction was expressed at the inadequacy of the consultations which preceded his being put forward for the office.

Once elected, the speaker is not normally opposed in his constituency by official party candidates at general elections. Since 1895 such contests have only occurred in 1935 and 1945, though unofficial candidates opposed the speaker in 1950 and 1955. If there is a contest the speaker refrains from any party political campaigning. It has been suggested, in order to spare the speaker any possibility of a contested election and to prevent his constituents being, in effect, disfranchised, that his constituency should be authorized to elect two members, that a fictitious constituency should be created, or that he should be enabled to sit in parliament without a constituency. However, a select committee of the house that reported on the matter in 1939 recommended that no change be made.

History of the Office.—Although the office can be traced back to 1258, when Simon de Montfort presided over the "Mad" parliament at Oxford, the first speaker to be so designated and the first known to have been chosen by the house of commons was Sir Thomas Hungerford, elected to the chair in 1377. Since that date, every parliament has elected its speaker. Although the choice of speaker was strongly influenced by the crown in the medieval period, several holders of the office showed considerable independence. A number of the early speakers met violent deaths, and the position of the speaker, between a strong king and a house concerned to increase its own power, could be an unenviable one. The custom whereby a newly elected speaker shows reluctance to accept office dates from this period. By 1642, after the constitutional struggle between Charles I and parliament, it was established that the first duty of the speaker was to the house, and when the king came to arrest five members, Speaker Lenthall replied to his questions with the memorable words: "May it please your majesty, I have neither eyes to see nor tongue to speak in this place, but as the house is pleased to direct me, whose servant I am here."

Since 1679, when Charles II refused to approve the re-election of Sir Edward Seymour, but failed to secure the election of his own nominee, royal approval of the commons' choice has been a formality, and by the end of the 17th century the speaker's independence of the crown was an established principle. In the 18th century dependence of the speaker on the crown threatened to be replaced by dependence on party. However, Arthur Onslow (speaker, 1728–61) raised the conception of the office to its present level by his resolute independence of all parties and his complete impartiality. His immediate successors failed to maintain his standards and once again the speakership was regarded as in the gift of the party in power, and few speakers took any pains to stand aside from politics. William Grenville, Pitt's nominee, was elected to the chair in 1789 when less than 30 years old, remained speaker for only five months, and later became home secretary and prime minister. Henry Addington, (afterward Viscount Sidmouth, speaker, 1789–1801), stepped straight from the chair to the premiership. However, during the 19th century the firm tradition of an independent and nonpartisan speaker was built up, a process especially associated with Charles Shaw-Lefevre's tenure of the office (1839–57). (H. J. Pa.)

UNITED STATES

The United States constitution (art. 1, sec. 2) provides that "the House of Representatives shall chuse their Speaker and other Officers . . ." In creating this office its authors had in mind the speaker in the colonial assemblies who had been both a presiding officer and a political leader. Early expectations were later realized, for U.S. speakers after 1789 were both parliamentary and political personalities.

The speaker of the house holds an office of dignity and honour, of great power and influence. Usage dictates that he shall always be a member of the house, and as such the constitution requires him to be seven years a citizen and at least 25 years of age. Candidates for the office are nominated by party caucuses prior to the convening of each congress and the speaker is customarily elected by the house on the opening day, barring a protracted contest. Competition for the place has often been keen and the final choice is influenced by knowledge of parliamentary procedure, previous legislative experience, sectional claims, private interests, personal qualities and political considerations.

History of Office.—From the beginning the speakership has generally been held by men with previous experience in state legislatures; many of the speakers of the house of representatives have been men with experience as speakers of their state legislative bodies. Fewer than one-fifth of the speakers who served during the first 170 years of the constitution's history were without any prior state legislative experience, and of these none held the office prior to at least his fourth term in the national legislature. Only three speakers were elected to the office in their first term in the house of representatives: Frederick A. C. Muhlenberg, the first speaker, who had been speaker of the Pennsylvania house and a delegate to the constitutional convention; Henry Clay, who had served several terms in the Kentucky legislature as member and speaker and had been a U.S. senator; and William Pennington, who had once served in the New Jersey assembly. With the exception of these three, every speaker had had prior service in congress, ranging from 1 to 16 terms.

Many famous men have served as speaker of the house—James K. Polk, Schuyler Colfax, James G. Blaine, Thomas B. Reed, Joseph G. Cannon, Champ Clark, Frederick H. Gillett, Nicholas Longworth and Sam Rayburn. Rayburn set a record for the longest service—nine terms, including two half-terms. After Rayburn the next longest term of service as speaker was that of Clay—six terms, five of them consecutive. Andrew Stevenson, Cannon and Clark served four terms each; Nathaniel Macon, Colfax, Blaine, Samuel J. Randall, John G. Carlisle, Reed, Gillett, Longworth and William B. Bankhead each had three terms in the chair.

More than one-third of the speakers have come from the south, about twice as many as from any other part of the country. Southern influence predominated in the organization of the Democratic party, for the "solid south" was overwhelmingly Democratic after the Civil War and repeatedly re-elected the same congressmen. Southern members thus acquired seniority and power in party councils and most Democratic speakers from 1865 to 1961 were southerners. In 1962 John W. McCormack, Democrat of Massachusetts, was elected speaker to succeed Rayburn. (For a list of speakers *see* UNITED STATES [OF AMERICA]: *Table VI.*)

Powers and Duties.—The parliamentary powers and duties of the speaker stem from four sources: the constitution and laws of the United States, the rules of the house, previous decisions of the chair and general parliamentary law. He presides at sessions of the house and opens and closes its sittings, acts as its spokesman and representative and authenticates documents, announces the order of business, puts questions and announces the vote.

He also decides questions of order, prevents obstruction of house business, refers bills and reports to the proper committees or calendars, chooses speakers pro tempore and appoints the chairmen of committees of the whole. These duties he performs impartially like the speaker of the house of commons, as evidenced by the fact that his decisions are rarely appealed; however, he also enjoys the privileges of an ordinary member of the house and may vote and take part in debate on the floor.

Prior to 1910 the speaker also possessed important political powers. He appointed, deposed and demoted the chairmen and members of committees; he served as chairman of the committee on rules, which enabled him to control the business of the house; and he recognized or refused to recognize members claiming the floor. These powers, gradually evolved by strong speakers like Reed and Cannon, reached a peak of perfection and omnipotence during the latter's regime. So dictatorial and despotic, indeed, did Speaker Cannon become that the house rebelled, and in March 1910 a coalition of Democrats and insurgent Republicans succeeded in drastically curtailing his powers. They removed the speaker from the rules committee, deprived him of the power to appoint the standing committees of the house and restricted his right of recognition.

Although the speaker was removed from the committee on rules in 1910, he sits on the unofficial steering committee which meets in his office and determines party strategy. Though he no longer has the power to appoint standing committees, he retains the power of appointing house conferees, the chairman of the committee of the whole and select committees. His power of recognition was reduced, but he still has a discretionary power of recognition over motions to suspend the rules and on days other than consent calendar days.

Place in Political System.—The Presidential Succession act of 1947 ranked the speaker of the house first in line of succession in case of the death or disability of both the president and vice-president. Relations between speakers and presidents have fluctuated between rivalry and harmony through the years. During periods of congressional supremacy the chair has overshadowed the White House, while popular-leader presidents have dominated congress. Despite the reduction in his powers, the speaker continues to be at once the presiding officer and titular leader of his party in the lower chamber, a private member and the most influential man in the house of representatives. (G. B. Gy.)

BIBLIOGRAPHY.—*Great Britain.* J. A. Manning, *Lives of the Speakers* (1850); E. and A. G. Porritt, *The Unreformed House of Commons,* 2 vol. (1903); A. I. Dasent, *The Speakers of the House of Commons* (1911); M. MacDonagh, *The Speaker of the House* (1914); Sir Thomas Erskine May, *Treatise on the Law, Privileges, Proceedings and Usage of Parliament,* 16th ed. (1957); Sir William Ivor Jennings, *Parliament,* 2nd ed. (1957). (H. J. Pa.)

United States. M. P. Follett, *The Speaker of the House of Representatives* (1896); C. W. Chiu, *The Speaker of the House of Representatives since 1896* (1928); F. M. Riddick, *The United States Congress: Organization and Procedure* (1949); G. B. Galloway, *History of the House of Representatives* (1961). (G. B. Gy.)

SPEARMAN, CHARLES EDWARD (1863–1945), British psychologist whose chief contributions dealt with the theory of intelligence and cognition, was born in London on Sept. 10, 1863. As an officer in the British army, he served in the Burmese war of 1885 and in the South African War, and did not obtain his doctorate in psychology at Leipzig until he was 40. His earliest researches were concerned with the improvement of psychophysical methods, especially as applied to visual perception. On returning to England, he joined a group at Oxford who at William McDougall's suggestion were applying Karl Pearson's correlational techniques to develop tests of intelligence for Francis Galton's projected surveys in British schools. His first results were reported in "General Intelligence Objectively Determined and Measured" (1904). Galton had maintained that intellectual processes depended both on a "general ability" and on a number of "special abilities"; Spearman sought, by simplifying Pearson's methods, to show that "general ability" alone sufficed, but later modified this extreme position; his animated controversies with Pearson and other critics led to a valuable series of factorial researches from both sides. He died in London on Sept. 17, 1945.

Spearman's books include *The Nature of Intelligence and the Principles of Cognition,* 2nd ed. (1927); *The Abilities of Man: Their Nature and Measurement* (1927); and, jointly with L. Wynn Jones, *Human Ability* (1950). (Cy. B.)

SPEARMINT (*Mentha spicata*), a species of mint (*q.v.*), used for culinary and other purposes, distinguished by its smooth sessile leaves and tapering pale-blue flower spikes. Spearmint, native to temperate Europe and Asia, has become widely naturalized throughout the temperate world. Propagated from seeds and cuttings, the plant thrives in a moist and sunny location. The volatile oil from spearmint is widely used for flavouring and as a carminative.

SPECIES. The nature of species is a question of considerable importance in general biology and the species is regularly employed as a classificatory unit in systematic botany, zoology and mineralogy. The word "species" has, however, an important history outside those sciences. It was the Latin equivalent of the Greek *eidos, idea* ("class"). This meant primarily shape or visible form, but was extended to cover the sensible character of things generally, and also the intelligible.

As intelligible character, *eidos* or species was the common essence of individuals or instances, as we say, of the same kind. As such it was an object of thought, not of sense; universal, not individual; and, unlike its instances, exempt from generation and decay. Its relation to its instances (which may be compared to that of a law to the events exemplifying it) was a subject of dispute both in Greek and scholastic philosophy. There were also disputes about the relation of species to their common genus, of which they are determinations, and to the properties connected and accidents conjoined with their specific nature in individuals of a species. In particular it was disputed whether species and genera were anything real independently of our minds, or only common names, or else notions of ours, the products of our classifying activity. A settlement of these questions is not necessary to the work of the systematic biologist, but they must be faced if we are really to understand what we mean by the evolution of species; for this evolution is a process of which no persisting material thing is the subject. (H. W. B. J.)

BIOLOGICAL SPECIES

The concept of species as a kind of animal or plant is an old one in biological literature. Aristotle (384–322 B.C.) used the term for groups of organisms with common characters and these in turn he subdivided into more species. John Ray (1627–1705) is usually given the credit for using the term for groups with mutual fertility. C. Linné (*see* LINNAEUS, CAROLUS; 1707–78) established a system of binomial nomenclature universally used in the 20th century. Each organism was classified as a member of a species and each species was included in a higher taxonomic group, the genus. The genera were also classified in still higher taxonomic categories, but the binomial name consisted of the names of the genus and the species together.

International congresses of botany and zoology respectively formulate, administer and interpret the international rules of botanical

nomenclature and the international rules of zoological nomenclature. The basic principle used for natural classification is homology (*q.v.*). In early modern times homology was a vague concept applied roughly to apparent similarities of structure. During the 20th century genetics, embryology, ecology and evolutionary analysis came to place a firm scientific foundation under the concept of homology with consequent clarification, correction and verification of systematic biology.

Linnaeus first believed in the fixity of species but later in his life suggested that the genus constituted the ancestral species at the time of creation, and that derived species subsequently multiplied by means of hybrid generation or intercrossing between species. During the 19th century, particularly through the work of J. B. P. Lamarck (1744–1829) and Charles Darwin (1809–82), the theory of the evolution of species was accepted by the large majority of biologists (*see* EVOLUTION, ORGANIC).

The importance of isolation in the origin of species was emphasized by M. Wagner (1813–87) and J. T. Gulick (1832–1923). Modern genetic theory, initiated by Johann Gregor Mendel (1822–84) by means of his investigations of domestic varieties of plants, was applied to both plants and animals in the early 20th century. Mendelian genetics of individuals led gradually to the establishment of the science of population genetics during the third and fourth decades of the 20th century. Population genetics substantiated the concept of the biological species and greatly improved the understanding of evolutionary processes.

The species is recognized as a prime biological unit with objective criteria. More than 1,000,000 species of plants and animals had been named by mid-20th century and probably at least twice this number would ultimately be recognized when the knowledge of the world faunas and floras approached completion. There are almost as many definitions of the species as there are biologists who have attempted to condense the major aspects of the concept.

A brief definition that includes the most generally recognized criteria of the biological species was given by A. E. Emerson in 1945: a species is an evolved or evolving, genetically distinctive, reproductively isolated, natural population. All of these attributes are necessary, and no others would seem to be essential. Each attribute is discussed separately in the following paragraphs.

Since the establishment of the principle of evolutionary change, species have been considered to be dynamic entities and not fixed or static. Evolutionary characterization is therefore necessary in the definition. Demarcation sometimes is somewhat arbitrary when transitional stages are encountered, but true evolutionary gradations are not often found, so that ancient continuities are seldom observable between living species and the paleontological record seldom bridges the gaps between species categories. Although nomenclature is easier when the ancient continuities are unknown, cases in which intergradations between species can be directly observed offer an opportunity for a more exacting study of the processes of speciation, and analytical biologists seek such material for investigation. Also the early stages in species divergence that have not progressed to the point of complete species division enable the scientists to analyze the dynamic factors more exactly and are consequently sought.

Genetic distinction seems to be the common denominator for all valid qualitative or quantitative species characters including cytological, physiological, embryological, ethological, ecological and morphological characters. Taxonomists, of course, seldom can experimentally investigate the genetics of the species they study and must rely upon structural or colour indications of genetic relationship. Anatomy and pigmentation often are sensitive to small genetic influences that are not easily detected by other techniques of observation. Heredity may be inferred through demonstrated germinal continuity in populations and through the experimentally established basis of similar characters in investigated organisms. Hybrids and intergrading populations give some genetic information. If striking structural characters are not correlated with genetic distinctions of reproductively isolated natural populations, the taxonomist usually does not use them for species separation and will correct errors that have been committed in ignorance of the genetic implications. For example, structural changes produced by various conditions at the time of growth, by differences in stages of development, by sexual dimorphism, by castes of social insects and by various types of polymorphism do not meet the criteria of genetic distinction and reproductive isolation. On the other hand, slight genetic distinction of natural populations that are reproductively isolated may be used for species separation, while gross genetic differences of individuals or of only partially isolated populations are not valid for species distinction.

Reproductive isolation (W. C. Allee *et al.*, 1949), or lack of gene flow between populations, may result from chronological separation, spatial separation, topographical or geographical barriers to dispersal, habitat or ecological restrictions, cyclic differences in breeding periods, mechanical inability to crossbreed, psychological or ethological barriers to interbreeding, physiological barriers to fertilization, genetic unbalance preventing fusion, inviability of hybrids, hybrid sterility and selective elimination of hybrids. Regardless of the exact type of isolation or the combination of isolating factors, anything that prevents genes from moving from one population to another has profound evolutionary consequences. Reproductive isolation is the dividing factor in the branching of the phylogenetic tree, and the two populations thus deprived of gene exchange will tend to drift apart through changes in gene incidence and gene or chromosome mutation. Because of its effect upon inbreeding, reproductive isolation also influences gene and character fixation. As most taxonomic characters are the result of complex genetic patterns rather than single genes or chromosomes, the consolidation of these patterns and their transmission as systems becomes a necessity for further adaptive evolution. Fixation of complex genetic patterns is a primary consequence of reproductive isolation. Reproductive isolation may take place gradually and give rise to many subdivisions of the species population before speciation is completed. Theoretically the point of species origin is the moment that genetically distinctive natural populations cease gene exchange. Practically it may not be possible to recognize this point in instances of contemporary speciation. However, the large majority of valid species have been isolated for thousands or millions of years, so that it is a rare case that presents difficulties when an abundance of evidence is accumulated.

An important group of biologists including Theodosius Dobzhansky (1951) and E. Mayr (1942) excluded extrinsic barriers to gene flow (geographical or ecological) in their definitions of reproductive isolation as applied to species and confined the concept to examples separated by intrinsic barriers (psychological, genetic and physiological). They did not give full species rank to genetically divergent natural populations completely lacking gene exchange but which have not lost their potential ability to interbreed if the extrinsic barriers are removed. Inasmuch as it is often impossible to judge the potentialities of interbreeding under natural conditions, these biologists had to rely upon a rough correlation between interbreeding potentialities and degrees of taxonomic divergence. When two related species occupy the same territory (sympatric species) and no intergrades occur between them, biologists may assume an intrinsic barrier to gene flow. Such an assumption is much less warranted in the cases of closely related but geographically separated distinctive populations (allopatric species), however, and there is a resulting confusion of taxonomic status of numerous groups. One school of taxonomists may refer certain populations to full species rank, while the other school will treat the same populations as subspecies.

If reproductive isolation includes the extrinsic barriers to gene flow, it is possible that species that have not evolved intrinsic barriers may ultimately interbreed with the occasional breakdown of temporary extrinsic barriers. The hybrid origin of a species from two preexisting species is thereby a possibility. The school of biologists that confine reproductive isolation between species to intrinsic prevention of gene flow only do not recognize full species status for the extrinsically isolated populations and would not allow for the possibility of hybrid origin of species except in special cases in which intrinsic barriers are also broken down (allopolyploidy in plants). In spite of these differences of opinion of the exact definition of reproductive isolation with consequent confusion over the taxonomic status of some natural populations, the large majority of plant and animal species prove to be intrinsically isolated. Therefore there is common agreement among biologists concerning most species limitations together with the criteria for the recognition of the species category.

The final characterization of the species is that it constitutes a natural population. The population is integrated by means of genetic continuity through reproduction, and it may also be integrated through sexual reproduction and various other adaptations for attraction and interaction between individuals of the same species. The species population is often subdivided into many subpopulations, but the larger species population is a real entity with a definite boundary. Differences in fundamental biological interactions occur within the species population in contrast with the biological interactions between species populations. The species is generally conceived to result from natural evolution. Artificial populations compounded by man through domestication or for scientific investigation are not given species rank even when they are genetically distinctive and reproductively isolated populations. Under natural conditions such man-produced populations would seldom maintain their genetic distinctiveness even if they were able to survive. The domestic dog is classified as a full species, but the separate breeds are not.

Numerous infraspecies categories relate to variations in the qualitative or quantitative attributes of the species. Each term places emphasis upon particular aspects of the species concept (G. du Rietz, 1930). Some of the terms are defined below and are useful in the study of speciation mechanisms.

Sexual Species.—This term is used for species that possess sexual adaptations and interfertility among the component individuals. Sexual recombination of genetic elements greatly increases genetic variability. Many definitions of the species confine the concept to the sexually interbreeding population (Dobzhansky, 1951).

Asexual Species.—These are species that have never evolved sexual fusion or have secondarily lost the ability to intercross. All reproduction is confined to division, parthenogenesis, apomixis or vegetative reproduction. This species category would include the primary asexual bacteria and many single-celled organisms and would also include the secondarily asexual parthenogenetic animals and apomictic plants (G. L. Stebbins, 1950). In asexual organisms, reproductive isolation between individuals of the same generation is absolute, and genetic continuity is only through descent. In some instances taxonomic distinction is maintained for periods comparable to the main-

tenance of taxonomic pattern in sexual species (species of flagellate protozoans inhabiting the hind gut of termites). In other instances, taxonomic distinction may vary over short periods of time and the species category is difficult to apply (virulent and nonvirulent strains of viruses and nutritive strains of bacteria).

Subspecies.—This category is a subdivision of a sexual species with all the attributes of the species except that reproductive isolation is partial rather than complete. Subspecies are given trinomial names by taxonomists and are recognized as constituting a valid category in the international rules of nomenclature. A species subdivided into subspecies or races is called a polytypic species.

Race.—This commonly used term is not separated by definition from the subspecies. However, it is either used as a synonym for the subspecies category or is given a rank below that of the subspecies, presumably because the degree of reproductive isolation is smaller than that separating the subspecies.

Deme.—G. S. Carter (1951) defined a deme as a communal interbreeding population within a species either wholly or partially isolated from other demes.

Supraspecies.—This term was used by J. Huxley (1939) for a collection of groups or subspecies which replace each other geographically or physiologically and in which the extremities only have evolved intrinsic isolation. Connecting groups between the extremities interbreed and intergrade. There are several known cases among snakes, birds and mammals in which intersterility occurs among some of the subspecies within the species.

Coenospecies.—This term was first used by G. Turesson (1922) for the sum total of possible combinations of a genotype compound as seen particularly through cultivation under artificial conditions. The term has also been used to include groups of species separated only by extrinsic barriers to gene flow.

Ecospecies.—G. Turesson (1922) used this term for the species as realized in nature under the limitations imposed by the natural environment. It emphasizes the limitations of potential variation produced by natural selection.

Biotype.—This term is used for a population of individuals with identical genotypical constitution under similar conditions. Biotypes are usually limited to asexual species or to populations within a sexual species that result from asexual reproduction from a single individual. Parallel mutation may possibly account for some biotypes.

Clone.—A group descended from a single individual without sexual combination. The individuals are presumably genotypically identical or isogenic. The term overlaps somewhat with the biotype.

Lineage.—G. C. Robson and O. W. Richards (1936) adopted this term for a racial complex of several lines of descent forming a meshed network of evolution within the species.

Variety.—This term is used with somewhat different meanings by different authors. J. Bequaert (1919) used the category for individuals or small groups of individuals that varied in a describable way from the typical individuals in the species. Varieties may be considered somewhat distinctive groups that fall within the range of species variation as a whole but do not necessarily have geographical or ecological correlations.

Cline.—J. Huxley (1939, 1942) introduced this term for a character gradient within a group—usually within a species. Numerous subdivisions of clines are named when there is a correlation with special factors. Ecoclines are correlated with ecological gradients. Geoclines are correlated with geographical gradients. Taxoclines are correlated with taxonomic gradations associated with hybridization. Chronoclines are correlated with time in paleontological sequences. Ontoclines are gradations in the appearance of characters during the life cycle of individuals.

Physiological Species or Races.—These terms are used to distinguish species or subdivisions of a species that are recognized by physiological characters rather than structural characters. Genetic distinction is usually manifest in structural or anatomical characters, but gene patterns may effect physiological characters without producing sufficient anatomical effects to be recognized. The term "biological race" is essentially synonymous with "physiological race."

Incipient Species.—This term is used for a diverging partially isolated population that may become a species when isolation becomes complete.

BIBLIOGRAPHY.—W. C. Allee *et al., Principles of Animal Ecology* (1949); J. Bequaert, "A Revision of the Vespidae of the Belgian Congo," *Bull. Amer. Mus. Nat. Hist.*, 39 (1919); G. S. Carter, *Animal Evolution* (1951); Theodosius Dobzhansky, *Genetics and the Origin of Species*, 3rd ed. (1951); G. du Rietz, "The Fundamental Units of Biological Taxonomy," *Svensk Bot. Tidsk.*, 24 (1930); A. E. Emerson, "Taxonomic Categories and Population Genetics," *Ent. News*, 56 (1945); J. S. Huxley, "Clines: An Auxiliary Method in Taxonomy," *Bijdr. tot de Dierk.*, 27 (1939), *Evolution, the Modern Synthesis* (1942); E. Mayr, *Systematics and the Origin of Species* (1942); G. C. Robson and O. W. Richards, *The Variation of Animals in Nature* (1936); G. L. Stebbins, *Variation and Evolution in Plants* (1950); G. Turesson, "The Genotypical Response of the Plant Species of the Habitat," *Hereditas*, 3 (1922). (A. E. E.)

SPECIFIC GRAVITY is the ratio between the density (mass per unit volume) of a body and the density of some substance assumed as a standard. For liquids and solids the standard is distilled water at 4° C. In the centimetre-gram-second system of units the numerical values for the density and the specific gravity of any substance are the same. In the English system the density of water is 62.4 lb. per cubic foot, while its specific gravity is still unity. *See* PHYSICS (ARTICLES ON).

SPECK, FRANK GOULDSMITH (1881–1950), U.S. anthropologist, an authority on Indian tribes of eastern North America, was born on Nov. 8, 1881, in Brooklyn, N.Y. He took his M.A. degree in 1905 at Columbia university under Franz Boas, and his Ph.D. at the University of Pennsylvania in 1908. In the same year he initiated a program of anthropological studies at the University of Pennsylvania, where he became professor in 1925 and continued teaching until his death. He devoted himself to intensive field research among the Indians of the eastern United States and Canada, particularly the Algonkians and Iroquois. Within the communities he visited, Speck was welcomed as a friend and confidant. Vitally interested in linguistics, he mastered several Indian languages and, in a few instances, recorded some known only to a few remaining speakers. He published such important monographs as *Ethnology of the Yuchi Indians* (1909), *Naskapi, the Savage Hunters of the Labrador Peninsula* (1935) and *Penobscot Man: the Life History of a Forest Tribe in Maine* (1940). He discovered and described the territorial hunting system of the northeastern Algonkian tribes. Speck dealt authoritatively with a wide range of topics—technology and decorative art, social organization, folklore and music. One of his perennial interests was religion and ceremonialism. He also contributed to ethnobotany and ethnozoology. He died on Feb. 6, 1950, in Philadelphia.

See A. I. Hallowell, "Frank Gouldsmith Speck, 1881–1950," in *Amer. Anthrop.*, 53:67–87 (Jan.-March 1951), with a bibliography by J. Witthoft. (A. I. H.)

SPECTACLES: *see* EYEGLASSES.

SPECTROCHEMICAL ANALYSIS, in its broadest sense, comprises all methods of chemical analysis that depend upon the measurement of the wave length and the intensity of spectrum lines or bands, and encompasses all frequencies in the electromagnetic spectrum. In its restricted and commoner usage two methods are implied: (1) ultraviolet and visible emission spectroscopy and (2) ultraviolet and visible absorption spectrophotometry. Both methods serve to characterize the qualitative and quantitative composition of matter.

Emission Spectrochemical Analysis.—In emission spectroscopy atoms are excited to higher electronic energy levels above their ground states by means of electrical discharges (arcs or sparks) or flames. Atoms of unknown identity in a specimen so excited emit light of characteristic frequencies when they return to lower energy states. These characteristic frequencies are separated into an ordered sequence (spectrum) by diffraction or refraction in a spectrograph for visual, photographic or photoelectric registration. The process therefore consists of three interdependent steps: (1) the vaporization and electronic excitation of atoms, (2) separation of the emitted light into its component frequencies and (3) measurement of the wave length and intensity of selected frequencies.

Each chemical element possesses its unique and characteristic collection of spectral frequencies, whose recognition in the spectrum of an unknown substance therefore provides unambiguous evidence of the presence of that element. Ordinarily emission spectrochemical analysis is applied to the qualitative and quantitative determination of metallic elements, but it is not restricted to them. The method is among the most sensitive of all analytical methods, a few milligrams of a solid sample usually sufficing for the detection of metallic elements present to the extent of only parts per million. Although a given element may provide many thousands of spectrum lines, depending upon its energy level system and its concentration in a sample, it is not necessary to identify each wave length. Normally the identification of three prominent lines in the spectrum of a sample serves to ensure the presence of an element with high reliability, while the absence of one of its most sensitive lines certifies the absence of the element

to the level of spectroscopic detection.

The physical state of the sample suitable for spectrographic analysis may vary widely. Metallic samples are usually fabricated into rods or disks for electrodes in an arc or spark discharge. Powders and nonconducting solids are conveniently placed in a crater formed in a rod-shaped carbon or graphite electrode of high purity for volatilization and excitation. Solutions, biological fluids and tissues are generally converted to dried inorganic residues before vaporization and excitation from graphite or metal electrodes.

Quantitative analysis by emission spectroscopy depends upon the fact that the quantity of light emitted in a given spectrum line is proportional to the number of atoms vaporized and excited. To establish the relationship between the concentration of an element and the quantity of light emitted in one of its spectrum lines, it is necessary to provide a series of standard samples in which the matrix composition is fixed, and the sought element varies in concentration in a known and graded manner from standard to standard. Semiquantitative analyses may then be carried out by visual comparison of the photographic density of a spectrum line in the unknown sample with the densities of the same line in the spectra produced from the standards. More refined quantitative analyses, precise to about 1% of the element concentration, may be made by measurement of the spectrum line densities with a microphotometer. Since it is exceedingly difficult to maintain constant excitation conditions during the excitation of a sample and to reproduce them from sample to sample, absolute measurements of spectrum line intensities are not made. Instead a relative method, called the internal standard method, is employed. Devised by W. Gerlach in 1925, the internal standard method relates the intensity of a spectrum line of the sought element to the intensity of a suitable spectrum line of an element whose concentration remains constant in all standards and unknown samples. Fluctuations in the excitation conditions then affect the intensity of the unknown and standard equally, and their intensity ratio is independent of the unavoidable variations in the excitation and photographic processes. Frequently, as in the analysis of nearly pure metals or alloys of definite composition, the internal standard line is chosen from among the spectrum lines furnished by the base element. When this is not feasible, it is often possible to add an internal standard element in equal concentration to standards and samples alike.

Although the emission spectrographic method has relied greatly upon photographic emulsions for the recording and measurement of spectrum lines, and continues to be useful, increasing use is being made of photoelectric methods for the direct measurement of spectrum line intensities. Particularly in routine metallurgical control analyses, in which speed and a large volume of determinations are of importance, direct-recording spectrometers are available. These instruments receive spectrum lines in electron photomultiplier tubes and record the analytical composition directly on digital counters or print the results on tape.

Absorption Spectrophotometry.—Quantitative analysis by absorption spectroscopy depends upon the fact that many substances characteristically absorb light of a given frequency or range of frequencies. Commonly in analysis the absorbing substance is an ion or molecule in unknown concentration in aqueous solution. Those frequencies of light are absorbed which electronically excite the molecule or ion to a higher energy level. The diminution in intensity of light passing through the solution is related to the length of the light path, l, and the concentration of the absorbing species, c, by the Beer-Lambert law:

$$I = I_0 \exp(-\epsilon c l)$$

In the Beer-Lambert law, I is the intensity of light transmitted by the solution; I_0 is the intensity of light transmitted by the solvent alone; and ϵ, the extinction coefficient, is a constant characteristic of the absorbing species. Many ions which alone do not absorb light strongly can be made to absorb strongly by causing them to form complex ions with other ions or molecules. The sensitivity and specificity of detection may often be greatly increased by appropriate selection of complexing agents.

See also SPECTROPHOTOMETRY; SPECTROSCOPY: *Applications.*

See G. R. Harrison *et al., Practical Spectroscopy* (1948); N. H. Nachtrieb, *Principles and Practice of Spectrochemical Analysis* (1950).

(N. H. N.)

SPECTROHELIOGRAPH, an instrument that utilizes the spectrum for photographing the sun with monochromatic light. In its simplest form it consists of a spectrograph in the focal plane of which is a second slit that transmits monochromatic light to a photographic film. This slit may be set at any wave length in the spectrum, usually on a strong spectral line produced by some particular chemical element; *e.g.,* the K line of calcium or the Hα line of hydrogen. An image of the sun is formed by a telescope objective on the collimator (first) slit of the spectrograph, and the photographic film is mounted almost in contact with the second slit. In operation, the spectroheliograph is moved parallel to itself, admitting to the first slit light from all parts of the sun as the slit moves across the sun's image. Simultaneously the second slit is moving across the photographic film, building up an image of the sun with the monochromatic light transmitted by the slit. Thus the distribution of calcium or hydrogen clouds in the sun's atmosphere may be recorded. Such clouds are invisible to the eye and are not shown on a solar photograph taken in the usual way. Similar solar photographs can be obtained, however, without a spectrograph by using filters that transmit nearly monochromatic light.

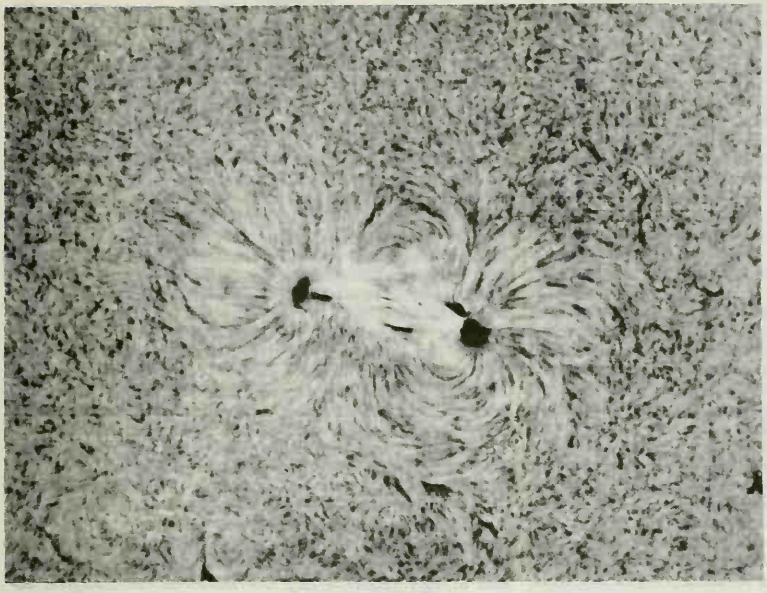
The spectroheliograph, originally designed for photographing only the solar prominences at the sun's edge, disclosed in its first application by George Ellery Hale at the Kenwood observatory (Chicago, Ill., 1891) new, unexplored features of the sun's atmosphere. Photographs of the solar disk showed extensive luminous clouds (flocculi or plages) vastly greater in area than the sunspots around and above which such clouds are always found. These clouds may also appear where no spots are visible. Although in the lower, denser vapour the clouds appear bright, in the rarer vapour at high levels the light from below is absorbed, giving rise to dark clouds. The relatively dark clouds are the prominences that appear bright at the sun's edge.

At about the time that Hale was devising his spectroheliograph, H. Deslandres at the Paris observatory introduced a velocity spectrograph which permitted motion in the line of sight to be measured on photographs of the K (or H) line in successive sections of the sun. Such a series of photographs reveals the same features as shown on spectroheliograms.

The first photographs of the sun in hydrogen light were made with the spectroheliograph at the Yerkes observatory (Williams Bay, Wis.) in 1903 using the blue and violet spectral lines of hydrogen. Later, in 1908, at the Mt. Wilson observatory (near Pasadena, Calif.), when methods for making photographic emulsions sensitive to red light had been developed, the red hydrogen line (Hα) was used. The long dark flocculi (filaments) shown on high-level calcium spectroheliograms and identified as prominences projected against the surface of the sun are very conspicuous on spectroheliograms made in the centre of Hα.

In the high region of the sun's atmosphere over sunspots the distribution of the small dark flocculi shown by the hydrogen photographs suggests the operation of definite forces. Surrounding large single isolated spots, these flocculi often form great vortices that resemble terrestrial cyclones or tornadoes. About 75% of these appear as huge spirals, clockwise in the southern hemisphere and counterclockwise in the northern. The detection of these vortices led, in 1908, to the discovery of magnetic fields in sunspots.

Spectrohelioscope.—The functions of the spectroheliograph and velocity spectrograph are combined in the spectrohelioscope, an instrument that renders visible to the eye many of the phenomena of the solar atmosphere photographed with the spectroheliograph and also permits their velocity in the line of sight to be measured. The principle of the spectrohelioscope was suggested by C. A. Young in 1870 and was used by him for observing prominences at the sun's limb. Hale, who revived and improved the instrument (1924), succeeded in observing with it the bright and dark hydrogen flocculi and in measuring their velocities.



HYDROGEN CLOUDS NEAR SUNSPOTS

These hydrogen clouds surrounding a large, magnetically bipolar group of sunspots were photographed at the Mt. Wilson observatory with the 13-ft. spectroheliograph using the red hydrogen line ($H\alpha$). The forms of such clouds resemble those of iron filings over a bipolar magnet or flow lines around double vortices rotating in opposite directions

Imagine a narrow slit, rapidly oscillating, between the eye and a telescopic image of the sun. Through persistence of vision, the image can be seen in white light. A spectroscope becomes a spectrohelioscope when placed in the optical path so that it transmits to the eye through a second slit only the light of the red hydrogen line. Instead of moving the slits, the sun's image may be moved. A method commonly employed is that of placing a rotating prism with a square cross section in the telescope beam in front of the first slit and viewing the second slit through another prism rotating at precisely the same rate as the first.

A "line shifter" consisting of a plane parallel-sided glass plate mounted behind the second slit displaces the spectral line toward the red or violet when rotated. The amount of rotation indicates the wave length of the light being utilized and the velocity of the observed features. In the spectroheliograph the displaced line would be off the slit, and the image of rapidly moving gas would not be recorded.

The spectroheliosceope affords a rapid means of detecting and studying many solar phenomena, since the entire disk and limb can be examined in less than a minute. These instruments have been widely used in the detection of flares that suddenly increase in brightness and then fade more gradually. Such flares are solar clouds and usually occur near sunspots. Flares affect the ionosphere directly, causing interference in radio transmission, and less directly in the production of aurora and geomagnetic storms.

See SPECTROSCOPY, ASTRONOMICAL: *Instruments;* SUN: *Spectroheliographic Investigation;* ECLIPSE: *Information Yielded by Solar Eclipses.* (S. B. N.)

SPECTROHELIOSCOPE: *see* SPECTROHELIOGRAPH.

SPECTROPHOTOMETRY is a subdivision of physics and chemistry which deals with the relative measurement of radiant energy (power) as a function of wave length, frequency or wave number. Thus, the radiant energy transmitted (reflected) by a body is compared with that transmitted (reflected) by some other system taken as a standard. Absorption spectroscopy and spectrometry are other terms often used synonymously.

Different types of modern instruments cover a wide range of wave lengths of the electromagnetic spectrum. Approximate limits of several spectral regions are: X-ray, less than 1 to 100 angstroms (Å); ultraviolet, 10 to 400 millimicrons (mμ); visible, 400 to 700 mμ; infrared, 0.7 to more than 500 microns (μ); and microwave, 0.1 to 1,000 cm. The relationship between these units is given by

$$1\text{Å} = 0.1\text{m}\mu = 10^{-4}\,\mu = 10^{-8}\,\text{cm}.$$

Two laws express the relationship of the absorption of radiant energy to the absorbing medium. According to Bouguer's law, each layer of equal thickness of the absorbing medium absorbs an equal fraction of the radiant energy traversing it. According to Beer's law, the absorptive capacity is directly proportional to the concentration of the solute in a solution. One expression of the combined laws is

$$\log_{10} P_0/P = \log_{10} 1/T = A = abc,$$

in which P_0 and P are the radiant power (flux) transmitted by the standard and by the unknown, respectively; T is the transmittance which is equal to P/P_0; A is the absorbance; and a the absorptivity, b the thickness and c the concentration of solute. Thickness and concentration are often expressed in centimetres and grams per litre, respectively. For a molar solution, the symbols ϵ and a_M are widely used for molar absorptivity. Deviations from Beer's law are common for solutions, but Bouguer's law is said to apply to all homogeneous systems.

Although spectrophotometers may vary widely in their spectrometric, geometric and photometric designs, they consist essentially of (1) a source of radiant energy, (2) a monochromator, (3) a photometer and (4) a detector for the unabsorbed radiant energy. An absorption cell holds the liquid or gas samples. With the majority of instruments a photometric value is obtained for each wave-length setting. For extensive scanning work, recording instruments greatly facilitate the accumulation of data. Most rapid of all is a model incorporating an oscillograph with attached camera for recording.

Occasionally monochromatic light sources are used, but generally continuous sources (for given regions) are employed. Examples of the latter are the hydrogen arc (ultraviolet), the incandescent tungsten filament (visible) and the Nernst glower (infrared). The monochromator isolates a narrow spectral band from a continuous source, and the photometer serves to measure P/P_0, or some related value, for this band. The detector varies from region to region, there being no universally usable means. Among those commonly used are photographic plates, photocells, thermopiles and bolometers.

A curve co-ordinating photometric values with spectral positions is known as an absorption spectrum. Usually ordinates are T, A, log A or log ϵ, and the common abscissas are wave length, log wave length, wave number or frequency. Often only a value at a single wave length is needed.

Ideally, the absorptivity, a, is a constant, characteristic of a given system. As such, it is comparable to a density or refractive index value. Unfortunately, because of lack of adequate control of instrumental and solution conditions, this ideal is not generally realized. Instrumental variables include spectral band width, cell thickness, stray radiant energy and general operation of the in-

strument. Various solution conditions may affect absorption data, among them hydrogen-ion activity, concentration, temperature, interfering ions, irradiation, redox potential, dissolution aids, component interaction, hydrolysis, light fading, fluorescence and plating.

Absorption data have many applications. One of the most important is the study of solution effects. The data are also widely used to follow the kinetics of chemical reactions and to determine dissociation constants and formulas of complexes. Analysts use absorption data to identify constituents or to determine their amounts. If analysis of a solution is the purpose of the spectrophotometric measurement, all relevant variables must be under control to prevent misinterpretation of the data.

As each pure substance which absorbs selectively has a characteristic absorption spectrum, the curves are important for qualitative analysis. The contour or form of the bands and their wave-length position are the important items. Ions, molecules and various groups such as the carbonyl group in organic compounds are involved. The infrared region is particularly useful for study of the last type, for revealing structures and purity. For mixtures, of course, a measured absorbance gives the combined absorbing power of the components. Application of such data in large quantities is made by means of machine-sorted punched cards.

In the visible region, absorption spectra are basic for applying the C.I.E. international system (Commission Internationale de l'Éclairage) of colour specification. A recording instrument, with attached automatic tristimulus integrator, greatly facilitates calculations.

Quantitative analytical uses of absorption data depend upon the fact that the magnitude of the absorbance is a function of the concentration of the absorber. For any solution conforming to Beer's law, the expression $c = A/ab$ is used, a and b being known and A being measured for the unknown solution. If the solution does not conform to Beer's law, a simple calibration curve coordinating A and concentration may be established.

The quantitative application to systems with many components assumes that the values of A and abc for the individual components are additive for the absorption at the wave lengths employed. The concentration of n components may be obtained from absorbances at n wave lengths, where no two curves coincide or intersect, by solving n simultaneous equations. Spectrophotometers passing narrow band widths are needed for such determinations.

See also SPECTROCHEMICAL ANALYSIS; SPECTROSCOPY: *Applications.* (M. G. ME.)

SPECTROSCOPY (from Lat. *spectrum*, an appearance, and Gr. σκοπεῖν, to see), pertains to the investigation of spectra, the phenomena observed when the radiations from a luminous source are separated into their constituent colours or wave lengths. Such separation results from refraction (as in a prism) or from diffraction (as in a grating). Instruments designed for this purpose are called spectroscopes if used for visual observation, and spectrographs if photography or other methods of recording the spectra are employed. The rainbow, formed by refraction of sunlight in rain droplets, is the most beautiful and common spectrum in nature.

HISTORICAL SPECTROSCOPY

The composite nature of white light was first demonstrated by Isaac Newton (1664) when he allowed sunlight entering a round hole in a shutter to pass through a glass prism and fall on a screen. This elongated and coloured image of the sun he called a spectrum. In 1800 W. Herschel studied the spectral distribution of heat from the sun with the aid of thermometers and found the maximum temperature beyond the red end, thus discovering the infra-red spectrum. In 1801 J. W. Ritter, studying the effect of spectral light upon silver salts, found this action extending beyond the violet, thus discovering the ultra-violet spectrum. The first connection between spectral colour and wave length appeared in 1802 when Thomas Young substituted his wave theory of light for Newton's corpuscular theory, explained the colours of thin films, and calculated the approximate wave lengths of the seven colours recognized by Newton. In 1814 Joseph von Fraunhofer modified Newton's solar-spectrum experiment by substituting a narrow slit for a hole and a telescope for a screen. Under these conditions he observed the continuous spectrum of the sun irregularly interrupted by many hundreds of dark lines, still known as "Fraunhofer lines." These lines were the first fiducial marks in spectra, and were promptly exploited as wave-length standards for the comparison of spectral dispersion of various optical glasses. Fraunhofer constructed the first diffraction gratings by winding a fine silver wire upon two fine parallel screws, or by ruling a glass plate with a diamond point, and with these he made surprisingly accurate determinations of wave lengths corresponding to the Fraunhofer lines. Although Fraunhofer and others observed that certain bright lines in the spectra of flames seemed to coincide with dark lines in the solar spectrum it remained for G. R. Kirchhoff in 1859 to enunciate the general law connecting absorption and emission of light and to emphasize the fact that each species of atom has a uniquely characteristic spectrum. Kirchhoff and R. Bunsen (1861), by systematically comparing the solar spectrum with the flame or spark spectra of the purest elements available, made the first chemical analysis of the sun's atmosphere and thus laid the foundation for spectrochemical analysis and for astrophysics (*q.v.*). In the course of these investigations they discovered spectroscopically two new chemical elements, caesium and rubidium. These demonstrations greatly stimulated spectroscopic research and increased the need for accurate standards of wave length. The first useful standards were provided in 1868 by A. J. Ångström who measured the wave lengths of about 1,000 Fraunhofer lines and expressed them in units of 10^{-10} metre—a unit now known as the angstrom, abbreviated A, and used for all spectroscopic measurements. Further refinements in standard wave lengths were made in 1887 by H. A. Rowland who invented and applied the concave diffraction grating. In 1892 A. A. Michelson invented the first interferometer. With this instrument he measured the wave lengths of cadmium lines relative to the standard metre and attained an accuracy far exceeding that of any previous work. The wave length of the red radiation from cadmium vapour was remeasured (1905) in terms of the standard metre by R. Benoît, C. Fabry and A. Perot, using the interferometer invented by Fabry and Perot, and since 1907 their value—6,438.4696 A—has been internationally accepted as the primary standard of wave length. Several thousand atomic radiations ranging in wave length between 2,100 and 10,200 A, have been measured relative to this primary standard by interference methods and many of these have been adopted internationally as secondary standards of wave lengths. The wave lengths corresponding to lines characteristic of any spectrum whatsoever can be obtained by interpolation between secondary standards impressed on the same spectrogram. In this manner spectroscopists have accumulated wave-length data for several million lines observed in atomic and molecular spectra, extending from the extreme ultra-violet to the far infra-red and embracing nearly 30 octaves as compared with the single visible octave that was first recognized in 1802. These data have served, and will forever continue to serve, in making chemical identifications; since 1885 they have contributed mightily toward the analysis of spectral structures, interpretation of spectra in terms of quantum theory and development of fruitful hypotheses concerning atomic and nuclear structure.

These developments and applications of spectroscopy have unobtrusively but profoundly changed world civilization and have both figuratively and literally expanded the universe (*see* LIGHT).

EXPERIMENTAL SPECTROSCOPY

Any experiment in spectroscopy involves (1) a source of light; (2) a spectroscope or spectrograph for forming the spectrum; (3) detectors for observing or recording details of the spectrum; (4) measurements of wave lengths and intensities; and (5) the interpretation of such measurements either as chemical identifications or as clues to the structure of atoms and molecules.

LIGHT SOURCES

The sun and lightning flashes are examples of bright sources of light occurring in nature. Artificial light sources studied spectroscopically may be divided into two groups according as they emit continuous or discontinuous spectra. A continuous spectrum is characterized by an uninterrupted gamut of wave lengths over a considerable range. Discontinuous spectra exhibit bright lines or bands of different colours or wave lengths on a dark background. Continuous spectra are emitted by incandescent solids and liquids (gas mantle, lamp filament, hot molten metal) or by certain electrical discharges (underwater sparks). These spectra are usually dependent only on temperature and are of no use for identifying the emitting substance; their principal use in experimental spectroscopy is to provide continuous backgrounds for the production of absorption spectra of gases, vapours, solutions, or solids. Discontinuous spectra are emitted by atoms, ions, or molecules in a gaseous or vaporous state in which the individual particles are excited to radiate uniquely characteristic spectra. The particles are excited by absorbing energy either from collisions with other atoms or electrical particles, or from incident radiation. Light sources producing discontinuous spectra are flames, furnaces, and electrical discharges in arcs and sparks at atmospheric pressure, or in lamps containing gases or metal vapours at reduced pressure. In the order named these sources form a rough sequence in which increasing temperatures or energies of excitation are responsible for producing different spectra.

Flames.—Since the days of Kirchhoff and Bunsen it has been customary in elementary chemistry courses to demonstrate spectrochemical identification by dipping a platinum wire into salt solutions, inserting it in a Bunsen burner and examining the spectrum of the coloured flame with a Bunsen spectroscope or Amici prism. Because the temperature of this flame hardly exceeds 2,000° C. the kinetic energy of atomic collisions is not sufficient to produce more than the strongest lines of easily excited atoms like alkalis and alkaline earths. These flames also emit band spectra of carbon compounds and of water vapour present in the gas and air and other bands due to oxides formed in the flame. Higher temperatures, about 2,500° C., are attained in an oxy-acetylene flame and atomic spectra are more fully developed. H. Lundegårdh found that when solutions of materials to be investigated were sprayed into such flames 34 elements could be determined spectrographically.

Furnaces.—An evacuated furnace, in the form of a carbon tube heated by forcing large electric currents through it, provides a source of metallic spectra in which the excitation is pure thermal energy. The temperature may be controlled between 1,500° C., where spectra begin to appear, and 3,600° C. where the carbon tube tends to fail. Small samples are placed in a porcelain boat in the tube and the luminous vapour is imaged on the slit of a spectrograph. If a carbon plug is inserted in one end of the tube it provides a continuous spectrum for the observation of absorption spectra of metal vapours. Such furnace spectra have been extensively investigated by A. S. King at the Mt. Wilson laboratory, in Pasadena, Calif., and the results have greatly advanced the understanding of atomic radiation processes, and aided the structural analyses of atomic and molecular spectra.

Electric Arcs.—Next to the Bunsen burner the simplest source of atomic spectra is the direct-current arc consisting of an insulated holder for electrodes, an applied potential of 100 or more volts and a resistance to limit the current to 5 or 10 amperes. The actual potential drop in such metallic arcs is 20 to 40 volts, most of which occurs immediately at the cathode. Atomic spectra are very fully developed in electric arcs because the temperatures range from 3,500° to 8,000° and all known substances are melted and vaporized. In addition to violent thermal agitation, many atoms collide, especially near the electrodes, with electrons or ions accelerated by the applied potential. For the study of metallic-arc spectra it is best to use solid rods of the pure metals themselves; other forms of material (powders, salts, precipitates, ashes, filings, foils, refractories, minerals, etc.) are readily volatilized by placing some of the substance on the lower (positive) pole of a pure carbon arc. The direct-current arc is characterized by high intensity and effectiveness in generating the first spectra of atoms, but it is subject to erratic wandering and flickering which impair its usefulness for accurate spectrochemical analysis unless standardized conditions are imposed. A high-voltage (2,200 volts) alternating-current arc, introduced in 1936 for quantitative analysis, reduces some of the irregularities of the direct-current arc at the expense of simplicity and sensitivity of spectral detection.

Electric Sparks.—Compared with the electric arc the high-voltage condensed spark supplies higher excitation and is easier to control. Initially the high voltage was produced by a Rhumkorff induction coil, and a Leyden jar served as a condenser. In its modern forms the spark is animated by discharges from suitable condensers connected, in parallel with the spark gap, to the secondary terminals of a transformer giving from 10,000 to 50,000 volts. The condenser is charged on every half cycle to the voltage at which the gap breaks down. An oscillating current then flows in the spark circuit with an initial value of $I = V\sqrt{C/L}$, where V is the condenser voltage, C the capacity in farads and L the circuit inductance in henrys. This initial current may be hundreds of amperes and the result is an energetic emission of excited or ionized atoms from the electrodes. Such sparks have higher effective temperatures than arcs. Indeed, in the highest attainable vacuum these sparks produce multiple ionization until atoms are stripped of all outer electrons and the corresponding temperature is estimated at several million degrees. The ordinary spark is somewhat variable owing to lack of constancy in breakdown potential or in the rate of decay of the oscillating discharges in each half cycle. Several devices have been introduced to control or stabilize sparks used for quantitative spectral analysis. Such devices are the synchronous auxiliary rotating spark gap and the air-blasted auxiliary gap. Any conducting solids may be used directly as spark or arc electrodes.

Other substances will serve as electrodes if they are packed in a hollow carbon rod or if they are pressed into solid form with some binder such as pure copper powder, but in such cases the spectrum of the substance is contaminated by that of the auxiliary electrode material.

Geissler Tubes.—An important type of light source is that produced by electrical discharges in gases at reduced pressures, first demonstrated and studied by J. Plücker in 1858 (fig. 1). It usually consists of a small-bore or capillary tube of glass (or silica) connecting two bulbs or larger bore tubes containing metal electrodes which can be attached to an electrical potential difference by means of wires leading through the glass. These light sources are prepared by thoroughly exhausting the lamp, then admitting some pure gas (or vapour) through a side tube and sealing it off with a blow-torch when suitable conditions are established. The optimum pressure of gas is usually that equivalent to a few millimetres to a centimetre of mercury. When electrical potential is applied to such tubes a sheath of ions near the cathode marks the edge of the cathode fall within which electrons released from the cathode by ion bombardment attain sufficient speed to produce new ions by collisions with gas atoms. Outside this negative glow is the Faraday dark space, and beyond it is the positive column extending nearly to the positive electrode. This positive column is responsible for most of the light emission, and being a region of relatively low potential gradient and electron speeds, it favours the first spectrum of the included gas. A glass blower named Heinrich Geissler, under Plücker's direction, prepared the first gaseous discharge tubes, and Plücker himself referred to them as "Geissler tubes"—a name that still clings to them.

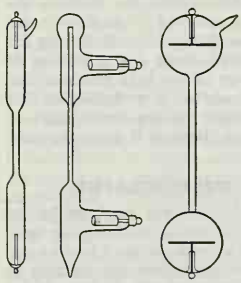


FIG. 1.—GEISSLER TUBES
(a) original Geissler, (b) modern Geissler, (c) NBS Geissler

About 1930 a modified form of the Geissler tube became a spectacular type of light source for luminous coloured signs and shortly thereafter helium, neon and mercury lamps became commonplace. A more recent modification of the Geissler tube is the fluorescent lamp consisting essentially of a low-pressure mercury vapour tube whose discontinuous spectrum is partially converted to continuous bands characteristic of the fluorescent materials lining the inner walls of the tubes.

Hollow-cathode Discharge.—Another type of gas discharge of great utility for spectroscopic research is that devised by F. Paschen. It consists of a hollow cylindrical cathode within a cylindrical anode, and a low pressure of pure noble gas. As the pressure is reduced the discharge retreats within the cathode until, at a pressure such that the mean free path of the electrons equals the cathode diameter, the negative glow fills the hollow cathode. The material of the cathode, or of any other metal within it, is vaporized by bombardment of the noble gas ions and excited by collisions with electrons whose maximum speeds are fixed by the ionizing potential of the sustaining gas. Since this source is operable at low pressures and low temperatures (the cathode may be cooled with liquid air) it yields atomic spectrum lines of extraordinary sharpness.

Electrodeless Discharges.—Gases or metal vapours may also be made luminous at low pressures in closed vessels without internal electrodes by placing the vessel inside a coil carrying a high frequency current generated by either a Tesla spark generator or a vacuum tube oscillator of ultra-high radio frequency. The oscillating high-frequency electric fields accelerate electrons sufficiently in rarefied gases and metal vapours to excite atomic and ionic spectra. These electrodeless discharges are especially convenient for minute quantities of pure materials (such as artificial elements), or for gases and vapours which attack electrodes or which may be contaminated by them.

Linelike Sources.—When viewed through a spectroscope any of the above mentioned light sources will be seen as a multitude of images, one for each monochromatic radiation characteristic of the excited atoms or ions in the source. Since all practical light sources are more or less wide these spectral images will overlap even if the spectroscope has enormous resolving power. In the case of electrical discharges in gases and metal vapours at reduced pressure, the light source is usually constructed with a long constriction to increase the intensity and this incidentally reduces the confusion of overlapping spectral images. However, in general, it is not possible to make all light sources linelike except by imaging the light source on a narrow slit with straight parallel jaws in which case the illuminated slit becomes the light source and the spectrum is a series of bright line images of the slit. The same effect may be obtained by substituting for the slit a cylinder of polished metal about one millimetre in diameter, and using the virtual image of any light source in this cylindrical mirror as a linelike light source. It is emphasized that spectral "lines" do not occur in nature but are consequences of using an artificial linear light source, because it gives the least confusion of overlapping images.

SPECTROSCOPES AND SPECTROGRAPHS

Prism Instruments.—The production of a spectrum by the use of a prism of glass or any transparent material depends upon the fact that light rays of different colours, or wave lengths, are refracted differently on passing obliquely from one medium to another of different density. A composite ray of light is accordingly dispersed as well as refracted in passing through a prism, the amount of refraction being usually greater for shorter wave lengths. The first prism spectroscopes employed by Newton and by Fraunhofer exhibited astigmatism, a type of image defect in which a point source is seen as two perpendicular image lines at different distances from the optical system. This defect was effectively eliminated in Kirchhoff and Bunsen's first spectroscope by using a collimating lens to illuminate the prism with parallel rays. Astigmatism is least and definition best in a prism spectroscope if the prism is traversed by parallel light, the slit is parallel to the prism edge, the light rays are parallel to a section perpendicular to the edge, and the rays pass through the prism symmetrically at minimum deviation. The essential parts of a simple prism spectroscope are sketched in fig. 2. Light from a source under examination enters the spectroscope through a narrow slit *S* located at the principal focus of the collimator lens *C*. Parallel rays of complex light thus fall on the prism, and rays of each component colour continue parallel in their passage through the prism and on emerging from it. The colours are refracted differently, and after passing through the telescope lens *T* are focused to form a spectrum which may be observed with an eyepiece or recorded on a photographic film.

The brightness of the spectrum depends upon the way in which the slit is illuminated by the light source, upon the focal lengths and apertures of the spectrograph lenses, and upon reflection and transmission losses within the instrument. The maximum illumination is attained if the light source is large enough or so close

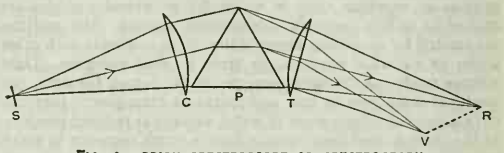


FIG. 2.—PRISM SPECTROSCOPE OR SPECTROGRAPH
S=slit, C=collimator lens, P=60° prism, T=telescope lens, RV=red to violet spectrum

to the slit that it subtends the same angle as the collimator lens. If the source is too small to fill the collimator with light the brightness of the spectrum may be increased by using a condensing lens of equal aperture ratio to form an image of the source on the slit. This type of slit illumination of a stigmatic spectrograph gives a point to point correspondence between a section of the source, the slit and the spectral images; it is preferred when the distribution of excitation in a source or of energy in an interference pattern is to be studied. When a perfectly uniform slit illumination is desired, as for the measurement of spectral-line intensity ratios by means of stepped filters, a condensing lens is placed just in front of the slit to form an image of the source on the collimator lens, in which case each point of the source sends light to every point of the slit and to every point of the spectral image. In either case a movable metal slide pierced with a series of openings (known as a Hartmann diaphragm) may be used to illuminate successive portions of a slit and thus record on a stationary photographic plate a number of spectra side by side for comparison and for measurement.

The light efficiency of a spectrograph, like that of a camera or telescope, depends chiefly on the ratio of the focal length, *F*, to the diameter, *D*, of the lenses, *F*/*D*. This ratio may be 100/1 or more in spectrographs employed for the study of bright sources like the sun, but extremely faint sources like phosphorescence or the night sky demand unit ratio or less. A ratio of about 15/1 represents a common compromise between desired speed and the necessity of using compound corrected lenses. In such spectrographs single plano-convex or meniscus collimator and camera lenses may be used if the curvatures are chosen so as to minimize spherical aberration and coma, and the plate is tilted and bent to fit the chromatic focus. The inclined plate obviously provides an increased spectral dispersion, and the use of single lenses ensures that light losses resulting from absorption and surface reflections are a minimum.

When all relevant factors are considered it appears that an isosceles prism with 60° angles is the best form for a spectrograph. As to prism and lens materials there is a wide choice of optical glasses for the study of visible and near infra-red spectra but for spectral regions absorbed by glass it is necessary to use other materials such as crystalline quartz or fluorite in the ultraviolet, and rock salt or other alkali halides in the infra-red. Unfortunately, no crystals, natural or artificial, are transparent to all wave lengths; quartz is opaque below 1,800 A, and fluorite below 1,200 A, whereas rock salt is opaque above 17,000 A, and potassium iodide above 31,000 A. Furthermore they have limita-

tions on account of size, homogeneity and cost. For ultra-violet spectrographs the most popular material is crystalline quartz. But quartz polarizes the light it transmits and produces double images both on account of its birefringence and its property of rotating the plane of polarization. The birefringent effect is eliminated by cutting quartz lenses and prisms so that the light travels along the optical axis of the crystal, and the rotation doubling can be compensated by using equal parts of natural left-handed and right-handed quartz. A simple quartz spectrograph, therefore, has a collimator lens of one variety of quartz, a camera lens of the other variety, and the 60° prism is made by combining two 30° prisms of the opposite varieties of quartz. The same effect is obtained with economy of material in a Littrow-type spectrograph consisting of a single quartz lens and a 30° quartz prism backed by a mirror. (*See* fig. 3.) Obviously the rotation doubling

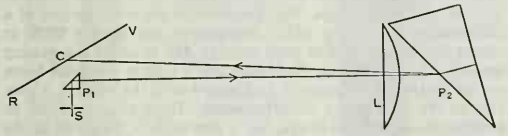


FIG. 3.—LITTROW QUARTZ SPECTROGRAPH

S=slit, P_1=totally reflecting quartz prism, L=autocollimating quartz lens, P_2=Littrow quartz prism, C=camera, RV=red to violet spectrum

experienced by light in its forward passage through the lens and prism is cancelled by its return along the same path after reflection from the mirror or metallic coating on the back of the prism.

Because of their portability and popularity for visual observations of spectra two types of compound prisms, the Amici and the Pellin-Broca, deserve brief descriptions. A form of "direct-vision spectroscope," invented in 1860 by G. B. Amici, consists of two prisms, ABC and CDE (fig. 4) of crown glass, having a low ratio of dispersion to index of refraction, cemented to a prism BCD of flint glass having a high ratio of dispersion to index. The prism angles are so chosen that the median spectral ray is undeviated upon emergence but all other rays are refracted and dispersed. Greater dispersion and resolving power are obtained by assembling trains of five or seven prisms.

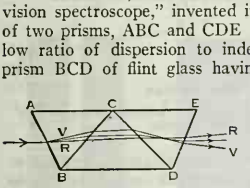


FIG. 4.—AMICI DIRECT-VISION SPECTROSCOPE

ABC-CDE=crown glass, BCD=flint glass, RV=red to violet spectrum

A type of "constant-deviation spectroscope," invented in 1899 by P. Pellin and A. Broca, has a four-sided prism ABCD (fig. 5) which is effectively a combination of two 30° refracting prisms ABE and DAC, and a totally reflecting right-angle prism BEC. Since the two refractions cancel and a right-angle deviation is introduced by the internal reflection the collimator and telescope axes are fixed 90° apart and different wave lengths are brought into the field of view by rotating the prism with a screw to which a wave length calibrated drum is attached.

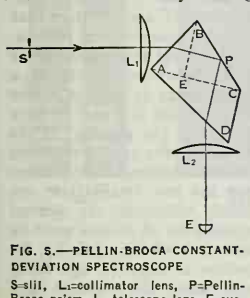


FIG. 5.—PELLIN-BROCA CONSTANT-DEVIATION SPECTROSCOPE

S=slit, L_1=collimator lens, P=Pellin-Broca prism, L_2=telescope lens, E=eye-piece

Dispersing Power of Prisms.—The power of any spectroscope to separate different wave lengths emerging at different angles and focused at different points is called its dispersion; it is defined as the rate of change of the angle of deviation, θ, with respect to a small change in wave length, λ, and represented mathematically by $d\theta/d\lambda$. Since θ varies with the refractive index, n, of the prism, and the refractive index varies with the wave length, λ, we may write

$$\text{Angular dispersion} = \frac{d\theta}{d\lambda} = \frac{d\theta}{dn} \cdot \frac{dn}{d\lambda}.$$

Dispersion thus depends upon two factors. The second factor, $dn/d\lambda$, is determined solely by the optical properties of the prism material. Since the refractive indexes of a glass prism are approximately represented by an equation of the (Cauchy) form $n=a+b/\lambda^2$, where a and b are constants obtained from measured values of n for two different wave lengths, $dn/d\lambda = -2b/\lambda^3$. The negative sign merely indicates increase of n with decrease of λ. This formula shows that the dispersive power of a glass prism varies inversely as the cube of the wave length. For example, the dispersion at 4,000 A is about eight times that at 8,000 A. The other factor, $d\theta/dn$, depends upon the angle of the prism, and on the angle it presents to the incident light. If, as in fig. 6a, the angles of incidence and emergence be represented by i_1 and i_2 respectively, the internal angles of refraction by r_1 and r_2, the angle of the prism by α, and the angle of deviation by θ, it is readily deduced from Snell's law of refraction, $n=\sin i/\sin r$, that

$$\frac{di_2}{dn} = \frac{\sin (r_1+r_2)}{\cos r_1 \cos i_2}$$

or, since $d\theta = di_2$, and $r_1+r_2=\alpha$

$$\frac{d\theta}{dn} = \frac{\sin \alpha}{\cos r_1 \cos i_2}.$$

This expression clearly shows the dependence of the dispersion upon the prism angle and the angle of incidence. A convenient equivalent expression was first deduced by Lord Rayleigh; namely,

$$\frac{d\theta}{dn} = \frac{t_2-t_1}{a}$$

where t_2 and t_1, in fig. 6b, are the lengths in the prism traversed by the extreme rays, and a is the breadth of the emergent beam. When the prism is at minimum deviation and completely covered by the beam, t_2-t_1 is simply the base length of the prism. The original expression now becomes

$$\text{Angular dispersion} = \frac{d\theta}{d\lambda} = \frac{d\theta}{dn} \cdot \frac{dn}{d\lambda} = \frac{t_2-t_1}{a} \cdot \frac{dn}{d\lambda}.$$

FIG. 6.—DISPERSING POWER OF A PRISM

When there is more than one prism t_2-t_1 is summed for all prisms, and in Amici prisms the algebraic sum is taken, the lower dispersions being regarded as negative. The angular dispersion may be expressed in radians or degrees per angstrom, but it is usually more convenient to measure linear distances, x, in the focal plane and express linear dispersion $dx/d\lambda$ in millimetres per angstrom. In general, if f is the focal length of the image forming lens, $dx/d\lambda = f \cdot d\theta/d\lambda$.

Resolving Power of Prisms.—It would appear from this equation that, by choosing a large focal length, it would be possible to separate two close slit images due to light of two wave lengths, however little the wave lengths differed. Such, indeed, would be the case if an infinitely narrow slit were imaged as an infinitely narrow line. However, owing to the wave nature of light any optical image is a diffraction pattern of finite size which depends both on the wave length and on the dimensions of the optical system. An instrument is thus limited in its power to resolve images of slightly different wave length, or closely adjacent spectrum lines. This resolving power, R, is expressed numerically as the ratio of a mean wave length λ to the least difference $d\lambda$ between two wave lengths that can be recognized as a double image, *i.e.*, $R=\lambda/d\lambda$. Each actual image of the slit is a diffraction pattern having most of the light concentrated in a central band of finite width flanked by secondary bands of negligible intensity. If f be the focal length of the camera lens, and a the effective breadth of the beam passing through a rectangular cross section, the half width, w, of the central diffraction band is given by the elementary theory of diffraction as $w=f\lambda/a$. This says that the image width increases with focal length and with wave length, but decreases with aperture increase. The condition for resolu-

tion is explained in fig. 7 where the spectrum is focused on a screen by a lens of focal length f. The beams of two equal-amplitude closely-adjacent wave lengths λ and $\lambda+d\lambda$ are separated by the prism an increment of angle $d\theta$. Each slit image appears as a diffraction band, and, according to Rayleigh's criterion, resolution is just effected when the central maximum of one image coincides with the first minimum of the other, *i.e.*, when the distance between centres of the two bands is w. Under these conditions the integration of the two bands gives a curve with two maxima, and if the intensity of these be rated as unity that of the minimum between is 0.81. Experience has shown that this intensity difference is sufficient to reveal duplicity of the image

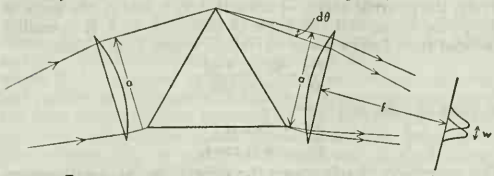


FIG. 7.—RESOLVING POWER OF A PRISM SPECTROSCOPE

and the condition for resolution is thus $d\theta = w/f = \lambda/a$; or $\lambda = a \cdot d\theta$. By definition, $R = \lambda/d\lambda$, and therefore $R = a \cdot d\theta/d\lambda$, or in Rayleigh's form,

$$R = (t_2 - t_1) \cdot dn/d\lambda.$$

When there are several prisms, the algebraic sum is to be taken as in the case of dispersion. An ordinary flint glass prism of one centimetre base is found to have a resolving power for sodium-yellow light of about 1,000 but this number increases rapidly with decreasing wave length since $dn/d\lambda = -2b/\lambda^3$ approximately.

An infinitely narrow slit was assumed in the derivation of the theoretical resolving power but such a slit would transmit no appreciable light. In practice it is necessary to use a slit of finite width and sacrifice some of the resolving power for the sake of intensity in the spectral images. The actual separating power with finite slits was called "purity" by A. Schuster who defined a "normal slit" as the product of a quarter wave length by the aperture ratio, $\frac{\lambda \cdot F}{4 \cdot D}$, and calculated that nearly 99% of the theoretical resolving power was conserved by using this normal slit. Thus, if $F/D = 20$ the normal slit for 4,000 A would have a width of 5λ or 2μ. If double the normal slit is used the resolving power is reduced to 94% of maximum but the intensity is doubled. Increasing the slit width to about four times normal reduces the resolving power to 80% but trebles the intensity. Still wider slits rapidly reduce the resolving power or spectral purity without appreciably intensifying the spectra of monochromatic images except where they overlap. A wide slit greatly increases the intensity of continuous background relative to that of monochromatic lines. The above definitions of dispersing power, resolving power and spectral purity are valid also for diffraction gratings discussed below.

Grating Spectrographs.—The use of diffraction gratings instead of refraction prisms for the study of spectra brings advantages of constant and usually greater dispersion and resolution, as well as opportunity to observe extreme ultra-violet and infra-red spectra for which no transparent prisms can be found. Modern gratings are made by ruling with a diamond point equidistant parallel lines on a polished plate of glass, of speculum (copper+tin) metal, or of glass on which a film of aluminum has been deposited. Those in most general use for visible and ultra-violet spectra have either 15,000, 20,000, 24,000, or 30,000 lines per inch and effective ruled areas ranging from about 2 to 6 inches. A grating with 15,000 lines per inch permits measurements in infra-red spectra to 30,000 A, but for the investigation of longer waves coarser gratings must be used. In 1910 R. W. Wood first ruled 2,000 to 3,600 lines per inch on copper plates and by proper choice of groove form succeeded in concentrating much of the visible light into relatively high orders. Such "echellete" gratings concentrate the greater part of infra-red energy in a first-order spectrum and they can be ruled to perform in this manner for almost any desired wave length. Thus, a grating with 900 lines per inch may be made to function well at about 300,000 A, 360 lines per inch around 900,000 A and 80 lines per inch around 1,500,000 A. These coarse echellete gratings have ruled surfaces up to 10x20 inches in area and are usually cut in a solder surface with a steel tool. A grating ruled on a sheet of transparent glass is called a "transmission grating," but it is impossible to rule large areas of glass without wearing down the diamond point. For this and other good reasons most gratings are ruled on opaque soft metal surfaces, and are called "reflection gratings." The fundamental principle is the same in both cases.

Theory of Diffraction Gratings.—The simple theory of the plane diffraction grating gives the wave length of light as a function of the grating space and of the angles of incidence and diffraction of the light rays. Fig. 8 represents the cross section of a transmission grating in which transparent and opaque lines or spaces alternate and each pair, such as AB, is called the grating space b. If light from an illuminated slit S falls as a parallel beam on the grating at an angle of incidence i with the normal, a portion of the light passes directly through the grating aperture to produce an undeviated image, but a part of it is diffracted in the direction AD making an angle θ with the normal. After drawing the perpendiculars CB and DB it is seen that the difference in

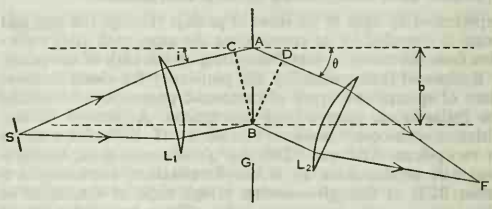


FIG. 8.—TRANSMISSION GRATING

S=slit, L_1=collimator lens, G=Grating, L_2=camera lens, b=grating space, i=angle of incidence, θ=angle of diffraction, F=focused spectrum

path travelled by corresponding rays from adjacent apertures, or the retardation as it is called, is given by the sum of the lengths CA and AD. When this retardation is equal to any integral number, n, of wave lengths, λ, bright images of S in colour corresponding to λ are produced where the diffracted rays are brought to a focus, F. This condition is simply expressed by the equation

$$CA + AD = n\lambda = b\ (\sin i + \sin \theta_n)$$

in which n represents the order number of the spectrum.

The law of reflection gratings is similar to that of transmission gratings. Thus in fig. 9 a beam of parallel rays falls on a plane reflection grating with grating space $AB = b$. The incident light makes an angle, i, with the normal, and part of the light is diffracted at an angle, θ. The perpendiculars AC and BD show that BC and AD are the retardations. In this case the total retardation is $AD - BC = b\ (\sin i - \sin \theta_n)$. Whenever this equals one or more whole wave lengths, $n\lambda$, a bright image is seen in colour corresponding to λ and we have the law of the grating. If the diffracted rays are on the same side of the normal as the incident rays the two retardations are added, wherefore the general equation for all spectra must be written $n\lambda = b\ (\sin i \pm \sin \theta_n)$, the positive or negative sign being used when the incident and diffracted rays are on the same or opposite sides of the normal, respectively. If the grating space is not too small the groove form of the ruled lines can be controlled by shaping the diamond point in such a way that the diffracted light is largely concentrated in

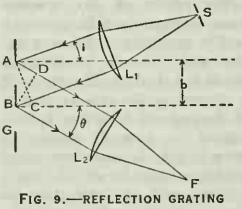
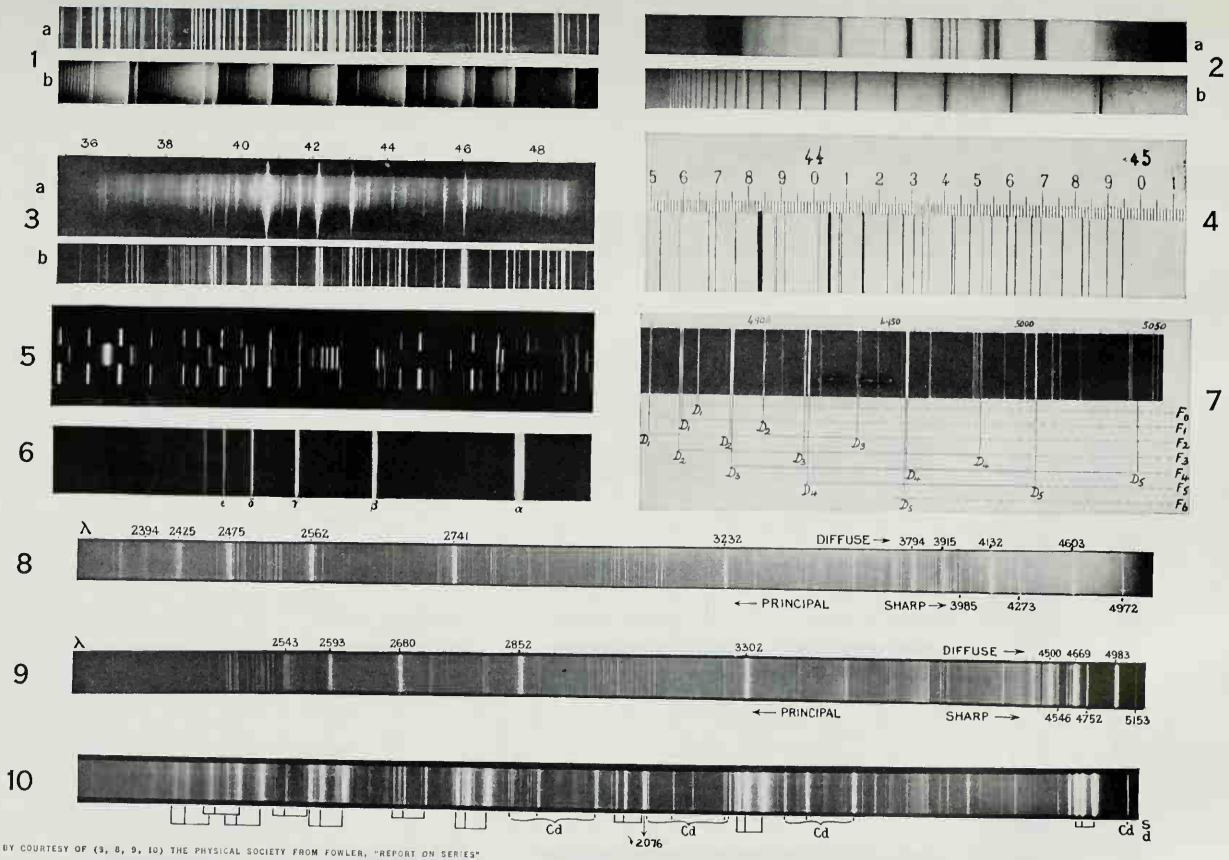


FIG. 9.—REFLECTION GRATING

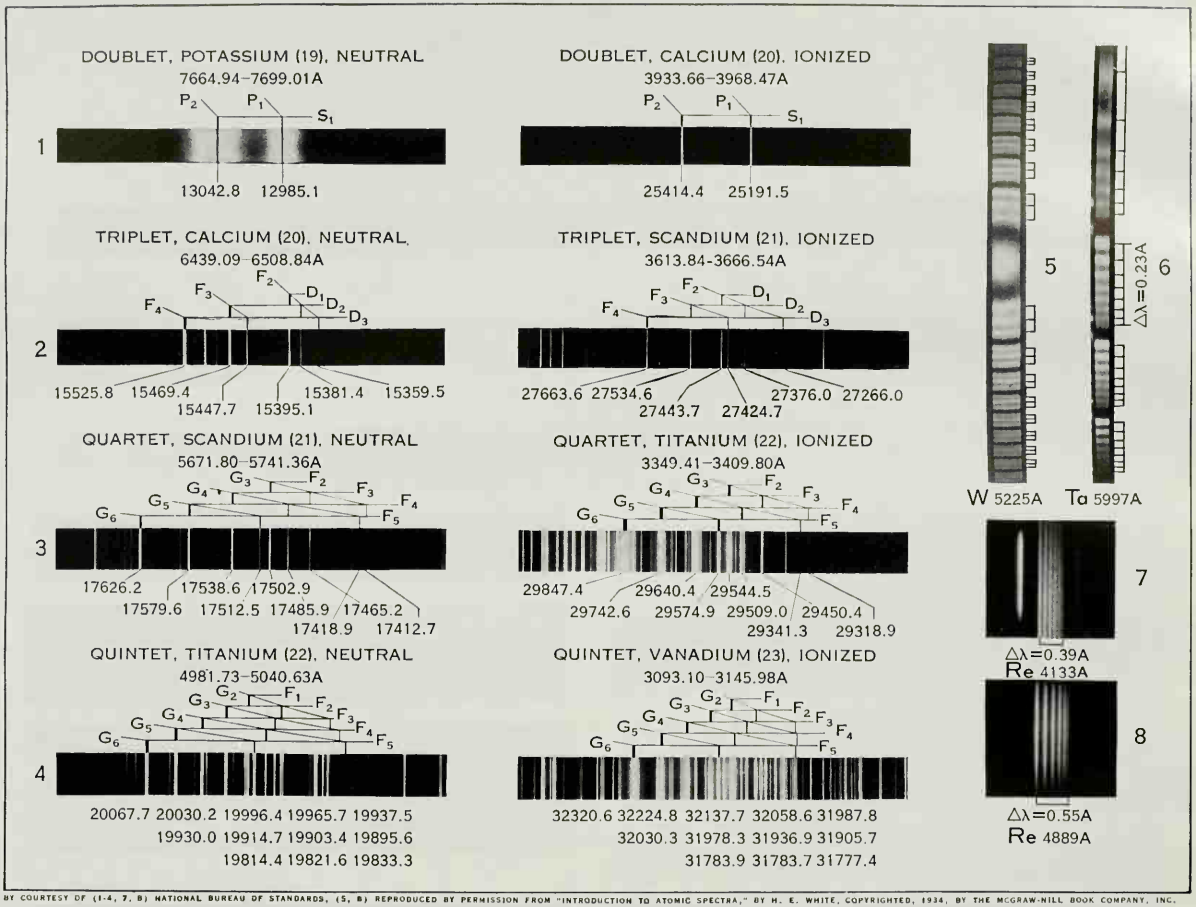
S=slit, L_1=collimator lens, G=grating, L_2=camera lens, b=grating space, i=angle of incidence, θ=angle of diffraction, F=focused spectrum



BY COURTESY OF (3, 8, 9, 10) THE PHYSICAL SOCIETY FROM FOWLER, "REPORT ON SERIES"

SELECTED PORTIONS OF TYPICAL SPECTRA

1. Spectra of nitrogen; (a) line spectrum, (b) band spectrum. 2. Absorption spectra; (a) neodymium glass, (b) potassium vapour. 3. Spectra of strontium; (a) spark, (b) arc. Flame lines are reduced in intensity in the spark spectrum, while "enhanced" or "spark" lines are intensified. 4. Portion of photographic map of iron-arc spectrum, from which approximate wave-lengths may be obtained. 5. Spectra of iron and oxygen (middle), placed in juxtaposition for comparison and determination of wave-lengths. 6. "Red" spectrum of hydrogen, showing Balmer series of lines peculiar to this element in the visible and near ultra-violet regions. 7. Arc spectrum of iron, under high dispersion, showing a multiplet of 15 lines and plan of its structure. 8. Arc spectrum of lithium, showing the constituent principal, sharp and diffuse series. The lines, apparently single, are very close doublets. 9. Arc spectrum of sodium, similar to that for lithium. The terms principal, sharp, and diffuse refer to the appearance of the lines composing the respective series. 10. Arc spectrum of zinc and cadmium, showing triplet systems



BY COURTESY OF (1-4, 7, 8) NATIONAL BUREAU OF STANDARDS, (5, 8) REPRODUCED BY PERMISSION FROM "INTRODUCTION TO ATOMIC SPECTRA," BY H. E. WHITE, COPYRIGHTED, 1934, BY THE MCGRAW-HILL BOOK COMPANY, INC.

TYPICAL STRUCTURE OF ATOMIC SPECTRA

1. Doublet-, 2. Triplet-, 3. Quartet-, 4. Quintet multiplets of neutral and of ionized atoms. (Reduce subscripts of even-multiplicity levels by ½ to obtain true J values.) The six-place numbers under each multiplet are wave numbers associated with the lines. The alternation law is illustrated vertically, and the displacement law horizontally. A septet multiplet of iron is shown in Plate I, fig. 7. 5. Fabry-Perot interferometer pattern of the tungsten line, 5225A, showing components for three principal isotopes, W^{182}, W^{184}, W^{186}. 6. Fabry-Perot pattern of tantalum line, 5997A, showing eight components of hyperfine structure. 7 and 8. Plane-grating images of rhenium lines, 4133 and 4889A, showing six components of hyperfine structure

one general direction or spectral region. Such gratings may for example, give an intense visible spectrum and weak ultra-violet, or vice versa. If light is incident normally on any grating the angle i is zero and the grating formula reduces to $\pm n\lambda = b \sin\theta$ which represents a central zero-order image flanked on either side by first, second, third, etc., order spectra according as $n = \pm 1$, ± 2, ± 3, etc. It is obvious that the absolute values of wave lengths may be determined from the measurement of the grating space and of one (or two) angles. This was the method employed by Fraunhofer, and by Ångström to establish standards or fiducial points in spectra.

Concave Gratings.—All plane gratings, like prisms, require the use of lenses or mirrors to collimate and to focus the incident and emergent light. To eliminate lens aberrations and limitations of light absorption, Rowland in 1881 first ruled a grating on a spherical mirror which collected light from a slit, dispersed the diffracted light into spectra and focused them on a circle the diameter of which equalled the radius of curvature of the concave grating. All concave gratings mounted on the Rowland-circle principle exhibit strong astigmatism, *i.e.*, a point slit is imaged as a line with a length dependent on wave length and spectral order. This astigmatism reduces the brightness of the spectra and precludes the use of diaphragms or sectors at the slit, and other observations requiring a one to one correspondence in the slit and its spectral images. Fortunately this defect can be avoided either by placing horizontal lines to be focused on the circle at a greater distance from the grating than the slit, or by illuminating the concave grating with parallel light, in which case the astigmatism is practically eliminated and the brightness of spectra greatly increased.

With both plane and concave gratings a great advantage is obtained by forming and observing spectra on the normal to the reflecting surface, for this gives the "normal spectrum" in which angular or linear distances measured either way from the grating normal are, for a considerable distance, proportional to the change in wave length observed. Since the concave reflection grating is free from chromatic aberration and focuses all spectra on the same focal surface, Rowland took advantage of the normality of spectra and of the coincidence principle in overlapping orders of spectra to determine the relative values of the "Rowland system" of standard wave lengths.

Dispersion.—The explanation of dispersion by a grating is easily seen since the formula shows that the angle θ increases of decreases with the wave length λ. When composite light falls on the grating it is obvious that the shorter violet waves will give an image closer to the central image than the longer red waves, and successive orders of spectra will increase in dispersion proportional to the order number.

If θ be the angle of diffraction, dispersion is defined as before by $d\theta/d\lambda$, and since, for a given position of the grating, the angle of incidence, i, is constant, the grating formula gives $d\theta/d\lambda = n/b \cos\theta$. This equation shows clearly how the dispersion varies directly with the order of the spectrum, and inversely as the grating space; the closer the rulings the greater the dispersion, irrespective of the total number of grating spaces. Furthermore the dispersion is smallest when $\cos\theta$ has its maximum value (which is unity for zero angle), *i.e.*, when the spectrum is observed normal to the plane of the grating. In this position also the dispersion is most nearly uniform throughout the spectrum, since it varies with $\cos\theta$, and in the proximity of $\theta = 0$ this trigonometric function changes very slowly. The spectrum given by a grating, unlike that given by a prism, is accordingly "normal" in so far as $\cos\theta$ can be considered constant. There is a certain amount of overlapping of successive orders since $n\lambda$ may acquire the same value in different ways. For example, $\lambda =$ 7,000 A in the first order will be coincident with $\lambda = 3{,}500$ A in the second order, and $\lambda = 4{,}000$ A in the third order will coincide with $\lambda = 6{,}000$ A in the second. When observation of a particular spectrum is desired it is usually possible to remove the overlapping spectra by using suitable colour filters in front of the slit or selective detectors at the focus.

Resolving Power.—As in the case of the prism, the resolving power of a grating is a theoretical quantity expressing the image-separating power when the slit is indefinitely narrow; it is represented symbolically by $R = \lambda/d\lambda$, and is given in simple form by $R = nN$, where n is the order number of the spectrum and N is the total number of rulings. Thus the resolving power of a grating depends only on the total retardation in wave lengths from the first to the last line, and not on the grating space. For example, a five-inch grating with 20,000 lines per inch would have a theoretical resolving power of 200,000 in the second order, surpassing an imaginary glass prism of two metres base length. At $\lambda = 8{,}000$ A, $d\lambda$ would be 0.04 A, while at $\lambda = 4{,}000$ A it would be 0.02 A. This theoretical resolving power is attained only under ideal conditions since actual resolution depends on various factors such as the relative intensities of spectral lines, their form or structure, the width and manner of illuminating the slit, the perfection and adjustment of the optical parts, graininess or contrast of photographic emulsions, etc.

Interferometers.—Since the resolving power of a grating is proportional to the product (nN) of the order number and the total number of elements, it may be increased in two ways. To increase the number of elements the number of lines per inch may be increased or a greater number of inches ruled, but the total number of useful lines is usually limited to about 100,000 by the practical difficulties of making all grooves exactly alike and equally spaced. The alternative method of using higher orders is limited in highly dispersive gratings by the great reduction in aperture when the grating is viewed tangentially. These limitations in ruled gratings are circumvented more or less in certain interferometers, known as Michelson's echelon interferometer, the Fabry-Perot interferometer and the Lummer-Gehrcke interferometer, all of which employ a relatively small number of elements (10 to 30) but very large spectral orders (10,000 to 100,000 or more). Since such high orders are very close together confusion due to overlapping of many lines must be avoided by crossing these interferometers with any stigmatic spectrograph. The Michelson transmission echelon (fig. 10) consists of a number of glass plates of equal thickness, t (*c.* 1 cm.), and refractive index, μ, but each shorter than its predecessor. These are wrung into optical contact with each other to form a series of steps, *i.e.*, an echelon. Parallel light from a collimator, with the slit parallel to the edges of the steps, is then passed through the larger end.

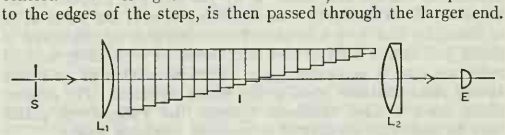


FIG. 10.—MICHELSON ECHELON INTERFEROMETER

S=slit, L_1=collimator lens, I=interferometer, L_2=achromatic lens, E=eyepiece

In each step the light is retarded an equal amount $(\mu - 1)t/\lambda$ and diffraction spectra of very high order are formed when all the transmitted beams are collected and focused by a lens. This interferometer was later improved as a reflection echelon which is not limited by the absorption of glass and has nearly four times the resolving power since the retardation per step is $2t/\lambda$. An echelon of 30 plates each 1 cm. thick has a theoretical resolving power at 5,000 A of $30 \times 2/.00005 = 1{,}200{,}000$. This instrument has been employed for the examination of spectral line structure and for the precise measurement of relative wave lengths.

When a beam of light passes through a transparent film or plate multiple reflections occurring at the surfaces produce interference colour effects seen for example in soap bubbles or in oil films on a wet pavement. Such interference bands are relatively broad because the reflecting power is so low that only the first reflection at each surface is effective. When the reflecting surfaces are exactly parallel and the reflection coefficient is increased, the multiple reflections produce very narrow bright fringes on a broad dark background. These conditions are met in the Fabry-Perot and in the Lummer-Gehrcke interferometer. In the former a large coefficient of reflection is obtained by depositing bright partly-transparent metal films on two glass or quartz planes. These planes are mounted face to face, separated some milli-

metres or centimetres, and adjusted to exact parallelism. The most common type of Fabry-Perot interferometer (fig. 11) is the fixed-etalon type in which the plates are maintained at a fixed distance by being pressed against a separating ring spacer or etalon. When illuminated with monochromatic light, and viewed axially, circular interference bands, known as Haidinger's fringes, are seen at infinity. When illuminated with complex light the overlapping of fringes belonging to neighbouring spectral lines is avoided by crossing the interferometer with a stigmatic spectrograph. The order of interference at the centre of the pattern for wave length λ is in general a whole number, n, plus a fraction, ϵ, determinable from the diameters of successive rings. The double distance between the interferometer plates is $2D=\lambda\,(n+\epsilon)$ for wave length λ, or $2D=\lambda'(n'+\epsilon')$ for another wave length, λ'. Thus the simple proportion $\lambda/\lambda'=(n'+\epsilon')/(n+\epsilon)$ offers an opportunity for the precise comparison of wave lengths, and this is in fact

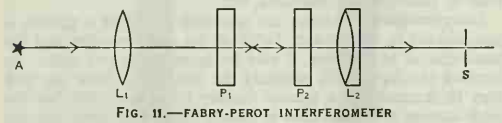


FIG. 11.—FABRY-PEROT INTERFEROMETER

A=light source, L_1=collimator lens, P_1 and P_2=interferometer plates, L_2=achromatic projection lens, S=spectrograph slit

the method employed for the establishment of secondary standards of wave length measured relative to the primary standard (red radiation from cadmium) which was itself found to be 6.438.4696 A by comparison with the meter. (*See* INTERFEROMETER.) This method of comparing wave lengths has been used extensively also for measuring the ratios of wave lengths in vacuo and in gases and thus determining the refractive indexes and dispersions of gases, notably air. In addition the Fabry-Perot etalon interferometer has proved the most effective instrument for the measurement of exceedingly small wave length differences which appear, for example, in hyperfine structure, isotope shifts, pressure shifts, etc. Besides simplicity, stability and freedom from defects this interferometer possesses a large advantage in the flexibility of its resolving power, since this is equal to the product of the spectral order, n, by the effective number of reflections, N. With a reflection coefficient of 0.75 the value of N is about 10, and with 0.90 the effective number of reflections is about 30. Thus a Fabry-Perot interferometer with silver mirrors reflecting 0.90 is equivalent to a reflection echelon of 30 plates having plate thickness equal to the etalon separator. The echelon plates have a fixed thickness whereas the Fabry-Perot plates may be separated as desired to increase resolving power.

In fact, the highest resolving powers on record have been attained with Fabry-Perot interferometers and the most detailed and reliable results on hyperfine structure must be credited to them.

The Lummer-Gehrcke interferometer consists of a long unsilvered plate with perfectly plane parallel surfaces in which a high reflection coefficient is obtained by directing the entering light beams so that they are reflected at nearly the critical angle (fig. 12). The greater part of the light is reflected at each internal reflection but the remaining fractions emerge with equal retardations as parallel refracted beams which may be collected and focused by a lens to form identical sets of interference patterns above and below the optic axis. The resolving power is

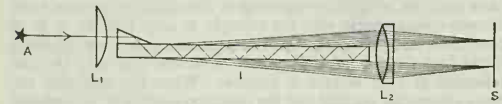


FIG. 12.—LUMMER-GEHRCKE INTERFEROMETER

A=light source, L_1=collimator lens, I=interferometer, L_2=achromatic projection lens, S=spectrograph slit

again $n\,N$ where n is the order of interference (retardation) and N the effective number of reflections or emergent beams, which usually ranges from 10 to 20. The practical resolving power of the Lummer-Gehrcke interferometer has been found to be about equal to the length of the plate measured in wave lengths. This instrument has been employed successfully for the observation of hyperfine structure, isotope shifts and Zeeman patterns.

The most famous of all interferometers is the very first one, invented by A. A. Michelson in 1892, and fully exploited by him in his classical investigations of "ether drift," sharpness of spectral lines, and evaluation of the international metre in wave lengths of cadmium radiations (*see* INTERFEROMETER).

METHODS OF OBSERVING OR RECORDING SPECTRA

There are in general four ways of observing spectra: visually, photoelectrically, radiometrically, and photographically; each having certain advantages and limitations.

Visual.—Although the average human eye is extraordinarily sensitive for green light (5,500 A), its sensitivity declines rapidly to zero for infra-red ($\lambda>7{,}700$ A) and ultra-violet ($\lambda<3{,}800$ A). Being highly selective, variable, localized and restricted to one octave, visual methods of observing are relatively unimportant in spectroscopy.

Photoelectric.—As a substitute for the human eye the "electric eye" or photoelectric tube may be employed for the investigation of spectra in limited ranges. Briefly, the photoelectric tube consists of two metal electrodes in an evacuated glass tube. The anode is maintained at a positive potential with respect to the cathode. When radiation falls on the cathode the liberated electrons are drawn across the evacuated space to the anode. This flow of electrons constitutes a photoelectric current which is directly proportional to the incident energy. Since these photoelectric currents are usually very small they are magnified by the use of thermionic amplifiers, or in the case of feeble radiation, use is made of a Geiger-Müller tube in which each electron emitted gives rise to a pulse of current. Since 1940 multiplier phototubes, in which enormous amplification is obtained by a sort of chain reaction, have become popular as detectors and intensity recorders of spectral lines. All photoelectron emitters (*e.g.*, Na, Zn, Cd, Mg) are highly selective and usually most efficient in the ultra-violet. By depositing thin films of heavy metals or compounds (*e.g.*, CsO, PbS) photoelectric sensitivity may be extended to near infra-red, but not much beyond 40,000 A.

Radiometric.—A unique and important class of radiation detectors, responding to the heating effect of the radiation, is found in blackened radiometers, radiomicrometers, thermopiles and bolometers. They are unique because they are nonselective, and important because their sensitivity extends from ultra-violet, through visible, to far infra-red spectra. The radiometer is a blackened vane supported some distance from the axis of rotation of a glass or quartz fibre in vacuo. When radiation is absorbed by the vane the supporting fibre is twisted through an angle proportional to the incident radiation. The radiomicrometer is essentially a moving coil galvanometer of the D'Arsonval type. A single thermocouple is suspended by a quartz fibre with the plane of the couple coil in the line joining the poles of a magnet. When radiation falls on the thermocouple junction the electric current generated causes the couple to turn in the magnetic field. The thermopile consists of one or more junctions of dissimilar metals, *e.g.*, bismuth and antimony or silver, iron and constantan (alloy of nickel and copper). If radiation falls on alternate junctions a galvanometer in the circuit will give a deflection proportional to the heating effects. All radiation detecting and measuring devices depending upon the initial conversion of radiant energy into heat energy and finally into mechanical or electrical energy are extremely inefficient because all heat engines are naturally inefficient. The feeble final effects of radiation may be magnified up to about 200 times by suitable amplifiers but thermopile currents causing galvanometer deflections of the same magnitude as Brownian-motion effects obviously cannot be amplified to advantage.

The bolometer, devised by S. P. Langley in 1880, consists of two similar very thin strips of metal, usually platinum, which form two arms of a Wheatstone bridge. The strips are blackened

on one side, one strip is shielded and the other is exposed to radiation. The absorption of radiation by one strip increases its temperature, thus changing its electrical resistivity and causing a deflection of the galvanometer. With his bolometer Langley, in 1888, mapped the solar spectrum to 55,000 A by using a spectrometer with rock salt lenses and prisms and passing the spectrum over the bolometer strip by rotating the prism. The galvanometer readings were recorded on photographic paper whose motion was co-ordinated with that of the prism. The same basic principles are today employed, with improved instrumentation, in all automatic recording of spectra scanned either by photoelectric or radiometric detectors.

Photographic.—The actinic action of radiation on silver salts is fairly effective throughout many octaves of spectra, from blue to extreme ultra-violet, but the maximum sensitivity is usually in the blue. Photographic image formation rests on a latent and lasting effect produced by light in microscopic silver halide crystals suspended in gelatin coated thinly on a glass, film, or paper base. The crystals affected by light are readily reduced to metallic silver by a suitable reducing agent or developer, and the unreduced silver halide is dissolved out with a fixer, usually "hypo," thus producing images composed of tiny silver grains, the density of which is a function of the intensity or total energy incident upon the emulsion. Although gelatin is highly transparent to waves longer than 3,000 A, it is strongly absorbent for shorter waves and practically opaque below 2,000 A. In order to record spectra below 2,000 A, in 1901 V. Schumann effectively eliminated the shielding effect of gelatin by depositing silver salts on glass plates with only enough gelatin to support them. Another way to circumvent the ultra-violet absorption of gelatin was discovered in 1921; it consists of coating an ordinary photographic plate or film with a thin film of oil or suitable organic substance that fluoresces in ultra-violet light and emits longer wave radiation not absorbed by the gelatin. (*See* PHOTOGRAPHY.)

In 1873 H. M. Vogel discovered that the light sensitiveness of silver salts could be extended to longer waves by bathing ordinary plates in certain dye solutions. By 1900 the incorporation of suitable photosensitizing dyes in emulsions produced panchromatic plates that were sensitive to the entire visible spectrum, and after 1930 the synthesis and application of an important series of polycarbocyanine dyes extended the sensitivity of photographic emulsions into the infra-red to about 13,000 A. Since large intervals of spectra and multitudes of spectral details could thus be recorded simultaneously it follows that more spectrographic observations have been made photographically than by all other methods. In addition to being applicable in an enormous range of wave lengths the photographic method provides a permanent, detailed record of a spectrum that can be conveniently examined or accurately measured by any number of persons as often as desired without repeating the experiment. Furthermore, photographic exposures are cumulative, and the photographic method thereby derives additional advantages in integrating the light from flickering or intermittent sources and in recording extremely faint images by giving long exposures. Although the absolute sensitivity of a photographic emulsion may vary considerably with wave length, it can often be regarded as constant over a limited range of spectrum, and in any case can be calibrated as a function both of wave length and intensity, thus providing a convenient method of measuring the relative intensities of spectral lines. This photographic-photometry method was developed about 1925 for the purpose of testing quantum rules for the relative intensities of certain lines, and since 1930 it has become standard practice in quantitative spectrochemical analysis.

It should be emphasized that any selective receiver (eye, phototube or photographic plate) employed for measuring the relative intensities of spectral lines requires calibration by comparison with nonselective receivers (thermopiles, etc.).

OBSERVATIONS AND MEASUREMENTS

Spectroscopic measurements are concerned essentially with energy distribution as a function of wave length. This function is characteristic primarily of the light source and secondarily of its operating conditions, assuming that the spectrograph is nonselective. If the energy changes slowly from point to point in large intervals of spectra, the spectra are called "continuous." They may be either emission spectra or absorption spectra and consist usually of one broad band.

If the spectral intensity changes rapidly from point to point and experiences sudden inflections the spectra are called "discontinuous." Such spectra also may be either emission spectra, or absorption spectra; they usually consist of hundreds or thousands of lines. The division of observable spectra into continuous and discontinuous, either in emission or in absorption, is a broad but useful classification. Each class contains an enormous number of examples, some of which may be grouped according to certain criteria into smaller classes. The following examples will illustrate the principal classes of spectra that are subjected to observation and measurement.

Emission Spectra.—All self-luminous sources yield emission spectra, which are of different types according to the nature of the light source. Spectrally dispersed light from glowing gas mantles, electrically heated metallic filaments, or hot molten metals are continuous over great ranges of wave lengths, and for "black body" radiation the wave length of maximum intensity is an inverse function of absolute temperature. It follows that such substances cannot be distinguished by means of their spectra so long as they remain in the solid or liquid state.

The effects are quite different when the substances examined are in a state of luminous gas or vapour. They then emit characteristic discontinuous spectra by which they can be uniquely identified, whether they are excited in the laboratory or in a celestial body far away. These spectra consist of bright lines or bands variously spaced on a dark background.

Line spectra are characteristic of atoms or ions, while band spectra originate in molecules. The former consist of a certain distribution of slit images of different intensity and usually of various types such as partially self-reversed lines, sharp lines, diffuse lines, unsymmetrical lines and complex lines with two or more components. Partially reversed or easily absorbed lines always indicate that the normal state or some low energy state of the atom or ion is involved. The relative sharpness or diffuseness of lines is determined by such factors as pressure, temperature, mass and excitation. Complexity of spectral lines is referred either to different isotopes of the same element (isotope shift) or to interaction of valence electrons with a spinning isotopic nucleus (hyperfine structure).

In a band spectrum a great number of fine lines are closely and regularly placed in each band, becoming more and more crowded near the band head. The band heads are on the short-wave side of the line groups in some band systems, and on the long-wave side in others. The differences between line and band spectra are illustrated in Plate I where fig. 1a and fig. 1b are portions of the line and band spectra respectively of nitrogen. Some bands are less regular in appearance, and others do not appear to be resolvable into lines. In pure gases the commonly observed band spectra characterize molecules consisting of two similar atoms, such as H_2, He_2, N_2, O_2. The line spectra of metals excited in flames, arcs or sparks operated in the free atmosphere are usually accompanied by band spectra characteristic of monoxides such as CaO, BaO, ScO, LaO, formed by the oxidation of the metals. Arc spectra frequently show band spectra identified with NO and CN as well as line spectra of N, O, A and C, the principal constituents of the atmosphere. Many dipole molecules unknown to ordinary chemistry but recognized spectroscopically are apparently formed only under special conditions existing in the light sources.

No two substances yield the same spectrum and consequently the chemical nature of substances can be determined spectroscopically. Thus glowing atomic hydrogen is characterized by a bright line in the red, besides progressively weaker ones of shorter wave length (*see* Plate I, fig. 6), and since these are exhibited by nothing but hydrogen, they serve to disclose the presence of atomic hydrogen wherever it occurs in the luminous state. Each of the known chemical elements has its distinctive family of

spectral lines, differing in distribution, relative intensity and total number, but always the same under the same conditions of excitation. The same element, however, under radically different excitation conditions will emit entirely different spectra. The most easily excited spectrum is that of neutral atoms in possession of all their valence electrons, but if the excitation is violent enough to remove one electron from each atom the resulting ions when excited radiate another characteristic spectrum. A third type of spectrum is emitted when the atoms are doubly ionized by removal of two electrons, and so on until only one outer electron remains. Thus each element or atom has as many characteristic spectra as it has electrons. In most arc or spark sources generating atomic spectra of a pure element two or more spectra appear simultaneously with one usually predominating (*see* Plate I, fig. 3a and fig. 3b). When two or more elements with comparable excitation characteristics are abundant in a light source each element emits its own spectra essentially the same as if all other elements were absent. The spectra of complex substances such as ores, minerals, metal alloys and chemical mixtures thus reveal simultaneously all the types of atoms and ions that are excited in the light source. Such supernumerary line spectra often have superposed on them band spectra of molecules formed and excited in the same light source as well as lines and bands of atmospheric elements, and in addition, a more or less intense continuous background which may be traced either to unquantized radiation from the atoms, to continuous spectrum from incandescent oxides in the light source, or to stray reflected or scattered light in the spectrograph, sometimes to all three. In such complicated cases it is obvious that only experienced spectroscopists can interpret the spectrograms correctly. Familiarity with the spectra of various elements often enables an experienced observer to identify lines or bands at sight for the distribution and relative intensities in certain groups are memorized and recognized in the same way that constellations are recognized among stars. In theory any unknown spectrum or mixture of spectra can be identified by comparing it with all the spectra of known elements excited under similar conditions, but in practice the endless labour of repeatedly observing the comparison spectra is avoided by measuring wave lengths corresponding to the unknown lines and comparing these with wave-length tables of identified lines compiled for this purpose.

For purposes of chemical identification the wave lengths and the relative intensities of lines in spectra may be regarded as constants of nature, but closer examination shows that in any given spectrum certain variations are produced by different excitation conditions. Thus spectral wave lengths are generally slightly increased as the total pressure of gas or vapour surrounding the radiating particles is increased. Such pressure shifts are further correlated with atomic diameters and with atomic energy levels.

Also the wave lengths are either decreased or increased, according to Doppler's principle (*see* DOPPLER EFFECT) when the radiating source approaches or recedes from the observer. In astrophysics this principle has disclosed the sun's rotation, the radial velocities of stars, and an apparently expanding universe, whereas in the laboratory it explains, when combined with the kinetic theory of gases, the intrinsic natural width of spectral lines as a function of $\sqrt{T/M}$ in which T represents absolute temperature and M molecular weight. When standard wave lengths in the solar spectrum are measured it is necessary to correct the observations for Doppler effect due to the relative motion of the sun and the spectrograph.

Again, with few exceptions, all atomic spectral lines are split into three or more components when the radiating particles are in strong magnetic fields (*see* ZEEMAN EFFECT) and similarly many lines either shift or widen or divide into resolved components when the radiating particles are subjected to strong electric fields. (*See* STARK EFFECT.)

From the general theory of relativity (*q.v.*) A. Einstein in 1917 concluded that wave lengths characteristic of atoms should increase in strong gravitational fields. This effect is quite inappreciable in the laboratory but it appears to be present in the solar spectrum, and especially in the spectra of very dense stars.

The most striking changes in relative intensities of lines belonging to a particular spectrum occur when excitation conditions favour the partial self-reversal of selected lines. These variations of spectral characteristics are highly objectionable in standard wave lengths and in spectrochemical analysis, but they are of primary importance in spectroscopic research since they yield information about atomic structure and the mechanism of radiation. Consequently, when truly invariable spectra are desired they may be produced by carefully specifying and reproducing the light source in such a manner that spectral variations are eliminated or minimized, but if atomic structure is being investigated it is desirable to effect the largest practicable changes in spectra and relate these to variations in the light sources. In any case it is evident that some means of accurately expressing the positions of lines in a spectrum is of primary importance.

Wave Lengths.—Measurements of spectra become directly comparable with each other only when they have been converted to the scale of wave lengths (or frequencies) of the radiations which produce the spectral lines. The credit for introducing this scale into spectroscopy is due to A. J. Ångström who described the spectrum of the sun in terms of wave lengths in 1868, and since that time the positions of practically all known spectral lines have been expressed in wave lengths. The unit of wave length in spectroscopy is a ten-millionth part of a millimetre or 10^{-10} m. It is called the "angstrom" and is symbolized by A. In these units, the visible spectrum ranges from about 3,800 A (violet limit) to about 7,700 A (red limit). Extremely short waves, in particular those observed in Roentgen or X-ray spectra which range from about 0.1 A to about 100 A, are more conveniently expressed in still smaller units, and for this purpose the X unit, approximately 1/1,000 A, is generally used. Similarly, very long optical waves, in particular the far infra-red heat waves, are more conveniently expressed in larger units, and for this purpose the micron, $\mu = 10{,}000$ A, is commonly used. Summarizing: 10^{-6}m. = 0.001 mm. = 1μ = 10,000 A = 10,000,000 X. Wave lengths ranging from 6 A to 5,000,000 A have been measured in spectra associated with valence electrons but these limits will surely be extended in both directions. Absolute wave lengths cannot be determined directly from prisms but have been measured with gratings and interferometers by methods outlined in the discussion of these instruments. Because such measurements are relatively difficult and tedious they have been made only for the purpose of setting up standards of wave length, and in all other spectroscopic work, whether with prisms or gratings, wave lengths are deduced by interpolation between adopted standards.

Wave Numbers.—In the interpretation of spectra as transitions between quantized energy states ($E_2 - E_1 = h\nu$), it is necessary to express the positions of the spectral lines on a frequency, ν, scale instead of a wave-length scale. Unfortunately, neither nature nor human ingenuity has disclosed a method of directly measuring the oscillation frequencies corresponding to light waves. These frequencies, defined as the number of waves that pass a given point in one second, must be derived from the relation $\nu = c/\lambda$ in which c is the velocity of light and λ is the observed wave length. In order to make these frequencies independent of the medium through which the light passes both c and λ must be evaluated in vacuo. Since the velocity of light is very great (29,977,600,000 cm. per second) and a light wave is very small (0.00000550 cm. for brightest visible light) the frequency is an extremely large and awkward number (545,000,000,000,000 in this example). In spectroscopy it has become common practice to replace the ungainly actual frequencies by "wave numbers," representing the number of waves in one centimetre of vacuum and symbolized by cm.$^{-1}$. If the wave length has been measured in air it must be multiplied by the refractive index of air to change it to its value in vacuo. In 1926 H. Kayser prepared his *Tabelle der Schwingungszahlen* for the convenient conversion of wave lengths in standard air to wave numbers in vacuo. For example 5,500,000 A = 18,176.781 cm.$^{-1}$. Such wave numbers of observed spectral lines are strictly proportional to the natural frequencies and to the corresponding changes in atomic or mo-

lecular energy.

Standard Wave Lengths.—In modern spectroscopy the primary standard of wave lengths is that of the red radiation from cadmium vapour first measured relative to the international metre bar by interferometer methods in a classical experiment by Michelson in 1892. (*See* INTERFEROMETER.) Benoit, Fabry and Perot, in 1905, found that 1 metre contained 1,553,164.13 of these waves and the reciprocal of this number has been internationally adopted as the definition of the primary standard: "the wave length of the red ray of light from cadmium is 6,438.4696 angstroms in dry air at 15° C. on the hydrogen thermometer, at a pressure of 760 mm. of mercury, the value of g being 980.67 (45°)." The probable error of this absolute value is of the order of 1 part in 16,000,000, *i.e.*, somewhat less than the least error in the intercomparison of two metre bars.

A half century of intensive study of the spectra of natural elements proved that the red radiation from cadmium vapour is probably the best primary standard that nature provides, but in 1940 W. Wiens and L. Alvarez demonstrated that a pure artificial isotope of mercury (Hg^{198}) obtained by bombarding gold (Au^{197}) with neutrons probably presents a superior primary standard. It is not unlikely that a progressive world will some day adopt the wave length of green radiation (5,461 A) from Hg^{198} as its most convenient, constant, indestructible, accurately definable, and reproducible fundamental standard of length.

A method of measuring, with high accuracy, the ratio of two different wave lengths was outlined in principle in our brief discussion of the Fabry-Perot etalon interferometer. After 1907 this method was used in several laboratories to determine, relative to the primary standard, the wave lengths of selected lines which might serve as secondary standards. The wave lengths of several thousand atomic radiations characteristic of helium, neon, argon, krypton, xenon, iron, titanium in laboratory spectra, and well over 1,000 lines in the sun's spectrum, have thus been measured, but only such mean values are accepted as secondary standards by the International Astronomical union as have been independently and concordantly measured in at least three laboratories. For example, in the iron spectrum interferometric measurements have been made on thousands of lines with wave lengths ranging from 2,101 A to 10,216 A but only 306 secondary standards ranging from 2,447 A to 6,677 A have been adopted. These iron secondary standards are believed to be accurate within ±0.001 A but to insure their reproducibility the source is carefully specified as "the Pfund arc operated between 110 and 250 volts, with 5 amperes or less, at a length of 12–15 millimetres used over a central zone at right angles to the axis of the arc, not to exceed 1–1.5 millimetres in width, and with an iron rod 6–7 millimetres diameter as the upper pole and a bead of iron oxide as the lower pole." The International Astronomical union has also adopted as tertiary standards the wave lengths between 3,371 A and 6,750 A of 312 iron lines derived from secondary standards by grating or interferometer interpolation. In the absence of official standards in the extreme ultra-violet, spectroscopists have been forced to use preliminary values obtained by the coincidence method of overlapping orders in grating spectra or by direct calculation from spectral series formulae or from spectral terms accurately determined from longer waves.

Interpolation of Wave Lengths.—The wave lengths corresponding to lines in any spectrum whatsoever are readily determined by interpolation between standards, provided that the spectrum of the international iron arc is recorded with the same spectrograph. For this purpose the spectrum of the source under examination is photographed through one portion of the slit, and the arc spectrum of iron through one or both of the adjacent portions, so that two spectra in juxtaposition, or slightly overlapping, are obtained as illustrated in Plate I, fig. 5. For accurate determinations the relative positions of spectral lines and standards are measured with a micrometer or comparator consisting of an accurate screw designed to translate the spectrogram under a microscope so that the lines and standards are brought successively into coincidence with a reticle. The positions of the lines are then read from a scale which indicates complete turns of the screw, and a divided drum head which registers fractions of a turn.

In prismatic spectra the dispersion increases rapidly as the wave length diminishes, and linear interpolation will not give correct results. Approximate values of wave lengths may be obtained by interpolation on a graph of dispersion versus wave length constructed by plotting scale readings of standard lines as abscissae and wave lengths as ordinates. The most accurate method of determining wave lengths in prismatic spectra is by mathematical interpolation with the Hartmann dispersion formula which gives a simple relation between wave length, λ, and linear scale reading, r,

$$\lambda = \lambda_0 + \frac{c}{r - r_0}$$

where λ_0, c, and r_0 are constants which can be evaluated from the solution of three simultaneous equations containing three known wave lengths or standard lines and their corresponding scale readings.

Relative wave lengths are most readily obtained from diffraction spectra since linear interpolation between standards is justified by the fact that the dispersion is constant or nearly so, in a given grating order or setting. If r_i and r_f are millimetre scale readings for initial and final standards, λ_i and λ_f, on a spectrogram, the dispersion $d = (\lambda_f - \lambda_i)/(r_f - r_i)$ A per mm. The product of this dispersion constant and r_i subtracted from λ_i gives a constant λ_o which, on any calculating machine, can be automatically added to any $d \cdot r_x$ and the final result is $\lambda_x = \lambda_o + d \cdot r_x$. Both in prismatic and in grating spectrograms intermediate standards should be measured, in the former case to test the fit of the dispersion formula and in the latter to test the normality of the spectrum. If the calculated values deviate from the true standards a plot of such deviations will reveal the proper corrections to apply to the wave lengths calculated for the other measured lines. Under the most favourable conditions such interpolated wave lengths are probably correct within a few thousandths angstrom. All wave-length measurements relative to iron standards are obviously facilitated by reference to a good photographic map of the iron-arc spectrum permitting instant identification of the standards. A small portion of such a map is reproduced in Plate I, fig. 4.

Intensity Measurements.—The exact determination of relative intensities of the various parts or lines of a spectrum is in most cases as important as exact wave-length measurement but this phase of spectroscopy had a tardy development, partly because of experimental difficulties and partly because the important applications were mostly unrecognized before 1925. Only heat-absorbing detectors show a linear and nonselective response to spectral intensity and since they are naturally inefficient heat engines their use has been restricted mainly to infra-red spectra. All other detectors, being nonlinear and selective, require calibration but this is a complicated and difficult operation, especially in the extreme ultra-violet where heat detectors are relatively insensitive. Consequently, even the most modern tables of identified spectral lines contain only rough estimates of line intensities on arbitrary scales which differ from spectrum to spectrum and frequently from region to region in the same spectrum. These arbitrary intensity scales usually represent visual estimates of photographic blackening and they range from 1 to 10 or 1 to 1,000 or 1 to 10,000. Such estimates have no definite meaning and no truly quantitative value.

After 1925 the need for accurate measurements of spectral line intensities was stressed by the new quantum theory of atomic spectra and by a rebirth of quantitative spectrochemical analysis. Mathematicians predicted important relationships between line intensities and multiplet structures or quantum numbers, and experimental spectroscopists promptly devised methods of testing the theory. Simultaneously in applied spectroscopy quick methods of accurate chemical analysis based on the measurement of spectral-line intensity ratios were empirically developed. Since almost all spectroscopic observing was done photographically the quantitative determination of intensities from spectrograms required

the construction and use of suitable microphotometers or microdensitometers to measure the photographic densities of spectral line images. By making a series of exposures of known ratio, either in intensity or in time, photographic density as a function of radiation intensity of a given wave length can be determined for the particular type of photographic emulsion employed, and from density measurements at different wave lengths in a comparison spectrum of known energy distribution true intensity ratios can be deduced. Standard sources of known energy distribution are found in the black body or in metal-filament lamps operated at specified colour temperatures. Photographic sensitivity and contrast may be regarded as constant in narrow spectral regions such as those encompassed by compact multiplets, hyperfine multiplets, Zeeman patterns or close line pairs, and in these cases relative intensities of component lines are determinable without spectral energy calibration.

Absorption Spectra.—In 1859 G. Kirchhoff gave mathematical and experimental proof of the following law: "The ratio between the powers of emission and the powers of absorption for rays of the same wave length is constant for all bodies at the same temperature." From this it follows that a transparent body cannot emit light, and one that radiates a continuous spectrum must be opaque. Further, a gas that radiates a line spectrum will absorb the lines which it radiates at the same temperature. For example, lines observed in laboratory arcs as emission spectra of 66 elements or compounds have been identified as absorption lines in the spectrum of the sun. As a consequence of Kirchhoff's law practically everything that has been said about emission spectra applies also to absorption spectra. There are, however, some experimental conditions that permit molecular identification and quantitative chemical analysis by means of absorption spectra that cannot be duplicated in emission spectra. Thus in the analysis of inorganic compounds (salts, minerals, etc.) by emission spectra the light sources (flames, arcs, sparks) commonly used for exciting the spectra completely dissociate the compounds revealing intense spectra of the metallic atoms and of unstable compounds formed in the light source, whereas the spectra of the halogens or radicals are either unobserved or masked. Thus, essentially the same atomic emission spectrum is observed with various metallic compounds (halides, oxides, carbonates, sulphates, etc.). Similarly organic molecules (tissues, dyes, etc.) are atomized by the heat and excitation of emission sources, revealing only spectra of metallic impurities that may be present, while the emission spectra of carbon, hydrogen and oxygen are indistinguishable from those emitted by atmospheric constituents, dissociated carbon dioxide and water vapour. In many cases, however, inorganic compounds in solution and organic compounds in general exhibit characteristic absorption spectra which distinguish them from each other. Since absorption spectra of compounds are observed at relatively low temperatures the molecules are stable and unimpaired, but since the substances are then usually in a liquid or solid state, rather than in a gaseous state, the absorption will depend on the energy content of the molecules and on the perturbing effects of surrounding molecules. These absorption spectra therefore usually consist of one or more broad bands in which the energy distribution is characteristic of the compound. If an unknown mixture contains two or more absorbing compounds their identity and relative amounts can be determined if the absorption bands are sufficiently different in wave length and accurately defined to permit an estimation of their relative intensities. Extraordinarily narrow absorption bands, almost linelike, are exhibited by solids, solutions and glasses containing atoms that possess one or more *f*-type electrons ("rare earths" with atomic numbers 58 to 71 and 91 to 96). This is illustrated in Plate I, fig. 2a, by glass coloured with salts of neodymium, atomic number 60. A typical example of true line absorption is shown in Plate I, fig. 2b, which represents the appearance of a continuous spectrum after passing through a column of atomic potassium vapour. Narrow lines belonging to molecular spectra of oxygen (O_2) and water vapour (H_2O) are readily observed in absorption but not in emission. Red and infra-red bands of these molecules invariably appear on spectrograms made in the laboratory with long focus spectrographs provided that the light source emits a continuous background, and they are outstanding features of the solar spectrum especially when the sun is near the horizon and shines through the thickest layer of terrestrial gas and vapour. Similarly, because of the great depth of absorbing vapours in the solar atmosphere the atomic spectrum of iron is much more fully revealed in the Fraunhofer spectrum than has been observed in laboratory emission spectra.

APPLICATIONS

The principal applications of spectroscopy are to (1) chemical and physical analysis, (2) research on atomic structure and the mechanism of radiation and (3) investigations of some properties of atomic nuclei.

Spectrochemical Analysis.—That a system of qualitative chemical analysis could be based on spectroscopic observations was first convincingly demonstrated by Kirchhoff and Bunsen who (1860) applied spectroscopy to the discovery of two alkali metals (Rb and Cs) and to the identification of chemical elements in the sun. Save for chemical analysis of meteorites that fall to earth there is no way to learn anything concerning the chemical and physical conditions of celestial bodies except from an analysis and interpretation of the light those remote bodies emit. Consequently the challenge and urge to apply spectroscopy to astronomy was irresistible, and an amazing accumulation of information on the radial motions, chemical compositions, and physical conditions of planets, comets, stars and nebulae now constitutes the science of astrophysics (*q.v.*).

In terrestrial laboratories the application of spectroscopy to chemical analysis has always had to compete with conventional chemical methods and progress was slow until precise photographic photometry was introduced about 1925, whereupon quantitative spectrochemical analysis quickly became common practice throughout the world. It is readily observed that when one element is progressively diluted in another the spectra of the diluted element become progressively weaker and simpler until they finally vanish when the concentration is less than the limit of spectroscopic detection. Under specified conditions of observing, these partial or residual spectra may be correlated with the known composition of synthetic mixtures or standards and thus provide criteria for the quantitative estimation of comparable unknowns. This is essentially the principle of spectrochemical analysis as first applied to solutions by W. N. Hartley about 1880, to fused salts by A. de Gramont about 1900 and to metal alloys after about 1920. In 1925 W. Gerlach showed that unavoidable fluctuations in light sources and spectrum photography could be largely eliminated by referring the intensity of an impurity line to a neighbouring line of the major component as standard. In the spectrum of any alloy or mixture of known composition, one or more pairs of lines can usually be found in which a strong (analysis) line of a minor constituent has the same intensity as a weak (internal standard) line of the major constituent. When such homologous pairs are found in spectrograms of samples containing different but known amounts of the analysis element, the concentration of this element in other samples is found by intercomparing the line pairs and selecting the one in which the two lines are of equal intensity. A further improvement of this internal-standard method resulted from the use of only one pair of lines the relative intensities of which in known samples are determined by photometric measurements, and plotted as a function of concentration of the analysis element. If the two variables are plotted on logarithmic scales, the resulting analytical curve is usually a straight line with slope about 1. After preparing such a calibration graph for each analysis element any sample of comparable composition is easily analyzed by measuring line-intensity ratios and reading concentrations from the graphs. An excellent description of the preparation of samples and standards, spark excitation of spectra and analytical procedure can be found in *Spark Spectrographic Analysis of Commercial Tin*, published (1942) by B. F. Scribner as Research Paper 1,451 of the national bureau of standards, U.S. department of commerce.

Similar procedures have been perfected for the analysis of iron alloys, aluminum alloys, magnesium alloys, brasses, bronzes, type metal, and practically all other types of metallic compounds. Careful chemical calibration and photographic photometry have reduced routine analytical errors to 1% or 2% of the amount present, and the time required for a complete analysis of a sample has been reduced to ten minutes or less. Since most of the time required for such an analysis is consumed in procuring and measuring a photographic spectrogram, elaborate spectrographs have been constructed with 12 photoelectron multiplier tubes to receive light from an internal standard line and from 11 different analysis lines. Radiant energies received by these phototubes are amplified to actuate indicators calibrated in percentage composition thus permitting simultaneous and almost instantaneous determination of 11 elements. For composition control of metals and alloys in production, spectroscopy has surpassed chemistry in speed, sensitivity and accuracy of analysis. It must be remembered, however, that no satisfactory absolute method of spectrochemical analysis has been devised; all the practical methods by 1947 were relative to synthesized standards or chemically analyzed samples which serve for spectrographic calibration.

Emission spectroscopy has also found extensive application to analytical problems in agricultural, biological and pathological chemistry, in mineralogy and geochemistry and in archaeology and criminology. In these problems the samples are often in liquid or powder form which can be vaporized and excited to emit spectra either in hot flames or in electric arcs or sparks between electrodes of pure carbon. Usually a definite amount of some suitable element is added to supply internal standards and then the calibration and analytical procedures are essentially the same as for solid metal samples.

Empirically it is found that operating with light sources at atmospheric pressure and with spectral detectors sensitive between 2,000 and 9,000 A at least 70 chemical elements can be readily determined by emission spectroscopy. This is related to the fact that these elements have ionization potentials less than ten volts and their strongest lines have wave lengths between the limits mentioned. Gaseous elements have higher ionization potentials; their strongest lines are in the far ultra-violet and their spectra are fully developed only at reduced pressure in Geissler tubes. In principle, every one of the 96 chemical elements can be detected and determined by means of emission spectra.

Structure of Spectra and of Atoms.—Because atomic spectra appear, generally, to consist of a random distribution of lines of different intensities, a century elapsed between the discovery of spectral lines by Fraunhofer and a fruitful clue to their physical interpretation by N. Bohr (1913). Then in little more than a decade an amazing development in analysis and theory of spectral structure made all spectra interpretable in terms of atomic properties and, conversely, permitted the deduction of atomic structure from observations on spectra.

Hydrogen is the lightest atom and possesses the simplest spectrum; in the visible it consists of a regular succession of lines with ever closer spacing and decrement of intensity in the direction of shorter wave lengths (*see* Plate I, fig. 6). In 1885 J. J. Balmer discovered an extremely simple mathematical relationship between these 1H lines. Expressed in wave numbers the Balmer formula gives $\nu = R/2^2 - R/n_1^2$, where the constant $R = 109{,}677.58$ cm.$^{-1}$, and n_1 represents the integers 3, 4, 5 . . . ∞. This Balmer series begins with 6,562 A and ends at 3,646 A. Similar 1H series were later found in the extreme ultraviolet, $\nu = R/1^2 - R/n_1^2$, by T. Lyman, and in the infra-red, $\nu = R/3^2 - R/n_1^2$, by F. Paschen. All possible H series can be accurately represented by one formula, $\nu = R/n_2^2 - R/n_1^2$, where n_2 and $n_1 > n_2$ are integers and n_2 is constant for a given series. Thus the wave number of any H line can be expressed as the difference of two quantities, the larger being a constant term called the limit of the series, whereas the other varies from line to line and is known as the running term.

In 1889 J. R. Rydberg reported finding series in the spectra of alkali (Plate I, figs. 8, 9) and alkaline earth elements (Plate I, fig. 10), and notwithstanding the imperfect data then at his disposal he discovered most of the important properties of spectral series. From certain characteristics of the lines he distinguished three species of series, principal, sharp and diffuse, superposed in the same spectrum, and he recognized that the members of each series were double for alkalis (7 Li, 11 Na, 19 K) and either single or triple for alkaline earths (12 Mg, 20 Ca, 30 Zn). Rydberg found that each series could be represented in wave numbers by a formula of the type $\nu = L - \frac{R}{(m+\mu)^2}$, in which L is the limit of the series, R the Rydberg constant (practically identical with R in the Balmer formula) and m has successive integral values to which a constant fractional part, μ, is added. A fourth type of series was discovered in 1907 by A. Bergmann, and because it was more hydrogenlike ($\mu \approx 0$) than the others it came to be called fundamental. Rydberg also indicated that the difference between the limit of the P series and the common limit of the D and S series is equal to the wave number of the first line of the P series, and C. Runge in 1908 pointed out a similar relation between the D and F series, so that all the different series were interconnected. An abbreviated notation was then adopted for the four basic types of series:

Principal,	$\nu = n\ \mathrm{S} - m\ \mathrm{P}$
Sharp,	$\nu = n\ \mathrm{P} - m\ \mathrm{S}$
Diffuse,	$\nu = n\ \mathrm{P} - m\ \mathrm{D}$
Fundamental,	$\nu = n\ \mathrm{D} - m\ \mathrm{F}$

where n stands for a simple integer, and m for appropriate integers plus μ. In order to distinguish the components of doublets and triplets from each other it was necessary to add subscripts to these term symbols. In addition to the regular series others were detected when the limits were changed and still others between singlet and triplet terms (Ritz combination principle). Thus it appeared that any given term might combine with many others, and although two terms were necessary for each observed line the total number of terms required to represent any spectrum completely might be much smaller than the total number of lines.

In 1914 A. Fowler found ionized helium atoms (2 He^+) characterized by a hydrogenlike series except that the Rydberg constant R required multiplication by 4. He also discovered, in the spark spectra of 12 Mg^+, 20 Ca^+, and 38 Sr^+, series of doublets resembling the P, S and D doublets of alkalis, but with 4 R in their formulas. By 1919 the regularities detected in various spectra sufficed to suggest two general laws: (1) the alternation law which states that even and odd multiplicities of terms alternate in successive columns of the periodic system; and (2) the displacement law which states that the spectrum of an ionized element resembles that of the neutral element preceding it in the periodic system (*see* Plate II). Term multiplicities of atoms or ions are thus determined solely by the number of electrons, whereas the atomic charge affects the position of the spectrum. These facts suggested a direct connection between spectral lines and the electrons and protons in atoms.

In 1913 N. Bohr proposed two fundamental postulates, viz., that atomic stationary states exist, and that the radiation frequency is equal to the energy difference of two states divided by Planck's constant ($h = 6.624 \times 10^{-27}$ erg sec.). The empirically established spectral terms must therefore be interpreted as quantized stationary energy states characteristic of atoms and ions. The rapid development of the quantum theory of radiation constitutes one of the most thrilling chapters in the history of science (*see* Quantum Mechanics).

Quantum Numbers and Atomic Energies.—Omitting further review of historical and experimental details the following discussion is restricted to a vector model of the Rutherford-Bohr atom (*see* Atom) which is assumed to consist of a tiny but massive nucleus (composed of protons and neutrons) around which circulate one or more electrons. The maximum number of electrons in any atom is expressed by its atomic number, Z (identical with the number of protons in its nucleus), and this number ranges from 1 for H, the lightest, to 92 for U, the heaviest element in nature. These electrons move in imaginary orbits about the nucleus at average distances represented by a principal quantum number, n, which in turn numbers the successive shells of the

periodic system and serves as a coefficient to the term symbol, S, P, D, F, etc. A sequence of n term values subtracted from a limiting term gives the wave numbers of lines in a spectral series. In most cases an electron jump from one shell or value of n to any other shell or value of n constitutes the largest change in energy that can occur in a given atom. If an electron is moved from its lowest value of n to $n = \infty$ the atom is ionized, and the energy required to accomplish the removal of an electron is called the ionization potential. This energy is usually expressed either in wave numbers (cm.$^{-1}$) or in electron volts (ev); in the case of 1H it is 109,677.58 cm.$^{-1}$ (the Rydberg constant) or 13.595 ev. (*See* Table VII.) Hydrogenlike spectra are also emitted by ionized helium ($2He^{+}$), by doubly ionized lithium ($3Li^{++}$), by trebly ionized beryllium ($4Be^{+++}$), etc., but the spectra are shifted successively toward higher frequencies, because the energy of electron orbits as a function of n is proportional to Z^2/n^2. It follows that for $2He^{+}$ the separation energy of the remaining electron from the lowest level (the ionization potential) will be 4 times that for the 1H atom; for $3Li^{++}$ it will be 9 times as great and for $4Be^{+++}$ 16 times.

The next largest change in atomic energy (after that due to a change in n) is usually that associated with a constant orbital angular momentum of the optical electron. This momentum is pictured as a vector of magnitude 0, 1, 2, 3 . . . in $h/2\pi$ units; it is called the azimuthal quantum number and is symbolized by l. The numerical values of $l = 0, 1, 2, 3 \ldots$ correspond respectively to the empirical term symbols, S, P, D, F . . . originally abbreviated from the different types of series first found in the spectra of alkali atoms. Electrons with $l = 0$ are called s-electrons; those with $l = 1$, p-electrons; those with $l = 2$, d-electrons; and those with $l = 3$, f-electrons. It may be interjected here that these four l values and the first seven n values suffice to describe the normal electron configurations of all possible atoms and ions. When two or more optical electrons are present their individual orbital momenta, $l_1, l_2, \ldots$ are added vectorially to form a resultant L, but quantum theory dictates that such a resultant, as well as the individual l, can only exist as an integral multiple of $h/2\pi$. The individual l_i can therefore be oriented only in certain discrete directions to one another. For the case of two electrons with l_1 and l_2 the possible resultant L values are: $L = (l_1 + l_2)$, $(l_1 + l_2 - 1)$, $(l_1 + l_2 - 2) \ldots |l_1 - l_2|$. The vector addition for this case is shown graphically in fig. 13. If all but one of the l_i are zero, the resulting L will be identical with that one l_i. When there are three electrons for which $l \neq 0$, the vector addition is simply performed by combining the l values of two electrons and then combining each of the resulting L values with the l of the third electron. The term types given by various simple electron configurations are shown in Table I.

TABLE I.—*L Values and Term Symbols Arising From Various Simple Electron Configurations*

Electron Configurations	L	Term Symbols
ss	0	S
sp	1	P
pp	0 1 2	S P D
pd	1 2 3	P D F
dd	0 1 2 3 4	S P D F G
df	1 2 3 4 5	P D F G H
ff	0 1 2 3 4 5 6	S P D F G H I
ppp	0 1 1 1 2 2 3	S P P P D D F

Quantum theory and observation prove that an important selection rule controls atomic energy changes resulting in radiation—viz., the greatest intensity of emission results when $\Delta L = \pm 1$. In addition, when the interaction of the electrons is small, only those quantum transitions take place for which only one of the electrons alters its l value in accordance with the selection rule $\Delta l = \pm 1$. Notwithstanding these simple rules, many spectra exhibit lines corresponding to $\Delta L = 0, \pm 2, \pm 3$, as well as to double-electron jumps.

A third contribution to the total energy of an electronic system comes from the rotation of each electron about its own axis. This axial angular momentum has the same magnitude for each electron, $s = \frac{1}{2}(h/2\pi)$; it is usually called the electron spin. When two or more electrons are present, the individual spin vectors s_i combine with each other to yield a resultant S (just like the l_i combine to form L). Analogous to L, the resultant spin S can take only certain discrete values, the maximum being obtained when all the s_i are parallel, and the minimum being one-half or zero according as the number of electrons is odd or even.

The spinning electrons, and their resultant spin vectors S, account for the splitting of most spectral terms into two or more components, and provide a physical meaning for the subscripts (inner quantum numbers) originally attached to components of polyfold terms. The total angular momentum J of an atom is the vector sum of the resultant orbital angular momentum L and the resultant axial angular momentum S. According to quantum theory, L and S can be oriented to each other only in certain directions, and therefore only certain discrete values of the resultant J are possible. The largest and smallest values of J result from simple addition and subtraction of L and S, while all intermediate values of J that differ by integral amounts are allowed. That is:

$$J = (L + S), (L + S - 1), (L + S - 2), \ldots |L - S|.$$

This rule is illustrated graphically for simple cases in fig. 14. When $L > S$, it is obvious that the number of permitted J values for a given value of L is $2S + 1$. On the other hand, if $L < S$, the number of permitted J values for a given L is $2L + 1$. Thus, for S terms ($L = 0$, $2L + 1 = 1$) there is only one value of J, viz., $J = S$. Similarly, for P terms ($L = 1$, $2L + 1 = 3$) the number of components can never exceed three. However, for a great majority of terms $L > S$ and for all these the possible number of components or term multiplicity r is equal to $2S + 1$. The empirically indicated alternation law follows directly from this result since the maximum multiplicity will be even or odd according as the number of electrons is odd or even.

In general, atomic energy states with different S are not expected to combine with one another because of a selection rule $\Delta S = 0$, but intersystem combinations are found to be fairly abundant and intense, especially as the atomic number increases. Table II displays the resultant S values and possible multiplicities for different numbers of electrons. (*See* also Plate II, figs. 1, 2, 3, 4.)

TABLE II.—*Resultant S Values and Possible Multiplicities*

Number of Electrons	S	Possible Multiplicities, r.
1	1/2	Doublets
2	0, 1,	Singlets, triplets
3	1/2, 3/2	Doublets, quartets
4	0, 1, 2	Singlets, triplets, quintets
5	1/2, 3/2, 5/2	Doublets, quartets, sextets
6	0, 1, 2, 3	Singlets, triplets, quintets, septets
7	1/2, 3/2, 5/2, 7/2	Doublets, quartets, sextets, octets
8	0, 1, 2, 3, 4	Singlets, triplets, quintets, septets, nonets

The total angular momentum of an atom can have integral or half integral $J(h/2\pi)$ values because $s = \frac{1}{2}(h/2\pi)$ for each electron. Furthermore, different sets of J values are associated with the same type of term (same L value) in systems of different multiplicity, because S is different. And finally, different sets of J values apply to different terms in the same multiplicity system because $J = L + S$. Since J represents the total angular momentum it is the most important quantum number for the fine-structure or multiplet analysis of atomic and ionic spectra. It has a perfectly definite physical meaning and quantitative value under all circumstances. Atomic energy changes are rigorously regulated by a simple selection rule, viz., $\Delta J = 0, \pm 1$, except that all 0 to 0 transitions are prohibited. Since the *a priori* probability that an atomic energy state exists is expressed by its statistical weight $2J + 1$, it is obvious that J is important for the relative intensities of spectral lines. The J values that belong to a few different spectral terms in different multiplicity systems are shown in Table III.

For any given spectrum in which energy levels have been established, relative values of J attached to the levels are readily determined from their combining properties and the selection rule, $\Delta J = 0, \pm 1$. In spectra resulting from terms of odd multiplicity

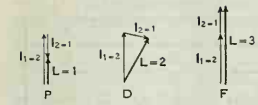


FIG. 13.—VECTOR ADDITION OF ORBITAL ANGULAR MOMENTA $l_2=2$ and $l_2=1$ to give resultants $L=3, 2, 1$

the absolute value of J is fixed by the absence of the transition $J = 0$ to $J = 0$ which is forbidden. In other cases, the absolute value of J can be deduced from the multiplicity and type of term if these are unambiguous, as, for example, in multiplets which are sufficiently law-abiding so that the terms can be identified from the sum rule (the sum of the intensities of all the lines of a multiplet which belong to the same initial or final state is proportional to the statistical weight $2J + 1$ of the initial or final state respectively), or from the interval rule (the interval between two successive components, J and $J + 1$, of a polyfold term is proportional to $J + 1$). Analysis of the more complex spectra is usually hopeless without the aid of Zeeman effect, *i.e.,* the splitting of the lines into polarized components when the source of light is placed in a strong magnetic field. The total angular momentum J is space-quantized in a magnetic field and can assume $2J + 1$ discrete values which are differentiated by magnetic quantum numbers M. The magnitude of the level splitting is the same for all singlet levels (normal Zeeman effect) but differs for all other terms according to the values of L and J (anomalous Zeeman effect). From any completely resolved Zeeman pattern the J values of the two combining levels are deduced at once from the number of components and the L values are simultaneously derived from the separations of the components since the splitting factors have been accurately given empirically and by the quantum theory (*see* ZEEMAN EFFECT).

Summarizing the above discussion of quantum numbers we find that any atomic energy level (spectral term) is completely specified by four quantities: (1) its principal quantum number n which indicates the electron shell or period which the optical electron occupies; (2) its type—S, P, D, F, etc. where the capital letters stand for azimuthal quantum numbers or orbital angular momenta $L = 0, 1, 2, 3$, etc., respectively; (3) its inner quantum number or total angular momentum J (written as a suffix or subscript to the term type symbol); and (4) its multiplicity number, $r = 2S + 1$, written as a superior prefix to the term type symbol. Thus 4^3D_2 represents the middle component ($J = 2$) of a D term ($L = 2$) in a triplet system ($r = 3$) for which the principal quantum number $n = 4$. Theoretically, the energy difference of two terms appears as intense radiation only when $\Delta L = \pm 1$, $\Delta J = 0, \pm 1$, $\Delta r = 0$, and the two terms are of opposite parity, *i.e.,* odd and even (Laporte rule). Typical energetic relationships of spectral terms and combining properties of the levels are shown graphically in fig. 15 which represents the spectrum of neutral mercury atoms.

Pauli Exclusion Principle.—Having found an electronic-quantum interpretation of spectral terms it is next in order to seek

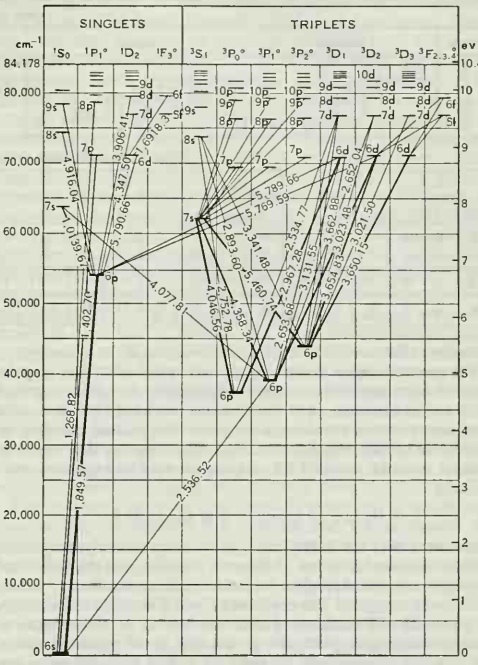


FIG. 15.—ENERGY-LEVEL DIAGRAM OF THE FIRST SPECTRUM OF MERCURY

the connection between spectral terms and atomic structure. In order to understand the building-up of the periodic system (*see* PERIODIC LAW, THE) and account for striking periodicities in atomic properties it is necessary to introduce a new assumption, known as the Pauli exclusion principle, which states that: "in one and the same atom, no two electrons can have the same set of values for the four quantum numbers, n, l, m_l and m_s." Table IV shows the electron shells and subgroups that are possible for atoms as a consequence of this exclusion principle. The n cells are divided by l and the l cells are further divided by m_l and by m_s, but instead of inserting the last quantum numbers the presence of an electron in a (n, l, m_l) cell is represented by arrows pointing up or down according as m_s is $+\frac{1}{2}$ or $-\frac{1}{2}$. Only two electrons can be in each such cell, and then only when they have antiparallel spin directions, since otherwise these two electrons would have the same four quantum numbers n, l, m_l, m_s. The maximum number of electrons that can have the same n and l is given by the number of arrows in the corresponding cells. This number is equal to $2(2l + 1)$, since $2l + 1$ is the number of possible m_l values for a given l. Thus in any one shell no more than 2 s electrons (s^2), 6 p electrons (p^6), 10 d electrons (d^{10}) and 14 f electrons (f^{14}), can exist, and summing these numbers successively gives 2, 8, 18 and 32, which are precisely the totals of elements found in successive groups of atoms exhibiting periodic properties when arranged in atomic number sequence. Naturally the electron configurations of atoms in their lowest possible energy states, and the corresponding spectral terms, will show a periodicity, since, after a certain number of electrons of a given type have been added, the

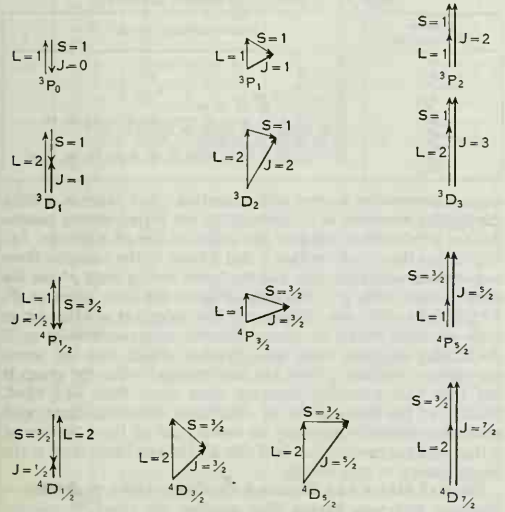


FIG. 14.—VECTOR ADDITION OF ORBITAL MOMENTUM L AND AXIAL MOMENTUM S FORMING RESULTANTS J FOR ^3P, ^3D, ^4P AND ^4D TERMS

TABLE III.—*J Values for Terms of Odd and Even Multiplicity*

Odd Multiplicities

Terms	Singlet	Triplet	Quintet	Septet
S	0	1	2	3
P	1	0 1 2	1 2 3	2 3 4
D	2	1 2 3	0 1 2 3 4	1 2 3 4 5
F	3	2 3 4	1 2 3 4 5 6	0 1 2 3 4 5 6
G	4	3 4 5	2 3 4 5 6 7	1 2 3 4 5 6 7
H	5	4 5 6	3 4 5 6 7	2 3 4 5 6 7 8

Even Multiplicities

Terms	Doublet	Quartet	Sextet	Octet
S	1/2	3/2	5/2	7/2
P	1/2 3/2	1/2 3/2 5/2	3/2 5/2 7/2	5/2 7/2 9/2
D	3/2 5/2	1/2 3/2 5/2 7/2	1/2 3/2 5/2 7/2 9/2	3/2 5/2 7/2 9/2 11/2
F	5/2 7/2	3/2 5/2 7/2 9/2	1/2 3/2 5/2 7/2 9/2 11/2	1/2 3/2 5/2 7/2 9/2 11/2 13/2
G	7/2 9/2	5/2 7/2 9/2 11/2	3/2 5/2 7/2 9/2 11/2 13/2	1/2 3/2 5/2 7/2 9/2 11/2 13/2 15/2
H	9/2 11/2	7/2 9/2 11/2 13/2	5/2 7/2 9/2 11/2 13/2 15/2	3/2 5/2 7/2 9/2 11/2 13/2 15/2 17/2

TABLE IV.—*Possible States of an Electron in an Atom*

	K	L				M									N															
n	1	2				3									4																
l	0	0	1			0	1			2					0	1			2				3							
	s	s	p			s	p			d					s	p			d				f							
m_l	0	0	−1	0	+1	0	−1	0	+1	−2	−1	0	+1	+2	0	−1	0	+1	−2	−1	0	+1	+2	−3	−2	−1	0	+1	+2	+3
m_s	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓

outermost electron will be once more, for example, an s-electron.

The spectral terms describing ground states of atoms may be deduced from assumed electron configurations by adding together the l and s moments of the individual electrons, but first it is necessary to make an assumption about the mutual coupling or interaction of the individual vectors. The coupling that seems to be most common is called *LS* coupling; it may be expressed symbolically:

$$(l_1, l_2, \ldots)\ (s_1, s_2, \ldots) = (L, S) = J$$

This means that the individual l_i vectors are strongly coupled to produce resultant L values of different energies, and the individual s_i vectors are also strongly coupled to produce resultant S values of different energies. The resultants L and S are then less strongly coupled with one another and their resultant is J. Other types of electron coupling, in particular *jj* coupling, occur relatively rarely, but coupling intermediate between *LS* and *jj* is found more frequently, especially among excited states of heavy elements. However, because *LS* coupling applies to all the lighter elements, predominates in many others, and is either accurately or very approximately valid for the ground states of all atoms and ions it forms the basis for the standardized notation for spectral terms.

The spectral terms arising from nonequivalent electrons (belonging to different n, l groups of Table IV) may be obtained, in simple cases, by adding the L values of Table I to the S values of Table II. For example, two nonequivalent p electrons, pp, yield ($L = 0, 1, 2$) S, P, D terms, and since the electron spins can be either antiparallel or parallel ($S = 0, 1$) there will be singlet and triplet terms ($r = 2S + 1$), *i.e.*, pp always produces a family of terms ^{1}S, ^{1}P, ^{1}D, ^{3}S, ^{3}P, ^{3}D. Examples of the term families invariably associated with certain pairs or triads of nonequivalent electrons are given in Table V.

When two electrons are equivalent, *i.e.*, have the same n and the same l, they must at least differ in their values of m_l or m_s. For example, when two equivalent p-electrons (p^2) have l with the same direction (giving a D term) m_l is the same (+1, 0, −1) for both and therefore, both cannot have $m_s = +\frac{1}{2}$, or both $m_s = -\frac{1}{2}$. This means that their spins can only be antiparallel for a D term, thus producing ^{1}D but not ^{3}D. Altogether two equivalent p-electrons (p^2) can produce only the terms ^{1}S, ^{1}D, ^{3}P, whereas two nonequivalent p-electrons (pp) yield ^{1}S, ^{1}P, ^{1}D, ^{3}S, ^{3}P, ^{3}D. Other examples of terms identified with equivalent electrons are displayed in Table VI.

TABLE V.—*Terms of Nonequivalent Electrons*

Electron Configurations	Terms (omitting J values)
ss	^{1}S, ^{3}S
sp	^{1}P, ^{3}P
sd	^{1}D, ^{3}D
pp	^{1}S, ^{1}P, ^{1}D, ^{3}S, ^{3}P, ^{3}D
pd	^{1}P, ^{1}D, ^{1}F, ^{3}P, ^{3}D, ^{3}F
dd	^{1}S, ^{1}P, ^{1}D, ^{1}F, ^{1}G, ^{3}S, ^{3}P, ^{3}D, ^{3}F, ^{3}G
df	^{1}P, ^{1}D, ^{1}F, ^{1}G, ^{1}H, ^{3}P, ^{3}D, ^{3}F, ^{3}G, ^{3}H
sss	^{2}S, ^{2}S, ^{4}S
ssp	^{2}P, ^{2}P, ^{4}P
ssd	^{2}D, ^{2}D, ^{4}D
spp	^{2}S, ^{2}P, ^{2}D, ^{4}S, ^{4}P, ^{4}D
spd	^{2}P, ^{2}D, ^{2}F, ^{4}P, ^{4}D, ^{4}F
etc.	

An important consequence of the Pauli principle is that closed shells (in which the maximum number of equivalent electrons is present) have $L = 0$ and $S = 0$, and therefore always present the ^{1}S$_0$ state. Closed shells can therefore be ignored when deriving the terms given by any electron configuration. Furthermore, ^{1}S$_0$ for a closed shell must result when the shell is divided into two parts, the term types for each part derived, and the resulting angular momentum vectors added together. For example, adding the angular momenta of p^2 electrons to the corresponding quantities for p^4 electrons will give the resultant for p^6 electrons, *i.e.*, ^{1}S$_0$. From this it follows that L and S must be the same for these two electron configurations, and the terms arising from p^4 are the same as those from p^2. Similarly, d^6 gives the same terms as d^4, f^{12} the same as f^2, etc. In general, any subgroup lacking one or more electrons to fill the group behaves spectroscopically as if the lacking electrons alone were present, except that the terms are regular (smallest J level has least energy) when the group is less than half filled but inverted when more than half filled. Finally, of the terms given by equivalent electrons, those with greatest multiplicity generally lie lowest, and of these the lowest is that with the greatest L. In Table VI the last listed term is the lowest energy in each family.

TABLE VI.—*Terms of Equivalent Electrons*

Electron Configurations	Terms (omitting J values)
s^2	^{1}S
p^2	^{1}S, ^{1}D, ^{3}P
p^3	^{2}P, ^{2}D, ^{4}S
d^2	^{1}S, ^{1}D, ^{1}G, ^{3}P, ^{3}F
d^3	^{2}P, ^{2}D(2), ^{2}F, ^{2}G, ^{2}H, ^{4}P, ^{4}F
d^4	^{1}S(2), ^{1}D(2), ^{1}F, ^{1}G(2), ^{1}I, ^{3}P(2), ^{3}D, ^{3}F(2), ^{3}G, ^{3}H, ^{5}D
d^5	^{2}S, ^{2}P, ^{2}D(3), ^{2}G(2), ^{2}H, ^{2}I, ^{4}P, ^{4}D, ^{4}F, ^{4}G, ^{6}S
f^2	^{1}S, ^{1}D, ^{1}G, ^{1}I, ^{3}P, ^{3}F, ^{3}H
f^3	^{2}P, ^{2}D(2), ^{2}F(2), ^{2}G(2), ^{2}H(2), ^{2}I, ^{2}K, ^{2}L, ^{4}S, ^{4}D, ^{4}F, ^{4}G, ^{4}I
etc.	

Normal States and Electron Configurations of Atoms.—Quantum principles having thus specified the types of spectral terms arising from certain electrons, it became theoretically pos-

sible in 1925 to determine from identified spectral terms the electron configurations of all atoms and their ions. By 1946 the ground states of 80 atoms had been uniquely determined from spectral structure, and since these data are of great importance in spectroscopy, atomic physics and chemistry, they are collected in Table VII. In addition Table VII lists all the spectral multiplicities experimentally detected, all the reported spectroscopic ionization potentials, and the wave lengths of the strongest spectral lines characteristic of neutral atoms. In general, the strongest lines result from $s \longleftrightarrow p$ electron transitions.

TABLE VII.—*Atomic Properties From Spectroscopy, Through Cerium (96)*

Period	Atom	Ground Level	Electron Configuration	Spectral Multiplicities	Ionization Potential ev	Strongest Line, Å
1	1 H	$^2S_{0\frac{1}{2}}$	$1s^1$	2	13.595	1,215.66
	2 He	$^1S_0$	$1s^2$	1, 3	24.580	584.33
2	3 Li	$^2S_{0\frac{1}{2}}$	$2s^1$	2	5.390	6,707.85
	4 Be	$^1S_0$	$2s^2$	1, 3	9.320	2,348.61
	5 B	$^2P^0_{\frac{1}{2}}$	$2s^2\ 2p^1$	2	8.30	2,497.78
	6 C	$^3P_0$	$2s^2\ 2p^2$	1, 3	11.264	1,657.01
	7 N	$^4S^0_{1\frac{1}{2}}$	$2s^2\ 2p^3$	2, 4	14.54	1,134.98
	8 O	$^3P_2$	$2s^2\ 2p^4$	1, 3, 5	13.614	1,302.19
	9 F	$^2P^0_{1\frac{1}{2}}$	$2s^2\ 2p^5$	2, 4	17.42	954.8
	10 Ne	$^1S_0$	$2s^2\ 2p^6$	1, 3	21.559	735.89
3	11 Na	$^2S_{0\frac{1}{2}}$	$3s^1$	2	5.138	5,889.95
	12 Mg	$^1S_0$	$3s^2$	1, 3	7.644	2,852.13
	13 Al	$^2P^0_{\frac{1}{2}}$	$3s^2\ 3p^1$	2	5.984	3,961.53
	14 Si	$^3P_0$	$3s^2\ 3p^2$	1, 3	8.149	2,516.12
	15 P	$^4S^0_{1\frac{1}{2}}$	$3s^2\ 3p^3$	2, 4	11.0	1,774.94
	16 S	$^3P_2$	$3s^2\ 3p^4$	3, 5	10.36	1,807.31
	17 Cl	$^2P^0_{1\frac{1}{2}}$	$3s^2\ 3p^5$	2, 4	13.01	1,347.2
	18 A	$^1S_0$	$3s^2\ 3p^6$	1, 3	15.755	1,048.22
4	19 K	$^2S_{0\frac{1}{2}}$	$4s^1$	2	4.339	7,664.91
	20 Ca	$^1S_0$	$4s^2$	1, 3	6.111	4,226.73
	21 Sc	$^2D_{1\frac{1}{2}}$	$3d^1\ 4s^2$	2, 4	6.7	5,671.80
	22 Ti	$^3F_2$	$3d^2\ 4s^2$	1, 3, 5	6.834	4,981.73
	23 V	$^4F_{1\frac{1}{2}}$	$3d^3\ 4s^2$	2, 4, 6	6.738	4,379.24
	24 Cr	$^7S_3$	$3d^5\ 4s^2$	1, 3, 5, 7	6.76	4,254.35
	25 Mn	$^6S^0_{2\frac{1}{2}}$	$3d^5\ 4s^2$	4, 6, 8	7.43	4,030.76
	26 Fe	$^5D_4$	$3d^6\ 4s^2$	1, 3, 5, 7	7.896	3,581.20
	27 Co	$^4F_{4\frac{1}{2}}$	$3d^7\ 4s^2$	2, 4, 6	7.86	3,453.50
	28 Ni	$^3F_4$	$3d^8\ 4s^2$	1, 3, 5	7.633	3,414.76
	29 Cu	$^2S_{0\frac{1}{2}}$	$3d^{10}\ 4s^2$	2, 4	7.723	3,247.54
	30 Zn	$^1S_0$	$4s^2$	1, 3	9.391	2,138.56
	31 Ga	$^2P^0_{\frac{1}{2}}$	$4s^2\ 4p^1$	2, 4	6.00	4,172.06
	32 Ge	$^3P_0$	$4s^2\ 4p^2$	1, 3	8.13	2,651.18
	33 As	$^4S^0_{1\frac{1}{2}}$	$4s^2\ 4p^3$	2, 4		1,890.5
	34 Se	$^3P_2$	$4s^2\ 4p^4$	3, 5	9.750	1,960.91
	35 Br	$^2P^0_{1\frac{1}{2}}$	$4s^2\ 4p^5$	2, 4	11.84	1,488.4
	36 Kr	$^1S_0$	$4s^2\ 4p^6$	1, 3	13.996	1,235.82
5	37 Rb	$^2S_{0\frac{1}{2}}$	$5s^1$	2	4.176	7,800.23
	38 Sr	$^1S_0$	$5s^2$	1, 3	5.692	4,607.33
	39 Y	$^2D_{1\frac{1}{2}}$	$4d^1\ 5s^2$	2, 4	6.6	5,466.47
	40 Zr	$^3F_2$	$4d^2\ 5s^2$	1, 3, 5	6.95	4,687.80
	41 Cb	$^6D_{\frac{1}{2}}$	$4d^4\ 5s^1$	2, 4, 6	6.77	4,058.94
	42 Mo	$^7S_3$	$4d^5\ 5s^1$	3, 5, 7	7.38	3,798.25
	43 Tc					
	44 Ru	$^5F_5$	$4d^7\ 5s^1$	3, 5, 7	7.5	3,498.94
	45 Rh	$^4F_{4\frac{1}{2}}$	$4d^8\ 5s^1$	2, 4	7.7	3,434.89
	46 Pd	$^1S_0$	$4d^{10}$	1, 3, 5	8.33	3,404.58
	47 Ag	$^2S_{0\frac{1}{2}}$	$5s^1$	2, 4	7.574	3,280.68
	48 Cd	$^1S_0$	$5s^2$	1, 3	8.991	2,288.02
	49 In	$^2P^0_{\frac{1}{2}}$	$5s^2\ 5p^1$	2, 4	5.785	4,511.32
	50 Sn	$^3P_0$	$5s^2\ 5p^2$	1, 3	7.332	3,175.04
	51 Sb	$^4S^0_{1\frac{1}{2}}$	$5s^2\ 5p^3$	2, 4	8.64	2,068.38
	52 Te	$^3P_2$	$5s^2\ 5p^4$	1, 3, 5	9.01	2,142.75
	53 I	$^2P^0_{1\frac{1}{2}}$	$5s^2\ 5p^5$	2, 4	10.44	1,830.4
	54 Xe	$^1S_0$	$5s^2\ 5p^6$	1, 3	12.127	1,469.62
6	55 Cs	$^2S_{0\frac{1}{2}}$	$6s^1$	2	3.893	8,521.10
	56 Ba	$^1S_0$	$6s^2$	1, 3	5.210	5,535.55
	57 La	$^2D_{1\frac{1}{2}}$	$5d^1\ 6s^2$	2, 4	5.61	6,249.93
	58 Ce	$^1H_4$	$4f^2\ 6s^2$			5,699.23
	59 Pr	$^4I^0_{4\frac{1}{2}}$	$4f^3\ 6s^2$			4,951.36
	60 Nd	$^5I_4$	$4f^4\ 6s^2$	5		4,924.53
	61					
	62 Sm	$^7F_0$	$4f^6\ 6s^2$	7, 9	5.6	4,296.75
	63 Eu	$^8S^0_{3\frac{1}{2}}$	$4f^7\ 6s^2$	6, 8, 10	5.67	4,594.02
	64 Gd	$^9D^0_2$	$4f^7\ 5d^1\ 6s^2$	7, 9, 11	6.16	4,225.85
	65 Tb					
	66 Dy					
	67 Ho					
	68 Er					
	69 Tm	$^2F^0_{3\frac{1}{2}}$	$4f^{13}\ 6s^2$	2		5,675.83
	70 Yb	$^1S_0$	$4f^{14}\ 6s^2$	1, 3	6.2	3,987.99
	71 Lu	$^2D_{2\frac{1}{2}}$	$5d^1\ 6s^2$	2	5.0	4,518.57
	72 Hf	$^3F_2$	$5d^2\ 6s^2$	1, 3, 5		3,682.24
	73 Ta	$^4F_{1\frac{1}{2}}$	$5d^3\ 6s^2$	2, 4, 6		2,647.47
	74 W	$^5D_0$	$5d^4\ 6s^2$	1, 3, 5, 7	7.98	4,008.75
	75 Re	$^6S^0_{2\frac{1}{2}}$	$5d^5\ 6s^2$	2, 4, 6, 8	7.87	3,460.47
	76 Os	$^5D_4$	$5d^6\ 6s^2$	1, 3, 5, 7	8.7	2,909.06
	77 Ir	$^4F_{1\frac{1}{2}}$	$5d^7\ 6s^2$	2, 4, 6	9.2	2,543.97
	78 Pt	$^3D_3$	$5d^9\ 6s^1$	1, 3, 5	8.96	2,659.44
	79 Au	$^2S_{0\frac{1}{2}}$	$5d^{10}\ 6s^1$	2	9.223	2,427.95
	80 Hg	$^1S_0$	$6s^2$	1, 3	10.434	1,849.68
	81 Tl	$^2P^0_{\frac{1}{2}}$	$6s^2\ 6p^1$	2	6.106	5,350.46
	82 Pb	$^3P_0$	$6s^2\ 6p^2$	1, 3	7.415	4,057.82
	83 Bi	$^4S^0_{1\frac{1}{2}}$	$6s^2\ 6p^3$	2, 4		3,067.72
	84 Po					2,449.99
	85 At					
	86 Rn	$^1S_0$	$6s^2\ 6p^6$	1, 3	10.745	1,786.07
7	87 Fa					
	88 Ra	$^1S_0$	$7s^2$	1, 3	5.277	4,825.91
	89 Ac					
	90 Th	$^3F_2$	$6d^2\ 7s^2$	3, 5		
	91 Pa					
	92 U	$^5L^0_6$	$5f^3\ 6d^1\ 7s^2$	5, 7	≈ 4.	5,915.40
	93 Np					
	94 Pu					
	95 Am					
	96 Cm					

Hyperfine Structure of Spectral Lines and Properties of Atomic Nuclei.—When examined with apparatus of adequate resolving power, many spectral lines have been found to consist of two or more components lying extremely close together. This hyperfine structure is caused by properties of the atomic nucleus with which optical electrons interact more or less according to their orbits. The actual line-splitting is exceedingly small (usually a fraction of a wave number) but nevertheless of very great importance because it provides a means of obtaining quantitative information about certain properties of isotopic nuclei. The spectroscopic influence of a nucleus may be due either to its mass or to an intrinsic angular momentum or nuclear spin, analogous to electron spin. The former is called isotope effect and the latter nuclear-spin hyperfine structure, or *hfs* for short. (*See* Plate II, figs. 5, 6, 7, 8.)

Isotope Effect.—Many natural elements consist of a mixture of two or more isotopes, each of which has an approximately whole number atomic weight. Different isotopes of an element have the same number and arrangement of outer electrons, and consequently have the same multiplet structure for their spectra. They are, however, distinguished from each other by their mass because an electron necessarily revolves about the common centre of gravity of itself and the nucleus, and the Rydberg constant therefore depends on the nuclear mass. Thus the Balmer lines of $1H^1$ atoms are accompanied on the short-wave side by very faint companions ascribed to $1H^2$ and the detection of these isotope-effect components in 1932 constituted the discovery of heavy hydrogen. Similar isotope effects have been observed in many other spectra, but as soon as several electrons are present the simple mass effect observed in the hydrogen spectra is complicated by other factors which were not completely understood in 1947.

Nuclear-spin Hyperfine Structure.—True *hfs* can be simply and quantitatively explained if it is assumed that the nucleus of the emitting atom possesses an angular momentum and an associated magnetic moment. According to quantum theory the nuclear spin can be only an integral or half-integral multiple of $h/2\pi$. It is written $I(h/2\pi)$, where I is the spin vector or quantum number of the nuclear angular momentum. The spectrum of the H_2 molecule shows that for the simplest atomic nucleus, the proton, $I = \frac{1}{2}$. The proton thus has the same mechanical moment as the electron, but various values may be expected for heavier nuclei since they contain more than one proton. As with electron spin, a magnetic moment is associated with the nuclear spin, since the nucleus is electrically charged and the rotation of electrically charged particles gives rise to a magnetic moment. Because a proton is the same size as an electron but 1,840 times heavier, the magnetic moment of an atomic nucleus is always 1,000 to 2,000 times smaller than that of an electron. For this reason the spectral term-splittings (energy differences) deduced from *hfs* are usually less than 1/1,000 those in multiplet structure.

In the same way that, for multiplet structure, L and S were combined to obtain the total angular momentum J of the optical electrons, it is necessary, for *hfs*, to combine J and I in order to obtain the total angular momentum F of the whole atom, including nuclear spin. The hyperfine quantum number F can take the following values: $F = J + I, J + I - 1, J + I - 2, \ldots, |J - I|$. This gives, in all $2J + 1$ or $2I + 1$ different values, according as $J < I$ or $J > I$. The vector-addition diagrams for J and I will be identical with those for L and S where $J = L$ and $I = S$ (*see* fig. 14). In general, the number of *hfs* components of a spectral term is different for different terms of the same spectrum. Terms with $J = 0$ are always single, but all other terms are split into two components if $I = \frac{1}{2}$. If I is greater than $\frac{1}{2}$, terms with $J < I$ have $2J + 1$ components, whereas those with $J > I$ have $2I + 1$ components.

The magnitude of hyperfine splitting depends on the nuclear magnetic moment but is also dependent on the type of electron. Thus, spectral terms arising from electron configurations containing one s-type electron will be split wider than those from a p

electron with the same principal quantum number, because the *s* electron approaches closer to the nucleus and experiences a stronger interaction. The transitions between *hfs* states are regulated by the same selection rule for F that holds for J, *i.e.*, $\Delta F = 0, \pm 1$, but 0 to 0 transitions are forbidden. Furthermore, the same intensity rule and the same interval rule is valid for both *hfs* and ordinary multiplet structure, except F replaces J.

After deriving, from spectroscopic observations and quantum theory, a satisfactory explanation of *hfs* it became possible conversely to deduce, from the results of spectroscopic-experiments, several properties of atomic nuclei. In 15 years following 1927, the mechanical moments or spins of 90 atomic nuclei were determined, the magnetic moments for 70 and the quadrupole moments for 20. The results are compiled in the *Journal of the Optical Society of America*, No. 36, page 438 (1946). Typical data are shown in Table VIII in which the exponent on the chemical symbol represents the mass of the atomic nucleus. In general, nuclear mechanical moments are derived either from the number of observed *hfs* components, or from their relative intensities or from the intervals between *hfs* levels. Magnetic moments are calculated from a theoretical formula containing the observed *hfs* intervals and other constants but the most precise values have been obtained from measured deviations of atomic or molecular beams in inhomogeneous magnetic fields. A positive moment indicates that the *hfs* terms are regular, *i.e.*, the level with smallest F value has the lowest energy. Deviations of *hfs* levels from the interval rule are referred to a quadrupole moment arising from a nonspherical distribution of protons in the nuclei. There are two possibilities: if the positive charge is spread in the direction of the mechanical moment the nucleus is elongated and has a positive quadrupole moment; otherwise the nucleus is flattened and has a negative quadrupole moment.

TABLE VIII.—*Mechanical, Magnetic and Quadrupole Moments of Atomic Nuclei*

Atomic Nucleus	Mechanical Moment $I(h/2\pi)$	Magnetic Moment in Nuclear Magnetons $\mu(he/4\pi M_p c)$	Quadrupole Moment $q(10^{-24}$ cm.$^2)$
$_0n^1$	1/2	−1.935	
$_1H^1$	1/2	2.7896	
$_1D^2$	1	0.855	0.00273
$_3Li^6$	1	0.820	
$_3Li^7$	3/2	3.2532	
$_4Be^9$	3/2	−1.176	
$_5B^{10}$	1	0.597	
$_5B^{11}$	3/2	2.682	
$_7N^{14}$	1	0.402	
$_7N^{15}$	1/2	±0.280	
$_9F^{19}$	1/2	2.625	
$_{11}Na^{22}$	3/2	2.215	
$_{13}Al^{27}$	5/2	3.628	
$_{19}K^{39}$	3/2	0.391	
$_{19}K^{40}$	4	−1.299	
$_{19}K^{41}$	3/2	0.217	
$_{31}Ga^{69}$	3/2	2.11	0.20
$_{31}Ga^{71}$	3/2	2.69	0.13
$_{83}Bi^{209}$	9/2	3.45	−0.4

The only regularity apparent in Table VIII is that the nuclear mechanical moment is a half-integer if the mass of the nucleus is odd, and a full integer if the mass is even. It is also seen that the algebraic sums of the mechanical moments and of the magnetic moments of the neutron ($_0n^1$) and proton ($_1H^1$) are exactly equal to the respective quantities for the deuteron ($_1D^2$). The manner in which the moments of the elementary particles must be combined to yield the observed values for the other atomic nuclei was a most tantalizing question in nuclear physics. *See* also Index references under "Spectroscopy" in the Index volume.

BIBLIOGRAPHY.—H. Kayser, *Handbuch der Spectroscopie* (Leipzig, 1900–34). The standard work in eight large volumes, containing all spectroscopic information to 1912 in the first six volumes, and further data on about half the elements, Ag to Nb (Cb) in the last two volumes; E. C. C. Baly, *Spectroscopy*, 3 vol. (London, 1924–27); F. Hund, *Linienspektren und Periodisches System der Elemente* (Berlin, 1927); A. Sommerfeld, *Atombau und Spektrallinien*, 5th ed. (1931); R. F. Bacher and S. A. Goudsmit, *Atomic Energy States* (1932). Compilation of spectral terms reported in 231 spectra of 52 elements; H. E. White, *Introduction to Atomic Spectra* (1934): E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge, 1935); H. Kayser and R. Ritschl, *Tabelle der Hauptlinien der Linienspektren aller Elemente* (Berlin, 1939). Contains 27,000 wave lengths, 90,850 A to 33 A, in spectra of 88 elements; G. R. Harrison (ed.), *M.I.T. Wavelength Tables* (1939). Contains 109,275 wave lengths, 10,000 A to 2,000 A, for 87 elements; G. Herzberg, *Atomic Spectra and Atomic Structure* (1944); R. A. Sawyer, *Experimental Spectroscopy* (1944); W. F. Meggers, "Spectroscopy, Past, Present, and Future," *Journal of the Optical Society of America*, 36, pp. 431–448 (1946). (W. F. M.)

SPECTROSCOPY, ASTRONOMICAL. Astronomical spectroscopy is that branch of the science of astrophysics which is built around information obtained by means of a prism, a diffraction grating or some other dispersing device.

This article is divided into the following sections:

I. Principles
II. Historical Development
III. Instruments
  1. Slit Spectrograph
  2. Slitless Spectrographs
  3. Spectrophotometric Instruments
  4. The Spectroheliograph
  5. Coronal Spectrographs
IV. Classification of Stellar Spectra
  1. Methods
  2. Interpretation of Spectral Classification
  3. Continuous Spectra
V. Radial Velocities
VI. Astrophysical Results of Astronomical Spectroscopy
  1. The Wolf-Rayet Stars
  2. Stars with Bright Lines
  3. T. Tauri Variables
  4. Spectra of Shells
  5. Novae
  6. Pulsating Stars
  7. Abundances of Elements
  8. Cor Caroli Stars
  9. Magnetic Stars
  10. Spectroscopic Binaries
VII. Line Profiles
  1. Radiation Damping
  2. Thermal Doppler Effect
  3. Turbulence
  4. Stark Effect
  5. Rotation
  6. Central Intensities
  7. Convection
  8. Centre to Limb Differences
VIII. Gaseous Nebulae
IX. Planets and Comets
X. The Sun

I. PRINCIPLES

When a ray of light from the sun is transmitted through a prism it is broken up into its component shades of colour and we observe a continuous spectrum. Closer examination shows that not all shades of colour are represented in the spectrum of sunlight. If the incident light beam is made very narrow by means of a slit which runs parallel to the refracting edge of the prism and if an image of the slit is formed on a screen by means of two lenses; one—the collimator—just in front of the prism, and the other—the camera lens—just behind it, the resulting band of the rainbow colours separates out the radiations so that the missing ones appear as black gaps or absorption lines. The spectra of most stars resemble that of the sun in that they show continuous spectra—some with greater intensity in blue light, others with more of the red light—upon which are superposed numerous absorption lines.

Occasionally we observe stellar spectral lines in the form of brilliant emission lines on top of the continuous spectra. The absorption and emission lines are produced by absorbing and radiating atoms in the atmospheres of the stars. Each chemical element or chemical compound has several characteristic sets of spectral lines, and each set is indicative of definite conditions of excitation, such as pressure and temperature. Hence, it is possible not only to identify the chemical elements in the atmospheres of the stars but to draw conclusions concerning the temperature and pressure in these atmospheres.

Some astronomical objects, such as the Orion nebula, produce only emission lines without a continuous spectrum or with one that is abnormally weak. Such spectra resemble those produced by the commercial neon signs used in advertising. Astronomical spectra are of great variety. They represent conditions which range in temperature from a few degrees above the absolute zero for interstellar particles to 100,000° K. (absolute temperature),

or even more, in the atmospheres of the hottest star, and 1,000,000° in the solar corona.

In density the range is from that of interstellar space, which contains about 10 atoms per cubic centimetre, to that of the atmosphere of a white dwarf, which exceeds the density of air at the surface of the earth.

Of all the results of stellar spectroscopy the most interesting is the uniformity of distribution of the chemical elements throughout the universe. It must have been a revelation to the earlier spectroscopists when they discovered in the sun the same familiar substances—hydrogen, iron, calcium and the rest—which they knew on the earth. Their belief in the principle of the uniformity of the chemical elements must have been strengthened when several mysterious spectral lines, observed by Norman Lockyer in 1869 at the rim of the sun, turned out, in 1895, to be produced by the newly discovered gas, helium. As recently as in 1941 the last great enigma of line identification was solved when B. Edlén in Sweden announced that the previously unidentified lines of the solar corona originate in the atoms of the common elements, iron, calcium, nickel and argon, excited to a degree of ionization not even dreamed of previously. Henry Norris Russell had remarked some years earlier, when Ira S. Bowen in a similar manner disposed of the so-called "nebulium" lines in certain nebulae, that the mysterious substances of the astrophysicists one after another literally disappeared in "thin air." As a matter of fact, Bowen found that the nebulae consist of oxygen, nitrogen, hydrogen and a few other gases; in other words, they are not very different in composition from air. The principle of the uniformity of chemical elements has now been extended to include even the distant galaxies whose light travels hundreds of millions of years at the rate of 186,000 mi. per second before it reaches the eye of the observer. The principle of uniformity of chemical elements means that the atomic building blocks of the universe are the same throughout space. It does not mean that the proportions of these elements are the same in all astronomical objects. The most pressing problem of stellar spectroscopy today is to determine accurately the relative abundances of the elements in different objects, to study the nuclear transformations of the elements in stars and to discuss the origin of the elements.

II. HISTORICAL DEVELOPMENT

Astronomical spectroscopy had its beginning in 1666 when Isaac Newton observed the separation of the colours of the rainbow in a prism. In 1802 William H. Wollaston placed a narrow slit in a window blind and with a prism of better quality than the one used by Newton, but without a collimator or a camera, noticed several dark lines in the spectrum of sunlight. Real progress in astronomical spectroscopy became possible when Joseph von Fraunhofer in 1814 made use of a telescope, in addition to the prism and a distant slit, and thereby constructed the first modern spectroscope. A collimator lens was first used by W. H. Simms in 1840. With his instrument Fraunhofer observed and mapped 754 dark lines in the spectrum of the sun. He also observed the spectra of several bright stars and found that although they all had dark lines upon a coloured continuum, the lines were not the same as those of sunlight. Fraunhofer noticed that the yellow emission line from a flame fed with table salt, which had already been observed by John Herschel, coincides in position with a conspicuous black line in the solar spectrum, which he had designated by the letter D. But he did not follow up this result.

It was only after Gustav R. Kirchhoff had advanced his famous three laws of spectroscopy that the connection between absorption and emission became obvious. Thereafter William Huggins (1824–1910) in England undertook a systematic survey of the spectra of all stars and nebulae within the reach of his instruments at Tulse Hill in London. This period was the richest in the number and importance of astrophysical discoveries. Many different types of stellar spectra were found; the Orion nebula was found to be gaseous, contrary to the opinion of other astronomers, while the Andromeda nebula showed a continuous spectrum similar to that of normal stars like the sun. Huggins concluded independently of Johann K. F. Zöllner that the solar prominences could be observed without an eclipse by means of a spectroscope, and that the planets shine by reflected sunlight. He was the first to try to measure the radial motions of stars from the Doppler displacements of their lines. The work of Huggins was of a pioneering character, and it was he who laid the foundations for all further investigations. His tradition in England was carried on by Norman Lockyer and F. E. Baxandall at the Solar Physics observatory in South Kensington, which was transferred to Cambridge after Lockyer's retirement in 1913, and was directed first by H. F. Newall and later by F. J. M. Stratton.

The problem of spectral classification was undertaken in the middle of the 19th century by Angelo Secchi in Italy and was later continued by H. C. Vogel in Germany who was the first to introduce the photographic method into spectroscopy and who succeeded in determining accurate radial velocities of many stars. Vogel's work, especially his measurements of radial velocities, stimulated several other observatories, near the turn of the century, to undertake systematic studies in this field. W. W. Campbell at the Lick observatory undertook the determination of the motions of all stars brighter than magnitude 5.5 in both hemispheres of the sky. E. B. Frost at the Yerkes observatory concentrated upon the motions of hot stars. J. S. Plaskett at Victoria, British Columbia, discovered many spectroscopic double stars and determined their orbits; he also extended to fainter stars the work of other observatories. W. S. Adams and his associates at Mount Wilson observatory, California, determined the velocities of hundreds of faint stars and increased the precision to about ±0.01 km./sec. in the case of the bright star Arcturus. A. A. Belopolsky at Pulkovo, Russia, and, more recently, G. A. Shajn at Simeis, in the Crimea, also made important contributions in this field. H. Spencer Jones, then at the Cape observatory, published numerous measures of stellar motions in the southern sky. C. Fehrenbach in France has developed a reliable method of determining stellar radial velocities on objective-prism (slitless) plates.

The work on spectral classification received a new impetus when E. C. Pickering at Harvard, assisted by Miss A. J. Cannon, Miss A. C. Maury and Mrs. W. P. Fleming, undertook the classification of several hundred thousand stars over the entire sky. This work, which culminated in the *Henry Draper Catalogue* by Miss Cannon, published 1918–24, provides the basis for all modern astrophysical work. The spectroscopic study of the sun was advanced by J. C. Janssen, of the Meudon observatory in Paris, who followed Huggins' suggestion and succeeded, in 1868, in observing the prominences of the sun in full daylight. Further advances were made by N. C. Dunér in Sweden, who determined the rotation of the sun from the Doppler effect (*q.v.*) and found it to be different in different latitudes of the sun, and by H. A. Rowland at Johns Hopkins university, who prepared a catalogue of solar wave lengths. At the turn of the century great progress in this field was made by G. E. Hale, who successively founded and directed the Kenwood observatory in Chicago, the Yerkes observatory in Williams Bay, Wis., and the Mount Wilson observatory at Pasadena, Calif. The spectroheliograph was invented independently by Hale and by Deslandres of the Paris observatory. Hale discovered the magnetic fields in sunspots and announced the existence of a 22-year period in the polarity of sunspots. One of his most striking discoveries was that of invisible sunspots, made in 1922; he pointed out that sunspots usually appear in pairs of opposite magnetic polarity. He concluded that single spots are single only in appearance, and that in all probability they are associated with invisible spots "in which the cooling due to expansion is insufficient to cause perceptible darkening of the sun's surface." He was able to confirm this hypothesis by means of the Zeeman effect (*q.v.*) which clearly demonstrated the existence of the invisible spots of opposite polarity to that of the visible single spots.

The connection between the temperatures of the stars and their colours was recognized as soon as differences in colour were

observed, but the reddening effect of interstellar dust was for many years a subject of controversy and was settled only in 1930 through the work of R. Trumpler at the Lick observatory. The recognition of spectroscopic differences between stars of high and of low luminosity goes back to the work of E. Hertzsprung, around 1905, but the systematic utilization of the method of spectroscopic parallaxes—the most powerful means of penetrating distance within our galaxy—was developed by W. S. Adams and A. Kohlschütter at the beginning of World War I.

III. INSTRUMENTS

1. Slit Spectrograph.—The stellar spectrograph is an instrument constructed like an ordinary laboratory spectrograph (*see* SPECTROSCOPY: *Spectroscopes and Spectrographs*). But since nearly all astronomical light sources are exceedingly faint, special precautions must be observed that all the available light of the object be utilized in forming the spectrum. The instrument most commonly used is a slit spectrograph attached at the focus of an astronomical telescope. It is essential that the angular aperture of the collimator be the same as that of the telescope. Otherwise, light is spilled over the edges of the collimator (collimator of greater focal ratio than telescope) or the optical parts of the spectrograph are unnecessarily large and wasteful (collimator of smaller focal ratio than telescope). If the instrument is used in good seeing for the observation of stars, which appear nearly as points on the slit of the spectrograph, then the efficiency is proportional to the area of the telescope objective while the purity is proportional to the length of the collimator. But if the instrument is used for the observation of luminous surfaces, such as nebulae, the moon, the planets, etc., the efficiency is independent of the linear or angular apertures of the telescope and depends only upon the aperture of the collimator.

In practice the astronomer is usually concerned with intermediate conditions: the stars have tremor disks caused by irregularities in the air, and while their images have a pronounced, sharp peak of light intensity in their contours, they cannot be regarded as mathematical points. Hence, it is advantageous to use a telescope of large aperture to gather in a large amount of light, and a long collimator to secure the maximum possible purity. In very "poor seeing," when the star images may be as large as ½ minute of arc in diameter (this would project as a disk of 4 mm. in diameter on the slit of the spectrograph of the 82-inch reflector of the McDonald observatory) the advantages of the large aperture are largely lost, while in good seeing, when the diameter of the image is less than one second of arc, the efficiency is almost exactly proportional to the linear aperture.

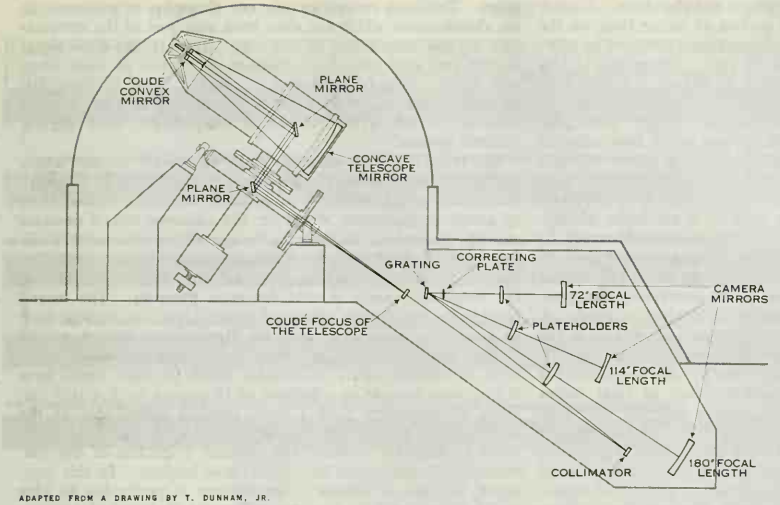
The length of the camera and the dispersion of the prism or grating are chosen to give the desired scale of the spectrum. For very bright objects long cameras and powerful dispersing units may be used. For example, the spectrum of the sun has been completely photographed for photometric purposes in the first and second orders of a diffraction grating having 600 lines per millimetre and installed in a 75-ft. spectrograph of a 150-ft. tower telescope at the Mount Wilson observatory. The linear dispersions of these plates were 0.33 Å/mm. and 0.67 Å/mm. For special purposes even higher dispersions can be used to advantage. For the brightest of the fixed stars, linear dispersions of one Å/mm. have been used at Mount Wilson. Naked-eye stars have been photographed with 2 to 3 Å/mm. The ordinary stellar spectrograph used for radial velocity work (see *Radial Velocities* below) has a dispersion of the order of 10 to 75 Å/mm. Finally, for the faintest objects, such as stars of apparent magnitude 16, or fainter, for galactic nebulae, external galaxies, etc., dispersions of 500 to 1,000 Å/mm. are frequently used.

Designing and constructing spectrograph camera systems are two of the most difficult tasks in optics. For the greatest efficiency it is desirable to make the ratio of collimator to camera lengths as large as possible, so that the slit can be opened without increasing the width of the projected image of the slit on the plate to more than the size of the plate-grain clusters. Thus, short-focus cameras are favoured, which require dispersing units of great power in order to secure the desired scale of the spectrum. For ordinary work with cameras of the order of f/3 to f/5, four-component lenses designed by F. E. Ross or three-component Cooke-type lenses give excellent results. For instruments of very short focus, Schmidt cameras of conventional design can be used down to about f/1; modified Schmidt cameras made from a solid block of glass, one side of which is figured to approximate a sphere and silvered on the outside while the other side is figured in such a way as to remove the spherical aberration, can be used at a speed of f/0.35. Schmidt cameras of improved design have been proposed by James Baker at Harvard. Microscope objectives for use as camera lenses were proposed by W. B. Rayton and by R. J. Bracey. A lens designed by the latter at f/0.35 has been used successfully at the Mount Wilson observatory for work on external galaxies.

Slit spectrographs are usually attached at the focus of a moving telescope and must meet two important requirements: they must be sufficiently rigid to show no perceptible change in flexure during a long exposure and they must be accurately controlled for temperature to prevent changes in excess of about 0.1° in the spectrograph throughout the night. Permanent stationary spectrographs of high dispersion are used for solar work where vertical towers with mirror systems (coelostats) are used for forming the solar image on the slit, and a vertical underground shaft serves to house the spectrograph. Installations of this type are used at the Mount Wilson observatory, Pasadena, Calif., at the McMath observatory of the University of Michigan at Pontiac, Mich., at the Potsdam observatory in Germany, at the Arcetri observatory near Florence, Italy, and elsewhere. For stellar work stationary spectrographs are used at the coudé focus of very large reflecting telescopes (fig. 1). The Mount Wilson and Palomar observatories and the McDonald observatory of the University of Texas, Fort Davis, Tex., have made excellent use of such installations.

The principal advantages of slit spectrographs consist of (1) greater purity than is possible with slitless instruments; (2) a comparison spectrum that can be photographed alongside the astronomical spectrum to serve in radial-velocity determinations; and (3) the elimination of the background of the night sky, which places a definite limit on the efficiency of slitless instruments. The spectral region covered with modern instruments ranges from λ 2,900 Å, where the terrestrial atmosphere becomes opaque, to about λ 13,500 Å, where even the special infrared photographic emulsions cease to be effective. In fact, for stellar work the practical photographic limit until 1945 was about λ 10,000. Beyond this limit special methods may be used, such as employing thin layers of a substance which vaporizes easily and is removed from regions exposed to infrared radiation, being deposited on the cooler, unexposed regions. In 1946, two sensational advances were made in extending the observable spectral region: physicists of the U.S. Naval Research laboratory at Washington, D.C., announced that with a spectrograph in a German V-2 rocket at White Sands, N.M., they had photographed the sun's radiation to about λ 2,100 at a height of 33.5 mi. Since then spectrographs mounted in high-flying rockets have yielded information on the spectrum of the sun in the region of Lyman alpha of hydrogen (λ 1,216 Å) and even in the X-ray region (to about 1 Å). In the infrared region the construction of special PbS photoelectric cells by R. J. Cashman permitted G. P. Kuiper and W. Wilson to observe the spectra of stars and planets to about 3 μ, with sufficient dispersion to record a mass of new absorption features. Radio telescopes have recorded the solar spectrum and the spectra of "radio sources" between about λ 1 cm. and λ 3 m.

The dispersing units of astronomical spectrographs are made in the form of prisms (glass, quartz, rock salt) or of diffraction gratings. The latter are particularly useful in the red and infrared regions, where the dispersion of glass is so low that a prohibitive thickness of the substance is required to obtain the same dispersion that a grating with 600 lines per millimetre gives in the second order. The grating is also efficient in the ultraviolet region, where ordinary glass is opaque and where quartz prisms are difficult to obtain in sufficiently large sizes for high-dispersion



ADAPTED FROM A DRAWING BY T. DUNHAM, JR.

FIG. 1.—THE 70-IN. COUDÉ TELESCOPE SPECTROGRAPH

instruments. Large quartz prisms of the Cornu type are used at the McDonald observatory. R. W. Wood at Johns Hopkins university ruled gratings on a thin aluminum film deposited on glass, using diamonds specially shaped to give the highest possible concentration of light in the desired region of the spectrum. Other fine gratings have been ruled by H. W. Babcock at the Mount Wilson observatory. An important source of error in a high-dispersion spectrograph, namely, the turbulence of air in it, was eliminated in a large vacuum solar spectrograph at the McMath observatory.

Under ordinary observing conditions a large amount of light is lost on the slit jaws of the spectrograph (where it is used for guiding the telescope on the star image). It is not possible to filter this light into the spectrograph in such a way that it will reinforce the spectrum of the central portions of the star image. But an ingenious device has been constructed by Ira S. Bowen—an instrument called the image slicer—in which small plane mirrors are used to place the originally wasted light in the slit, alongside the main spectrum. In this manner a widened spectrum is obtained without letting the star trail on the slit, and this results in a substantial reduction of the exposure time.

Very large luminous areas, such as diffuse gaseous nebulae, the zodiacal light, etc., can be observed without a telescope if the collimator is sufficiently long so that its angular aperture is smaller than the diameter of the object. A powerful spectrograph employing a slit of 2 to 3 cm. in width and 1 m. in length, at a distance of 46 m. from two quartz prisms, and a f/1 Schmidt camera of 94 mm. aperture, was used at the McDonald observatory for recording the spectra of exceedingly faint nebulous regions in the Milky Way. This instrument has no collimator lens because the rays are sufficiently nearly parallel when they enter the prism.

2. Slitless Spectrographs.—The image of a star is approximately a point. Hence, if a prism is placed in front of the lens or mirror of a telescope, the spectrum is recorded as a narrow continuous band with interruptions in place of spectral lines. If during the exposure the star is permitted to trail slightly in a direction parallel to the refracting edge of the prism, a widened spectrum is obtained in which the star serves as its own slit. The great advantage of slitless spectrographs consists in the fact that they record a large field of stars in a single exposure instead of being limited to a single star on the slit of the instrument. The principal disadvantage lies in the fact that in a large telescope, such as would be required to record the spectra of very faint stars, the images of the stars are tremor disks of appreciable diameter, the latter being very sensitive to the conditions of seeing in the atmosphere.

Hence, spectra obtained in this way are not as pure as are spectra obtained with slit spectrographs. Another difficulty consists in the absence of a comparison spectrum for determining the radial motions of the stars. This can be circumvented by placing a glass container of a solution of neodymium chloride in the path of the rays. This substance produces one sharp absorption band in the photographic region, which can serve as a reference point for measuring the displacements of the stellar lines. Nearly all work on the spectral classification of faint stars is being carried on with objective prisms.

Various modifications of the objective prism have been invented for special purposes. The light rays can be made parallel by means of a negative lens just before they reach the focus of the telescope, and a prism or grating can be used to produce slitless spectra of stars and small nebulae, which are then brought to a focus by means of a positive lens. Instruments of this kind are being used in several observatories, one of the most successful being the slitless spectrograph of the 36-inch Crossley reflector of the Lick observatory, with which W. H. Wright obtained some of the finest spectra of small planetary nebulae.

The wide field and the critical definition of star images at relatively short focal lengths, obtained with large Schmidt cameras, render these instruments especially suitable for objective-prism work. Among the best instruments of this type are the 24-inch Schmidt reflector of the Case Institute of Technology in Cleveland, O., and the 18-inch telescope on Mount Palomar in California.

3. Spectrophotometric Instruments.—For the measurement of the light intensities in continuous spectra and in spectral lines, microphotometers of various designs are used. Usually the instrument is employed to measure the photographic density on a plate, but occasionally direct measures of the light at the telescope are preferable. Instruments for measuring photographs usually consist of a light source and a narrow slit; the latter—or its image—is placed in contact with the plate, parallel to the spectral lines, so that only a narrow band of light passes through the plate. This band falls upon a photoelectric cell which activates a galvanometer whose mirror produces a displaced light signal upon a sheet of photographic paper. As the plate is slowly moved across the slit, the drum carrying the paper is turned and the signal produces a continuous record of the galvanometer deflection of the original beam as modified by the continuous spectrum and by the lines on the plate. Tracings of spectra obtained in this manner must be calibrated by means of marks produced on the original plate by a step wedge, a tube sensitometer or some similar device which changes intensities by known amounts. This is required in order to express the displacements of the signal on the tracings in terms of relative intensities.

Since this conversion is often laborious, a number of devices have been perfected for reading directly the intensities on the tracing, or for converting the original tracing automatically into one giving relative intensities. Instruments of this type have been used by R. C. Williams and W. A. Hiltner at the University of Michigan, by M. Minnaert and his associates at Utrecht, by

Theodore Dunham, Jr., at Mount Wilson and elsewhere. A much higher precision in measuring the profiles of stellar lines, or the relative intensities in the continuous spectrum is obtained by scanning the spectrum with a photoelectric cell. Instruments of this type have been used by A. Code and O. C. Wilson at Mount Wilson, and by L. Spitzer, Jr., at Princeton and Mount Wilson.

4. The Spectroheliograph.—Invented almost simultaneously in 1890 by G. E. Hale in Chicago and by H. Deslandres in Paris, this is a large spectrograph into which a narrow slit has been built in front of the focal plane of the camera so that only the light of a narrow range of wave lengths falls upon the plate. This slit is so adjusted that the light of a particular absorption line in the sun's spectrum is transmitted. It must be remembered that the lines originate in the uppermost regions of the sun's atmosphere and are not completely dark but contain light of smaller intensity than that of the continuous spectrum. The continuous spectrum is formed by light which is produced, on the average, at a depth of several hundred kilometres in the atmosphere. The image of the sun is projected upon the first, the ordinary, slit of the spectrograph and the entire telescope is slowly moved by means of an electric motor at right angles to the slit. Simultaneously the plateholder carrying the sensitive emulsion is moved past the second slit. In this manner an image of the sun is obtained on the plate which is produced by the residual light of a particular spectral line. Since not all regions on the surface of the sun are equally bright in such monochromatic light, it is possible to study features like calcium eruptions (or flocculi), hydrogen tornadoes (or vortexes) and other structures which cannot be seen in the brilliant light of the continuous spectrum which alone is strong enough to register in the extremely short exposures required for obtaining a direct photograph.

A modification of the spectroheliograph, the spectrohelioscope, was invented by Hale for visual observation. Some phenomena, such as brilliant eruptions in hydrogen light (bombs), last only a few minutes and can best be detected and followed by means of prolonged visual observation.

R. R. McMath of the University of Michigan has improved the spectroheliograph by combining it with a moving picture machine so that a continuous record of a solar phenomenon may be obtained. Since radial motions in the solar atmosphere sometimes shift a spectral line outside the limits of the second slit, a picture obtained in two dimensions is necessarily distorted and is inadequate for many purposes. To correct this feature, McMath has built an instrument referred to as a radial-velocity-spectroheliograph, which resembles an earlier instrument by Deslandres. By its means it is possible to study the radial motions, as well as the tangential motions, of such features as prominences, eruptions, etc.

5. Coronal Spectrographs.—B. Lyot built on the Pic du Midi in France an instrument with which he succeeded in photographing the spectrum of the solar corona on any clear day. The principal new feature of this instrument consisted of devices for the elimination of scattered light in the optical parts. All glass parts were chosen for freedom from bubbles and striae and the surfaces were made and are kept free from scratches, minor irregularities and dust. In addition, the direct image of the sun was eliminated by an occulting disk, and special precautions were taken to eliminate diffracted light from the image forming lens. Other coronal spectrographs are used at Arosa in Switzerland by M. Waldmeier, at the Climax station of the University of Colorado by W. O. Roberts, at Sacramento Peak by J. Evans, among others.

IV. CLASSIFICATION OF STELLAR SPECTRA

1. Methods.—Nearly all of the 500,000, or more, stellar spectra which have been examined at several observatories can be arranged empirically in a two-dimensional sequence in which the intensities of selected absorption lines or bands change gradually. The system of classification of stellar spectra is based principally upon the *Henry Draper Catalogue* of the Harvard observatory where spectral types of nearly 300,000 stars are given. This work recognizes only one dimension or parameter in the classification, which has since been identified as the temperature of the atmosphere of the star. In this one-dimensional system two stars may have identical intensities of some lines, and slightly different intensities of others. Most of these discrepancies were removed when it was recognized that stars of different luminosity, but of similar temperature, show slightly different spectra.

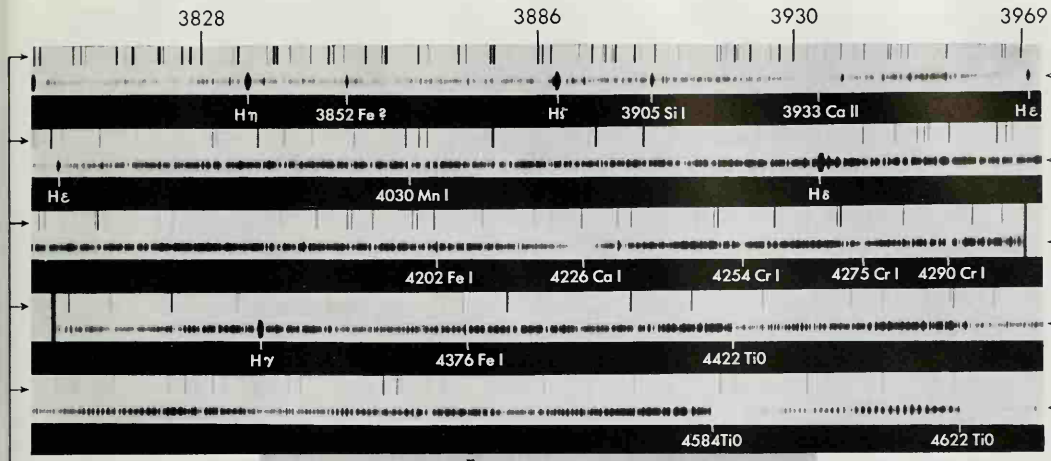
The temperature co-ordinate can be empirically divided into a number of steps which are designated, following Miss Cannon's procedure, by the letters O, B, A, F, G, K, M, R, N and S, or by actual temperatures, which for the sequence named decrease steadily from about 30,000° K. for type O, to about 3,000° K. for types M, R, N and S. Since the exact correspondence between temperature and empirical spectral class is not known, the empirical scheme is used for most purposes. The second co-ordinate can also be expressed by arbitrary designations such as I, II, III, IV and V, used by Morgan and Keenan at the Yerkes observatory, or by actual luminosities expressed in the form of absolute stellar magnitudes, which the stars would have if they were brought to a distance of 10 parsecs, or 3×10^{14} km. Stars of luminosity class I have *absolute* magnitudes of the order of -5, those of class V of the order of $+5$. The latter value corresponds to about the intrinsic brightness of the sun, while the former would be 10,000 times brighter. In this case, again, the use of arbitrary designations is preferable as long as we do not have an accurate calibration in terms of absolute magnitudes.

The hottest stars, class O, have spectra dominated by the lines of ionized helium; those of class B have hydrogen and neutral helium. In class A hydrogen is very strong, helium is weak or absent, and ionized metals, such as Fe II, Ti II, Mg II, are prominent. In types F, G and K the lines of neutral metals, Fe I, Ti I, etc., become gradually stronger, and the lines of Ca II, Ca I and Na I are conspicuous. Near type G we first observe bands of certain diatomic molecules, CH and CN, which also increase in strength as we pass toward the cooler stars. Types M, R, N and S are characterized by very strong molecular bands of TiO in class M, C2 and CN in classes R and N and ZrO in class S. These four types do not form a well-defined temperature sequence, and the exact interpretation of the empirical classes in terms of physical parameters is incomplete.

The luminosity criteria which are used within each spectral class differ greatly. In the hottest stars, classes O, B and A, the hydrogen lines are much narrower in the more luminous stars (supergiants). In the intermediate and cool stars, lines of relatively high ionization potential, such as Sr II in classes G and K, are enhanced in the more luminous stars. The molecular bands of CN, and others, can also be used as luminosity criteria.

An accurate classification of stellar spectra is important for three widely different problems: (1) the study of the physical properties of stellar atmospheres; (2) the statistical study of stars according to their principal characteristics; and (3) the study of the spatial arrangement of the stars in our galaxy by means of their luminosities. It is important to have a clear understanding of the purpose of an investigation before the methods and the criteria are chosen. Thus, certain spectral features may be useful in empirical determinations of luminosities but may be almost useless in studies of physical characteristics because they consist of unresolved lines or bands of different sources.

The principal catalogues of stellar spectra are those of the Harvard observatory—the *Henry Draper Catalogue*, which covers the entire sky and is complete to about apparent magnitude 8.5, and its extensions, which carry the work to much fainter stars in selected regions of the sky. Approximately 100,000 faint stars located in the selected areas of the southern sky, chosen by J. C. Kapteyn, were classified by F. Becker and H. Brück at the Potsdam observatory from photographs taken at La Paz, Bolivia. A similar study for some of the northern selected areas, containing approximately 68,000 stars to apparent magnitude 13, was made by A. Schwassmann at the Bergedorf observatory near Hamburg. Spectral types of several thousand other stars located

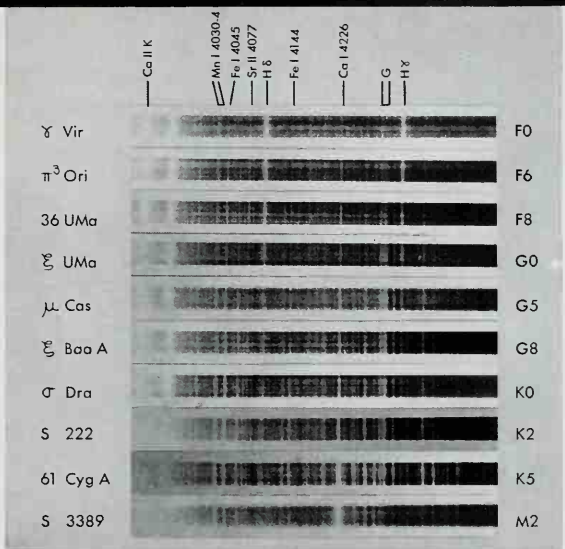


COMPARISON LINES

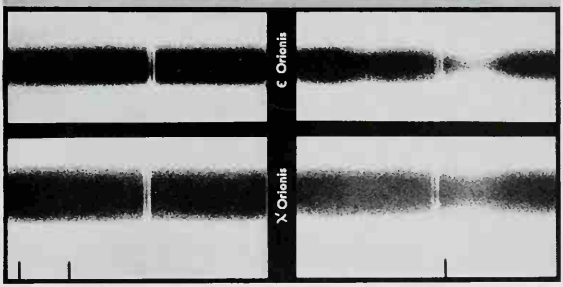
STELLAR SPECTRA

1

2



3



PHOTOGRAPHS, (1, 3) W. S. ADAMS WITH THE COUDE SPECTROGRAPH OF THE MOUNT WILSON OBSERVATORY, (2) FROM "AN ATLAS OF STELLAR SPECTRA" BY MORGAN, KEENAN AND KELLMAN

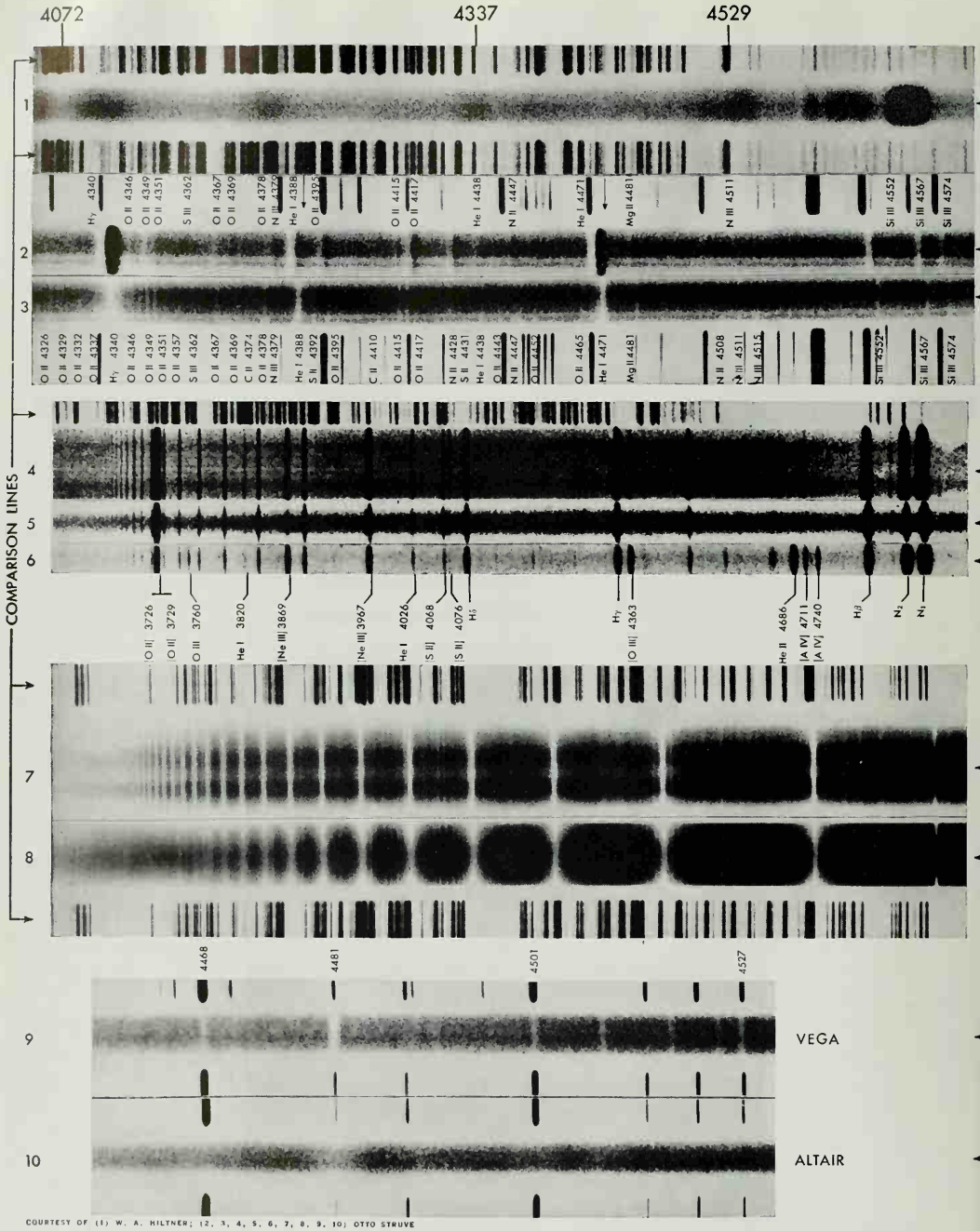
Stellar spectra. All spectra are shown as negatives: the absorption lines appear white on a black background

1. Mira Ceti, a long-period variable star, showing strong absorption lines of atomic and molecular origin. There are also a number of strong emission lines. The emission line Hε is abnormally weak. The comparison spectrum is Fe I. The star lines of Fe I are shifted to the red by the Doppler effect

2. A part of the spectral sequence showing the gradual changes in line intensities for several subdivisions in spectral class between F0 and M2 in stars of moderate luminosity

3. Multiple interstellar absorption lines in two stars: Ca II K in (a) and (c); Ca II H in (b) and (d). The very broad and diffuse line in (b) and (d) is the stellar line Hε

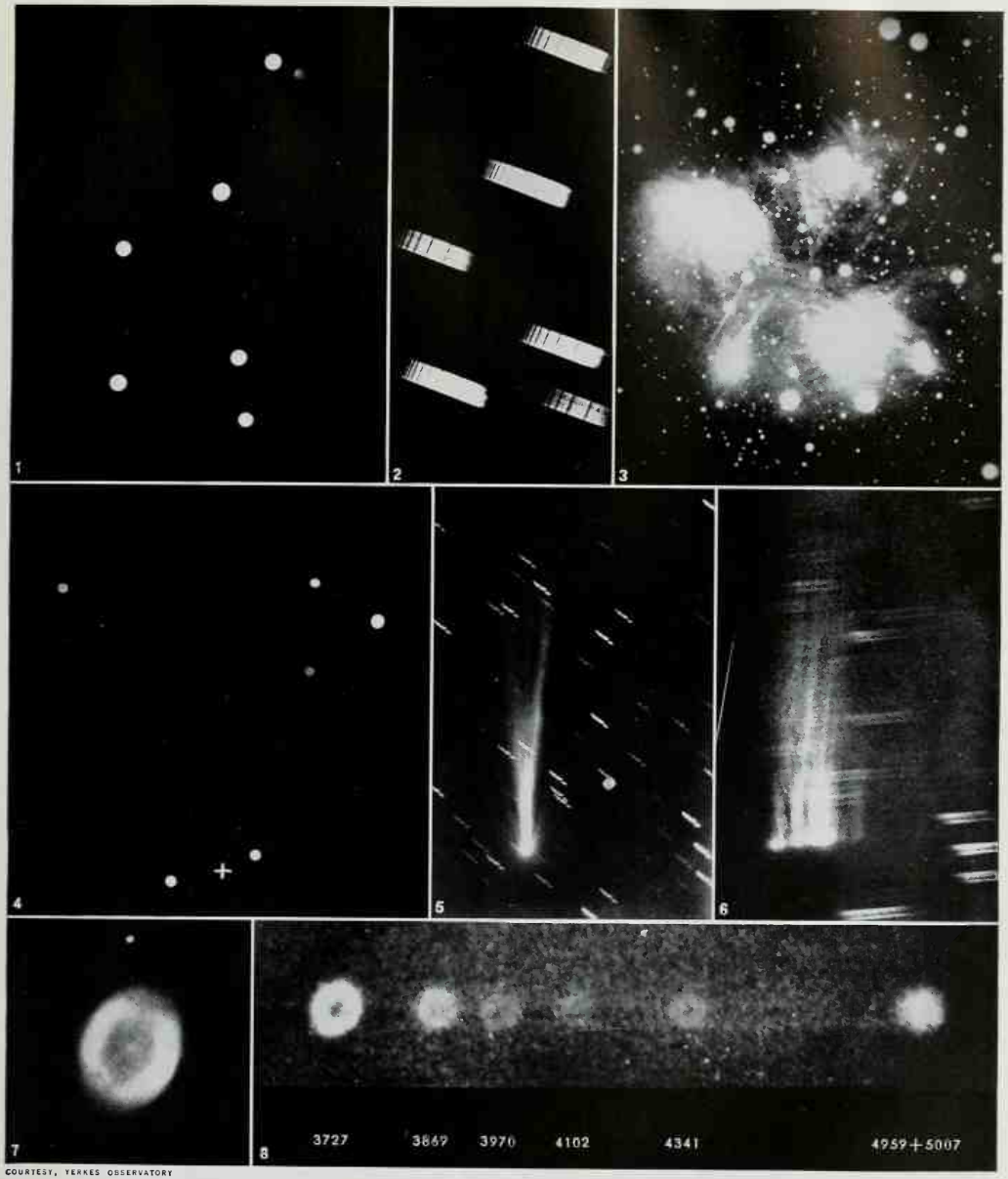
SPECTROSCOPY, ASTRONOMICAL



COURTESY OF (1) W. A. HILTNER; (2, 3, 4, 5, 6, 7, 8, 9, 10) OTTO STRUVE

Stellar spectra. All spectra are shown as negatives. The comparison spectra above or below the star spectra are due to Fe I or to Ti II

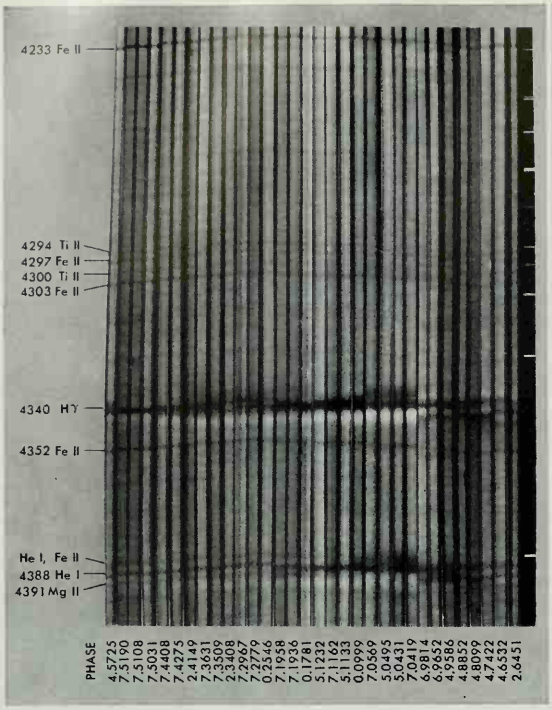
1. A Wolf-Rayet star showing broad emission bands, some of which have violet absorption borders
2. The expanding star P Cygni with emission lines in their normal positions and with violet-displaced absorption lines
3. The normal B-type star τ Scorpii whose absorption lines agree in position with the emission lines of P Cygni
4. 5, 6. Spectra of three gaseous nebulae: the Orion nebula (low excitation), IC 418 (low excitation), and IC 2165 (high excitation). The continuous spectra are produced in the nebulae. The lines N1 and N2 are forbidden lines of [O III]
7. 8. The Balmer series in absorption in two stars; the upper has an atmosphere (or shell) of very low pressure and the lines are sharp; the lower has lines broadened by Stark effect
9. 10. Spectra of Vega and Altair, showing the broadening of spectral lines by axial rotation in Altair



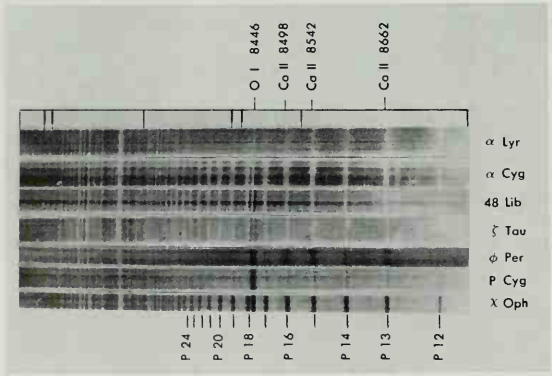
COURTESY, YERKES OBSERVATORY

1. Extrafocal photograph of the seven brightest stars in the cluster of the Pleiades
2. Slitless spectra of six of the seven brightest stars in the Pleiades. These spectra were obtained by placing a prism in front of the camera lens and the stars were permitted to move slightly in an east-west direction, thereby causing the spectra to be widened along the spectral lines; these spectra correspond in position to the six brightest stars of photograph 1. The faintest star is hidden by the spectrum of the bright star near it
3. A direct photograph of the cluster of the Pleiades, obtained with a long exposure. The cluster contains hundreds of stars enmeshed in a diffuse nebulosity
4. A small scale photograph of the constellation Lyra. A cross marks the position of the so-called *Ring Nebula;* the brightest star is Vega or α Lyrae
5. A direct photograph of Comet Morehouse, obtained on Oct. 29, 1908
6. A slitless spectrum of the same comet. Because of the motion of the comet among the stars, the camera was trained upon the moving comet and the star images in the direct photograph appear as beady trails, while in the objective-prism photograph they are drawn out into irregular spectra. The spectrum of the comet is discontinuous; some radiations are strong in the nucleus of the comet and weak in the tail, while others are very strong in the tail
7. A direct photograph of the *Ring Nebula*, showing a faint central star and a ring-like gaseous structure around it
8. A slitless spectrum of the *Ring Nebula*, obtained by placing a prism in front of the camera lens; the light of the nebula is discontinuous and each radiation forms a distinct image of the nebula; the lines at 4959 and 5007 are caused by forbidden [O III]. The strong image at 3727 is caused by forbidden [OII]

SPECTROSCOPY, ASTRONOMICAL

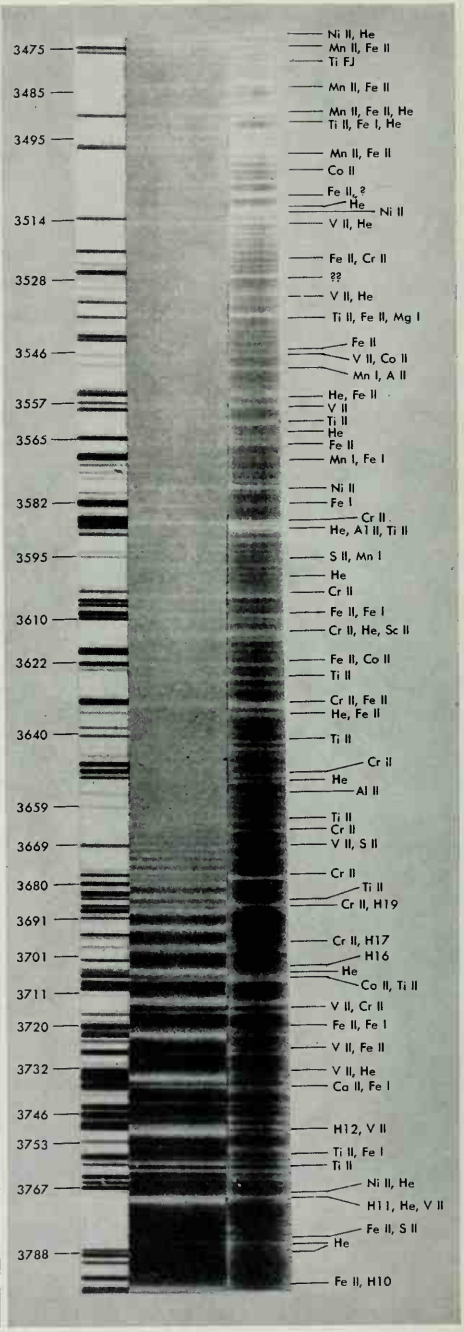


Spectra of the eclipsing binary β Lyrae arranged according to phase



Stellar spectra showing the Paschen series of hydrogen in absorption and in emission

Spectrum of the sun. Positive print: the absorption lines are black



Spectra of (left) the normal supergiant α Cygni, (right) the hydrogen-deficient star υ Sagittarii

PHOTOGRAPHS TOP LEFT CENTRE LEFT AND RIGHT TAKEN AT MCDONALD OBSERVATORY, FORT DAVIS, TEX.

in regions of special interest, such as star clusters, nebulae, Milky Way clouds, etc., have been determined by A. A. Wachmann at Bergedorf, and by numerous observers at the observatories of Uppsala and Stockholm in Sweden and elsewhere.

In most of these studies only the temperature parameter of the classification has been recognized. At the Dearborn observatory O. J. Lee made a search for stars of very low temperature over the northern sky. At the McCormick observatory of the University of Virginia, Charlottesville, Va., A. N. Vyssotsky made a systematic classification of many stars in 1,200 areas of the sky to magnitude 12. Classifications of several hundred faint stars have been obtained with slit spectrographs by M. H. Humason at Mount Wilson and by W. W. Morgan at the Yerkes observatory.

The latter uses what is known as the M-K (Morgan-Keenan) system, regarded as the most accurate classification based upon the estimated intensities of stellar absorption lines. An even more precise classification was developed by B. Strömgren at the McDonald observatory. It is based upon photoelectric measurements of the intensities of certain lines which are especially sensitive to absolute magnitude and of other features (such as the Balmer discontinuity at λ 3,647 Å) which are especially sensitive to temperature.

2. Interpretation of Spectral Classification.—The theoretical interpretation of the two-dimensional classification of stellar spectra rests upon the theory of thermal ionization proposed by M. N. Saha in 1920 and developed in later years by H. N. Russell, E. A. Milne, R. H. Fowler and others. Some of the ideas underlying this theory were already present in the pioneering work of Norman Lockyer, but the exact mathematical formulation was first given by Saha. In a gas exposed to high-temperature radiation, neutral atoms are constantly being ionized by quanta of light, and free electrons are being captured by ions. An equilibrium is set up by which the number of ionization processes is exactly balanced by an equal number of recombinations. The relative numbers of ions (n') and neutral atoms (n) depend upon the temperature of the radiation (T), the ionization potential of the atoms (I), and the partial pressure (p_e) of the free electrons, many of which may result from the ionization of some other element than the one we are considering:

$$\log_{10}\frac{n'}{n} = -I\,\frac{5040}{T} + \frac{5}{2}\log_{10} T - 0.48 - \log p_e$$

The observed intensity of a neutral line is directly dependent upon n, that of an ionized line upon n'. As the temperature increases, n decreases while n' increases, by virtue of the first factor containing T, which is more important than the second factor. The ionization is smaller when I is large. But the ratio n'/n also depends upon p_e and increases as the pressure decreases.

Consider two stars of the same temperature but of different pressures and suppose that we observe in their spectra lines of two elements having different ionization potentials. In the dense star the ionization is less than in the more tenuous star. Moreover, the element of higher potential will be less ionized than the one of lower potential. In practice we do not choose the stars according to temperature but according to spectral class. The latter rests upon a compromise of several slightly conflicting criteria, corresponding to different ionization potentials; for example, two stars classified as F do not have exactly the same temperature if their pressures are not alike; the more tenuous star has the lower temperature. It must also be remembered that many elements can be ionized more than once, so that as T increases or p_e decreases, n' at first increases at the expense of n, reaching a maximum at definite pairs of values of T and p_e, after which it again decreases, while n'', representing the next stage of ionization, begins to increase. Ionization temperatures are determined for a sequence of stars, for which it may be assumed that p_e is constant, by observing at which spectral class lines of different ionization potential reach their greatest intensity. The condition for maximum intensity is obtained by differentiating the ionization formula with respect to T.

The theory of ionization suffices to explain many of the observed absolute magnitude effects in stellar spectra. For a given temperature high luminosity is equivalent to large size (giants) and therefore to small surface gravity and, in turn, small surface gravity indicates low pressure. The tendency of ionized lines to become stronger in giants is explained by the fundamental formula. But other effects require a more complete theory which has been developed by Milne and which involves the interplay of two tendencies: (a) the increased effective mass of the atmospheres of giants as compared with dwarfs, and (b) the effect of altered pressure, consequent on change in surface gravity, on ionization.

3. Continuous Spectra.—The continuous spectra of the stars show conspicuous differences in the distribution of energy as a function of wave length. The latter are closely related to the observed colours of the stars. The O and B stars are bluish, while the M, N, R and S stars are red. Observations of the absolute energy distributions of selected stars have been made by Vogel, G. Müller and others at Potsdam, by Greaves and his associates at Greenwich, by R. C. Williams at the University of Michigan, by D. Chalonge and D. Barbier in France, and by others. From these observations the absolute colour temperatures are derived by adjusting T in the Planck function (*see* HEAT) to give the best possible fit with the observed distribution. These absolute determinations have yielded a number of standards with which it is possible to compare other stars. Differences in the energy distributions of two stars are usually given in the form of "relative gradients":

$$\Delta\phi = \frac{c_2}{T_1} - \frac{c_2}{T_2},$$

where c_2 is the constant of the Planck function. These considerations rest upon the empirical fact that the energy distributions of many stars, in regions not affected by absorption lines, resemble fairly closely the black-body curves. However, accurate observations have shown a number of important departures:

(1) At the limit of the Balmer series the spectra of moderately hot stars (class A) show a discontinuity, usually described as the "Balmer jump": the continuous spectrum is abnormally weak on the violet side of λ 3,647. (2) In the spaces between the strong hydrogen absorption lines on the violet side of H γ of many A-type stars the continuous spectrum is depressed because of the enormous extent of the wings of these lines, this phenomenon being especially conspicuous in white dwarfs, like O^2 Eridani B. (3) In some stars the hydrogen lines are complicated by emission, and the "Balmer jump" appears as an abnormal strengthening of the spectrum on the violet side of λ 3,647. (4) In a few unusual stars the colour temperature is different for different regions of the spectrum: for example, in P Cygni, J. Dufay found colour temperatures of the order of 6,000° to 7,000° in the region of λ 5,000 Å, with a tendency to increase in the shorter wave lengths. The spectral type suggests that the temperature is about 18,000° in the region λ 500 to λ 1,000 Å, but this depends upon an indirect conclusion. (5) Many distant stars appear greatly reddened by selective absorption in interstellar clouds. This type of reddening is not in accordance with Rayleigh's λ^{-4} law but is more nearly proportional to λ^{-1}. (6) In stars which have heavy absorption lines and molecular bands, there must be a pronounced distortion of the continuous spectrum because of the "blanketing effect" of the absorption features. This phenomenon has been predicted theoretically by S. Chandrasekhar and observed indirectly in the central intensities of He absorption lines by A. Unsöld. (7) The continuous spectrum of the sun is weaker and redder at the edge than in the centre, and the combined light is therefore a mixture of different energy distributions. This phenomenon of limb darkening is present also in a number of eclipsing variables. (8) A strange inconsistency between spectral class inferred from the absorption lines and colour temperature inferred from the surface brightness is observed in several close double stars. This effect is attributed to the "reflection effect" which causes lines of relatively high excitation to be produced in the atmosphere of a cool binary component by the radiation of a hot

component.

All of these phenomena are reasonably well understood. The theory of the continuous spectrum, and especially of the effect of limb darkening, is closely related to the problem of the physical mechanism which causes the gases in the outer layers of stars to be more or less opaque to continuous radiation. Continuous absorption at the series limits of H and He, and to a lesser extent of other elements, scattering of light by free electrons, and, in the cooler stars, continuous absorption from negative ions of hydrogen and probably of other elements are known to contribute to the continuous absorption coefficient (as distinct from the absorption coefficient within a spectral line). The energy distributions of the stars in the extreme ultraviolet region are not observed because of the oxygen and ozone bands in the upper air. Considering the very high absorption of H at the Lyman limit and the probable strength of the Lyman absorption lines, this energy curve must depart greatly from that of a black body.

V. RADIAL VELOCITIES

The Doppler principle gives the displacement of a spectral line ($\Delta\lambda$) as a function of the wave length (λ) and the relative radial velocity of the source and the observer (v):

$$\Delta\lambda = \frac{\lambda \cdot v}{c},$$

where c is the velocity of light. The displacement can be measured on plates furnished with a comparison spectrum and, accordingly, v can be determined. The quantity thus obtained must be corrected for the components of the Earth's motion around the sun and the Earth's rotation around its axis. The precision of the determination depends upon the dispersion of the spectrogram, the precision of the laboratory wave lengths used for star and comparison lines, the sharpness of the star lines, the quality of the optical parts used, the uniformity of the illumination of the slit and the collimator by the star image and a number of other factors. W. W. Campbell and J. H. Moore at the Lick observatory obtained a precision of about 0.1 km./sec. for the best stars photographed on many nights. They investigated all systematic errors and established an accurate system of radial velocities free of any appreciable systematic errors.

With the highest attainable dispersion of the Mount Wilson Coudé spectrograph, Adams obtained the velocity of α Bootis with a probable error of ± 0.005 km./sec. The largest velocity of a single star in our galaxy is more than 500 km./sec. The average velocity of the stars is of the order of 10 to 15 km./sec. It varies with spectral class; the B stars have smaller individual motions than the stars of lower temperature. The most complete catalogue of radial velocities is by R. E. Wilson and was published by the Carnegie Institution of Washington in 1953. It contains 15,106 stars. An earlier catalogue by J. H. Moore in 1932 gave the radial velocities of 133 gaseous nebulae, 18 globular clusters and 90 galaxies.

Among the results obtained with the aid of radial velocities are the following: (1) The determination of the velocity (about 20 km./sec.) and the direction in space of the motion of the solar system, with respect to the nearer stars. (2) The recognition that the entire galaxy rotates around a distant centre located in the direction of the constellation Sagittarius, with a velocity of between 200 and 300 km./sec. in the region where the sun is located. (3) A method for determining the distances of groups of stars from the effect of galactic rotation, because the differential motion is directly proportional to the distance. (4) The discovery of the so-called K-effect, which can be described as a slow expansion (about 4 km./sec.) of the system of relatively near stars of spectral class B but which has not been adequately explained; it may be caused by the accidental distribution of the motions of only a small fraction of the stars. (5) The discovery that all rapid stars tend to be moving in the same direction with respect to the sun; this phenomenon can be satisfactorily explained in terms of the theory of galactic rotation if it is remembered that the orbits of individual stars may be highly eccentric, instead of circular. (6) The measurement of red shifts (inferred from the existence of otherwise unaccountably large recessional velocities) amounting to about 20 km./sec. in very dense and massive stars, such as the white-dwarf companion of Sirius, the hot Trumpler stars in clusters and the group of O stars in the Orion nebula; it is possible that these red shifts are caused by the Einstein effect (*see* GRAVITATION). (7) The existence of a peculiar connection between the average speed of a group of stars and certain peculiarities in their spectra; for example, the so-called subdwarf stars (less luminous than real dwarfs) have fast motions. (8) The law of linear increase in the velocity of recession of external galaxies with distance; the largest velocity known in 1960 was about 140,000 km./sec. (9) The determination of the distance of the sun from measurements, throughout the entire year, of the relative velocity, with respect to the Earth, of a star of constant radial velocity; in this problem, worked out by H. Spencer Jones at the Cape of Good Hope observatory and by W. S. Adams at Mount Wilson, the variable correction of the star's velocity to the sun is regarded as the unknown, and the distance appears through its connection with the orbital velocity of the Earth. (10) The discovery of numerous stars showing periodic variations in radial velocity—close double stars, pulsating stars, etc. (11) The study of eruptions in novae and in other expanding shells of gas which surround such stars as P Cygni and others. (12) The discovery of stationary lines in spectroscopic binaries and their interpretation as absorption produced by interstellar gas.

VI. ASTROPHYSICAL RESULTS OF ASTRONOMICAL SPECTROSCOPY

Although the great majority of the stars can be arranged within the two-dimensional framework of the temperature-luminosity classification there are some important exceptions.

1. The Wolf-Rayet Stars.—Less than 100 out of about 500,000 well-observed stars, these show continuous spectra upon which are superposed broad emission bands of atoms of a high degree of ionization, such as He II, C III, C IV, N III, N IV, N V, etc. These bands are often flanked on their violet sides by more or less diffuse absorption lines—a phenomenon which is believed to indicate that the lines and bands originate in shells expanding in all directions with velocities of several hundreds or even several thousands of km./sec. No forbidden lines have been found among the various features, which suggests that the shells cannot be very great in extent. A large fraction of the known Wolf-Rayet stars are close binaries, and in some the spectra of the companions turn out to be normal stars of class O or B. In these cases the Wolf-Rayet stars invariably have much smaller masses. O. C. Wilson, working at Mount Wilson, called attention to the absence of a lag between the observed light minimum of a Wolf-Rayet star which happens to be also an eclipsing binary, and the computed time of conjunction as inferred from the velocities of the emission lines produced by atoms which must have required some time to reach that portion of the shell where conditions are suitable for their radiation. C. S. Beals, at Victoria, discovered that there are two kinds of Wolf-Rayet stars—one group in which carbon lines are strong and another in which nitrogen predominates.

2. Stars with Bright Lines.—Some stars that are otherwise normal in spectrum have bright lines of hydrogen and of other elements, in addition to the usual absorption lines. Sometimes these lines are broad, or even double; in other stars they are sharp and narrow. The broadening appears to be caused by Doppler effect, but the nature of the motions involved is not the same in all cases. In a large group of bright-line stars of classes B and A the motions are rotational in character, a tenuous shell or ring of gas revolving around a normal star which itself has such a rapid axial rotation that it approaches (or perhaps even reaches) the condition of rotational instability. Several stars of this kind happen to be members of close eclipsing double stars. The occulting star in its path eclipses first the approaching lobe of the rotating shell, then the star itself and, finally, the receding lobe of the shell. In other stars the motions are those of expan-

sion, as in Wolf-Rayet stars and in novae, and in still others they are of an irregular or turbulent character. In many stars the bright lines vary in the course of years.

A separate group is formed of various stars of the cooler classes which also show bright lines of H, Ca II, Fe I, Mg I, Si I, etc. Even the sun has weak central emissions in the lines designated as H and K of Ca II and in other stars, like Arcturus, they are fairly strong. The origin of these emission lines is somewhat obscure. O. C. Wilson has shown that the widths of the Ca II emission lines are strongly correlated with the luminosities of the stars. F. Hoyle and Wilson have suggested that the turbulent motions which give rise to the line broadening is produced in the convective layer just below the photosphere. In a large subgroup of these stars, the long-period variables, several hydrogen emission lines are weakened by absorption lines in gases which lie at still higher levels. This is especially striking in the case of Hϵ in Mira Ceti, where this particular hydrogen line is greatly weakened by overlying vapour of Ca II.

An interesting but small group of stars have spectra containing forbidden emission lines, *i.e.,* lines which are not produced in the laboratory because the corresponding transitions are not permitted by the rules of the quantum theory. Some of these stars have weak continuous spectra upon whose background are seen numerous bright lines of H, Fe II, [Fe II], Fe III, [Fe III], [Fe V], [Fe VI], [Fe VII], and probably even [Fe X] and other elements. The bracketed designations of elements stand for forbidden lines, and the Roman numerals designate the state of ionization. Since the ionization potentials of these substances differ widely, by Saha's formula of ionization they could not all be observed in a gas of uniform temperature and pressure. The conclusion is that they come from different regions of a highly complex nebulous structure. P. W. Merrill found that most of these stars also show absorption bands of TiO. It is believed that they consist of three sources: a cool star, a hot companion and a gaseous mass in which the radiation of the hot star causes ionization and excitation. In some cases, like the star Antares, the binary can be observed directly. Its bright component is a cool giant, but it also has a faint, hot companion whose spectrum reveals forbidden [Fe II]. Swings and Struve have found that in these stars the forbidden lines of [Fe II] and those permitted lines of Fe II whose energy levels are low are abnormally strong, so strong, in fact, that an impossibly low excitation temperature results when the intensities are interpreted by means of the Boltzmann exponential relation. The conclusion is that the mechanism of excitation consists in collisions between atoms and electrons, and not in the process of line absorption or ionization and that the emission lines are produced in a small nebulosity, about 3 sec. of arc in diameter, which surrounds the hot star. A. J. Deutsch has found evidence of rotation in this nebula. Most of these spectra undergo rapid variations, sometimes in the course of a few days.

3. T Tauri Variables.—The spectra of these stars also show emission lines superposed upon fairly normal G or K type absorption spectra. According to A. H. Joy and G. Herbig, they are always found in or near large cosmic dust clouds. Herbig suggests that they are very young stars, of small mass, which are still contracting and have not yet reached the main sequence in the H-R diagram. Their absorption lines are broadened by rapid axial rotation, as would be expected in young stars which have not experienced the braking action of magnetic fields or which have not had time to transfer their angular momenta to a system of planets. The spectra and brightnesses of the T Tauri variables undergo irregular variations.

4. Spectra of Shells.—In 1938 the spectrum of the star Pleione, one of the brighter members of the cluster of the Pleiades, underwent a sudden change. For about 30 years the spectrum had been that of a normal B star with very diffuse absorption lines of H and He. It now developed emission lines of H and a set of sharp absorption lines of H, Fe II, Ti II, Ca II and other ionized elements. In the course of several years these sharp lines became very strong and indicated a tendency toward lowered ionization. A number of other stars which possess similar spectra are known, though they have not often been observed in the process of formation. The source of these spectra is designated as shells because it is inferred from the reduced intensities of a few characteristic lines that the gases in which they are formed are not at the surface of the star but are removed above it to several times its radius. The phenomenon producing these lowered intensities is designated as "dilution of radiation." To obtain normal intensities the populations of the different atomic levels must correspond to Boltzmann's relation. If the mechanism of excitation is the absorption of radiation this happens only when the radiation corresponds to conditions of thermodynamic equilibrium. When a gas is illuminated by a distant star the radiation is "diluted" and those excited energy levels which possess transitions to lower terms are depopulated in the same ratio that the apparent area of the star as seen from the gas bears to the complete sphere. The phenomenon of dilution causes the lines of Mg II 4481 and Si II 4128, 4131 to appear abnormally weak. It also produces remarkable changes in the relative intensities of the He I lines. The formation of shells seems to be related to the process of axial rotation of the stars. In some shells (ζ Tauri and ϕ Persei) the shell is stratified and different layers of it revolve at different speeds around the main body.

5. Novae.—Few astronomical phenomena can compare in spectacular interest with the outburst of a bright nova. The spectra of these stars show very rapid changes. Prior to its outburst the spectrum is as inconspicuous as the star itself. But in a few hours, when the light suddenly increases several thousandfold, the spectrum becomes that of a rapidly expanding shell.

At first the shell shows mostly absorption lines which are greatly displaced toward the violet, because in the early stages the exploding gases are still largely projected upon the disk of the star, and we observe only the absorbing action of those regions which are being ejected in our direction. Gradually numerous broad undisplaced emission lines appear in the spectrum and rapidly increase in intensity. Usually several distinct expanding shells are formed, each having its own velocity of expansion. The bright lines of the metals which appear early in the history of a nova gradually give place to those of the lighter elements, like H, O and N. The forbidden lines of [O III] become particularly strong and often cause the star to appear conspicuously green when observed through the telescope. D. B. McLaughlin has shown that all normal novae follow a definite pattern of spectral evolution, the character of which depends upon the duration and the violence of the outburst. The amount of matter lost by a nova during its outburst is a very small fraction of the total mass of the star.

One of the most interesting objects of this type is Nova Herculis which exploded in 1934. According to M. Walker, it is a close eclipsing double star with an orbital period of four and one-half hours. One of the two stellar components varies in brightness—perhaps as a result of pulsation in a very dense star—with a period of about one minute, the shortest known at the present time.

A totally different series of changes is presented by the supernovae, whose visual brightnesses may suddenly increase by a factor of 100,000,000 times, while ordinary novae vary about 100,000 times. R. Minkowski has investigated the spectra of several of these objects, many of which have been discovered by F. Zwicky on photographs of remote galaxies. Statistical considerations show the outburst of a supernova in our galaxy may be expected once every 200 or 300 years. The spectra of the most luminous supernovae show unresolved features which may be emissions or absorptions, or both. No satisfactory identifications of these features have been made. In the less luminous supernovae, F. L. Whipple and Mrs. C. H. Payne-Gaposchkin have found evidence of some of the same bands which are present in ordinary novae but which are broadened by expansion to a much greater extent than in the latter (about 5,000 km./sec., as against about 1,000 km./sec.). There are many unexplained features, such as a variable red shift discovered by Minkowski, but it is difficult to study these stars because of their small apparent brightnesses. The so-called Crab nebula has re-

sulted from a supernova in our galaxy which exploded in the year 1054.

6. Pulsating Stars.—Among the stars whose radial velocities undergo periodic changes are the variables of the Cepheid and the long-period groups. The former have velocity curves which may be described as the mirror images of their respective light curves. The variations in velocity are not caused by orbital motion, as in ordinary binaries, but are believed to be produced by periodic pulsations. The periods vary from a small fraction of a day to about 50 days. In some short-period variables two, three or even more periods combine to produce beat effects in the radial velocity curves and the light curves. Some of these periods represent the higher modes of the fundamental period. The spectral classes also undergo periodic variations, being hottest near maximum light, or near minimum radial velocity. There is conflicting evidence from the luminosity criteria but the hydrogen lines give maximum luminosity near maximum light; this correspondence is not physically obvious because the observed change in brightness (of the order of one stellar magnitude) is fully accounted for by the variation in temperature, and the elementary pulsation theory would predict maximum radius and, hence, minimum atmospheric pressure halfway from minimum toward maximum radial velocity. Evidently the hydrogen atmosphere is expanded at the time of maximum light, and the phenomenon is only indirectly related to the pulsation. The long-period variables are probably also pulsating stars. Their spectroscopic changes have been described principally by P. W. Merrill and A. H. Joy at Mount Wilson. All of these stars show a fairly close relation between period and average spectral class, and a close relation between period and luminosity. Probably related to the long-period variables are several other groups of variable stars, designated as the irregular and the semi-regular variables. The spectra of these stars show interesting changes. Many of them, like the long-period variables, have strong, variable emission lines. The Cepheid variables show emission lines in very few cases. A special group of pulsating variables, the β Canis Majoris stars, are of spectral classes B I to B III. Their periods, ranging from $3\frac{1}{2}$ hours for γ Pegasi to 6 hours for β Canis Majoris, are strongly correlated with the luminosities of the stars. These stars form a sequence which departs from the main sequence of the Hertzsprung-Russell diagram (*see* Astrophysics). Their periods appear to increase at an average rate of one second per century; this may be the result of "post main-sequence" evolution of these stars.

7. Abundances of Elements.—The question has often been asked whether the two-dimensional spectral classification outlined previously is sufficient to classify every normal star within the precision of modern spectroscopic measurements. The importance of at least one additional parameter, namely, abundance of atoms of different elements, is strongly indicated by a number of data. The question is difficult because nearly all the observed differences in the line spectra of the stars are caused by differences in temperature and pressure. But there are a number of striking cases where these parameters appear to be insufficient. Thus, R. H. Curtiss, and later H. N. Russell, showed that the predominant appearance of TiO in classes K and M, and of compounds of C in classes R and N is probably caused by real differences in the abundances of O and C.

Russell's calculations found an interesting application in the work of L. Berman on the spectrum of the irregular variable R Coronae Borealis. In this star the H lines are much weaker than is consistent with its class, while the features due to C are too strong. Berman found that among the elements present in its atmosphere carbon contributes 69%, hydrogen 27%, nitrogen less than 0.3%, and the metals about 4% by volume. In the sun D. H. Menzel finds 82% of hydrogen, 18% of helium, 0.003% of carbon and 0.01% of nitrogen.

Another interesting case of abnormally low hydrogen abundance was observed by J. L. Greenstein in υ Sagittarii. The compromise class of this star is A, which is usually characterized by very strong lines of H. In reality these lines are very weak. It turns out that He is about 100 times as abundant as H, and H is about 100 times as abundant as Fe. But the great majority of the stars and nebulae have abundances similar to that of the sun, with hydrogen about 8,000 times more abundant than all the metals taken together.

In addition to differences of abundance there are several other small factors, as yet unexplained, which cause a certain amount of uncertainty in the spectral classification. These factors are quite unimportant when the purpose of the classification is to serve in the study of the structure of the galaxy and of its statistical properties. But they assume great interest in purely astrophysical studies. Among them are the following: (a) among the hotter stars (class B) there are some, like the star Maya in the Pleiades, which have abnormally weak lines of all elements; (b) among the stars of class A the spectral sequence divides out into several parallel branches; one branch is characterized by strong lines of Si II, others are strong in Mn II or in Sr II, still others have weak Ca II or strong metallic lines. It is not possible to reconcile these differences with the theory of ionization; of special interest are the very old giant stars in several globular star clusters. Their spectra have abnormally weak metallic absorption lines and relatively strong absorption lines of hydrogen. Thus, from the usual criteria they would be classified as F stars; but the relative intensities of Fe I lines of different excitation potentials show that their excitation temperatures are of the order of 4,000° K—as though their real spectral types were K. In this respect they resemble some old "high-velocity" stars in the neighbourhood of the sun. Presumably the original abundances of the metals in the gaseous medium out of which the stars in globular clusters were formed were lower than in the nebulae out of which the younger stars of the Milky Way were condensed; J. L. Greenstein and others have discovered a number of very blue, hot stars with strong lines of He and weak lines of H. These are believed to be old stars which have exhausted much of their hydrogen by having converted it into helium.

Altogether, it is no longer possible to speak of a single table of cosmic abundances in the universe. Presumably most of the stars in the neighbourhood of the sun were formed by the gravitational contraction of interstellar gas clouds of fairly uniform chemical composition. In the first stage of their evolution—that of contraction—few nuclear transformations take place. Only the elements lithium, beryllium and boron (if they were present in the original interstellar gas) would be destroyed by thermonuclear processes. During the long-enduring main-sequence stage of a star, hydrogen is fused into helium in its interior, but convection is so slow that the chemical abundances in the atmospheres of the stars are little affected by this process. After a large fraction of the internal hydrogen of a star has been converted into helium and a very hot contracting helium core inside the star has been formed, the outer envelope of the star begins to expand; the star becomes larger in radius, cooler at the surface and much hotter at the centre. It departs more and more from the main sequence (τ Scorpii becomes a β Canis Majoris variable and finally a yellow or red giant). Its evolutionary track in the H-R diagram carries it from the main sequence to the right side. In this stage we could expect appreciable changes in the abundances of the elements in stellar atmospheres; hydrogen should become less abundant and helium more abundant, but the observable effects would be complicated by the varying amounts of general atmospheric opacity. Only extreme cases of hydrogen depletion (as in υ Sagittarii) would be easily discerned.

When the central temperature of a star rises from about 20,000,000° (as in main-sequence stars) to several hundred million degrees, helium would be fused into heavier elements, but again it is doubtful that this effect would produce readily observable changes in the atmospheric spectrum. It is reasonable to suppose that in Greenstein's hydrogen-deficient stars the chemical composition of the atmosphere reflects only those processes that occurred near their centres about 1,000,000,000 years earlier.

It is probable that all massive stars ultimately become white dwarfs with masses of the order of one or two solar masses. They lose their excess mass in the form of more or less violent nova-type explosions. It is possible that in these explosions the material inside a star becomes quite thoroughly mixed, and we may

expect the nova shells, and especially the supernova shells, to give us a fairly reliable indication of the average chemical composition of a very old star. The evidence on this point is disappointing: the spectrum of the Crab nebula shows strong lines of hydrogen, as do also the spectra of the expanding shells of novae, P Cygni type stars and planetary nebulae. But we must remember that the conditions of excitation in a nebula are quite different from those in a normal stellar atmosphere; it is therefore possible that a moderate deficiency in hydrogen would not be easily discovered. M. Hack has also raised the question whether in a violent nova explosion hydrogen might be regenerated in some sort of fission process.

The spectra of the white dwarfs might also be expected to show very weak hydrogen lines. But some white dwarfs have strong hydrogen lines. It has been suggested that this is due to a thin surface layer of the gas accreted by the star from its surrounding interstellar medium.

We conclude that internal thermonuclear processes of the stars usually do not produce any marked changes in the chemical abundances of the atmospheres. In those relatively rare stars whose atmospheres show abnormal abundances, mixing of layers by convection may be faster than in normal stars (rapid axial rotation has been suggested as a mechanism for accelerating the mixing process). It is also probable, as G. Burbidge has suggested, that rapidly changing magnetic fields in some stars may result in nuclear transformations of chemical elements on their surfaces. This may account for the great strength of lithium and technetium lines in certain types of stars and for the anomalous abundances of europium and other rare earths in others.

8. Cor Caroli Stars.—A remarkable group of stars of intermediate class shows abnormally strong lines of rare earths, especially Eu II and Gd II. In some of them the lines of the rare earths are variable in intensity, while in others there are periodic variations in the intensities of lines of Cr II and of other lines commonly present. The best-known representative of this group is Cor Caroli, whose period is five and one-half days. P. Swings has found strong lines of doubly ionized rare earths, in addition to the singly ionized elements. The variations are not easily explained by the theory of ionization as changes in temperature and pressure, and the velocity curves, which are different for different elements, are neither those of binary motion, nor those of ordinary pulsation. A. J. Deutsch has found that the equatorial rotational velocities of these "spectrum variables" are correlated with the periods of their light curves and velocity curves. Presumably the latter periods are the periods of axial rotation: the chemical elements are distributed unevenly over the surfaces of these rotating stars, alternately bringing into view a region of heavy europium concentration and a region of heavy chromium concentration. Other elements, such as magnesium or hydrogen, may be distributed more uniformly. The regions of heavy concentration may be identified as the magnetic poles of the stars, under the assumption that the magnetic axis is greatly inclined to the axis of rotation.

9. Magnetic Stars.—H. W. Babcock has discovered approximately 100 stars with strong general magnetic fields. The observations were made by means of the Zeeman effect on spectrograms obtained with the 100-in. Mount Wilson and 200-in. Palomar telescopes. The magnetic field intensities may be as large as 10,000 gauss. In many magnetic stars the field strengths vary between about +10,000 and −10,000 gauss. Most of these are at the same time spectrum variables, and the periods of magnetic variation coincide with the periods obtained from the changes in the line intensities, the radial velocities and the brightnesses of the stars.

10. Spectroscopic Binaries.—Unusual spectroscopic phenomena have been observed in a number of close double stars not otherwise identified with one of the preceding groups. Among these are the following: (a) In several eclipsing variables consisting of supergiant components which eclipse one another, the light of one component is observed to shine through the extended atmosphere of the other (ζ Aurigae). In one case (ϵ Aurigae) the light of the brighter and smaller component shines through a considerable thickness of the outermost, nebulous but semitransparent envelope of the companion star whose physical properties remain unknown. The eclipsing envelope is ionized partly by radiation of the visible supergiant primary star and partly by that of the invisible companion which, according to M. Hack, may be a small but hot underluminous B star. The period of ϵ Aurigae covers 27 years. (b) The absorption lines of several eclipsing variables show changes in intensity and distortions of their velocity curves, which indicate the presence in these systems of streams of gas which do not strictly belong to either component but which flow in complicated orbits through the system. The gaseous streams in binaries often form rings around one of the two stars and are observed as double hydrogen emission lines superposed over the normal stellar absorption lines. (c) A combination of these streams with a spiral of outflowing gaseous material satisfactorily accounts, according to Struve and Kuiper, for the complicated changes in emission and absorption lines of β Lyrae. (d) There is a well-known tendency among components of close double stars to have more nearly similar spectra than is predicted from other considerations. This is attributed to the reflection of light of one component by the other. The reflection effect probably causes asymmetry in the profiles of spectral lines (fainter component of α Virginis). (e) There are peculiar periodic variations in the relative intensities of components of double-lined spectra, which have not been explained. (f) Of special interest is the problem of evolution of close double stars. The period of β Lyrae, now about 13 days, has steadily increased since its discovery in 1784 at a rate of about 19 seconds per year. This is undoubtedly due to a loss of mass of the system amounting to approximately 5×10^{22} grams per second (or 5×10^{16} tons per second). Since the mass of the heavier star is roughly 10^{35} grams, the entire mass of it would be consumed at the present rate in $10^{35}/5 \times 10^{22} = 2 \times 10^{12}$ sec., or 70,000 yr. Evidently, this system is rapidly evolving into one of smaller mass. The spectrum of β Lyrae shows emission and absorption lines produced in gaseous streams which are ultimately dispersed in interstellar space. Other binaries show similar though less conspicuous effects. The evolutionary theory of close binaries has been studied by S. S. Huang, J. Sahade, O. Struve, Z. Kopal, D. McNamara and others.

VII. LINE PROFILES

By means of a spectrophotometer it is possible to measure the intensity of a star's radiation at successive wave lengths, from one edge of a spectral line to the other. When the results of these measurements are expressed in units of the intensity of the continuous spectrum in the immediate neighbourhood of the line and are plotted as ordinates against the wave lengths as abscissae, the resulting curve is designated as the profile (or contour) of the line. The unit can, of course, be given in ergs per cm^2. per second if the effective temperature and size of the star are known and if the approximately correct assumption is made that the star radiates as a black body. The profiles of faint lines are strongly affected by the finite resolving power of the spectrograph. For these it is customary to integrate the profiles and to express them in units of one angstrom of complete absorption. These values are designated as the equivalent widths of spectral lines. But for strong lines, like those of Ca II in the sun, the hydrogen lines in the A stars, etc., it is possible to derive physically significant profiles whose interest consists in the information they yield concerning the variation of the line-absorption coefficient with wave length and the character of the propagation of light through the atmospheres of the stars. Important information can be obtained from a curve in which the equivalent widths of the absorption lines are plotted as ordinates against the corresponding numbers of atoms effective in producing each line, as abscissae. Such curves are designated as curves of growth because they indicate the manner in which the intensity of a line grows as the numbers of atoms are increased.

1. Radiation Damping.—Many lines, like those of Ca II in the sun, have broad profiles with very deep, narrow pointed centres, where the residual light is only from 1% to 10% of the continuous spectrum, and with wings which extend very far to

both sides (fig. 2). These profiles are well represented by an absorption coefficient which is inversely proportional to the square of the distance from the centre of the line. Physical theory suggests that this is the common form of the absorption coefficient of a single, undisturbed atom, where the gradual damping of the radiation in consequence of the loss of energy is the sole broadening influence. Since the mathematical form of this expression is known for a single atom of Ca II, it is possible from the observed profile to derive the number of absorbing atoms of

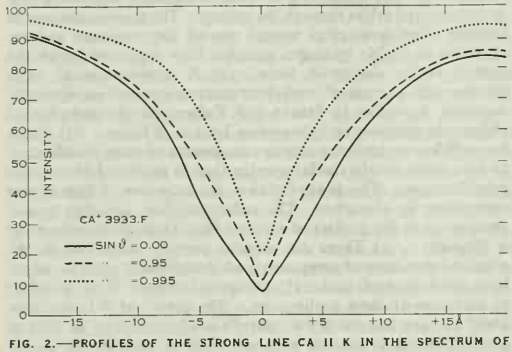


FIG. 2.—PROFILES OF THE STRONG LINE CA II K IN THE SPECTRUM OF THE SUN:

Solid line: near the centre of the sun's disk; dashed line: at a point intermediate between centre and limb; dotted line: near the extreme edge of the sun. These profiles are broadened by the effects of radiation dampinn and collisional damping. The general weakening of the line from centre to limb is explained by the theory of transfer of radiation through a stellar atmosphere

Ca II per cm². in the atmosphere of the sun. This was first done by Unsöld, who used earlier measurements by K. Schwarzschild. In some spectral lines, for example those of Mg I in the sun or some lines of He I in B stars, the profiles are similar in shape to those given by radiation damping but their widths are much greater—just as though the constant of the formula were several times as great as that predicted by atomic theory. This phenomenon is explained by the effect of collisional damping, which was first introduced into physical theory by H. Lorentz and later extended by Weisskopf and Wigner to include processes where atoms do not actually collide but experience close approaches to one another. This form of broadening is important in dwarf stars of low luminosity and is unimportant in giants of high luminosity. In both types of broadening the equivalent width of a line increases as $\sqrt{N}$, where N is the number of absorbing atoms. For example, we know that the line K of Ca II is formed by twice as many atoms as the line H. In consequence the equivalent widths of the two lines in the sun are in ratio $\sqrt{2}$ to 1.

2. Thermal Doppler Effect.—In a stellar atmosphere the atoms are never quite undisturbed. They are moving about in all directions because of the high temperatures prevailing in these regions, and these motions are presumably in accordance with statistical theory. Although every single atom radiates in accordance with radiation damping, some atoms are approaching while others are receding. The fraction which moves with a velocity v is, in accordance with Maxwell's law (*see* KINETIC THEORY OF MATTER), proportional to $e^{-(v/v_0)^2}$, where v_0 measures the average speed of the atoms. But if an atom moves with a velocity v in the line of sight it absorbs not in the normal position of the line but at a distance of $\Delta\lambda = \frac{\lambda \cdot v}{c}$. If v_0 were very large, then radiation damping would be relatively unimportant. The profiles would be bell-shaped, in accordance with Maxwell's formula, and the equivalent widths would increase as N, and not as $\sqrt{N}$. For normal stellar temperatures radiation damping predominates for all but the very faintest lines. We therefore have a gradual transition in the relation between equivalent width and N: at first it varies as N, then as $\sqrt{N}$. The corresponding bending of the curve of growth, as this relation is called, was observed in the sun by Minnaert, by Allen and by others. (Fig. 3).

3. Turbulence.—The curves of growth of many stars, such as α Persei or ϵ Aurigae, show a transition from the N to $\sqrt{N}$ relation at an equivalent width which is much too large to be explained by thermal motions. Moreover, these curves show a region of transition between the two major branches of the curve of growth, where the equivalent width changes little with N and remains almost horizontal. This phenomenon, first noticed by Struve and Elvey, is explained in terms of turbulent motions in the atmospheres of the stars. It is not known how large the volumes of gas are which are subject to these motions but they must be considerably smaller than the entire depth of the atmosphere, because otherwise the phenomenon would manifest itself differently in the curve of growth and would suggest convection, and not turbulence. The largest average turbulent velocity measured is about 67 km./sec., for the star 17 Leporis. The best analogue on the Earth are the thermals which form on sunny days, but it must be remembered that the causes of the two phenomena are not the same.

4. Stark Effect.—In the presence of an electric field the spectral lines of many atoms are disturbed. In hydrogen the lines are split into several components, in helium some lines are little affected while some are displaced toward the red and others are displaced toward the violet and completely new lines, never seen without electric fields, appear in the spectrum. It has never been possible to detect even the slightest evidence of a uniform electric field in the sun or stars. But the atmospheres of all stars contain vast numbers of charged particles, namely, free electrons and ions, each of which carries its electric charge and may in its thermal flight through the atmosphere come close enough to a radiating atom to produce an appreciable Stark effect ($q.v.$). We observe a combination of countless cases of this sort, with all possible distances between neighbouring particles and all orientations. The observed line should therefore have a blurred appearance, and its form can be predicted by a combination of statistical and physical theories.

The profiles which result from an application of the theory by Holtsmark differ greatly from those produced by damping or by Doppler effect. The wings are exceedingly broad and the centres

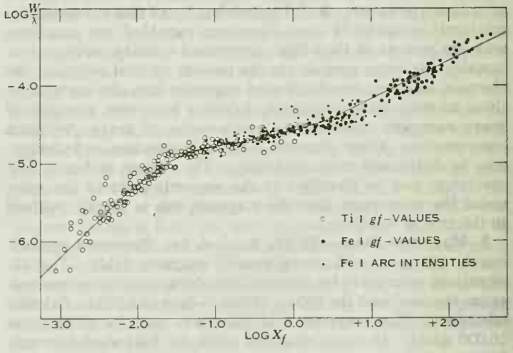


FIG. 3.—THE CURVE OF GROWTH OF THE SUN'S SPECTRUM ACCORDING TO K. O. WRIGHT

The abscissas are logarithms of the relative numbers of etoms (here designated as X_f) effective in producing the line and which can be inferred from multiplet and supermultiplet relations in spectroscopy, or from laboratory measurements. The ordinates are logarithms of the equivalent widths (W) of the lines divided by their wave lengths (λ). The curve is proportional to X_f in its lower left portion and is proportional to $\sqrt{X_f}$ in its upper right portion. The bending over is the result of the thermal motions of the atoms in the sun's atmosphere

of the lines are appreciably raised, so that the residual intensities may be as great at 50% of the continuous spectrum. Such profiles are observed in the hydrogen lines of dwarf stars, but because of the symmetrical form of the Stark effect splitting in H this was not sufficient to prove the existence of Stark effect

in stars. The proof came when the lines of He I in B stars were found to show appreciable broadening in those lines which are sensitive to electric fields and no broadening in those lines which are not sensitive to it. Stellar spectra also show several of the new lines forbidden by the quantum theory in the absence of fields; and the permitted lines show systematic shifts in accordance with the theory. The Stark effect is appreciable in ordinary dwarfs and is very large in white dwarfs. It can be used to determine the luminosity of a star, because of the close relation between this quantity and the atmospheric density.

5. Rotation.—In many stars, especially of classes O, B, A and F, all lines appear blurred, even those which are not sensitive to Stark effect, while at the same time the curve of growth shows no appreciable turbulence. The equivalent widths of the lines are not altered by this broadening influence and the great widths are obtained at the expense of the central intensities. In some cases it has been shown that the widths of the lines are proportional to the wave lengths, if corrections are made to reduce them to the same equivalent width. Since the Doppler effect is proportional to λ and since no physical alteration of the absorption coefficient can fail to influence the curve of growth, or the equivalent widths of the lines, the conclusion is justified that we are dealing with some form of Doppler effect. Axial rotation, radial expansion and convection suggest themselves. The effect of expansion is to produce an unsymmetrical profile which is deepest at the point of greatest negative displacement $\Delta\lambda_0$, and merges with the continuous spectrum at $\Delta\lambda=0$. Since we can see only one hemisphere of an expanding star, the line does not extend into positive values of $\Delta\lambda$. Neither the asymmetry nor the preference for negative displacements agrees with the observations. We are left with rotation and convection.

To decide between them Struve and Shajn made numerous determinations of line profiles in close spectroscopic binaries. In these stars the components are almost in contact and it is reasonable to suppose that their rotational periods are equal to the period of orbital revolution. All have very broad and shallow lines, and this phenomenon is the more conspicuous the shorter the period. Since no such relation is expected in the case of convection, the "dish-shaped" line profiles can be attributed to axial rotation. In a few eclipsing binaries, for example in Algol, the line profiles become unsymmetrical during the partial phases of the eclipse and, as Rossiter and McLaughlin have shown, it is possible to derive from them the rotational velocity at the equator. The observed values range from 0 to about 300 km./sec., with a few which exceed this value.

A statistical study of the frequencies of different rotational velocities shows that the stars do not all rotate uniformly, and the observed distribution is not simply a result of random orientations of the axes. The largest values are observed in several bright-line B stars. When the bright-line stars are considered alone, the distribution does seem to be due to the geometrical factor alone. It is probable that the rotations of all of these stars are so rapid that they may actually lose material at their equators. In a general way, according to W. W. Morgan, the broadest rotational lines occur in stars of low luminosity and the lines become progressively narrower with increasing luminosity. In all close spectroscopic binaries observed prior to 1960 the direction of axial rotation is the same as that of orbital motion, but in a few close systems (U Cephei) the rotational period is shorter than the orbital period.

6. Central Intensities.—It has been previously stated that even in those lines which are strong enough and broad enough not to be seriously affected by the finite resolving power of the spectrograph the central intensities are not exactly zero, and this remains true after instrumental corrections have been applied. Very accurate central intensities in solar lines have been measured by C. D. Shane at the University of California, who used an interferometer to increase the purity of the spectrum. The origin of these central intensities is interesting; according to physical theory the line-absorption coefficient never becomes infinitely great. Yet, the value predicted by radiation damping is so great that the ordinary absorption formula $I/I_0 = e^{-kl}$, where k is the absorption coefficient and l is the length of the path in the absorbing medium, would require that the central intensity to be so small that it could not be observed to differ from zero.

We are here concerned with a problem of great importance, namely, the manner in which the radiation is propagated through the star's atmosphere. The corresponding theory has been developed by Karl Schwarzschild, Edward A. Milne, Arthur S. Eddington, Bengt Strömgren, A. Unsöld, S. Chandrasekhar and others. The fundamental difference from the ordinary formulation of this problem comes from the fact that the atoms not only absorb but also re-emit radiation, and that this re-emitted radiation is thrown in all directions, while the absorption takes place in radiation directed outward from the star. The central intensities in a star which is devoid of an extended chromosphere come from this re-emission. In a shell, like that of Pleione, where the absorbing gas is very far from the source of the continuous spectrum, the photosphere, only a small fraction of the re-emitted radiation is projected upon the apparent disk of the star, as would be seen by the observer if he had a sufficiently large telescope. Hence the absorption process is essentially that described by the exponential formula, and the central intensities are so close to zero that they have not been measured. But in a normal star of approximately the same temperature the central intensities may be as high as 50% of the continuous spectrum. There are two possibilities: an atom may, after a process of absorption, immediately re-emit the same line; this is the case of pure scattering. Or the atom may after a process of absorption lose track of the quantum, allowing it to be absorbed again and re-emitted a number of times, so that the energy which finally is re-emitted by the original atom is a kind of statistical average over all possible processes and is then determined by the Kirchhoff-Planck function. This is the case of pure absorption.

In reality, both cases may be realized in the stars. Lines of high excitation potential, like those of H and He, probably approach the second case, while resonance lines, like H and K of Ca II, approach the first case. Schwarzschild has shown that for pure absorption the central intensity is that which corresponds to the temperature of the outermost, and hence, the coolest, layers of the stellar atmosphere. This, incidentally, is also the intensity of the continuous radiation at the extreme limb of a star's disk. Hence, if we observe the central intensities of H and He I in a B-type star we obtain a measure of the darkening of the star's continuous radiation from centre to limb.

Unsöld found that in the star τ Scorpii the central intensities, which are almost 50%, are still too small to accord with the usual theory of the limb darkening. He would have expected an even higher central intensity. The discrepancy can be used to improve the theory of the continuous absorption coefficient, which determines very largely the phenomenon of limb darkening.

Some puzzling differences in the central intensities of lines of Fe I were found by L. Spitzer, Jr., in α Orionis and several similar cool supergiants. There is a pronounced tendency for the central intensities to be greater as the excitation potential increases, and there are several abnormally shallow lines. The latter are members of multiplets whose other members behave normally, but they all happen to have identically the same upper level. There is good reason to believe that this particular level is overpopulated because it is fed by a special process pointed out by A. D. Thackeray—there is a strong line of Mg II 2795.52 which almost coincides with the Fe I line 2795.00. If the former were in emission—it cannot be observed because it lies too far in the ultraviolet—this would stimulate the absorptions in the Fe I line and overpopulate the required level. Such abnormal re-emissions are designated as fluorescence. In some late-type variables the abnormal Fe I lines appear actually in emission, while the other members of their multiplet are present only in absorption.

7. Convection.—Adams and his co-workers at Mount Wilson have found that in many stars—especially in giants and supergiants—there are appreciable differential radial velocities between different sets of lines. In some M-type supergiants many lines are double, while others are single. In α Orionis the violet components are sharp and strong, while the red components are

weak. All the double lines arise from very nearly the ground state of the neutral or ionized atoms. This phenomenon is undoubtedly caused by differential motions within the enormously extended atmosphere of the supergiant star. Similar motions may account for the differences in displacement between absorption and emission lines in M-type variables, as observed by Merrill, Joy and others.

8. Centre to Limb Differences.—The most powerful method for testing theories of line formation consists in the study of line profiles at different distances from the centre of the sun. K. Schwarzschild showed that if a line is produced by the re-emission process designated as pure absorption, it should completely disappear at the extreme limb of the sun; in the case of pure scattering, on the other hand, the line should persist at the limb. From the measurements by Schwarzschild, H. H. Plaskett, Miss M. G. Adam, J. Houtgast, M. Minnaert and others, it is known that all strong solar lines become somewhat weaker at the limb but they do not disappear (fig. 2). Very faint lines, however, not only do not show this weakening but even become slightly stronger. The interpretation of these observations is difficult and involves a number of physical data which are not available. The most extensive work on the interpretation of solar line-profiles is by B. Strömgren. In the stars we usually observe line profiles integrated over the entire disk. But in a few eclipsing variables in which a large, dark star centrally eclipses a small, bright star there is a chance of observing the spectrum from the limb of the small star immediately before and immediately after totality. In this manner R. O. Redman, and later O. Struve, found some evidence that the lines of hydrogen almost disappear at the limb of the hot component of U Cephei.

9. Later Results.—The following represent results obtained to about mid-1960:

a. Electronic image intensifiers had been developed by A. L'Allemand in France, J. D. McGee in London, W. Baum and J. S. Hall at Mount Wilson and W. Hiltner at Yerkes. Especially successful observations of spectra were obtained with an intensifier at the focus of the Coudé spectrograph at the 120-in. telescope of the Lick observatory. In the case of the nucleus of the Andromeda galaxy, a small "superglobular cluster" of stars was shown to rotate around its axis with a velocity of 100 km./sec.

b. In addition to the emission spectrum of the nebular filaments of the Crab nebula, a large amorphous patch of nebulosity produced a continuous spectrum which is strongly polarized and is caused by the radiation of relativistic electrons in a magnetic field of the order of 10^{-3} gauss. Nonthermal radiation by relativistic electrons had also been observed in the Crab nebula and in other remnants of supernovae, in the radio frequency range of the spectrum.

c. Richard Tousey of the U.S. Naval Research laboratory obtained high-dispersion spectrograms of the sun, showing hundreds of emission lines produced in the solar corona down to a wave length of approximately 500 Å.

d. H. Friedman photographed the sun in X-rays. H. E. Hinteregger in Colorado measured the flux of the radiation of the sun in the X-ray region.

e. Under the auspices of the National Aeronautics and Space administration (NASA) several U.S. universities planned to build telescopes with spectrographs to be placed in artificial satellites and other space probes.

f. Scientists at the U.S. Naval Research laboratory discovered a new type of nebulosity (*e.g.*, around the star α Virginis), whose most intense radiation has a wave length of approximately 1,300 Å. The physical processes producing this radiation had not been fully explained.

g. Intensive work was being conducted at several radio astronomy observatories to measure the radiative flux in different frequencies of the radio spectrum of a large number of radio sources. Important advances had also been made (especially in the Netherlands), by measuring the Doppler displacements and the intensities of the 21 cm. line of hydrogen, in different sources.

VIII. GASEOUS NEBULAE

The spectra of the gaseous nebulae consist of a weak continuous spectrum and a set of strong emission lines. The former shows a conspicuous discontinuity at λ 3,647 and is therefore, in part, caused by the process of recombination of protons with free electrons. But the continuous spectrum of some of the nebulae in the ordinary photographic region is too strong to be entirely due to the Paschen continuum. It is probable that most, if not all, of it is caused by diffuse scattering of starlight by small solid particles in the nebulae.

The line spectra contain many features which were not identified previous to 1927. In that year I. S. Bowen attributed all the principal "nebulium" lines to forbidden transitions between metastable levels of some of the more common atoms, [O II], [O III], etc.

From the work of E. Hubble it is known that the nebulae derive their light from hot stars located within their confines. The ultraviolet continuous radiation of these stars, especially that on the violet side of the Lyman limit of hydrogen and the corresponding limits of He I and He II, is absorbed by the atoms of the nebula. In this manner the principal permitted lines are excited because they are radiated when the excited atoms fall back, either directly or by cascading, to their ground levels. In the process of cascading some atoms will find themselves in the low-energy metastable levels, where they are caught for a long time because there are no further permitted transitions downward.

The atoms can get out only by one of three processes, all requiring long intervals of time: they can swallow another quantum of energy and return to a higher level; they can collide with another particle and make use of the kinetic energy of the collision to escape the trap; or they can patiently await their release in the form of a forbidden transition which, despite its name, is not entirely impossible but occurs, on the average, once for every 10^8 permitted transitions.

In all laboratory experiments collisions are too frequent to allow enough forbidden transitions to take place. In the atmospheres of the stars the radiation is so intense that quanta fly about everywhere in large numbers and atoms which find themselves in the trap have plenty of opportunity to get out by absorption. But in the nebulae, collisions are very rare and the density of quanta is low. Hence, large populations of atoms accumulate in the metastable levels, and despite the relative infrequency of a forbidden transition the forbidden lines appear with great strength.

Bowen has shown that while the process of cascading will furnish appreciable populations in the metastable levels, the forbidden lines are relatively so strong, compared with the permitted lines of the same ions, that another process must also be postulated. It consists of collisions of unexcited ions with free electrons. Although collisions are insufficient to remove an appreciable fraction of atoms from an excited level to a higher level, with a large change in energy, they are very efficient in producing small energy changes and, therefore, account satisfactorily for the observations. Bowen has discovered a third process of nebular excitation, namely, fluorescence. By a chance coincidence the wave length of one of the strongest (though unobserved) emission lines produced by the primary process of recombination, He II 303.780, agrees almost precisely with the permitted line O III 303.799. In consequence, the emission of He II stimulates an abnormally large number of atoms of O III to absorb and to jump to the upper level of the O III line. Hence, they cascade downward, producing strong lines of certain members of various multiplets, while other members remain invisible. These mutilated multiplets have been observed by W. H. Wright.

The energies contained in the emission lines of a nebula are the result of conversion by the atoms of ultraviolet continuous radiation of an exciting star. Since the former can be measured, we obtain a measure of the unobservable ultraviolet light of a hot star; on this basis H. Zanstra has developed a theory for determining the temperature of the exciting stars. This method in the hands of D. H. Menzel and his associates gave excellent results and led to the realization that while the temperatures of the stars range from 3,000° to 100,000°, the electron temperature of the nebular gas is usually of the order of about 10,000°. The farthest ultraviolet continuous radiation of the star is absorbed in its immediate vicinity. Hence the lines of He II are limited to a small volume around the star. On the other hand, elements of lower ionization potential, like He I and H, occur at greater distances. This gives rise to an apparent stratification in the nebulae. Of great interest are the absorption lines of He I observed by O. C. Wilson in the Orion nebula. All these lines

have metastable lower levels. According to D. Osterbrock, the densities of the diffuse gaseous nebulae range between about 10^2 and 10^4 hydrogen atoms (or the same numbers of free electrons) per cubic centimeter.

Planetary nebulae, of which the ring nebula in Lyra is an example, consist of slowly expanding shells of gas whose spectra resemble those of the diffuse galactic nebulae. The central stars of the planetaries are small and hot. A few have properties resembling those of the white dwarfs. Presumably the descendants of all planetaries are ordinary white dwarfs. Herbig and Shklovsky have investigated the evolution of the planetaries; the former has suggested that their ancestors are variable stars of R Coronae borealis type whose spectacular variations in spectrum and brightness occur at irregular intervals. Shklovsky believes that the evolution of the planetaries is so rapid that within our Milky Way galaxy three new planetary nebulae are produced each year, and three disappear as planetaries and become white dwarfs.

A subject closely related to the nebulae is that of interstellar gas. It manifests itself in two forms: as absorption lines of abundant elements in their ground levels, observed on the background of the continuous spectra of distant stars, and as emission lines of H, [O II], [N II], etc., in many large areas of the Milky Way. The emission lines, as well as the absorption lines, are produced through excitation by the combined light of the stars.

The density and the temperature of this field of radiation can be computed for the neighbourhood of the solar system. Since we observe interstellar lines of Ca I and Ca II, we can form the ratio n'/n in Saha's ionization formula, appropriately amended for the effect of dilution in the radiation. We then have only one unknown quantity, the electron pressure. This turns out to be equivalent to about 10 electrons per cubic centimetre, a value which is consistent with the observed surface brightness of luminous hydrogen over large areas of the Milky Way.

C. S. Beals, R. F. Sanford, W. S. Adams and G. Münch have found many stars with multiple interstellar absorption lines showing that the gas consists of clouds having individual motions. Andrew McKellar, Theodore J. Dunham, Jr., Walter S. Adams and G. Herzberg have found and identified several interstellar molecules, such as CH, CH^+ and CN, but several broad interstellar lines remain unidentified.

A theory of interstellar emission has been developed by B. Strömgren. Because of the gradual exhaustion of the ionizing stellar radiation in the gas he added a term to Saha's formula to take care of this exponential absorption. It turned out that the boundary between the region where H is ionized and the region where it is un-ionized is fairly sharp.

IX. PLANETS AND COMETS

The planets shine by reflected light and their spectra show the usual solar spectrum; the distribution of continuous energy is modified by the colour of their surface structures or by the scattering of their atmospheres. The solar lines are displaced by the orbital and rotational motions of the planets and, in addition, there are absorption bands produced in these atmospheres. Measurements of radial velocity have been extensively used to determine the periods of rotation, especially in the case of Neptune, and the law of motions of the rings of Saturn. Differential radial velocities between the planet and the Earth are also important to distinguishing between absorption bands of water vapour and of oxygen, which are strong in our own air.

Adams concluded in 1941 that water vapour, if at all present on Mars, constitutes not more than 5% of that in air, and probably very much less. Nor is there any free oxygen in excess of that in air. But the red colour of the desert areas on Mars suggests the existence of large quantities of oxygen in the form of compounds, such as rust. There may also be enough water vapour to give rise to thin sheets of ice at the polar caps.

Adams and Dunham found two new bands of carbon dioxide on Venus, while A. Adel and V. M. Slipher estimated that the amount of carbon dioxide above the cloud layer of Venus is several hundred metre-atmospheres. The spectra of Jupiter, Saturn, Uranus and Neptune show conspicuous bands in the red and infrared regions. They were identified by R. Wildt in 1932 as ammonia and methane. Several successful attempts have been made by Dunham and by Adel to determine the temperatures from the distribution of intensity in the various bands. Thus, for Venus, Adel found 50° C., from the carbon dioxide bands. The ammonia band is strongest in Jupiter and becomes progressively weaker in Saturn, Uranus and Neptune. The methane bands, on the other hand, increase greatly from Jupiter, with 0.5 mile-atmosphere (amount of gas whose absorption of light is equivalent to a mile-long column of the same gas at standard atmospheric pressure), to Neptune, with 25 mile-atmospheres, according to Slipher and Adel. In 1944, G. P. Kuiper found the bands of methane on the largest satellite of Saturn, Titan. They are a little weaker in the satellite than in the planet. None of the other satellites in the solar system, nor Pluto, shows any of the bands with the spectrograph used by Kuiper.

The spectra of the comets are characterized by emission bands of CN, C2, CH, OH, NH, CH^+, CH2 in the heads, CO^+ and $N2^+$ in the tails. P. Swings has shown that the bands are excited by solar radiation. Since the latter is not uniform, like that of a black body, but is broken up by many strong absorption lines, the excitation to various molecular levels in the comet is quite abnormal, and the resulting intensities of the individual lines within the bands are different from those observed in the laboratory.

X. THE SUN

Since a separate article on the sun (*q.v.*) contains a description of its spectrum, only the principal features have been briefly mentioned in the foregoing sections to the extent to which they are valuable in furnishing independent information to the study of astronomical spectroscopy. There remains, for completeness, the problem of the spectrum of the corona.

The corona consists of an inner zone having a continuous spectrum, with emission lines, and of an outer zone where the continuous spectrum is broken up by the solar absorption lines, and rendered increasingly diffuse as the inner zone is approached, as though the scattering particles were in very rapid turbulent motion.

W. Grotrian accepts the idea that this scattering is produced by free electrons. The emission lines have been identified by Edlén as forbidden transitions of [Fe X, XI, XIII, XIV, XV], [Ni XII, XIII, XV, XVI], [Ca XII, XIII, XV] and [A X, XIV]. The corresponding ionization potentials are exceedingly high: 814 v. for [Ca XV]. In contrast, the ionization level in the sun's atmosphere is about 8 v. If we insert these values, one after the other, in the ionization equation and use the same value of $T = 5040°$ C., which is approximately correct for the sun's atmosphere, we obtain $log_{10}\, p_e/p_e' = 806$. This is impossible, since there would not be enough gas left to show any lines. Evidently we must adopt for the corona a very much higher temperature, perhaps one of the order of 1,000,000°, to obtain a reasonable pressure. The origin of this high temperature is not known. Because of the extremely tenuous nature of the corona, this high temperature does not make itself felt on the Earth; what we experience as the heating action of the sun is the radiation of the photosphere whose temperature is more nearly 6,000°. The strongest coronal lines, λλ 5,303, 6,375 of [Fe XIV] and [Fe X] and some others, were found by Adams and Joy in the peculiar star RS Ophiuchi. Thus, there is a link between the spectra of the stars and that of the solar corona.

BIBLIOGRAPHY.—History of Astronomical Spectroscopy: H. Kayser, *Handbuch der Spectroscopie*, vol. i (1900). Efficiency of Slit Spectrographs: J. E. Keeler, *Sideral Messenger*, vol. 10, p. 433 (1891).

Photometry of solar spectrum: "Photometric Atlas of the Solar Spectrum," The Utrecht Observatory (1940).

Wave lengths of solar lines: "Revision of Rowland's Preliminary Table of Solar Spectrum Wave Lengths" (Mount Wilson Observatory), *Carnegie Institution of Washington Publications*, no. 396 (1928).

Methods for observing the extreme infrared region of the spectrum: P. Swings, *Publications of the Astronomical Society of the Pacific*, vol. 56, p. 220 (1944), and vol. 57, p. 16 (1945).

Construction of astronomical spectrographs: G. Z. Dimitroff and J. G. Baker, *Telescopes and Accessories* (1945).

Spectral classification: The Harvard system is described by A. J.

Cannon in the introduction to each volume of the *Henry Draper Catalogue* (*Harvard Annals*, vol. 91–99, 1918–24); the method of Morgan and Keenan is described in *An Atlas of Stellar Spectra* (1943).

Interpretation of spectral classification: A. Unsöld, *Physik der Sternatmosphären* (1955); S. Rosseland, *Theoretical Astrophysics* (1936); *Handbuch der Physik*, vol. 50–53 (1958–59). (O. Sr.)

SPECTROSCOPY, X-RAY. A beam of short-wave electromagnetic radiation called X-rays is scattered when it strikes a crystal, much in the same way a beam of light is diffracted by a ruled grating. The science that deals with the study of X-ray diffraction is X-ray spectroscopy. Its practical applications are discussed in the concluding section of X-RAYS.

History.—Max von Laue discovered in 1912 that radiations in the X-ray wave-length range of 10 to 0.1 Å are diffracted by natural crystals whose layers of atoms are only a few Ångström units apart (1 Å = 10^{-8} cm.). Ruled gratings that work well in registering ordinary optical spectra with wave lengths down to about 2,000 Å are of no value in diffracting X-rays. To register the region between 2,000 and 10 Å, other methods were subsequently developed.

In one of his first papers dealing with X-rays, Röntgen noted that X-rays produced in an ordinary X-ray tube differ in quality when the vacuum of the tube and particularly the voltage on the tube are varied. It was shown early that the X-radiations became more penetrating as the voltage on the tube was raised. The degree of penetration (usually measured in terms of the thickness of aluminum necessary to reduce their intensity by half) was used as a means of characterizing the quality of the radiation.

It was by this means also that C. G. Barkla was able to show that the different elements when excited so as to give off X-rays all have their own characteristic radiations. For instance, for an element such as silver Barkla showed the existence of two characteristic radiations called the K- and L-radiation having very different penetrating powers.

These two characteristic radiations, the K-series and the L-series, were experimentally verified for a great number of elements. The measurements of their penetration in aluminum showed that their hardness (ability to penetrate) increases regularly for both series as the atomic weight of the emitting element increases.

Basis of X-ray Spectroscopy.—The diffraction patterns which W. Friedrich and P. Knipping obtained by allowing a fine beam of X-rays to pass through a crystal as suggested by Laue induced W. H. and W. L. Bragg to perform the experiments which were the first step in the development of the X-ray spectroscopy. As the Braggs showed, a monochromatic X-ray beam is reflected by a cleavage face (or any other atomic planes) of a crystal according to the ordinary laws of optical reflection; *i.e.*, the incident and the reflected beams are in the same plane, perpendicular to the reflecting face; further, the angles between these two beams and the reflecting face are equal. In addition to these laws the following condition must be fulfilled if reflection is to occur, namely:

$$n\lambda = 2d \sin \psi_n \qquad (1)$$

where λ is the wave length of the monochromatic radiation, d is the distance between two adjacent atomic layers parallel to the reflecting plane, ψ_n is the angle between the beam and the plane and n indicates the order of the reflection (fig. 1).

This equation, generally known as Bragg's law, forms the basis for measuring the wave lengths of X-rays. This law needs a small correction, due to the fact that the wave length is slightly different in a vacuum (or air) than in a crystal.

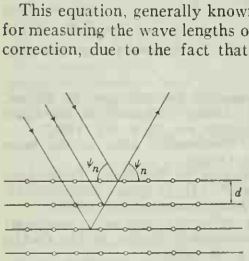


FIG. 1.—REFLECTION OF X-RAYS ACCORDING TO BRAGG'S LAW (*see* TEXT)

Spectrometric Methods.—The wave-length region of ordinary X-rays extends over the range 0.1 to 10 Å. The rays show different penetrating properties, and are termed hard (penetrating) or soft X-rays. The region of hard X-rays extends from 0.1 Å to about 1.5 Å, the wave length where the absorption in air becomes important. In the soft X-ray region the great absorption in air makes it necessary to enclose the spectrometer in a vacuum container. The X-ray region from 10 Å and up to ultraviolet is called the ultra-soft region. Ultra-soft X-rays can be studied by using ruled gratings.

In most spectrographs the X-ray beam incident on the crystal will be reflected by the crystallographic planes that are parallel to the crystal surface. This so-called reflection mounting of the crystal is best suited for the soft X-ray region (fig. 2A). In the hard X-ray region the angle ψ is small and the X-ray beam therefore sweeps almost parallel to the crystal. In this case the transmission mounting is the most suitable (fig. 2B). The short wave-length radiation can easily penetrate the crystal plate and the reflection occurs at crystal planes that are perpendicular to the crystal surface.

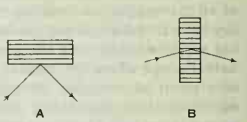


FIG. 2.—REFLECTION OF X-RAYS BY CRYSTALLOGRAPHIC PLANES: (A) REFLECTION MOUNTING FOR SOFT X-RAYS, (B) MOUNTING FOR HARD X-RAYS

The first X-ray spectrographs used a plane crystal as is shown diagramatically in fig. 3. The different monochromatic rays which constitute the beam coming from the X-ray source and defined by a slit are reflected by the crystal at different angles according to Bragg's law (equation 1). If the angles may vary from ψ_1 to ψ_2 only such rays are reflected whose wave lengths have values between λ_1 and λ_2 where

$$n\lambda_1 = 2d \sin \psi_1$$
$$n\lambda_2 = 2d \sin \psi_2$$

The photographic plate therefore registers a spectrum ranging from λ_1 to λ_2. The region of wave lengths can be varied by turning the crystal. The wave length of any spectral line on the photographic plate can be computed by equation 1 from its position on the plate.

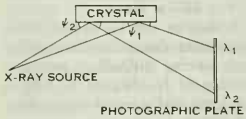


FIG. 3.—PRINCIPLE OF EARLY X-RAY SPECTROGRAPH USING PLANE CRYSTAL

The resolving power (and the sharpness of the lines) of a crystal spectrometer depends on the slit width and the distance between slit and crystal. In the two-crystal or double-crystal spectrometer the slit is replaced by another crystal (fig. 4). The first crystal selects a wave-length region which is analyzed by the second crystal. This crystal can be put in two different positions. At the parallel position the outgoing beam is parallel to the incoming beam throughout the wave-length region (fig. 4A). If the second crystal is turned to the antiparallel position the beam is deflected twice (fig. 4B). The crystal has thereby been turned through an angle of $180° - 2\psi$. The wave length is then computed by equation 1. Each setting of the second crystal allows only a very narrow wave-length band to be reflected. The width of this wave-length band depends on the quality of the crystals and is usually narrower than the width of a spectral line. In using the instrument the crystal is gradually turned through the wave-length region to be investigated; the reflected intensity is measured in each setting of the crystal by means of some counting device, *e.g.*, a Geiger-Müller counter. Photographic registration is never used at this instrument.

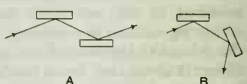


FIG. 4.—PRINCIPLE OF THE TWO-CRYSTAL SPECTROMETER: (A) CRYSTAL AT PARALLEL POSITION, (B) CRYSTAL AT ANTIPARALLEL POSITION

The spectrometers mentioned above, especially the double-crystal spectrometer, have the disadvantage of low luminosity. A great improvement of luminosity was achieved by the principle of bent-crystal spectrometers, first formulated by J. W. M. DuMond and P. Kirkpatrick in 1930. A reflecting type spectrograph with bent crystal was first constructed by H. H. Johann in 1931. A thin crystal plate, C in fig. 5A, is bent cylindrically to a radius

r and a circle is drawn with r taken as its diameter. This circle, which corresponds to the Rowland circle in optical spectroscopy, causes rays coming from a point A on the circle to be reflected and focused on another point B on the circle. Because the centre of curvature of the crystal O also lies on the Rowland circle,

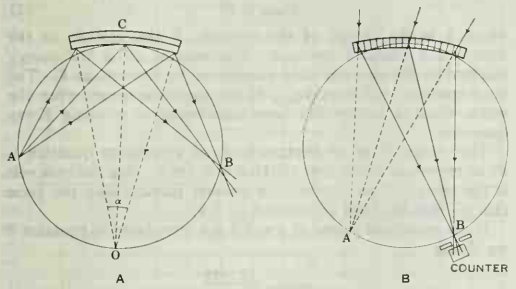


FIG. 5.—PRINCIPLE OF BENT-CRYSTAL SPECTROGRAPHS AS USED IN (A) A REFLECTING SPECTROGRAPH AND (B) A TRANSMISSION SPECTROGRAPH

Bragg's law is met for the reflection. The focusing is not absolutely perfect, however, since the crystal does not follow the Rowland circle smoothly. The focusing defects are not severe if the crystal aperture α is small and the Bragg angles great (long wave lengths).

A transmission type of bent-crystal spectrograph was first built by Y. Cauchois in 1932. Rays from the focal spot directed toward a virtual object point A are focused at B (fig. 5B). The rays are reflected by the internal crystal planes and consequently this spectrometer is best suited for short wave lengths. As in the Johann spectrometer the radiation is detected by a photographic plate or by means of a counter with a slit.

Detectors for X-rays.—In spite of its poor sensitivity, especially for hard radiation, the photographic plate is the most important X-ray detector and the only one which simultaneously registers all lines in the spectrum.

Modern precision spectrometers are often equipped with counting detectors, however, where the single quanta produce voltage pulses, counted and registered by electronic devices. Geiger-Müller counters or proportional counters are used for soft X-rays and scintillation counters preferably for hard X-rays. The proportional counter and the scintillation counter require more elaborate electronic devices than the Geiger-Müller counter but they offer the great advantage of giving output pulses that are proportional in voltage height to the energy of the radiation. By electronic circuits it is possible to discard pulses of wrong height arising from stray radiation, thus diminishing the background disturbances. Fig. 6 shows an X-ray spectrum detected by a curved-crystal spectrometer with a scintillation counter.

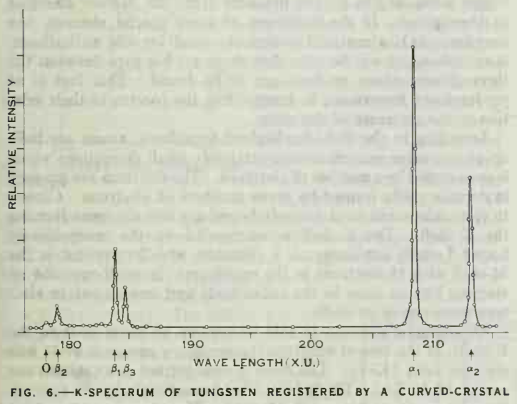


FIG. 6.—K-SPECTRUM OF TUNGSTEN REGISTERED BY A CURVED-CRYSTAL SPECTROMETER

Exciting X-ray Spectra.—To excite the X-ray spectrum of a substance, a small piece of it is placed as target material on the anticathode of an X-ray tube. After the tube has been exhausted to a suitable vacuum, a high potential (10 to 150 kv.) is applied to the electrodes of the tube, causing the anticathode to be bombarded by electrons from the cathode. The kinetic energy of the electrons imparted to the bombarded atoms is partially transformed into heat and light, and partially to X-rays.

There are two main ways of producing a bombardment team of electrons with the X-ray tube. In the first method, the tube has a vacuum of 0.01 to 0.001 mm. Hg. The high voltage causes the gas remaining in the tube to ionize, with the positive ions being thrown against the aluminum cathode. The collision between the positive ions and the cathode sets free a number of electrons. The electrons travel in the opposite direction—from the cathode to the anticathode—and bombardment of the anticathode gives rise to X-rays characteristic of the substance of the anticathode.

In the second method, the vacuum of the tube is much lower and there is no gas ionization of any importance. The bombarding electrons are supplied by the heated filament (usually of tungsten) in the cathode. This is the most commonly used type of X-ray tube because the electron beam can readily be controlled by adjusting the heating current in the cathode.

When analyzing materials, it is often necessary to change the substance on the anticathode; hence the tube must be built so the anticathode can be detached easily. This can be done by mounting

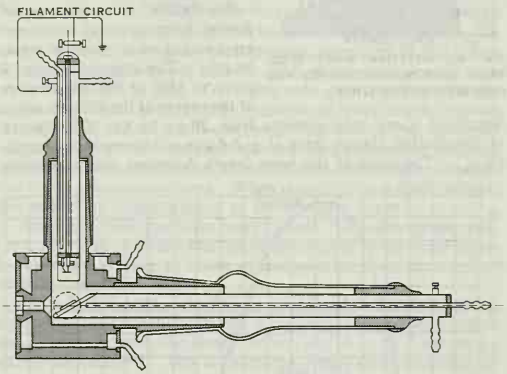


FIG. 7.—METAL X-RAY TUBE WITH HOT CATHODE FOR SPECTROSCOPIC USE

the anticathode in a special joint. A tube which permits rapid change of both cathode and anticathode is shown in fig. 7. The tube itself is built of metal with water cooling to permit a very high output of energy. A small window of aluminum or other low-absorbing material permits the study of radiation of longer wave lengths, which would be absorbed by the glass walls of an ordinary tube.

Another way of exciting X-ray spectra is to irradiate the substance with an intense beam of X-rays. In this case the substance emits secondary rays which, with a few exceptions, are identical with the X-radiation sent out by the substance when used as an anticathode in an X-ray tube. This method does not, however, give the same intensity as the former.

X-ray Wave-length Units.—In early X-ray work the wave lengths were given in Ångström units. Since this unit is too big for convenience, the X-unit X.U. (10^{-11} cm.), was introduced.

The X-ray wave lengths are determined by crystal spectrometers and this requires an exact knowledge of the grating constant d of the crystal (*see* equation 1). As experimental technique improved, the spectrometric measurements of X-ray lines became definitely more accurate than the measurements of the grating constants, which had to be determined from measurements of density and

dimensions of cubic crystals. For these reasons the X-unit has been defined by the lattice constant of the common plane in calcite set equal to 3,029.04 X.U. at 18° C. Later, E. Bäcklin, F. Tyrén and other investigators succeeded in measuring some soft X-ray lines using ruled gratings. A direct connection between the Ångström and X-units was thus established and by agreement 1,000 X.U. = 1.00202 Å.

Different Kinds of X-ray Spectra.—The X-ray spectrum emitted from any substance is made up of two kinds of radiation, one of which shows a continuous distribution over a wide range of wave lengths, the other consisting of a few monochromatic rays overlapping the former (fig. 8). The first kind of radiation corresponds to the "white" light of optics, whereas the second is analogous to the line spectra. The continuous spectrum contains the greater part of the energy of radiation and is therefore the most important part for medical and other applications of X-rays. This part of the X-ray spectrum shows qualitatively no dependence on the radiating substance, the intensity only being different from various anticathodes.

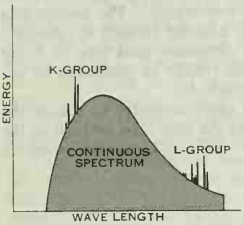


FIG. 8.—IDEALIZED X-RAY SPECTRUM SHOWING CONTINUOUS AND LINE-SPECTRA RADIATION

The line-spectra on the other hand are characteristic of the emitting substance. They correspond to the K- and L-radiation of Barkla, but analysis has shown that each of these radiations consists of a group of monochromatic spectral lines.

The Continuous Spectrum.—The "white" or continuous radiation from an X-ray tube covers a rather wide region of wave lengths. The diagram of fig. 9 gives an idea of the distribution of the energy of the different wave lengths at X-ray tube voltages from 20 to 50 kv. The curves of distribution always start at a definite minimum wave length (λ_{min}). The value of this wave length decreases with increasing voltage. The wave length (λ_{min}) or its corresponding frequency

$$(\nu_{max} = \frac{c}{\lambda_{min}})$$

obeys the Einstein law

$$h\nu_{max} = eV \qquad (2)$$

where e is the charge of the electron, V the voltage on the tube and h Planck's constant. The amount eV is the energy received by the electron when accelerated by the voltage V. The most energetic radiation $h\nu_{max}$ in the spectrum occurs when the whole electron energy has been transferred to a single X-ray quantum.

The energy eV of an electron or $h\nu$ of a radiation quantum is often expressed in the unit electron volt (ev). One electron volt is the energy of an electron (or another particle with the same charge) that has been accelerated by 1 v.

If the numerical values of e and h are introduced in equation 2 we obtain

$$\lambda \min = \frac{12.3978}{V} \qquad (3)$$

where the wave length is expressed in Ångström units and V in kilovolts.

The total amount of energy, that is the integral energy included by the distribution curves in fig. 9, is proportional to the square of the voltage on the tube, as has been verified by several investigators. Further, it has been found that for different elements used as anticathodes the total radiation is proportional to the atomic number of the element. For this reason, heavy elements have long been used as anticathodes in medical X-ray tubes.

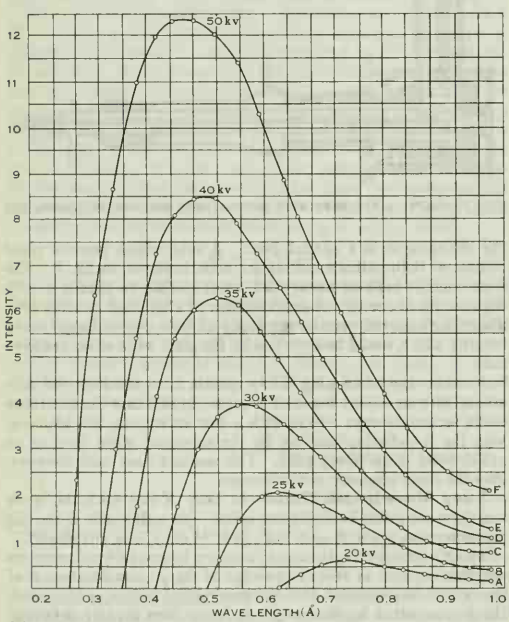
The Line-Spectra.—In 1913 and 1914 the young English physicist H. G. J. Moseley published two remarkable papers in the *Philosophical Magazine* entitled "The High-Frequency Spectra of the Elements." Moseley showed that X-ray spectra include line spectra of the same type as known in ordinary optics, and that these new spectra were built up in a more simple and regular way than is generally the case in spectra previously studied.

Two groups of spectral lines were found. One of these groups was identified with the K-series of Barkla; the other, of longer wave length and of more complex structure, corresponds to the L-series. In 1916 M. Siegbahn discovered a new series of still greater wave lengths called the M-series. The existence of series outside the K- and L-group had been suspected earlier by reasons of analogy. In the later development of X-ray spectroscopy, higher series, N, O, etc., have been found.

A general scheme of the X-ray spectra is given in fig. 10, which contains the strongest lines of the three groups, K, L and M for every third element from sodium (11) to uranium (92). The diagram shows how all three series are regularly displaced to longer wave lengths as one proceeds from the heavier elements to the lighter. If the spectrum of some special element, say tungsten (74), a material commonly used for the anticathode, is considered, it will be seen that there are big gaps between the three groups where no lines are to be found. This fact is of predominant importance in interpreting the spectra in their relation to the structure of the atom.

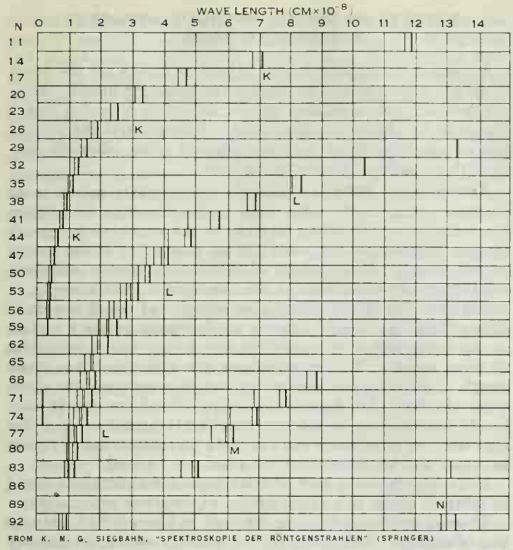
According to the Bohr-Rutherford hypothesis, atoms are built up of a positive nucleus of comparatively small dimensions which is surrounded by a number of electrons. The electrons are grouped in definite shells formed by given numbers of electrons. Closest to the nucleus and most strongly bound are two electrons forming the K-shell. The K-shell is surrounded by the more loosely bound L-shell, consisting of 8 electrons, and farther out is the M-shell with 18 electrons at the maximum. In an X-ray tube an electron hits an atom in the anticathode and knocks out an electron from one of its shells.

An energetic electron is able to remove an electron from the K-shell; in the case of tungsten the necessary energy is 69.51 kilo electron volts (Kev). The atom is now excited and contains the energy 69.51 Kev. The hole is filled immediately by an electron from an outer shell; *e.g.*, the L-shell. The atom is now lacking



FROM K. M. G. SIEGBAHN, "SPEKTROSKOPIE DER RÖNTGENSTRAHLEN" (SPRINGER)

FIG. 9.—INTENSITY DISTRIBUTION OF THE CONTINUOUS RADIATION FROM A TUNGSTEN ANTICATHODE AT VARIOUS VOLTAGES



FROM K. M. G. SIEGBAHN, "SPEKTROSKOPIE DER RÖNTGENSTRAHLEN" (SPRINGER)

FIG. 10.—THE STRONGEST LINES OF THE SEVERAL X-RAY SERIES ARE SHOWN FOR EVERY THIRD ELEMENT

one of its L-electrons and contains the corresponding energy of 10.20 Kev. The energy difference 69.51 − 10.20 = 59.31 Kev is transmitted as an X-ray quantum $h\nu$ and the wave length corresponding to 59.31 Kev can be calculated from equation 3 to be 0.209 A (208.6 X.U.). If the hole in the K-shell is filled by other electrons, other X-ray lines are obtained. They all have energies of the order of 60 to 70 Kev corresponding to 0.18 to 0.21 Å (178 to 213 X.U.). These lines form the K-series of tungsten (fig. 6). A hole in the L-shell involves an energy of about 10 Kev and when this hole is filled up by transitions from the M, N or other outer levels the L-series lines are emitted. The corresponding wave lengths are 1.2 to 1.6 Å (*see* fig. 10).

The transitions between the various energy states of the atom where different electrons are missing can be illustrated by an energy level scheme (fig. 11). The different levels are characterized by quantum numbers. The main quantum number n determines the various shells: K-shell has $n = 1$, L-shell has $n = 2$, etc. With one L-electron missing the atom is excited to the L-level. An analysis shows, however, that the energy of the atomic excitation is dependent on which of the eight L-electrons is ejected; in reality there are three different L-levels, L_I, L_{II} and L_{III}. Another quantum number l is now introduced. There are two L-electrons with $l = 0$. If one of these electrons is missing, the atom is excited to the L_I-level. The six other L-electrons have $l = 1$ and the corresponding levels are L_{II} and L_{III}. The M-shell with $n = 3$ gives rise to five levels: M_I ($l = 0, 2$ electrons) M_{II}, M_{III} ($l = 1, 6$ electrons) and M_{IV}, M_V ($l = 2, 10$ electrons). The building up of the electron configurations for some atoms (the rare gases) is given in the Table.

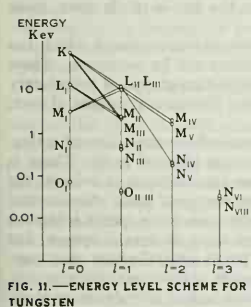
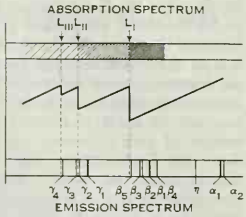


FIG. 11.—ENERGY LEVEL SCHEME FOR TUNGSTEN

In fig. 11 some of the various transitions in the atom are represented by lines connecting the levels. The lines originating from the K-level represent the K-group, etc. (The four lines from the K-level in fig. 11 correspond to the lines α_2 α_1 β_3 β_1 in fig. 6).

The electrons do not jump haphazardly from one level to another; the transitions are regulated by selection rules. The most easily occurring transitions, which give rise to the strongest lines, pass between levels where the quantum number l changes by one unit. The transitions noted in fig. 11 are some of the possible "dipole" lines. Other transitions are not strictly forbidden, but they give rise to very weak lines called "quadrupole" lines. Such a line is the KM_{IV}, $_V$-line.



FROM K. M. G. SIEGBAHN, "SPEKTROSKOPIE DER RÖNTGENSTRAHLEN" (SPRINGER)

FIG. 12.—ABSORPTION AND EMISSION SPECTRUM SHOWING THE RELATIVE POSITIONS OF DISCONTINUITIES AND LINES

The Absorption Spectra.—In addition to the emission spectrum of X-rays there is also an absorption spectrum analogous to the absorption spectra of ordinary optics. Such spectra are obtained when a thin foil of some substance is placed between the X-ray source and the detector (*e.g.*, photographic plate). Instead of a continuous blackening from the white X-radiation the photographic plate shows one or more sharp edges where there is rapid change in the blackening—the so-called "absorption-edges."

These absorption discontinuities are explained as follows: An X-ray quantum $h\nu$ impinging on an atom may transfer its energy to a K- or L-electron, which will be ejected from the atom. To eject a K-electron from a tungsten atom, an energy of 69.51 Kev is required, as mentioned before. This energy corresponds to a wave length of 178 X.U. Obviously only quanta with an energy higher than 69.51 Kev (that is, radiation of wave lengths shorter than 178 X.U.), can eject K-electrons. Radiation of wave lengths longer than 178 X.U. may pass relatively easily through a tungsten foil, but wave lengths shorter than the limit are strongly absorbed due to the process mentioned. When registered by a spectrometer, this results in a discontinuity in the intensity curve.

Since the L-shell has three energy levels, the L-absorption spectrum shows three discontinuities corresponding to the three levels L_I, L_{II} and L_{III}. Fig. 12 shows an L-absorption spectrum together with the L-emission lines.

Ultra-Soft X-rays.—As already mentioned, the crystal grating method has its limitations in that wave lengths longer than double the distance between the atomic layers, say 10 or 20 Å cannot be selectively reflected. This is clear from Bragg's law (equation 1) which shows that λ cannot be larger than $2d$.

Fortunately it was found that ruled gratings when used in nearly grazing incidence gave nice spectra of the radiation with wave lengths down to the region where the crystals had their limits. This method was first used by R. A. Millikan for the optical spectra and then by A. H. Compton and R. L. Doan, M. Siegbahn and others for the study of the ultra-soft X-ray spectra. An important result of this work was that it suggested a direct comparison between optical and X-ray wave lengths, resulting in a determination of atomic distances and the previously mentioned relation between the Ångström and X-units. Subsequently in this way a very exact determination of the value for the fundamental electric unit —the electronic charge—was worked out by Bäcklin. It was further possible to extend the different X-ray spectral series into the new wave-length region and to find lines belonging to the N- and O-series.

Miscellaneous Influences on X-ray Spectra.—The most

Electrons in Rare Gases

	K	L_I	L_{II}, L_{III}	M_I	M_{II}, M_{III}	M_{IV}, M_V	N_I	N_{II}, N_{III}	N_{IV}, N_V	N_{VI}, N_{VII}	O_I	O_{II}, O_{III}	O_{IV}, O_V	P_I	P_{II}, P_{III}
2 Helium	2	..	..	..	..	..	..	..	..	..	..	..	..	..	..
10 Neon	2	2	6	..	..	..	..	..	..	..	..	..	..	..	..
18 Argon	2	2	6	2	6	..	..	..	..	..	..	..	..	..	..
36 Krypton	2	2	6	2	6	10	2	6	..	..	..	..	..	..	..
54 Xenon	2	2	6	2	6	10	2	6	10	..	2	6	..	..	..
86 Radon	2	2	6	2	6	10	2	6	10	14	2	6	10	2	6

striking difference between the ordinary optical spectra and the X-ray spectra is the fact that the former show a marked resemblance between elements with the same chemical character and change considerably from one group of chemically analogous elements to another, whereas in the X-ray spectra no indication of the chemical nature of the elements is to be found. This is readily explained by the Bohr-Rutherford concept of the structure of the atom. From this interpretation, the optical spectra arise from the surface of the atom, which is also the seat of the chemical bonds, while the X-ray spectra are given off from the inner parts of the electronic atmosphere of the atom. It must, however, be expected that a small influence may be observed if an atom emitting an X-ray is acted upon by neighbouring atoms. Determining the degree of this influence depends on the sensitivity and accuracy of methods used in studying X-ray spectra. By increasing the precision of study techniques, it has been possible to detect a number of such influences. It was first found by J. Bergengren that the K-absorption edge for phosphorus was displaced slightly when the element was modified. The line-spectra are also influenced to a certain degree by the chemical bonds, as shown by A. Lindh, O. Stelling and others. *See* also QUANTUM MECHANICS; RADIOACTIVITY; RADIOLOGY; SPECTROSCOPY; X-RAYS.

(M. SIE.)

SPEE, MAXIMILIAN, COUNT VON (1861–1914), German naval officer, who commanded German forces at the battles of Coronel and the Falkland Islands (*qq.v.*), was born in Copenhagen on June 22, 1861.

Spee entered the German navy in 1878 and from 1887 to 1888 was commander of the port in German Cameroon. In 1908 he was made chief of staff of the German North sea command, and at the end of 1912 took over the far eastern squadron. On Nov. 1, 1914, he was engaged by Adm. Christopher Cradock off Coronel, on the Chilean coast, and defeated the British squadron. On Dec. 8, 1914, however, he was attacked by Adm. Sir Frederick Sturdee's cruiser squadron, the Germans being heavily defeated in the battle, usually known as the battle of the Falkland Islands. Spee himself went down with his flagship, the "Scharnhorst."

SPEECH DISORDERS.[1] One may be considered to have a disorder or defect of speech when his speech deviates sufficiently from the norm that: (1) what he says is partially or completely unintelligible, (2) his manner of speaking distracts from what he is saying, or (3) his manner of speaking is a source of embarrassment or effort to him or to his listener.

The concept of disorder or defect of speech usually connotes that the learning of speech was inadequate or that the physical mechanism for producing speech is inadequate. Thus, a foreign or regional dialect is not usually considered a defect of speech because the listener generally recognizes that the difference is not the result of inadequate learning or of an inadequate physical mechanism. Should the dialect result in sufficient deviation that the speech fails in its role of communication, however, the preceding definition of disorder of speech may apply. Faulty pronunciations also should be differentiated from defects of speech. A mispronunciation represents a wrong choice from an otherwise adequate repertory of speech sounds while a defect of speech represents an inadequate repertory of speech sounds.

The difficulties included under the general heading of speech disorders range from slight slurring of speech or faulty enunciation of one or two sounds, to complete unintelligibility or complete lack of speech. Surveys and estimates indicate that between 5% and 10% of the U.S. population over six years of age have defects of speech. There are more persons handicapped by defects of speech than there are crippled, deaf and blind combined.

This article considers, in turn, speech disorders without apparent physical basis (lisp, tongue-tie, baby talk and stuttering) and speech disorders with physical causes (those of the deaf and hard of hearing; cleft lip and palate; defective laryngeal function and structure; aphasia; and cerebral palsy).

[1]This article is adapted from a chapter, "Speech Therapy" in *Principles and Technics of Rehabilitation Nursing* by Florence J. Terry, Gladys S. Benz, Dorothy Mereness, Frank R. Kleffner and Deborah Jensen, edited by Deborah Jensen (The C. V. Mosby Company, St. Louis, Mo., 1957).

DEFECTS OF SPEECH WITHOUT APPARENT PHYSICAL BASIS

A common assumption is that defects of speech are the result of defects in the physical mechanism for producing speech. Actually, however, among children 70% to 80% of all defects of speech appear to be without physical cause. Among adults the proportion of defects without apparent physical cause, although substantial, is smaller than among children.

Articulatory Defects.—Literally, to articulate means to fit together. When used in describing speech, the term articulation has reference to the fitting together of the structures of speech (tongue, palate, lips, jaw, teeth) in the production of the sounds of oral speech. Most defects of articulation involve omissions, substitutions or distortions of the consonants. For example, in saying "tee" for "teeth" there is an omission of the *th* sound; in saying "teef" for "teeth" there is a substitution of the *f* for the *th;* in saying the *th* in "teeth" in such a way that it resembles no other sound in the language there is a distortion. Any given sample of defective articulation may include a variety of the above types of errors. For example, in saying "I wef my tote at koo" for "I left my coat at school" there is a substitution of *w* for *l* in "left," an omission of the *t* in "left," a substitution of *t* for *c* in "coat" and an omission of the *s* and *l* in "school."

Inconsistency in the type of error made for a particular sound in different words is a characteristic of defective articulation.

There is great variability in the type and severity of articulatory disorders. Judgments of the severity of an articulatory defect are based primarily upon the intelligibility of the speech. A slight defect may involve a substitution or distortion of only one or two sounds resulting in speech which is deviant but intelligible. A severe defect may involve errors on many sounds resulting in speech which is partially or completely unintelligible.

Lisp.—This term is used to designate articulatory defects involving specific errors on one or more of the sounds *s, z, sh, ch,* or *j.*

Tongue-Tie.—A term often used by laymen in referring to defective speech. True tongue-tie results from shortness of the frenum, the attachment between the underside of the tongue and the floor of the mouth. Tongue-tie is relatively rare as a cause of articulatory defects. Sufficient shortness of the frenum to cause defective speech usually causes feeding problems in infancy and may be alleviated surgically before speech begins to develop. If the tongue can be protruded past the lower lip or elevated to touch the roof of the mouth when the jaw is partly open, there is no true tongue-tie.

Baby Talk.—A layman's term for defective speech which implies, often incorrectly, that the defect will be outgrown. There are no definite criteria by which any particular type of articulatory defect can be classified as baby talk. The term is seldom, if ever, used by professional persons.

Causes.—Causes cannot be determined for most cases of articulatory defects which have no apparent physical basis. There is a tendency to assume that since no organic cause is evident, the defect must have a psychological or emotional basis. Serious and prolonged emotional upset during the years when speech is learned might well interfere with the acquisition of normal speech, but there is no justification for the broad generalization that nonorganic defects of speech are the result of emotional disturbances. Other causes frequently suggested are lack of stimulation or motivation to speak, bilingual environment, poor speech models to imitate and poor auditory memory span. All of these should be viewed as possibly contributing to, rather than as causes of, articulatory defects.

Therapy.—Therapy for articulatory defects is usually successful. Proper teaching techniques plus adequate co-operation from the patient can result in complete correction. Generally, the later in life the correction is attempted, the more difficult it is to accomplish. Early correction also lessens the extent to which the defective speech may affect adversely psychological, social and educational development.

Stuttering.—The terms stuttering and stammering are usually

considered synonymous and the term stuttering is the most frequently used. It is estimated that approximately 1% of the total U.S. population or 10% of those with defects of speech are stutterers. Stuttering is characterized by spasmodic interruptions of the fluency of speech in the form of hesitations, blockings, prolongations and repetitions of words, syllables or sounds. These interruptions may be accompanied by excessive muscular tension, facial contortions, irregularities of breathing and various other accessory or secondary characteristics. Stuttering occurs in the form of clonic (repeating) or tonic (complete blocking) interruptions of speech. Both of these occur in the speech of most stutterers.

The severity of stuttering ranges from an occasional slightly noticeable interruption, to severe blocking with every attempt at speech. The severity of stuttering varies greatly from stutterer to stutterer and from situation to situation. It is said that no two stutterers stutter alike. When a stutterer talks to himself, to babies or to pets, or when he sings or reads aloud in chorus with one or more other persons, he is generally free of his difficulties. Stuttering has its onset in the preschool years. Most surveys indicate that there are three to eight times as many male as female stutterers. For some, stuttering may decrease or even disappear completely over a period of weeks, months or years with or without therapy. For others, stuttering may continue and even increase in severity both with and without therapy.

Causes.—While there is no definitive answer to the question of the cause of stuttering, theories and investigations concerning its cause have been abundant. Stuttering has been the object of more study and speculation than any other defect of speech. Theories regarding the cause of stuttering may be roughly classified into four groups.

The first group suggests that stuttering is physiologically caused. Among the factors studied in the search for physiological differences between stutterers and nonstutterers are heart rate, blood chemistry, allergies, metabolism, handedness and various neuromuscular abilities. The accumulated evidence has failed to establish or verify that there are physiological differences between stutterers and nonstutterers.

The second group suggests that stuttering stems from some psychological difference between stutterers and nonstutterers. Most such theories explain stuttering as a form of reaction to frustration or emotional conflict. Evidence indicates that while stutterers as a group do tend to have personality deviations, they do not have severe psychological problems. Investigation is complicated by the knowledge that the stuttering itself can cause psychological or personality problems. Psychological problems are probably better viewed as related to, rather than as causes of, stuttering.

The third group suggests that stuttering is the result of a combination of physiological and psychological factors. Such theories propose that there is some physiological predisposition or weakness which causes excessive nonfluency in speech (particularly in preschool years) but that true stuttering develops only when this nonfluency becomes the source of anxiety or frustration.

The fourth group categorically denies that stutterers are in any way different from nonstutterers except in speech. This group feels that most stuttering has its beginning in the hesitancies and repetitions that are characteristic of the normal speech of 3- or 4-year-old children. The primary causal factor is felt to be in the attitudes and reactions of the parents to the way the child speaks. It is reasoned that stuttering is induced by parents who are overly concerned or perfectionistic about nonfluency in the child's speech.

Probably the most defensible viewpoint is that stuttering may result from various single or combined causes. The principle of multiple causation appears to be the most accepted among professional speech therapists.

Therapy.—There is considerably more agreement in the approach to therapy for stutterers than there is concerning its cause. Most therapy for stutterers, in one way or another, involves: (1) convincing the stutterer to accept objectively his stuttering and to place it in a proper perspective in his life; and (2) teaching the stutterer to speak in a more controlled, less deviant fashion. While therapy for stutterers is reasonably successful, speech therapists probably experience more failure and frustration in work with stuttering than in work with other disorders of speech. Generally, the younger the stutterer when therapy is started, the greater the chances of success. Some stutterers overcome their difficulty without therapy; others continue to stutter even after years of therapy. Many adult stutterers have made the rounds of therapies and therapists without significant or lasting success.

DISORDERS OF SPEECH WITH PHYSICAL CAUSES

Any anatomical or physiological defect affecting the physical mechanism may result in defective speech. There are certain physical inadequacies which because of frequency of occurrence and direct importance to speech are usually considered the more common organic causes of defective speech.

Impaired Hearing

Hearing is important in learning and maintaining normal speech. In order to develop speech normally the infant must be able to hear and imitate the speech around him. In order to maintain normal speech one must be able to monitor his own speech through hearing. Any deficiency in hearing may be reflected in a failure to learn or a failure to maintain normal speech. The effects that a hearing loss may have on speech are dependent upon (1) the severity and type of the loss, (2) the age of the onset of the loss and (3) whether the loss affects one or both ears. (*See* also Deaf and Hard of Hearing, Training and Welfare of; Deafness and Impaired Hearing.)

The Deaf.—The term deaf refers to those for whom hearing is completely inadequate for the activities of life in which hearing is important—particularly oral communication. For the deaf, the primary mode of communication is visual supplemented by auditory. Some persons deaf from birth or infancy may not be able to communicate at all through speech. Such persons usually communicate through the language of signs, manual alphabet, reading and writing. Many persons deaf from birth or infancy are taught to speak and lip read. The speech of such persons is distorted and may even be difficult to understand. When deafness does not occur until after speech has developed the speech may be fully intelligible, often near normal in many respects, even when hearing is functionally inadequate. The speech of deaf persons is usually characterized by (1) a lack of natural vocal volume, quality, pitch and inflection; (2) a lack of natural rhythm and stress patterns; and (3) distortions and omissions of the sounds of speech, both vowels and consonants.

The Hard of Hearing.—For the hard of hearing the primary mode of communication is auditory supplemented by visual. Some hard of hearing persons may have speech which is completely normal. Others may have speech which differs from the normal but without significant effect on intelligibility. When the speech of the hard of hearing person is noticeably different from the normal it is usually characterized by monotone voice and deviant production of certain consonants, particularly *s, sh, ch, j, t,* and *k*.

Defective Oral Structure

Any defect in the tongue, lips, jaw or palate as well as various dental abnormalities (malocclusions, missing teeth, poorly spaced or aligned teeth) may cause defective speech. All such defects of oral structures do not cause defective speech, however, for humans are remarkably capable of compensating for deviant structure in achieving normal speech. Speech may be normal in spite of serious structural defects although in some instances slight defects may cause defective speech.

Cleft Lip and Palate.—Clefts of the palate or of the lip and palate are the most serious of all defects of oral structure. In the U.S. one child in every 650 to 800 is born with cleft of the lip, palate or both.

Clefts of the lip and palate are congenital deformities which result from the failure of certain tissues of the face to join during the first three months of prenatal development. Clefts of this type may occur in the lip and palate. Clefts of the lip are

located on the upper lip continuous with the nostril on either or both sides. Clefts of the palate occur anywhere along the midline of the soft and hard palate to one or both sides of the premaxilla (a small forward section of the upper jaw where the four central teeth are located). About 50% of oral clefts include both lip and palate while the remaining 50% are equally divided between clefts of lip only and clefts of palate only.

Effect on Speech.—In cases of untreated cleft palate, speech always will be seriously impaired and probably unintelligible. Frequently in cases of treated (closed) cleft palate, speech is defective and in serious cases may be unintelligible.

Clefts of the palate are of much greater seriousness than clefts of the lip. Clefts of the lip should affect little more than the *p*, *b*, and *m* sounds which require bilabial closure. When the palate is cleft, the oral structure is seriously inadequate for the production of speech. At least 40% of the sounds of the English language require the generation of air pressure within the oral cavity and its controlled release through the oral cavity. In order to build up such pressures within the oral cavity, the person with normal structures closes the soft palate against the pharyngeal wall (back of the throat) thus preventing air or sound from escaping into the nasal cavity. The individual with an unclosed cleft of the palate or with a closed but short or poorly functioning palate cannot prevent air or sound from escaping into the nasal cavity and consequently all sounds which require the generation of pressure in the oral cavity are defective. The unclosed, or inadequately closed, palate also creates excessive nasal resonance causing additional distortion of speech.

Causes.—Whatever causes clefts of the lip and palate occurs before the end of the third month of pregnancy. Among the causes most frequently suggested are heredity, nutritional or metabolic factors and mechanical causes such as intrauterine crowding. The possibility of clefts being inherited frequently causes parents and relatives of persons with cleft lip or palate to fear repeated occurrence through the family strain. While the familial histories of persons with clefts show a slightly greater incidence of clefts than do the familial histories of persons without clefts, the majority of cases occur without clear evidence of any hereditary factor. In experimental studies of mice and rats, conditions resembling cleft palate have been produced in the offspring by manipulations of nutrition and metabolism in the parent. Similar experiments are not possible with humans. The actual cause or causes of cleft lip and palate remain unknown, for no cause so far proposed has been supported by conclusive evidence.

Treatment.—Successful treatment for the person with cleft lip and palate requires the integrated efforts of medical, surgical, dental and speech specialists. The goals of treatment for cleft lip and palate are (1) to create as near normal appearance as possible and (2) to create as near normal function as possible, particularly for speech.

Plastic surgery is the only procedure used for closing clefts in the lip and the most frequently used procedure for closing clefts in the palate. The lip usually is closed before the child is six weeks old, often during the first few hours of life. Palatal surgery usually is performed when the child is 18 to 36 months of age. Great progress has been made in developing surgical techniques which accomplish the goals of treatment. The problems of growth retardation in the palate following surgery and of the frequent failure of the surgically closed structures to function adequately for speech are of great concern and are the focus of continuous research and evaluation.

Another approach to the closing of clefts of the palate is through the use of a prosthetic (artificial) appliance that is designed to replace or supplement the missing or inadequate structures. The design and execution of these prosthetic appliances are carried out by dental specialists. Before anesthesia made possible the development of surgical techniques the artificial palate or obturator was widely used. The greatest problem encountered in the prosthetic approach to closing the palate was the need to design a device which would enable the wearer to have functional control over the opening and closing of the posterior passage into the nasal cavity. Various designs of obturators with hinged and movable sections were unsuccessful. The design which finally proved successful uses a stationary appliance with an extension reaching well back into the mouth, nearly touching the pharyngeal wall. The opening and closing of the posterior passage into the nasal cavity is made possible by the action of muscles in the pharynx (upper throat). When these muscles contract, the tissues of the pharynx close around the extended portion of the prosthesis and prevent air from escaping into the nasal cavity. When the pharynx is relaxed, there is space for the passage of air through the posterior opening of the nasal cavity. This type of appliance is often referred to as a "speech-aid," for it makes possible the functional control essential for speech. The dental speech-aid may be used instead of, or as a supplement to, surgery for palatal closure. The prosthetic approach is most frequently used to supplement unsuccessful or inadequate surgical results.

In addition to the design and construction of the speech-aid, dental specialists contribute importantly to the over-all appearance and function of oral structures by replacing missing teeth, supplementing portions of dental or facial structure where growth has been retarded, and by straightening and aligning teeth.

Speech Therapy.—Many persons with cleft palate need speech therapy; others do not. Surgery or prosthesis alone enable some to achieve normal speech. Others never achieve normal speech even after intensive speech therapy. Generally, the more functionally adequate the structure with which the child learns to speak, the more normal will be his speech. Speech will be increasingly defective if the structures with which speech is learned are inadequate. When defective anatomical structure causes faulty habits of speech to develop, such habits will usually persist even after treatment makes structures more adequate. In such cases speech therapy is necessary in order to alter habits and to make the best possible use of the new structure. Generally, regardless of age or palatal sufficiency, some improvement can result from speech therapy. When therapy is indicated, the earlier in life it is begun the better the result is likely to be.

Defective Laryngeal Structure

Defects in the larynx, particularly of the vocal cords may interfere with the vibration of the cords and result in unsatisfactory vocal tone. The vocal tone can be defective while the articulation of all vowels and consonants is correct. Such defects are perhaps better designated as disorders of voice than as disorders of speech, but if such disorders meet the criteria in the definition of defective speech stated earlier then the resultant speech is defective.

The Larynx.—Two common but temporary forms of vocal defect are laryngitis which accompanies viritic inflammation of the larynx and the hoarseness which follows the vocal abuse associated with excessive shouting. Persistent hoarseness, raspiness or weakness of voice calls for careful medical examination, for such defects are often the warning signs of serious conditions calling for immediate medical attention.

The Vocal Cords.—Paralysis of one or both vocal cords may result in weakness or absence of voice depending upon the severity of the paralysis. Such paralysis can be detected only by medical examination. When there is paralysis involving the vocal cords, the resultant vocal tone may range from a mere whisper in severe cases to a breathy, "stage whisper" tone in less severe cases. Occasionally voice therapy seems to restore function to apparently paralyzed cords. Usually, however, voice therapy can result only in some strengthening of tone without full restoration of former vocal function.

Various growths on the vocal cords themselves can cause a hoarse, harsh or raspy voice. They include benign and malignant tumours, local infections and nodules (localized tissue thickenings resulting from irritation). Nodule type growths on the vocal cords are often the result of poor use of the vocal mechanism and occasionally disappear after voice therapy. When nodules or other growths are surgically removed, postoperative voice therapy is frequently indicated. As long as the vocal structures retain essentially normal function, the results of voice therapy can be good.

Occasionally it is necessary for one vocal cord to be removed surgically. The resulting voice usually has both a weak, whispered quality and an uneven fluttering or gargling tone. Voice therapy for such cases can increase the strength of the voice and can reduce vocal fatigue but cannot restore the voice to its preoperative level of function.

Laryngectomy.—The removal of the larynx results in a very severe anatomic deficiency for speech. The effect of the laryngectomy is to leave the laryngectomized individual completely without speech. The ability to articulate all sounds is still intact, but the removal of the mechanism for producing vocal sound means that the individual's articulation will be silent.

There are a number of additional physical changes which follow laryngectomy. After the operation the patient breathes through an opening in the throat; inhaled or exhaled air no longer passes through the nose or mouth. The sense of smell is thus quite ineffective; foods may seem to have less taste; the activities of smoking, sneezing and coughing are greatly changed. Of all these various problems, however, the inability to speak is by far the most serious.

As far as communication is concerned, the patient who has had a laryngectomy has three alternatives available to him. The first, and the least satisfactory, is the use of a pad and pencil. The second is the use of an artificial larynx. The third is the development of esophageal speech.

The pad and pencil is at best clumsy and slow, and its use is limited to only the most essential messages. In addition, the laryngectomized person often finds when he uses a pad and pencil to communicate with strangers, that they will either write out their replies to him or shout at him as if they questioned his hearing.

The artificial larynx is functionally a very adequate device. The most widely used variety of artificial larynx is a small cylinder, 3 to 4 in. long and an inch in diameter, containing an arrangement of reeds which can be set into vibration to produce a voicelike sound. There is a 4- to 6-in. tube extending from each end of the cylinder. The cylinder is held so that one tube is placed over the opening in the throat, the other tube is placed in the mouth. The individual then exhales air from the lungs, into the tube held over the opening in the throat and through the cylinder, causing the reeds to vibrate. The sound from the vibrating reeds then continues on into the mouth where it can be articulated into words. There are also electrically powered devices which operate on the same principle. The use of the artificial larynx can be learned quickly and little practice is necessary to achieve adequate speech. Many laryngectomized individuals dislike the artificial larynx because it calls attention to itself; because of the inconvenience of having the device handy; and because of the necessity of having one hand occupied in its use. Others use the artificial larynx happily and effectively.

Esophageal voice is in many respects the most desirable of the alternatives available to the laryngectomized person. Esophageal voice is produced by trapping air in the esophagus and then forcing that air out through the contracted opening at the upper end of the esophagus. The resultant belchlike sound may then be articulated into words.

Some persons are able to master esophageal voice quickly, others have considerable difficulty in learning the technique, and some never learn to produce a functionally adequate esophageal voice. Often laryngectomized persons who have learned to use esophageal voice can do a capable job of teaching the technique to others. In the larger metropolitan areas, clubs of laryngectomized persons are organized, primarily for rehabilitative purposes.

Inadequate Function of The Central Nervous System

Aphasia.—One of the most serious problems of speech and language resulting from inadequate function of the central nervous system is aphasia. The term aphasia designates certain losses in ability to use speech and language which follow injury or disturbance in the brain. The most common causes of aphasia are head injuries, strokes, cerebral hemorrhage, cerebral thrombosis (blood clots) and brain tumours. Of course such conditions also may result in visual disturbances, hearing difficulties or muscular paralysis.

Aphasia is a deficit in the ability to use or understand language and is the result of an injury or disturbance in the brain rather than the result of a defect in the organs of speech, defect in hearing, defect in intelligence or severe emotional disturbance.

In aphasia the disturbances of ability to use language may involve any combination of partial or complete losses of the skills of speaking, writing, reading or understanding of speech. Aphasia may range in severity from slight difficulty with an occasional word to a complete inability to use or comprehend written or spoken language. There is an almost infinite variety of actual manifestations of aphasia. Some aphasics may be able to converse about objects or persons and yet be unable to name the objects or persons when asked (nominal aphasia); some cannot speak at all except for a pattern of syllables, a word or phrase which is repeated with each attempt to speak (motor aphasia). Those who talk in a free flow of corrupted speech and language with words coming in disconnected or meaningless sequences, often interspersed with nonsense jargon, suffer from syntactical aphasia. Others talk in a fragmentary or telegraphic style. Often the aphasic may be able to repeat after someone else words which he cannot initiate spontaneously, and occasionally he can read aloud or write certain words which he cannot say. Some aphasics recognize their errors; others do not. Frequently automatic sequences such as counting, reciting the alphabet, swearing or singing familiar songs can be accomplished when more communicative speech is impossible. All of the disabilities just described are accompanied by varying degrees of disturbance of reading, writing and understanding speech.

Whatever the manifestations of aphasia are in a particular case, the difficulty must be viewed as a disability in the use of speech or language and not as an intellectual defect. The fact that the aphasic cannot say the names of common objects does not mean that he has forgotten or lost the names or that he does not know the uses or functions of the objects. The fact that he may be unable to say the names of friends or relatives does not mean he no longer recognizes or knows these persons. His deficiency is primarily an inability to communicate through speech and language about the things he knows.

Aphasia has stimulated interest in the nature of the function of the brain. Although there are many unsettled issues concerning aphasia and the function of the brain, there is agreement on a few points. Certain parts of the brain are considered to be somewhat more critical for certain skills of language than are other parts of the brain. Disturbances in the left hemisphere of the brain are more likely to produce aphasia than are disturbances in the right hemisphere. Disturbances in different parts of the brain may produce similar effects, while disturbances in similar parts of the brain may produce different effects. Seldom, if ever, is an aphasic disturbance limited to only one skill or aspect of language. The impairment of one skill may be much greater than the impairment of others, and some skills will often appear to be unimpaired. It usually is found, however, that when one aspect of language is impaired, all other aspects are impaired to some extent.

Therapy.—Since aphasia may be accompanied by other disabilities the total rehabilitation program may include physiotherapy, occupational therapy, vocational guidance and speech therapy. The aphasic individual is usually most concerned about his difficulties in communication. Probably no single aspect of the total rehabilitation of the aphasic adult is more important than speech therapy.

For most cases of aphasia, there is a period during which spontaneous recovery of impaired skills and abilities takes place. For some, this spontaneous recovery may be nearly complete; for others, it may be so slight as to be quite insignificant. Speech therapy for aphasics proves to be of considerable success. Rarely is it possible to restore the patient to his premorbid level of ability in speech and language. It is possible, however, in nearly all cases to make definite improvement. Improvement may range from small gains in the aphasic's ability to communicate with

family and friends, to nearly complete recovery of skill in speech and language.

Congenital Aphasia.—Most writings on aphasia deal with aphasia in adults and many definitions of aphasia refer to it as an acquired disability or as a "loss of" ability to use language. There are among children, however, certain failures in the development of language, which can be properly classified as congenital aphasias. The condition of congenital aphasia in children and acquired aphasia in adults may be described in the same manner. The differences between the manifestations of congenital and acquired aphasia are probably the result of the differences in the time of onset of the disorder.

Congenital aphasias fall into two major categories: (1) motor or expressive aphasia and (2) sensory or receptive aphasia. The motor or expressive aphasic child cannot speak at all although he hears and understands speech, has good intelligence and good muscular ability. He is unable to reproduce the patterns and sequences of sounds which are meaningful to him when spoken by others. He knows the words he wants to say but he cannot negotiate them in motor speech. Generally the vocalizations of such children are limited to a pattern of one or two syllables which is repeated whenever there is an attempt to speak. The motor aphasic child uses gestures effectively to convey meanings to others.

The sensory or receptive aphasic child has no speech and no understanding of speech although he is intelligent and has sufficient hearing to have learned speech. Often these children are mistakenly thought to be deaf or mentally retarded. Some receptive aphasic children are difficult to differentiate from deaf children, for they do not respond to speech and often ignore sound. Others are described as "word deaf" because words which they can hear and repeat have no meaning for them. Receptive aphasic children are usually able to communicate through gestures.

Therapy.—Aphasic children by mistake may often be placed in schools for the mentally retarded or in schools for the deaf. Placement in schools for the retarded is unfortunate because full use is never made of the child's intelligence. Placement in schools for the deaf is usually unsatisfactory for aphasic children because the teaching procedures are inappropriate and the child learns poorly if at all. The aphasic child has a good future if his condition is correctly recognized early and appropriate education is available.

Cerebral Palsy.—Cerebral palsy is a general term which includes various types of defective muscular control resulting from disturbances in the central nervous system. The lack of adequate control of muscles can affect speech if the poor control involves the muscles used in speaking. Cerebral palsy is differentiated into several specific types on the basis of the manner in which muscular control is deficient. The spastic and athetoid types of muscular involvement make up 80% to 90% of the total group classed as cerebral palsied. The spastic type is characterized by fairly stereotyped postures and positions. In spasticity the limbs are generally in flexed positions. The athetoid (wormlike) type is characterized by a constant, jerky muscular activity without rhythm or pattern. The incidence of cerebral palsy in the U.S. is estimated at about 2 per 1,000. Cerebral palsy is generally congenital and most cases appear to result from birth injuries.

Effects on Speech.—Spasticity may involve all or part of the body. Defective speech may not occur when the spasticity is slight or when it does not involve the muscles used in speech. About 50% of all spastics have defective speech. The athetoid condition almost always involves the whole body and consequently speech is defective in nearly 90% of the cases. About 70% of all those with cerebral palsy have defective speech. Speech may be normal, slightly defective or completely lacking. The speech defects of the cerebral palsied may involve any or all of the following: (1) slow, laboured and uneven rate, (2) clumsy or slurred articulation, (3) poorly controlled voice pitch and volume and (4) irregular and poorly timed breathing.

Therapy.—Physical therapy and speech therapy for the cerebral palsied are both based upon similar principles: (1) to inhibit the involuntary muscular tensions and activity, (2) to teach specificity of muscular control, (3) to develop patterns and sequences of controlled muscular acts and (4) to improve the accuracy and speed of sequences of muscular patterns. Most cases can be improved with therapy but ultimate prognosis is an individual matter. The muscular difficulties of cerebral palsy can never be cured nor overcome. Therapy for the cerebral palsied aims to teach the individual to function as effectively as possible within the limits imposed by the handicap of muscular control.

BIBLIOGRAPHY.—Charles Van Riper, *Speech Correction,* 3rd ed. (1956); M. F. Berry and John Eisenson, *Speech Disorders* (1956); Wendell Johnson *et al., Speech Handicapped School Children,* rev. ed. (1956); R. West, M. Ansberry and A. Carr, *The Rehabilitation of Speech,* 3rd ed. (1957); L. E. Travis (ed.), *Handbook of Speech Pathology* (1957); E. F. Hahn, *Stuttering,* 2nd ed. (1956); Hallowell Davis (ed.), *Hearing and Deafness* (1947); M. E. Morley, *Cleft Palate and Speech,* 3rd ed. (1954); W. G. Holdsworth, *Cleft Lip and Palate,* 2nd ed. (1957); M. C. L. Greene, *The Voice and Its Disorders* (1957); J. L. Bangs *et al.,* "Speech after Laryngectomy," *J. Speech Disorders,* vol. ii, pp. 171–176 (1946); Henry Head, *Aphasia and Kindred Disorders of Speech,* vol. i (1926); T. H. Weisenberg and K. E. McBride, *Aphasia* (1935); M. C. Longerich and J. Bordeaux, *Aphasia Therapeutics* (1954); M. A. McGinnis, F. R. Kleffner and R. Goldstein, "Teaching Aphasic Children," *Volta Rev.,* vol. 8, pp. 239–244 (1956); J. F. Pohl, *Cerebral Palsy* (1950); W. M. Cruickshank and G. M. Raus (eds.) *Cerebral Palsy* (1955). (F. R. KR.)

SPEED, JOHN (1552–1629), English historian and cartographer, settled in Moorfields, where he was enabled to devote himself to antiquarian pursuits through the kindness of Sir Fulke Greville. His important works are: *Theatre of the Empire of Great Britaine* (1611), a series of 54 maps (with descriptive matter) of different parts of England, which had already appeared separately, and in which he was helped by Christopher Saxton, John Norden and William White, and *History of Great Britaine under the Conquests of the Romans . . . to . . . King James* (1611). Speed brought some historical skill to bear on the arrangement of his history; in preparing it he received help from Sir Robert Cotton, William Smith, Henry Barkham and Sir Henry Spelman. Although he repeated many of the errors of older chroniclers he added on the whole a substantial contribution of valuable material for the history of his country. He died in London on July 28, 1629.

Other maps of his, beside those in the *Theatre,* are in the British museum. Another edition of the *Theatre* is *Theatrum Magnae Britanniae latine,* redditum a P. Holland (London, folio, 1616). He wrote *Genealogies Recorded in Sacred Scriptures* (1611), and a similar work, *A Cloud of Witnesses* (1616).

SPEEDOMETER, an instrument that indicates the speed of a vehicle; usually, it is combined with an odometer, which records the distance traveled. The speed-indicating mechanism of the speedometer is actuated by a permanent magnet rotated at a speed of 1,000 revolutions per mile of vehicle travel by a flexible shaft driven by gears at the rear of the transmission. The magnet turns within a "speed cup" made of a light, nonmagnetic metal such as aluminum, and the magnetic circuit is completed by a circular stationary field plate surrounding the speed cup.

The rotating magnetic field passing through the metal of the speed cup induces eddy currents that increase with the magnet speed and set up a proportional magnetic field attracting the speed cup to the rotating magnet, tending to rotate it against the restraint of a coiled hairspring. The torque acting on the speed cup is proportional to the eddy currents and thus to the rotational speed of the magnet. The speed cup turns until the restraining torque of the hairspring, increasing with the angle through which the cup turns, equals the applied torque. The angular position at which the speed cup comes to rest varies directly with the rotational speed of the magnet and hence with the speed of the vehicle. The speed cup may have a scale of miles (or kilometres) per hour on its periphery that is visible through a small window, or it may have a pointer attached to it that moves across a graduated scale.

The odometer, which registers the distance traveled by the vehicle, consists of a set of figure wheels, usually from four to six, arranged along a cross shaft and turned by a series of star pinion gears located inside the drum formed by the figure wheels. A train of intermediate gears connects the figure wheel at the right to the magnet shaft at a gear ratio of 1,000:1, causing it to complete

one turn per mile. Its numerals thus indicate tenths of a mile. The figure wheels are so geared together that, as one wheel completes a revolution, it turns the figure wheel next to the left one-tenth of a turn, thus summing up the miles traveled. A "trip odometer" is sometimes included that is provided with a reset mechanism to return it to zero. (O. C. C.)

SPEEDWELL, the common name applied to the herbaceous species of the genus *Veronica* (snapdragon or figwort family, Scrophulariaceae), which contains about 250 species, temperate, alpine and arctic; many species are shrubs or trees with handsome spikes of flowers, and are often cultivated. About 15 speedwells occur in North America, including several that have been naturalized from the old world. Culver's root (*Veronicastrum virginicum* or *Veronica virginianum*), found across the continent, was once used in medicine. The woody species (some are transferred to the genus *Hebes* by some botanists), rarely known as speedwell, are especially characteristic of the mountains of New Zealand.

SPEKE, JOHN HANNING (1827–1864), English explorer who solved the mystery of the Nile sources by his discovery of Lake Victoria, was the first European to visit Uganda. He was born May 3, 1827, at Orleigh court, Bideford, Devon. Commissioned in the Indian army in 1844, he served in the Punjab and traveled in the Himalaya and Tibet. In 1854 he joined an expedition to explore Somaliland under the command of Richard Burton (*q.v.*). This venture was wrecked at the outset in an attack by Somalis outside Berbera in April 1855 during which Speke was severely wounded. In 1856 Speke joined Burton on a Royal Geographical society expedition to investigate the great lake said to lie at the heart of Africa and to be the source of the Nile. They arrived in Zanzibar in Dec. 1856 and left for the interior from Bagamoyo on June 27. They reached Tabora (or Kazé) by the Arab trade route on Nov. 7 where they learned of the existence of three great lakes (Nyasa, Tanganyika, Victoria) and although in ill-health they pressed on to Ujiji on Lake Tanganyika where they arrived two months later. Back at Tabora in June, Speke struck north alone to find the lake he guessed was the source of the Nile, and this he reached and named Victoria on July 30. Speke returned to England first and announced his discovery of the Nile source; his conclusions were disputed later, and Tanganyika suggested as the source. Sir Roderick Murchison, president of the Royal Geographical society, supported Speke who received the Founder's medal. He was invited to lead another expedition to confirm his findings, and chose James Grant to accompany him. Leaving Zanzibar in Sept. 1860 they reached the southwestern shore of Lake Victoria, via Tabora, a year later. Speke pressed on ahead up the west shore into Buganda arriving at King Mtesa's capital (near Kampala) on Feb. 19, 1862. He marked the Nile's exit from the lake by naming the Ripon falls on July 28. Speke and Grant then tried to follow the course of the Nile but were checked by civil war in Bunyoro. They reached Gondokoro on Feb. 15, 1863, and met (Sir) Samuel and Florence Baker to whom Speke gave directions for finding the lake in Bunyoro through which he guessed the Nile must flow. He described his adventures in *Journal of the discovery of the Source of the Nile* (1863) and *What led to the discovery of the Source of the Nile* (1864).

On his return to England, Speke's claims were again challenged. A discussion between him and Burton was arranged to take place at the British association at Bath in Sept. 1864, but while the audience waited news was brought that Speke had been found killed accidentally by his own gun while shooting at Neston park, near Corsham, Wiltshire, on Sept. 15, 1864.

See also Index references under "Speke, John Hanning" in the Index volume.

BIBLIOGRAPHY.—R. F. Burton, *First Footsteps in East Africa* (1856), *The Lake Regions of Central Africa* (1860), (with J. MacQueen), *The Nile Basin* (1864); J. A. Grant, *A walk across Africa* (1865); S. W. Baker, *The Albert N'yanza, Great Basin of the Nile* (1866); H. H. Johnston, *The Nile Quest* (1903); D. Middleton, *Baker of the Nile* (1949); J. N. L. Baker, "Sir Richard Burton and the Nile Sources," *English Historical Review* (Jan. 1944). (D. MN.)

SPELMAN, SIR HENRY (*c.* 1564–1641), English antiquary and historian whose *Concilia, Decreta, Leges, Constitutiones, in re Ecclesiarum orbis Britannici* was the first systematic compilation of church documents, was born about 1564 at Congham, Norfolk. Educated at Trinity college, Cambridge, and Lincoln's Inn, London, he was M.P. for Castle Rising in 1593 and 1597, sheriff of Norfolk in 1604, M.P. for Worcester in 1625 and served on several royal commissions. A student of antiquity, he belonged, with William Camden and Sir Robert Cotton, to the Society of Antiquaries, which he attempted to revive in 1614, and he endowed a lectureship in Anglo-Saxon at Cambridge. He died in London in 1641.

Spelman was the author of ecclesiastical and legal histories, most of which were published posthumously. His first religious work, *De non temerandis ecclesiis; a Tracte Of the Rights and Respect Due Unto Churches* (1613) was the foundation of his well-known *History and Fate of Sacrilege* (1698); but his chief work was his history of the church, *Concilia, Decreta, Leges, Constitutiones, in re Ecclesiarum orbis Britannici* (vol. 1, 1639; vol. 2, ed. by C. Spelman, 1664). Among his polemical writings were *De Sepultura* (1641) and *Tithes Too Hot to be Touched* (1646), and he also made a glossary of ancient legal terms, *Archaeologus* (part 1, 1626; part 2, ed. by W. Dugdale, 1664). (RD. C. G.)

SPEMANN, HANS (1869–1941), German embryologist and Nobel laureate, was born at Stuttgart on June 27, 1869. In 1891 he began the study of medicine and attended the universities of Heidelberg, Munich and Würzburg, but finally graduated in zoology, botany and physics. From 1894 to 1908 he worked at the Zoological institute of Würzburg and was then appointed professor of zoology and comparative anatomy at Rostock. In 1914 he became director of the Kaiser Wilhelm Institute for Biology in Berlin and from 1919 until his retirement in 1935 he held the chair of zoology at Freiburg.

Spemann was awarded the Nobel prize for physiology and medicine in 1935 for the discovery of what is known as the "organizer effect" in embryonic development (*see* EMBRYOLOGY, EXPERIMENTAL). As Spemann himself stated, "the beginnings of a human individual do not differ essentially from those of an echinoderm." In actual fact his concept of organizers was based upon a lifetime of research into the early development of the common newt. Much of his success was due to his intense concentration on this seemingly narrow field; but apart from this he had the controlled imagination of the true scientist and was a master of the techniques of microsurgery. His great achievement was to show that different regions of the embryo have not the same value for development. Within certain limits the fate of a given cell depends not upon itself but upon an influence emanating from an "organizing centre." Spemann's experiments showed that, in the earliest stages, the fate of the embryonic parts has not been determined; thus if a piece of presumptive skin tissue is excised and transplanted into an area of presumptive nervous tissue it will form nervous tissue, not skin. At a later stage the fate of various regions has been determined. Thus a presumptive eye region, if placed in the abdominal wall, becomes an eye, not skin or muscle. It appears that the organizing stimulus is chemical in nature, but it defies analysis. These researches threw new light not only on the normal processes of development and growth, but also on the genesis of various congenital abnormalities of the human body. Spemann gave a masterly summary of his researches in his Croonian lecture before the Royal society in 1927 and in his Silliman lectures at Yale in 1935.

Spemann died at Freiburg-im-Breisgau on Sept. 12, 1941.

(W. J. BP.)

SPENBOROUGH, an urban district in the West Riding of Yorkshire, Eng., 6 mi. S.E. of Bradford. Pop. (1961) 36,412. Area 12.9 sq.mi. It is a manufacturing district in the Spen valley. The textile industry is the chief one, including carpets, rayon and woolen materials. The manufacture of machinery and sheet-metal goods is important. The urban district was formed in 1915 and combined the three former urban districts of Cleckheaton, Gomersal and Liversedge. In 1937 its area was enlarged by the addition of the urban districts of Birkenshaw and Hunsworth and the rural district of Hartshead and part of Clifton. Brighouse and Spenborough were formed into a parliamentary borough in 1948, returning one member.

SPENCER, ANNA GARLIN (1851–1931), U.S. clergywoman, was born in Attleboro, Mass., on April 17. She early discovered the power and magnetism of her speaking voice and while on the staff of the *Providence Journal* (1868–78) lectured on social problems. After her marriage (1878) to the Rev. William H. Spencer a Unitarian minister, her husband encouraged her to preach in liberal pulpits. In 1891 she was ordained as minister of the Bell Street chapel (Independent), Providence, R.I., where she officiated for 14 years. From the beginning interested in women's rights, Mrs. Spencer participated actively in the Rhode Island Women's Suffrage association and in the national women's rights conventions, where she emphasized the contributions women could make to society, the theme of her book, *Woman's Share in Social Culture* (1913). Her own contributions were of wide scope in local and national charities and reforms. She was professor of sociology and ethics at the Meadville (Pa.) Theological school (1913–18) and lecturer at the University of Chicago (1918) and thereafter occasionally at Teachers college, Columbia university. Besides many articles and pamphlets, she wrote several books, among them the *History of the Bell Street Chapel Movement* (1903) and *The Family and Its Members* (1923). While director of the Family Relations division of the American Social Hygiene association, she died on Feb. 12, 1931. (L. A. G.)

SPENCER, HERBERT (1820–1903), English philosopher, achieved an influential synthesis of knowledge, advocating the preeminence of the individual over society and of science over religion. He was born in Derby on April 27, 1820. His father, William George Spencer, was a schoolmaster, and his parents' dissenting religious convictions inspired Herbert with that spirit of nonconformity which continued active in him after he had abandoned the Christian faith. He declined an offer from his uncle, the Rev. Thomas Spencer, to send him to Cambridge, and his higher education was largely the result of his own reading, which was chiefly in the natural sciences. After a few months as a schoolteacher he became a railway engineer in 1837 and worked on the Birmingham and Gloucester railway until its completion in 1841.

In 1842 he contributed some letters (republished as a pamphlet, *The Proper Sphere of Government*, 1843), to the *Nonconformist*, in which he argued that it is the business of governments to uphold natural rights and that they do more harm than good when they go beyond this. After some association with progressive journalism through such papers as the *Zoist* (devoted to mesmerism and phrenology) and the *Pilot* (the organ of the Complete Suffrage union), Spencer became in 1848 a subeditor of the *Economist*. In 1851 he published *Social Statics*, which contains in embryo most of his later views. He here argues in favour of an extreme form of economic and social *laissez faire* and describes progress as "not an accident but a necessity." About 1850 Spencer made the acquaintance of G. H. Lewes and Marian Evans (George Eliot). His philosophical conversations with the latter led some of their friends to expect that they would marry; but in his *Autobiography* (1904) Spencer says that there was never any such desire on his part, much as he admired Miss Evans' intellectual powers. Other friends were T. H. Huxley and J. S. Mill. In 1853 Spencer received a legacy from his uncle and, relying on this, resigned his position with the *Economist*.

Having published the first part of *The Principles of Psychology* in 1855, Spencer in 1860 issued a prospectus and obtained promises of subscriptions for a comprehensive work, *The Synthetic Philosophy*, which was to include, besides the already published *Principles of Psychology*, volumes on first principles and on biology, sociology and morality. *First Principles* was published in 1862, and between then and 1896, when the third volume of the *Principles of Sociology* appeared, the task was completed. In order to prepare the ground for *The Principles of Sociology*, Spencer, in collaboration with David Duncan and others, started in 1873 a series of works called *Descriptive Sociology*, in which information was provided about the social institutions of various societies, both primitive and civilized. The series was interrupted in 1881 because of lack of public support. Spencer was a close friend of Richard Potter, chairman of the Great Western Railway company, and thus became a friend and adviser of his children, one of whom was Beatrice Potter (who later became Beatrice Webb). She frequently visited Spencer during his last illness and left a sympathetic but sad record of his last years in *My Apprenticeship* (1926). Spencer died on Dec. 8, 1903, at Brighton, leaving a will by which trustees were set up to complete the publication of the *Descriptive Sociology*. Further volumes were written by such authorities as Sir W. Flinders Petrie and J. P. Mahaffy, so that the series eventually amounted to 19 parts (1873–1934).

Spencer was one of the most argumentative and most discussed English thinkers of the late Victorian period. He vigorously advocated the scientific or naturalistic view of the world against supernaturalism. He was, too, the prophet of evolution and of progress. As a liberal, he upheld the doctrine of *laissez faire* in a most uncompromising form. As a student of society, he urged the importance of examining social phenomena in a scientific way and endeavoured to do this himself. All these aspects of his thought were believed by him to form a coherent and closely ordered system. Science and philosophy, he held, gave support to individualism and progress. It is natural to cite him, therefore, as the great exponent of Victorian optimism, but it should also be noticed that he was by no means unaffected by the pessimism which from time to time clouded the Victorian confidence. Evolution, he taught, would be followed by dissolution, and individualism would come into its own only after an era of socialism and war.

The Synthetic Philosophy in Outline.—According to Spencer, philosophy is a synthesis of the fundamental principles of the special sciences, a sort of scientific *summa* to replace the theological systems of the middle ages. "Science is *partially-unified* knowledge; philosophy is *completely-unified* knowledge" (*First Principles*, 37). Spencer thought of unification in terms of development, and his whole scheme was suggested to him by the evolution of biological species. In an article entitled "The Development Hypothesis" (1852–54) he rejected the notion of special creation, and in "Progress: its Law and Cause" (1857) he applied the idea, borrowed from K. S. von Baer, that biological development in the individual is from the homogeneous to the heterogeneous to the evolution of the solar system, of animal species, of human society and of industry, art, language and science. It should be noticed that Spencer published his idea of the evolution of biological species before the views of Charles Darwin and Russell Wallace were made known, but that Spencer thought at that time that this evolution was caused by the inheritance of acquired abilities whereas Darwin and Wallace attributed it to natural selection. Spencer later accepted the theory that natural selection was one of the causes of biological evolution and himself coined the phrase "survival of the fittest" (*Principles of Biology* [1864], vol. i, p. 444). But he continued to hold that alterations caused in the organism of a particular animal by forces external to it may be passed on by it to its descendants (pp. 269–70). Furthermore, he did not think that the inheritance of acquired characteristics or the survival of the fittest were ultimate principles of evolution. In his view the principle of change from homogeneity to heterogeneity was more important than either and was itself a consequence of principles still more fundamental. In *First Principles* his argument is that there is a fundamental law of matter, which he calls the law of the persistence of force, from which it follows that nothing homogeneous can remain as such if it is acted on, because any external force must affect some part of it differently from how it affects other parts and hence cause difference and variety to arise. From this it follows that any force that continues to act on what is homogeneous must bring about an increasing variety. This law of the multiplication of effects is, in Spencer's view, the clue to the understanding of all development, cosmic as well as biological. He holds that an unknown and unknowable absolute force is continuously operating on the material world and producing variety, coherence, integration, specialization and individuation. Gases concentrate to form planets; the earth becomes more variegated and gives birth to such simple animals as the amoeba; man evolves from less complex species and at first lives in undifferentiated hordes; various social functions are developed so that there are priests, kings, scholars, workers, etc.; poetry splits off from music and painting

from drama so that a diversity of arts arises; and knowledge itself is differentiated into the various sciences. But when the final stage of integration has been reached, Spencer holds, dissolution must necessarily ensue, so that the variety and coherence of the world will lapse into quiescence once more. "And have we thus to contemplate," he asks, "as the outcome of things, a boundless space holding here and there extinct suns, fated to remain for ever without further change?" (*First Principles*, 1862). His answer is that this may be, but that perhaps a new form of integration will then commence which, after reaching an equilibrium, will itself decay and be followed by new evolutions and dissolutions in an infinite alternation.

Sociology and Social Philosophy.—Spencer applies his general evolutionary scheme to human society. Indeed, it was from reflection on human society that he conceived it, as can be seen in *Social Statics* where social evolution is held to be a process of increasing "individuation." Human societies, then, evolve from undifferentiated hordes, by means of increasing division of labour, into complex civilizations. Spencer believed that primitive men were smaller, less intelligent and more emotional than civilized men; that light was thrown on their mentality by study of children in civilized societies; that their religion arose from belief in ghost souls seen in dreams, so that worship was originally directed toward the spirits of dead ancestors; that civilized religions were all more or less elaborate variations on this primitive theme; and that the fundamental sociological classification was between military societies, in which co-operation was secured by force, and industrial societies, in which co-operation was voluntary and spontaneous. The word "industrial" is used here not in the sense of "technologically developed"—for Spencer gives instances of extremely simple societies such as the Todas to illustrate his meaning—but to designate societies chiefly concerned with producing for their members' needs by peaceful and voluntary co-operation. Military societies are ruled by warlike chiefs who maintain themselves in power by arms and superstition; in them women are depressed in status and often forced to do the most burdensome work; and there is an hierarchical organization in which each individual is expected to know and to keep his place. Industrial societies are the reverse of all this, although, if they become involved in war, they have to adopt the military, hierarchical organization.

Evolution is not the only biological conception that Spencer applies in his sociological theories. He also made a detailed comparison between animal organisms and human societies. In both there is a regulative system (the central nervous system in the one, government in the other), a sustaining system (alimentation in the one case, industry in the other) and a distributing system (veins and arteries in the animal body; roads, telegraphs, etc., in the social organism). The great difference, however, between an animal and a social organism is that whereas in the former there is one consciousness relating to the whole, in the latter consciousness exists in each member only and there is no consciousness relating to the whole. Society as a whole has no mind. At this point Spencer abandons description and theory in order to conclude that society exists for the benefit of its members, and not they for its benefit.

This individualism is the key to all of Spencer's work. His contrast between military and industrial societies is a contrast between types of order which he considered bad and those which he considered good. In industrial society coercion and regimentation are at a minimum, and the order achieved, although—indeed because—not planned by anyone, is delicately adjusted to the needs of all parties. In *The Man versus the State* (1884) Spencer says that the Tories support a military and Liberals an industrial social order, but he thought that the Liberals of the latter half of the 19th century were, with their legislation on hours of work, liquor licensing, sanitation, education, etc., developing a "New Toryism" and preparing the way for a "coming slavery." "The function of Liberalism in the past was that of putting a limit to the powers of kings. The function of true Liberalism in the future will be that of putting a limit to the powers of Parliaments."

In his emphasis on variety and differentiation Spencer was unwittingly repeating, in a 19th-century idiom, the metaphysics of liberalism which Spinoza and Leibniz had adumbrated in the 17th century. Spinoza had maintained that "God or Nature" has an infinity of attributes in which every possibility is actualized; and Leibniz had argued that the perfection of God is exhibited in the infinite variety of the universe. Neither of these philosophers believed that time was an ultimate feature of reality. Spencer, however, combined a belief in the reality of time with a belief in the eventual actualization of every possible variety of being. He thus gave metaphysical support to the liberal principle of variety according to which a differentiated and developing society is preferable to a monotonous and static one.

Spencer's attempt to synthesize the sciences showed a sublime audacity which was not repeated because the intellectual specialization which he welcomed and predicted increased beyond his expectations. His sociology, although it gave an impetus to the study of society, was superseded as a result of the development of social anthropology since his day and was much more concerned with providing a rationale for his social ideals than he himself appreciated. Primitive men, for example, are not the childlike emotional creatures that he thought them to be, nor is religion to be explained only in terms of the souls of ancestors. When T. H. Huxley said that Spencer's idea of a tragedy was of "a deduction killed by a fact," he called attention to the system-building feature of Spencer's work which led him to look for what confirmed his theories and to ignore or to reinterpret what conflicted with them. Spencer died on Dec. 8, 1903, in Brighton. His *Autobiography* (2 vol.) was published posthumously in 1904.

Spencer's *Synthetic Philosophy* finally comprised the following books: *First Principles* (1862) rev. ed. (1900); *The Principles of Biology*, 2 vol. (1864–67) rev. ed. (1898–99); *Principles of Psychology*, 2 vol. (1855, 1872), 4th ed. (1899); *The Principles of Sociology*, 3 vol. (1876–96); and *The Principles of Ethics*, 2 vol. (1892–93). Other important books not mentioned above are: *Essays: Scientific, Political and Speculative*, 3 series (1857–74), final ed. 3 vol. (1891); *Education: Intellectual, Moral, Physical* (1861); *The Study of Sociology* (1873); *The Nature and Reality of Religion* (1885), withdrawn from publication; *Facts and Comments* (1902). *First Principles* (6th ed.) was reissued in 1937.

BIBLIOGRAPHY.—Hugh Elliot, *Herbert Spencer*, with bibliography (1917); David Duncan, *The Life and Letters of Herbert Spencer* (1908); J. Rumney, *Herbert Spencer's Sociology: a Study in the History of Social Theory* (1934); E. Diaconide, *Étude critique sur la sociologie de Herbert Spencer* (1938); for a discussion of Spencer's ethical theory, G. E. Moore, *Principia Ethica*, sec. 31–35 (1903). (H. B. A.)

SPENCER, JOHN CHARLES SPENCER, 3RD EARL (1782–1845), English statesman, better known by the courtesy title of Viscount Althorp, which he bore during his father's lifetime, was the son of George John, 2nd earl (1758–1834), who was first lord of the admiralty 1794–1801 and home secretary 1806–07. John Charles was born at Spencer house, London, on May 30, and was educated at Harrow and Trinity college, Cambridge. He represented Okehampton (1804–06), St. Albans (1806), Northamptonshire (1806–32) and Northamptonshire South (1833–34). When Earl Grey's administration was formed in Nov. 1830 the chancellorship of the exchequer combined with the leadership of the house of commons was entrusted to Lord Althorp. His first budget was a failure, but this was soon forgotten in the struggles over the Reform bill, which he and Lord John Russell twice piloted successfully through the commons. "Althorp carried the bill," said a leading opponent, Sir Henry Hardinge, "his fine temper did it."

Althorp was neither eloquent nor nimble-witted; he owed his ascendancy over the commons to his courage, integrity and sound judgment. He was able to gauge feeling among the squires who predominated in the house of his day, for he shared their tastes and habits of mind. In Sir Robert Peel, then leading the opposition, he faced a redoubtable parliamentarian, and his own following of Whigs, Radicals and Irish was ill-mixed and hard to manage; yet his success as leader was complete. "He became," wrote the diarist Greville, a severe contemporary judge, "the very best leader of the . . . Commons that any party ever had. . . . The whole House liked him, his own party followed him with devoted attachment."

Lord Althorp retired from politics when called to the house of lords in Nov. 1834 by the death of his father. His removal from the

commons was the occasion of William IV's dismissal of Viscount Melbourne's ministry. He was the first president of the Royal Agricultural society (founded 1838), and a notable shorthorn cattle breeder. He died at Wiseton on Oct. 1, 1845.

Bibliography.—Sir Denis Le Marchant, *Memoir* (1876); W. Bagehot, *Biographical Studies* (1881); E. J. Myers, *Lord Althorp* (1890); "Earl Spencer," *Quarterly Review*, vol. 283 (Oct. 1945). For contemporary opinions *see* Lord John Russell's (anonymous) obituary, *Edinburgh Review*, vol. 83, pp. 240–73 (Jan. 1846); C. C. F. Greville, *Memoirs*, 1st series, iii, 106 (1875) and 2nd series, ii, 296 (1885); M. Van Buren, *Autobiography*, pp. 476–7 (1920); A. Aspinall (ed.), *Three Early Nineteenth Century Diaries* (1952). (M. G. B.)

SPENCER, SIR STANLEY (1891–1959), English painter of religious subjects, landscapes, portraits and social satires, was born at Cookham, Berkshire, on June 30, 1891, son of W. Spencer, organist. He studied at the Slade School of Art and began to exhibit at the New English Art club. During World War I he served in the royal army medical corps in Macedonia and painted "Travoys Arriving with Wounded" for the Imperial War museum. In 1932 he completed "Resurrection of the Soldiers" and scenes from army life in the chapel at Burghclere, Hampshire. "Christ Carrying the Cross" (1920), "Resurrection" (1927), representing the dead rising from their tombs in Cookham churchyard, and a later "Resurrection" (1950) are at Tate gallery, London. He also painted observed and highly finished landscapes and humorous comments on the absurdity of human behaviour. He was elected a royal academician in 1950 and knighted in 1959. He died at Taplow, Bucks., on Dec. 14, 1959.

Bibliography.—R. H. Wilenski, *Stanley Spencer* (1924); E. Rothenstein, *Stanley Spencer* (1945); Eric Newton, *Stanley Spencer* (1947); *Stanley Spencer. Resurrection Pictures* (1951); *Stanley Spencer Retrospective Exhibition Catalogue*, Tate Gallery (1955). (M. Ch.)

SPENDER, JOHN ALFRED (1862–1942), British journalist, was born at Bath, the son of a doctor, and educated at Bath college, and at Balliol college, Oxford. Adopting the career of a journalist, he was editor of the *Eastern Morning News*, Hull, from 1886 to 1890 and then was with the *Pall Mall Gazette* in London. When the *Westminster Gazette* was established in 1893, he became an assistant editor and from 1896 to 1922 was its editor, leaving just after it became a morning paper. As editor of the *Westminster* Spender won general respect. He held a high place in the councils of the Liberal party and was a member of the royal commission on divorce and of the Milner mission to Egypt. A slight volume, *The Comments of Bagshot* (1907), revealed Spender at his best. Other works include *The Indian Scene* (1912); *The Foundations of British Policy* (1917); *Life, Journalism, and Politics*, 2 vol. (1927); *A Short History of Our Times* (1934); *Men and Things* (1937). He died on June 21, 1942.

SPENER, PHILIPP JAKOB (1635–1705), German theologian, called "the father of Pietism," was born on Jan. 13, 1635, at Rappoltsweiler in Upper Alsace. He won his degree of master at Strasbourg (1653) by a disputation against the philosophy of Hobbes. In 1666 he was chief pastor in the Lutheran Church at Frankfurt am Main.

There he published his two chief works, *Pia desideria* (1675) and *Allgemeine Gottesgelehrtheit* (1680), and began that form of pastoral work which resulted in the movement called Pietism (*q.v.*). In 1686 he accepted the invitation to the first court chaplaincy at Dresden. But the elector John George III was offended by his chaplain's candour. Spener refused to resign his post, and the Saxon government hesitated to dismiss him. The difficulty was solved by his promotion (1691) to the rectorship of St. Nicholas in Berlin with the title of *Konsistorialrat*. The university of Halle was founded under his influence in 1694. Spener was exposed to the abuse of orthodox Lutheran theologians. In 1695 the theological faculty of Wittenberg formally laid to his charge 264 errors. He died on Feb. 5, 1705. His last important work was *Theologische Bedenken*, 4 vol. (1700–02), to which was added after his death *Letzte theologische Bedenken*, with a biography of Spener by C. H. von Canstein (1711). Spener was a voluminous writer. A new edition of his chief writings was published by P. Grünberg in 1889.

See W. Hossbach, *Philipp Jakob Spener und seine Zeit*, 3rd ed. (1861); A. Ritschl, *Geschichte des Pietismus*, ii (1884); P. Grünberg, *P. J. Spener*, 3 vol. (1893–1906).

SPENGLER, OSWALD (1880–1936), German philosopher, was born at Blankenburg am Harz on May 29, 1880. Mathematics and natural science were the main subjects of his university education, but he wrote a thesis on Heraclitus for his doctorate, taken at Halle in 1904. Spengler earned his living as a schoolmaster until 1911, when he went to live in Munich on a small inheritance. It was about this time that he began to work on *Der Untergang des Abendlandes* (Eng. trans., *The Decline of the West*, London, 1926–28), of which he had completed a draft by the outbreak of World War I. The first volume of this powerful book was published in Munich in July 1918, a few months before the final collapse of the German forces. It won its author immediate acclaim from the general public, if not from professional historians and philosophers. The second volume followed in 1922, and a revised version of the first a year later. Spengler tried from 1919 onward to turn his reputation to account by assuming the role of political commentator, but he met with little success in this field. He became increasingly isolated, and his difficulties were multiplied by criticism of his work by the National Socialists, despite a certain affinity between his political ideas and theirs. (*See* National Socialism.) He died in Munich of a heart attack on May 8, 1936.

Spengler's reputation rests entirely on *Der Untergang des Abendlandes*, which is in effect a study in the philosophy of history or in historical "morphology." Beginning, as G. B. Vico had done before him and as Arnold Toynbee was to do a few years later, by comparing modern western civilization with the civilization of the Greco-Roman world, Spengler claimed to be able to discern the outline of a life cycle through which, he believed, all such civilizations must pass. Possession of this key idea, he argued, would enable us both to reconstruct the past and, more significantly, to predict "the spiritual form, duration, rhythm, meaning and product of the still unaccomplished stages of our western history." Spengler's own view of the prospects of western civilization was a gloomy one. Since the west, he believed, had already passed through the creative stage of "culture" into the stage of reflection and material comfort (which constituted "civilization" proper in his special terminology), the future could only be a period of decline; nor was there any prospect of reversing the process, for civilizations blossomed and decayed like natural organisms, and true rejuvenation was as impossible in the one case as in the other.

Spengler's book won scant approval from professional scholars, who were scandalized by his unorthodox methods and contemptuous of his errors of fact. There could, however, be no denying his moments of genuine historical insight (as in his imaginative treatment of Greek mathematics) or the merits of his work as a contribution to social theory.

Other works by Spengler include *Preussentum und Sozialismus* (Munich, 1920), *Politische Pflichten der deutschen Jugend* (Munich, 1924), *Neubau des deutschen Reiches* (Munich, 1924), *Der Mensch und die Technik* (Munich, 1931), Eng. trans., *Man and Technics* (London, New York, 1932) and *Jahre der Entscheidung* (Munich, 1933), Eng. trans., *Years of Decision* (London, New York, 1934). Some of his miscellaneous writings are collected in *Politische Schriften* (Munich, 1932) and *Reden und Aufsätze* (1937).

Bibliography.—R. G. Collingwood, "Oswald Spengler and the Theory of Historical Cycles," *Antiquity* (1927); M. Schroeter, *Metaphysik des Untergangs* (Munich, 1949); P. A. Sorokin, *Social Philosophies of an Age of Crisis* (Boston, 1950; London, 1952); H. Stuart Hughes, *Oswald Spengler: a Critical Estimate* (New York, London, 1952). (W. H. W.)

SPENSER, EDMUND (?1552–1599), "the prince of poets in his time," was born in London—Oldys says, in East Smithfield—probably in 1552, possibly in 1549. (The date depends on whether Sonnet 60, written when Spenser was 41, belongs to 1593 or to 1590: *v. infra.*) His mother's name, he tells us, was Elizabeth: his father has been conjecturally identified with John Spenser, a journeyman clothmaker. Later, the poet claimed kindred with the Spencers of Althorpe in Northamptonshire, and had his claim allowed; but the poverty of his home is shown by the grants made to him out of the Nowell bequest as a poor scholar of Merchant Taylors' school. There, under the great Mulcaster, he laid the foundations of his wide, if inexact, scholarship; best of all, he learned from Mulcaster to "worship the English," and to believe it as capable of great poetry as any language.

Spenser was still at school when, in 1569, he supplied some verse translations from Marot and du Bellay to Van der Noodt's *Theatre of Worldlings*. Van der Noodt, it is true, took the credit of these verses to himself: but they were included, with some alterations, in the *Complaints* volume of 1591 (*v. infra*), and we need not doubt that they are Spenser's. The rhymed stanzas after Marot at least bear his mark: the blank verse renderings from du Bellay are inferior, and were considerably altered and tagged with rhymes for the reissue of 1591.

In this same year he entered Pembroke hall, Cambridge, as a sizar, matriculating on May 20. He remained at Cambridge for seven years, as was then not unusual, proceeding B.A. in 1572 and commencing M.A. in 1576. His health was poor; but he read

widely, especially in philosophy and rhetoric, studying Italian as well as Latin, Greek and French, and training himself for poetry. His studies profited from the friendship of Gabriel Harvey, a fellow of his college, a conceited pedant, but a real scholar, and sincerely attached to Spenser, though his friendship may not always have been quite disinterested nor his literary advice wise. Literature, however, stirred the mind of Cambridge less than theology. In the ecclesiastical controversy with which the university had seethed ever since Cartwright's brief tenure of the chair of divinity (1569–70), Spenser took the Puritan side. But he was neither sectary nor ascetic. He had a hatred of popery (inflamed by the Bartholomew massacre) and contempt for the hirelings who "for their bellies' sake" had climbed into the Anglican fold.

What he did on leaving Cambridge is still a mystery. He had missed his fellowship, and must look about him for a profession or a patron. The view that he went to live with relatives in Lancashire, though buttressed by appeals to family tradition and to the language and scenery of the *Calender,* rests at bottom on no ascertained fact.

But in the "North-parts" (Drayton brings Rosalind to the Cotswolds; Aubrey says that she was related to Sir Erasmus Darwin's lady, who was a Wilkes of Hodnell in Warwickshire), Spenser fell in love with the Rosalind of the *Shepheardes Calender,* of whom again we can say only that she was a lady of higher rank than his, who enjoyed, but did not reward, her young poet's devotion. In 1577 we seem to get a glimpse of Spenser in an unexpected quarter. In the *View of the Present State of Ireland* Irenaeus tells how he witnessed the execution of Murrogh O'Brien, which took place at Limerick in July, 1577. In the rest of that dialogue Irenaeus represents Spenser himself, and it is natural though not inevitable to conclude that he does so in this place also. But no other evidence has been found for this visit; Phillips's statement that Spenser was secretary to Sir Henry Sidney may be discounted, since he has probably confused Sidney with Grey. Yet Irenaeus's account of this hideous incident reads like the words of an eye-witness, and the visit, if it occurred, would point to a connection with the Sidney-Leicester circle earlier than 1579.

After this we are on firmer ground. We know that in 1578 Spenser was secretary to John Young, bishop of Rochester, who had been master of Pembroke hall while Spenser was a student. The bishop of Rochester (*Episcopus Roffensis*) is the Roffyn of the *Shepheardes Calender:* the fable of Roffyn's dog and the wolf may glance at the old Roman Catholic bishop, Thomas Watson, who was committed to Young's custody in Feb. 1579. Much of the *Shepheardes Calender* was apparently written at Rochester, where also Spenser first conceived the idea of the wedding of Thames and Medway. If Spenser at this time thought of entering the Church, he changed his mind on a nearer view of its condition, for by Easter 1579 he was in Westminster, sharing rooms with Harvey and holding "long conference" with him. There was a third party to their discussions, one "E.K.," to be identified probably with Edward Kirke, once a fellow-sizar of Spenser's at Pembroke. The notion that E.K. is a mask for Spenser himself has been disproved by Dr. Herford. It is not unlikely, indeed, that the "Mistress Kerke" in Westminster, who took charge of letters for Spenser in October of this year, was E.K.'s mother, and that all three were living in her house. Among other things they discussed Spenser's *début* as an author. He had several pieces ready or nearly ready—*Dreams, Legends, Court of Cupid, Pageants, The English Poet* (in prose), and the *Shepheardes Calender.* It was decided to bring out the last, and to bring it out in classical style with introduction, notes and glossary by E.K., even as Muretus edited the first book of Ronsard's *Amours.* E.K. wrote the introduction forthwith—it is dated from London, April 10, 1579—and got to work on the notes.

Harvey's Letter-book gives us an amusing picture of Spenser that summer. He is reading some law, evidently with an eye to diplomatic or administrative work; he has blossomed into a courtier and a gallant, bearded and moustachioed, Italianate, Frenchified. At what date, and in what sense, he entered Leicester's service we do not know precisely. At all events on Oct. 5 he writes from Leicester house to tell Harvey that he expects to go abroad in a week, "most what" at Leicester's charges—the "most what" shows that he was not entirely dependent on Leicester's bounty—and that "gentle Mr. Sidney" has proposed that they two should correspond. The tour apparently never took place, for his next letter of ten days later says no word about it, but is all concerned with quantitative verse and with doubts about the publication of his *Slumbers.* He has been to Court and expects to go again; Sidney and Dyer have him "in some use of familiarity." The *Shepheardes Calender* was licensed on Dec. 5, and published anonymously; the dedication to Sidney is by "Immerito."

The Shepheardes Calender.—The *Calender* consists of 12 pastoral eclogues in the artificial style of Virgil, Mantuan, Sanazzaro and Marot, which admits real persons and current events in a pastoral guise. From all these precursors Spenser borrowed ostentatiously, above all from Marot, whom E. K. nevertheless belittles, perhaps because the Pléiade had eclipsed him. Derivative and conventional as it is, the *Calender* made an epoch in English literary history, not so much by naturalising pastoralism as by showing (under cover of the pastoral convention) what English could do in many kinds and metres. The couplets of the satires and fables are intentionally rude, and the rhymeless sestina is mere virtuosity; but the elegiac stanzas of the complaints recaptured the metrical secret of Chaucer, and the lyric staves of the paean and the dirge extended the range of English song. In all these measures he made free use of alliteration, dear to English ears, though abused by his contemporaries and frowned on by the classicists. For his reform of poetic diction Spenser had before him the example and precepts of the Pléiade. Like them, he sought to embellish the beggarly vocabulary of contemporary verse, partly by foreign loan-words, but mainly by drawing on native sources, by archaisms, pseudo-archaisms and dialect words. He was in search of a vocabulary fit for the heroical poem that he already contemplated. But the new poet had to care not only for words but for the order of his words and the structure and juncture of his sentences. Here Spenser scored his greatest success, eschewing obscurity and looseness, and giving his syntax a movement, too copious perhaps, but admirably fluent, in lucid, easy yet well-knit sentences. Syntactically, Spenser is one of the simplest of poets.

The *Shepheardes Calender* may not have fulfilled all Spenser's hopes, but it went into a second edition in 1581. Sidney praised it judiciously, or judicially, but durst not approve its rustic diction. However, Spenser had many other arrows in his quiver. In April 1580 he is full of projects for the immediate publication of his *Dying Pelican* and *Dreams;* his Latin *Stemmata Dudleiana* will need "more advisement" before it can be "sent abroade," but the quantitative *Epithalamion Thamesis* shall be "shortly set forth." Under the influence of Sidney, Dyer and Drant, he had come for the moment to take quantitative verse more seriously. But he is more eager to proceed with the *Faerie Queene,* of which he sent Harvey specimens. Harvey thought little of them, preferring the (lost) *Nine Comedies* in the manner of Ariosto. Moreover, the poet is in love again, this time with better hopes. Harvey's compliments to *Mea Domina Immerito, Mea Bellissima Collina Clouta,* taken with an obscure passage in *Daphnaida* (11. 64–66), which reads like consolation tendered by one widower to another, have led some scholars to believe that Spenser married this "second Rosalind." But there is no other evidence for the marriage, nor is the lady ever heard of again. These letters of April 1580 give no hint of any friction with Leicester. But now Spenser seems to have made a false step. The queen had dallied for years with the project of marrying the duke of Alençon. In Jan. 1579 his agent Simier came to England and Alençon himself paid a flying visit in August to press his suit in person. He was ugly, dwarfish and half her age, but the queen smiled on him, called him her "grenouille" and Simier her "petit singe." The Puritans took alarm. Sidney remonstrated in such plain terms that the queen forbade him the Court. He retired to Wilton to write his *Arcadia,* and Spenser probably never saw him again. This was in Jan. 1580.

It was at this juncture apparently that Spenser drew his pen. He had begun a *fabliau* of an ape and a fox, satirising the Parson

Trullibers and other humbugs of the day. Now, catching at Elizabeth's trick of animal nick-names, he brought his ape and fox to a court of beasts. The ape becomes Alençon-Simier, the fox is Burghley, who was believed to favour the marriage. Their misrule when they have stolen the lion's skin foreshadows the fate of England under a French king-consort. But Spenser misjudged the situation. Leicester had concluded that the wind stood fair for France, and trimmed his sails accordingly. Spenser was snubbed, and the satire, which was circulating in manuscript as *Mother Hubberds Tale*, was called in, but not before it had come to Burghley's ears. The tale of Spenser's discomfiture is told in *Virgils Gnat*, killed by the man whose life it had saved. Such is the most plausible explanation of this obscure episode. At all events, abandoning all his projects of publication, Spenser accepted a secretaryship to Lord Grey of Wilton, the newly appointed Lord Deputy of Ireland.

In Ireland.—Grey landed in Dublin in Aug. 1580 and set himself at once to the task of suppressing Desmond's rebellion. He was a stern Puritan, to whom the pope was Anti-christ and the Papist rebels were traitors at once to God and their queen. Spenser soon had proof of his quality. A band of Spanish and Italian filibusters had landed in Kerry to support the Munster rebels and occupied a fort near Smerwick. Grey marched against them, forced them to surrender unconditionally, held the officers to ransom, and put the men to the sword. Then he ravaged Munster, leaving famine to complete the work. On campaigns Spenser saw "such wretchednesse as that any stonye harte would have rued the same." With Grey's recall in 1582 Spenser lost his secretaryship; but in the previous year he had acquired a clerkship for the faculties (*i.e.*, licences and dispensations) in the Dublin chancery; and he had begun to obtain leases of forfeited houses and lands at modest rents which he did not always pay. At one of these houses, New Abbey in Co. Kildare, he seems to have resided occasionally in 1583 and 1584. But his headquarters were still in Dublin. We get a pleasant glimpse of him there in Ludovick Bryskett's *Discourse of Civil Life*. Bryskett tells us how a distinguished company, including the primate of Armagh, and soldiers like Norreys and St. Leger, met at his cottage near Dublin. Spenser suggested that their host should read them his translation of Giraldi da Cinthio's *Discourse of Civill Life;* which Bryskett accordingly did. The reading lasted for three days, being interspersed with discussions. On the third day Spenser interposed again, urging (to our astonishment) Aristotelian objections to the immortality of the soul. But Bryskett is simply putting into his mouth the arguments given by Giraldi to Torquato. It is doubtful, indeed, if this famous discussion ever took place at all, for Bryskett imitates Giraldi not only in the substance of his dialogue but in its setting. But if it never actually occurred it must be such as might have occurred.

In 1585 the centre of Spenser's interests began to shift to Munster. In that year he is found acting as deputy to Ludovick Bryskett in the clerkship of the council of Munster. In 1586 he held the prebend of Effin in Co. Limerick; and in that same year, under the Government's scheme for the plantation of Munster, he obtained a perpetual lease of Kilcolman Castle in Co. Cork with 3,028 acres of land. His resignation of his clerkship for the faculties in 1588 probably means that he then began to live at Kilcolman, where tradition says that his sister Sarah kept house for him. Here in 1589 he received a momentous visit from Raleigh. He had been engaged on the *Faerie Queene* intermittently for nearly ten years, receiving a fresh stimulus from Tasso's *Gerusalemme Liberata*, which reached him probably about the time of Lord Grey's recall. By 1589 three books were completed. Raleigh at once perceived their superlative merits and carried Spenser off to England to lay his poem at the queen's feet. It was licensed on Dec. 1, 1589 and published in 1590 with a dedication to the queen. Its reception can only be described as reverential. The great English epic so long awaited had appeared.

Spenser remained in England for more than a year, enjoying his fame, making friends with brother-poets, entertained at country houses, and acknowledged by his kinswomen of Althorpe, Lady Carey, Lady Strange, and Lady Compton and Mountegle. But Burghley had not forgiven him; the substantial preferment for which he hoped was whittled down to a pension of £50 a year, and in 1591 he returned to Ireland a disappointed man. Before leaving, he arranged for the publication of some of his minor poems, which appeared in 1591 as *Complaints*.

Complaints.—Of the *Complaints*, the *Visions of Petrarch* and *Visions of Bellay* are refurbishings of his schoolboy translations. *The Ruines of Rome*, a rendering of du Bellay's *Antiquitez de Rome*, is clearly a juvenile production also. In the same kind, but much superior, is the original *Visions of the Worlds Vanitie*, in which, from Harvey's remarks, we may fairly recognize a fragment of the *Dreams* of 1580. The *Teares of the Muses*, mentioned in *A Midsummer Night's Dream*, V. i. 52, where Theseus dismisses it as "some satire, keen and critical," is an unconvincing complaint of the neglect of poetry, more excusable in 1580 than in 1590. The *Ruines of Time* laments various members of the Dudley family, including Sidney and Leicester, who had died in the previous decade: its present form belongs to 1590; but it probably incorporates some material collected for the Latin *Stemmata Dudleiana;* and the "Pageants" at the end may be another fragment of the *Dreams*. *Mother Hubberds Tale* was strengthened by the addition of a brilliant and bitter denunciation of the wretchedness of the suitor's state; and Spenser bated no jot of his satire on Burghley. *Virgils Gnat*, a free rendering of the *Culex*, has already been described. The one entirely new poem in the volume is *Muiopotmos*, which clearly belongs to 1590. Its theme recalls Sonnet 71, and may have been suggested simply by the sight of Lady Carey at her drawn-work. In spite of the grim ending the tone is so light-hearted that it seems perverse to read an inner meaning into this "airy trifle," with Burghley as the spider. Yet no reader would have suspected an inner meaning in *Virgils Gnat* but for the dedication to Leicester.

On reaching home, or so he would have us believe, Spenser wrote *Colin Clouts Come Home Againe*, dedicating it to Raleigh "from my house of Kilcolman the 27. of December 1591."[1] It was not published, however, till 1595, when it appeared in one volume with *Astrophel*.

Colin Clout; Astrophel.—*Colin Clouts Come Home Againe* is the most charming of Spenser's poems. He is again in Arcady, telling his fellow-shepherds of his voyage with the Shepherd of the Ocean, his reception by the mighty Cynthia, her court, her ladies and brave poets. Then the note changes to the old complaint of courtly falsity and praise of the shepherd's life. The whole poem is exquisitely written, in "an easy running verse with tender feet." There is nothing more attractive in Spenser than Colin's chivalrous defence of Rosalind.

Astrophel is the prelude to a set of pastoral elegies on Sidney by several hands. The second "number," the *Lay of Clorinda*, though credited to Sidney's sister, is demonstrably from Spenser's pen. If *Astrophel* seems conventional and frigid we must remember that Spenser had already lamented Sidney in the *Ruines of Time*, and that the Sidney whom he now laments is not the hero of Zutphen but the author of the *Arcadia*.

Colin's praise of the shepherd's life prepares us to find that Spenser, though disappointed of preferment in England, and harassed by law-suits with his Irish neighbour Lord Roche, had begun to love his Munster home and to weave its legends and scenery into his verse. Towards the close of 1592, on the orthodox view of the *Amoretti*, he fell in love with Elizabeth Boyle and married her on June 11, 1594. He celebrated his wooing in the *Amoretti* and his wedding in the *Epithalamion*, which were entered together at Stationers' Hall in November 1594, and published together in 1595. It is highly probable that some of the *Amoretti* had already done service to express an *amour courtois* for Lady Carey during his English visit. It is probable that Elizabeth Boyle was the orphan daughter of Stephen Boyle, of Bradden in Northamptonshire, who had accompanied her brother

[1]Yet the dedication of *Daphnaida* is dated "London this first of Januairie, 1591" (*i.e.*, 1592 n.s.). No convincing explanation of this discrepancy has been offered. On the whole it is easiest to believe that Spenser, writing on New Year's Day, used the new style, *i.e.*, that his 1591 means our 1591.

Alexander when he went to Ireland to seek his fortune under the protection of their kinsman Richard Boyle, afterwards earl of Cork. The view that she was a widow, having been married to one Tristram Peace in 1588 or 1589, at present lacks documentary confirmation, and is hard to reconcile with Burke's statement that she bore seven children to her third—on this view her fourth—husband, Sir Robert Tynte, after 1612 or 1613.

Amoretti; Epithalamion.—The *Amoretti* are love-sonnets, all but one in a form which Spenser made his own, consisting of three *linked* quatrains and a couplet. It had already been used in Scots, but Spenser may have got it direct from Marot. Though it lacks the pyramidal strength of the Petrarchan form, and the freedom of the Shakespearean, its "linked sweetness" suits Spenser's style. Many of these sonnets are imitated from French or Italian, especially from Desportes and Tasso: for one a Spanish original has been found. They are all graceful, but only the famous sonnet on Easter can be called great. The *Epithalamion*, on the other hand, is by common consent the greatest of all wedding-songs; rich in poetic allusions and in echoes from his own earlier poems, as if Spenser had gathered up all the fruits of his study and all the flowers of his fancy to present them to his bride. Here for once his whole nature speaks, flesh and spirit reconciled in the sacrament of marriage. The stanza, suggested by the Italian canzone, is magnificent far beyond anything previously heard in English. The structure of the whole poem is as perfect as that of its parts.

Towards the end of 1595 Spenser, having finished three more books of the *Faerie Queene*, came to London to publish them. During this year he also published his *Prothalamion* and *Fowre Hymnes*, and composed or revised his *Veue of the Present State of Ireland*, which, though entered at Stationers' Hall in 1598, was not allowed to be printed without further authority, and, in fact, did not see the light till 1633.

Prothalamion; Hymnes; Veue of the Present State of Ireland.—The *Prothalamion* graced the wedding of two of the Earl of Worcester's daughters, who were married together from Essex house. Though it lacks the glow and sweep of the *Epithalamion*, it is, if possible, even more perfect in metre and diction.

The hymns of *Heauenlie Love* and *Heavenly Beauty* were written in 1596 to propitiate two pious noblewomen and counteract the hymns in *Honour of Love* and in *Honour of Beautie* written in his youth. So Spenser avers; but to print the earlier hymns was scarcely the way to ensure their oblivion.

In the *Veue of the Present State of Ireland* Irenaeus expounds to Eudoxus the causes of the Irish troubles and propounds a cure. Irish laws, customs and religion must all be reformed on the English model. But subjugation must precede reform. Vacillation has been the curse of the Government. Let them now bring over 10,000 foot and 1,000 horse, plant these in six convenient garrisons, give the rebels 20 days in which to surrender, and then hunt down relentlessly all who stand out. Two winter campaigns will break their spirit. Let a fresh offer of pardon then be made, and rebellion will be at an end. There follows a detailed scheme, supported by statistics, for the administration of the pacified areas, ending with a proposal for the appointment of a lieutenant-general, Essex being clearly indicated. Political antagonism and racial antipathy combined with religious hatred to blind Spenser to the Irish cause.

During this second visit to England Spenser's hopes of preferment were centred in the Earl of Essex, but again he sued in vain. He returned to Kilcolman, probably in 1597, and resumed the *Faerie Queene*. In Sept. 1598 he was recommended for the sheriffdom of Cork. But preferment came too late: in October Tyrone's rebellion had broken out, the Munster Irish rose, Kilcolman Castle was burned and Spenser fled with his family to Cork. From Cork he was sent to London with a dispatch which bears date Dec 9, 1598. Along with it he brought a brief note of his own in which he reiterated the policy of the *View*. On Jan. 16, 1599 he died. He was buried in Westminster Abbey close to Chaucer; many nobles attended his funeral, and his fellow-poets brought elegies which they threw into his grave with the pens that had written them. Spenser's tragic reversal of fortune and sudden death gave rise to a crop of legends. Ben Jonson told Drummond that one of Spenser's children perished in the flames of Kilcolman Castle and that he himself died "for lack of bread." It is only too likely, after all he had gone through, that Spenser collapsed on reaching London, and was a dying man before his friends could learn of his condition and come to his aid. Two more cantos of the *Faerie Queene* with two stanzas of a third were published in 1609; if more was written, as is probable, after 1596, it is irrecoverably lost.

The Faerie Queene.—Spenser's place among the great poets depends on the *Faerie Queene*, a fragment of a great poem which was to have been "disposed," as the title informs us, "into XII. bookes fashioning XII. morall vertues," each virtue embodied in a knight, and the whole designed "to fashion a gentleman or noble person in vertuous and gentle discipline"—at once a chivalric romance and a handbook of morals and manners; nay more—for Spenser would emulate Virgil as well as Ariosto—a national epic to the glory of England's Elizabeth. Such a threefold cord is not easily twined. Yet the task seemed feasible to Spenser, because to him the eternal war of good and evil, which was his essential theme, was embodied in the struggle of Protestant England against her Catholic foes. Book I. indeed may be read not inconsistently either as a mere romance, or as the spiritual experience of an elect Christian, or as a history of the English Church in the sixteenth century. But in Book II. the story stands still for nearly two whole cantos; the tale of Hellenore in Book III. is told with a gusto ill calculated to fashion a gentleman in virtuous discipline; in Book IV. the poet grows so careless of his characters that he lets the long-parted lovers come together without recognizing one another; the recent history of Book V. has little allegorical or romantic interest; and Book VI. ends in a burst of sheer pastoralism in which the poet forgets his allegory as completely as the hero forgets his quest. Only Spenser's contemporaries could fully appreciate the historical element, and some of it even to them can only have had the interest of a *roman à clef;* yet it strengthens the texture of the poem, and gives figures like Arthegall and Satyrane some firmness of outline.

The allegory was treated with respect for two centuries; Milton thought "our sage and serious poet" a better teacher than Scotus or Aquinas; modern criticism has too often followed Hazlitt in commending the poem to readers with the assurance that the allegory won't bite them. But to ignore the allegory is to ignore the informing purpose of the whole, without which the story would be a mere series of ill-joined episodes, and much of the imagery meaningless, some of it grotesque or even repulsive. The ethical, as distinct from the structural, value of the allegory must depend on the worth of the ideas it embodies, as its poetical value depends on the imaginative force of their embodiment. Spenser was not an original nor a systematic thinker. His philosophy is a blend of Platonism refracted through Ficino, Aristotelianism of the scholastic tradition, and Christianity with a Calvinistic bias. And the elements do not perfectly combine. Thus the doctrine of love and beauty, by which Ficino thought to reconcile the love of woman with the love of God, this doctrine, brilliantly expounded in the first two Hymns, is fundamentally irreconcilable with the scheme of salvation through the atoning sacrifice of Christ which is expounded in the third Hymn. The poet was a sincere and militant Christian, but also a man sensitive to the allurements of the world and the flesh; he knew self-distrust, and weariness, and regret for evanescent youth, and is never so moving as when he writes out of such experiences or turns from this unstable life in longing for the eternal rest.

Of the romance itself, with its wealth of imagery and melody, there has never been but one opinion. Like all Spenser's work, it is highly imitative. He rifled romantic literature for incidents and situations. The descriptions in which he excels constantly recall works of art—picture, tapestry, pageant or masque. But he owed much also to real life; he was recommended for his sheriffdom as "not unskilful or without experience in the wars"; and his journeys through the wild woods of Munster and his intercourse with soldiers like Grey and Norreys yielded many a hint for the background and characters of the *F.Q.* Its diction.

archaic but not rustic, and rich in strange coinages and terms of chivalry, is in keeping with its remote and old-world air. Not less appropriate is the stanza which he invented, *not* by "extending" the ottava rima or the rhyme royal, but by adding to the linked quatrains which he had used in the *S.C.* an Alexandrine such as Ferrars had employed to tip his sestets. The elements are not new; the miracle lies in their combination. It is not a perfect stanza for narrative, nor indeed is Spenser a perfect storyteller; but its amplitude fits the slow *tempo* of his thought, as he unrolls his leisurely pageants or ponders on time and change. Its power must not be judged by single stanzas, but by the cumulative effect as stanza after stanza rolls in, each ninth wave breaking higher than the rest.

Spenser has not Chaucer's genial breadth, nor Milton's art, nor Wordsworth's vision; but in the purely poetic gifts he is inferior to Shakespeare alone. The "school of Spenser" was small and short-lived, quite eclipsed in the next generation by Jonson and Donne; but in a wider sense all later English poets have been his scholars. Beeston described Spenser to Aubrey as a little man, with short hair, little band and little cuffs—a description that fits the Pembroke College portrait better than the betuffled exquisite of the Dupplin portrait. These two portraits, whether genuine or not, do aptly illustrate the two sides of Spenser, scholar and Puritan on the one hand, courtier and man of affairs on the other; fundamentally a poet, sensitive and fastidious, yet responsive to the claim of affection or of honour. "Entire affection hateth nicer hands"; "No service loathesome to a gentle kind"; "The noblest mind the best contentment has"—in such lines we hear his authentic voice.

BIBLIOGRAPHY.—Dates of first issues are given above. Folios in 1609 (*F.Q.* only), 1611–13, 1617, 1670. Editions by J. Hughes (1715), H. J. Todd (1805), F. J. Child (1855), J. P. Collier (1862), R. Morris and J. W. Hales (1869), A. B. Grosart (1882–84), R. E. N. Dodge (1908), J. C. Smith and E. de Sélincourt (1909–10). *F.Q.* only by R. Church (1758), J. Upton (1758), K. M. Warren (1897–1900). *S.C.* by C. H. Herford (1895). *Fowre Hymnes* by L. Winstanley (1907). Commentaries and criticisms by T. Warton (1752), R. W. Church (1879), J. J. Higginson (*S.C.* only) (1912), E. Legouis (1923), W. L. Renwick (1925). Carpenter's *Reference Guide* (1923), Osgood's *Concordance* (1915) and Whitman's *Subject Index* (1918) are indispensable for detailed study. (J. C. SM.)

SPENSERIAN STANZA, a verse form so called from its invention by Edmund Spenser for his *Faerie Queene* (1590–1609). Its origin is disputed; Jakob Schipper derives it from the Old French ballade (*q.v.*), others from the Italian ottava rima (*q.v.*). Probably, however, its origin is Chaucer's eight-line *Monk's Tale* stanza, to which Spenser added a ninth line with an extra foot (alexandrine), thus making eight iambic pentameter lines and a ninth of six iambic feet. The rhyme scheme is *ababbcbcc*. Spenser's sonnet scheme is somewhat similar (*see* SONNET).

The Spenserian stanza was seldom used in the 17th century though Giles and Phineas Fletcher made adaptations of it. In the 18th century it was revived by Mark Akenside in "The Schoolmistress," James Thomson in *The Castle of Indolence* and James Beattie in *The Minstrel.* In the 19th century the Spenserian stanza was revived again by English poets of the romantic movement: Burns in "The Cotter's Saturday Night," Byron in *Childe Harold's Pilgrimage,* Keats in "The Eve of St. Agnes," Shelley in *Adonais* and "The Revolt of Islam" and Tennyson in "The Lotus-Eaters." Oliver Wendell Holmes aptly described the music and structure of the Spenserian stanza in these verses:

> The sweet Spenserian, gathering as it flows,
> Sweeps gently onward to its dying close,
> Where waves on waves in long succession pour,
> Till the ninth billow melts along the shore; . . .

(G. W. A.)

SPERANSKI, MIKHAIL MIKHAILOVICH, COUNT (1772–1839), Russian statesman who renovated the structure of administration under the emperor Alexander I, was born at Cherkutino in the Vladimir region on Jan. 12 (new style; Jan. 1, old style), 1772, the son of the village priest. After rising to be professor of mathematics and physics in the ecclesiastical seminary in St. Petersburg, he entered the government service in 1797 and quickly attracted the attention of the tsarevitch Alexander. When the latter had become emperor as Alexander I (*q.v.*), Speranski was appointed secretary of state (1801) and put to work in the ministry of the interior, where he remained until 1807. In 1808 he accompanied Alexander to Erfurt for the conference with Napoleon, who described him as "the only clear head in Russia."

From 1809 to 1812 Speranski was all-powerful in Russia; he replaced the earlier favourites at court and was practically sole minister. Intelligent, honest, hardworking and blessed with a lucid style of writing, he started many projects (including one of a constitution, 1809) and gave concrete administrative shape to Alexander's early visionary ideals. Speranski's strength lay in his mastery of bureaucratic operation, in his genius for organization and administration. He streamlined procedure and clarified objectives already adopted; he put final touches on the organization of the ministries, the council of state and the Siberian administration. Yet at the same time he gave a direction to planning and furthering the development of a systematic national policy. Liberal in that he desired spiritual, moral and economic progress, Speranski willingly supported enlightened bureaucratic absolutism as the means of effecting the necessary reforms. An introvert who bridled the expression of his own opinions, he was thwarted in many of his projects as Alexander's liberalism faded.

Speranski's absorption in work and his social isolation left him exposed to the intrigues of enemies at court. On the eve of Napoleon's invasion of Russia, the grand duchess Catherine, N. M. Karamzin, F. V. Rostopchin and G. M. Armfelt involved him in a charge of treason. Although Alexander did not credit the charge, he had begun to fear that Speranski's realization of some of the measures that he himself had conceived might undermine his autocratic authority, and on March 29 (N.S.; 17, O.S.), 1812, he dismissed Speranski and exiled him (first to Nizhni Novgorod, then to Perm). In 1816, however, Speranski was reinstated in the public service as governor of Penza, and in 1819 he was appointed governor general of Siberia, for which he drew up a new scheme of government. In 1821 he was recalled to St. Petersburg and entered the council of state. The major productions of his last years were the complete *Collection of the Laws of the Russian Empire,* 45 vol. (1830), and the systematically co-ordinated *Digest of Laws,* 15 vol. (1832–39). Created count by the emperor Nicholas I in Jan. 1839, Speranski died a few weeks later, on Feb. 23 (N.S.; 11, O.S.), in St. Petersburg.

See Marc Raeff, *Michael Speransky, Statesman of Imperial Russia, 1772–1839* (1957) and *Siberia and the Reforms of 1822* (1956).

(G. A. LN.)

SPERMACETI, a wax that separates from the oil of the sperm whale and other cetaceans. It is used in making candles, in the dressing of fabrics, in medicine and surgery, and in cosmetic preparations. *See* OILS, FATS AND WAXES; WHALE OIL.

SPERMOPHILE, a name applied to a group of squirrellike rodents (*Citellus*), widely distributed in the northern hemisphere. Most of them hibernate and all have internal cheek pouches. In Russia they are known as susliks and in the U.S. as ground squirrels or gophers. *See* GROUND SQUIRREL; SUSLIK.

SPERMOPHYTA (SPERMATOPHYTA), the name often used in botany to denote the seed plants, including the true flowering plants or angiosperms (*q.v.*) and the conifers and their allies or gymnosperms (*q.v.*).

SPERM WHALE (*Physeter catodon*), the largest of the toothed whales, called also cachalot, its size being about equal to those of the arctic right whale. The head is about one-third of the length of the body, massive, high and truncated in front. It owes its size and form to the accumulation of a peculiar form of fatty tissue in the large hollow on the upper surface of the skull. The oil contained in cells in this cavity, when refined, yields spermaceti, and the sperm oil of commerce comes principally from the thick covering of blubber which envelops the body. Ambergris, used in perfumery, is a secretion formed in the intestine, and found floating on the sea. *See* WHALE; SPERMACETI.

SPERRY, ELMER AMBROSE (1860–1930), U.S. inventor, whose most important inventions were based upon the application of the gyroscope (*q.v.*), was born at Cortland, N.Y., on Oct. 12, 1860, and studied at Cornell university, Ithaca, N.Y. He

began early to experiment with electric arc lights and in 1879 established a factory in Chicago to manufacture an improved model which was highly successful. He then invented the electric rotary and chain undercutting mining machines, and other mining apparatus and established another factory to manufacture this machinery. He next designed an electric industrial locomotive and motor transmission machinery for street cars. After 1900 he established a research laboratory in electrochemistry at Washington, D.C., where he invented the chlorine detinning process and a process for producing white lead from impure lead. In 1915 he produced his high-intensity arc searchlight.

Chief of Sperry's inventions related to the gyroscope is the gyrocompass (*q.v.*), first installed on the battleship "Delaware" in 1911. This compass, unaffected by iron and steel, entirely superseded the magnetic compass on submarines and battleships and was placed likewise on merchant ships. Sperry also designed and manufactured electrically sustained gyros for torpedoes. During World War I he developed aerial torpedoes with automatic gyrocontrol which proved effective at a range of 35 mi. Further experiments yielded the gyropilot for the steering of ships, the automatic gyropilot for stabilizing airplanes, the gyroscopic roll and pitch recorder for the testing of ships and, finally, a gyroscopic ship stabilizer that was installed on a number of ships in the U.S., British, Italian and Japanese navies. Sperry died at Brooklyn, June 16, 1930.

See "The Engineering and Scientific Achievements of Elmer A. Sperry," *Mech. Engng., N.Y.*, vol. xlix (1927). (S. C. Hr.)

SPES, the deified abstraction of hope (properly *bona spes,* "good expectation") worshiped at Rome at a comparatively early date. She had a temple in the Forum Holitorium built during the first Punic War. It was destroyed by fire in 31 B.C. and if not restored by Augustus was substantially remodeled by Germanicus in A.D. 17. It is possible that she was originally an agricultural deity; *i.e.*, the hope of good harvest. In imperial times Spes was emphasized as the hope of the future in the younger generation. She was represented as a maiden holding in her right hand an opening bud, while her left hand lifts her skirt as she speeds along.

See H. L. Axtell, *Deification of Abstract Ideas,* p. 18 ff. (1907); G. Wissowa, *Religion und Kultus,* 2nd ed., p. 329 ff. (1912); Platner-Ashby, *Topographical Dictionary,* s.v. (R. B. Ld.)

SPEUSIPPUS (d. 339/338 B.C.), the successor of Plato as head of the Greek Academy (*see* Academy, Greek), was a nephew of Plato and had played a great part in his uncle's relations with Dionysius II of Syracuse and with Dion (*see* Plato). An extant letter from Speusippus to Philip of Macedon, in which he violently attacks Isocrates (his rival for Philip's favour), is further evidence of his political interests.

Of Speusippus' philosophical writings very little more than the titles survives, so that for knowledge of his doctrines we have to rely on reports and scanty fragments, the longest being from a work *On Pythagorean Numbers* (explaining such matters as the "perfections" of the number ten). He followed the lines of what Aristotle presented as Plato's doctrines; viz., the timeless derivation of all reality, intelligible and sensible, from two opposite principles, often called "the One" and "the indeterminate dyad" (the modern terms "constant" and "variable" perhaps convey somewhat similar notions), which were assumed to explain the presence of both unity and multiplicity in the universe. To this doctrine Speusippus gave a peculiar turn. Whereas other Academics assumed that "the One" and "the dyad" would also be principles of good and evil respectively and be so by being themselves good and evil, Speusippus denied that "the One" was good or "the dyad" evil; in the same context he also denied the identification of "the One" with the divine mind. Specifically, good and even being itself would appear only in the later stages of the derivative process mentioned above, so that "the One" would be no more than the germ of the good and of being and thus less than either. Such a doctrine would make Speusippus a kind of evolutionist (with the good as an end product). Yet the reports of Aristotle on which such an interpretation is based are perhaps ambiguous and misleading. There is indeed some evidence that what Speusippus meant was that "the One" is non-being in the same sense in which the Plotinian "One" is non-being; viz., by being the source of all being and thus superior to it. Another aspect of the same doctrine is the derivation of successive spheres of reality—with numbers (which exist independently from sensibles and which in Speusippus' doctrine take the place of Plato's ideas) and geometrical magnitudes as the first sphere or spheres and with sensibles as the last. "The One" and "the dyad" appear in each sphere in an appropriately different form, which Aristotle criticized as making the spheres mutually independent, so that Speusippus' universe looks like an incoherent tragedy consisting of episodes and lacking dramatic unity. According to Aristotle, between the spheres of mathematicals and of the sensible (body) Speusippus inserted the sphere of the soul. Iamblichus, however, reports that Speusippus defined the (world?-) soul as the idea (in the sense of "form") of what extends in all dimensions, which makes the soul a mathematical entity (geometrical form). In any case Speusippus in some way seems to have intimately connected the soul with mathematicals—as did Plato in the *Timaeus* and Xenocrates in defining the soul as self-changing number. As to individual souls, Speusippus asserted the immortality of the whole soul, not only of its reasonable parts.

On the whole, Aristotle is highly critical of Speusippus. Nevertheless, Speusippus' *Similitudes,* a kind of comparative anatomy or physiology of plants and animals, are comparable to Aristotle's *History of Animals.* Whether the *Similitudes* were ultimately empirical or logico-ontological is dubious; if the latter interpretation is the right one, the *Similitudes,* together with his classification of names into synonyms, homonyms, etc., may somehow have been connected with Speusippus' doctrine that no single thing can be defined unless all have been, since classification and definition are closely related.

In introducing the concept of "knowledgelike sensation," a concept related to problems in Plato's *Theaetetus,* Speusippus perhaps tried to soften the strict ontologico-epistemological dualism of Plato. (*See* Knowledge, Theory of: *Plato.*) Many scholars see in Speusippus the representative of strict antihedonism, discussed in Plato's *Philebus.* His definition of happiness as the perfection of one's natural condition tends to derive the norm of conduct from nature.

Bibliography.—For the fragments and relevant testimonies *see* P. Lang, *De Speusippi academici scriptis* (1911). *See* further D. S. Margoliouth, "Book of the Apple . . . ," *Journal of the Royal Asiatic Society* (1892); E. Hambruch, *Logische Regeln der platonischen Schule in der aristotelischen Topik* (1904); E. Frank, *Plato und die sogenannten Pythagoreer* (1923); E. Bickermann and J. Sykutris, "Brief an König Philipp," *Sächsische Akademie der Wissenschaften: Verhandlungen: philologisch-historische Klasse,* 80 (1927); J. Stenzel, "Speusippus," Pauly-Wissowa, *Realencyclopädie der classischen Altertumswissenschaft,* 2nd series, iii, part 2 (1929); P. Merlan, *From Platonism to Neoplatonism* (1953); also Proclus' commentary on Plato's *Parmenides* as ed. by R. Klibansky and C. Labowsky, *Plato latinus,* iii (1953). (Pp. M.)

SPEY, river in the highlands of Scotland, rising from Loch Spey (1,143 ft.) in the Corrieyairack forest, Invernesshire, about 10 mi. S.S.E. of Fort Augustus. It runs 107 mi. N.E. and enters the Moray firth at Kingston. It is famous for its salmon fishing, and is fed by waters from the Monadhliath hills, the Grampians and the Cairngorms.

SPEYER or Spires, a town and episcopal see of Germany, in Rhineland-Palatinate, situated on the left bank of the Rhine, at the mouth of the Speyerbach, 16 mi. S. of Mannheim by rail. Pop. (1950) 31,841. Speyer, known to the Romans as Augusta Nemetum or Nemetae, and to the Gauls as Noviomagus, appears under the form of Spira about the 7th century. Captured by Julius Caesar in 47 B.C., it was repeatedly destroyed by the barbarian hordes in the first few centuries of the Christian era. The town had become an episcopal seat in the 4th century; but heathenism supervened, and the present bishopric dates from 610. In 830 Speyer became part of the Frankish empire. In 1294 it rose to be a free city. From 1801 until 1814 it was the capital of a *département* of France, but was restored to Bavaria in 1814.

A sandstone basilica, the cathedral is probably the earliest Romanesque basilica in which the nave as well as the side arcades

was vaulted from the first. Built in 1030–1061 by Conrad II and his successor, this church has had a chequered history, its disasters culminating in 1689, when the soldiers of Louis XIV burned it to the bare walls and scattered the ashes of eight German emperors. Restored in 1772–84 and provided with a vestibule and façade, it was again desecrated by the French in 1794; but in 1846–53 it was once more thoroughly restored. Its manufactures include paper, tobacco and cigars, sugar, machines, lead, vinegar, beer and musical instruments.

SPEZIA (La Spezia), a city of Liguria, Italy, 49 ft. above sea level. Pop. (1957 est.) 117,914 (commune). It is the chief naval harbour of Italy; having been adopted as such in 1861 though Napoleon I had entertained the idea. The Bay of Spezia (ancient Portus Lunae) is sheltered from all except southerly winds, and on its western shore are numerous openings, which afford perfectly safe anchorage in all weathers. Its important arsenal dates from 1870. The museum contains a unique collection of menhirs from the neighbourhood cut into the form of human figures. Spezia probably arose after the destruction of Luna. It was sold by the Fieschi in 1276 to Genoa. The town was fortified by its new possessors and made the seat of a governor of some importance. It became a city in the 16th century.

SPHALERITE (Blende or Zincblende), a sulfide of zinc, is the chief ore of that metal. (For production and uses *see* Zinc.). Sphalerite, the commonest zinc mineral, and galena, the most abundant ore of lead, are intimately associated in most of the important deposits of these metals. The name derives from a Greek word for "treacherous," in allusion to the ease with which the dark-coloured, opaque varieties are mistaken for galena. The synonyms blende and zincblende, from the German word for "blind," similarly allude to the fact that it does not yield lead. Associates, in addition to galena, include chalcopyrite and pyrite among metallic minerals, and quartz, barite, dolomite, calcite and siderite as gangue minerals.

In the United States the most important deposits of sphalerite are found in the Mississippi valley region. There it occurs chiefly with chalcopyrite, galena, marcasite and dolomite in solution cavities and brecciated zones in limestone and chert. Similar deposits occur elsewhere—in the region of Silesia in Poland, the Moresnet district in Belgium and north Africa. Sphalerite also has world-wide distribution as an ore mineral in hydrothermal vein deposits, in contact metamorphic zones and in high-temperature replacement deposits. Crystals of outstanding beauty and perfection are found in cavities in a dolomite rock in the Binnenthal, Switz., and a transparent variety of a golden colour occurs at Picos de Europa, Santander, Spain. Fine specimens also come from Cornwall, Derbyshire and Cumberland in England, and from Bottino in Tuscany, Italy.

The crystal structure of sphalerite is based on a tetrahedral, four-co-ordinated framework of zinc and sulfur atoms in which the bonding is essentially covalent. The structure is similar to that of diamond, and may be imagined as derived therefrom by the alternate substitution of zinc and sulfur atoms for carbon. The unit cell dimension is 5.40 Å. The common crystal forms are the tetrahedron and the dodecahedron. The crystals frequently occur as simple or multiple contact twins, or as complicated lamellar intergrowths, in which the twin axis is [111]. Sphalerite ordinarily occurs in massive granular form, less commonly as fibrous aggregates with a layered structure or as dense microcrystalline masses. It has a well-developed dodecahedral cleavage, and the fracture is conchoidal. The hardness is 3.5 to 4, and the specific gravity of pure material is 4.0. The colour varies widely. Generally it is a shade of brown, but some sphalerite is green, yellow or, in crystals of high purity, almost colourless. Material containing iron is brownish-black or black. Sphalerite may fluoresce when irradiated with ultraviolet light or X-rays, and some specimens emit light when scratched with a sharp point (triboluminescence). According to the formula, ZnS, sphalerite contains 67.1% Zn and 32.9% S. Usually, however, it contains iron in solid solution up to a limit of about 26%, and both manganese and cadmium may be present in small amounts. (Cl. F.)

SPHENE (Titanite) is a titanium-bearing mineral consisting of calcium titanosilicate. Sphene crystals vary considerably in habit, but are generally thin and wedge shaped; hence the name, from a Greek word for "a wedge." The colour is green, yellow, brown or black, and the lustre resinous to adamantine. Crystals, which are monoclinic, are transparent to opaque. The composition is $CaTiSiO_5$; hardness is 5.5; specific gravity 3.5. Sphene is cut as a gem stone, though it is too soft to stand much wear; because of its high dispersive power it gives brilliant flashes of prismatic colours. As small crystals, it is found in igneous rocks (granite, syenite, trachyte, phonolite, etc.) and in gneiss, schist, crystalline limestone and pegmatite.

SPHENISCIDAE: *see* Penguin.

SPHENODON: *see* Reptiles.

SPHENOPSIDA, a class (or subphylum) of early spore-bearing vascular plants whose sole living genus, *Equisetum,* comprises the horsetails. The group flourished as trees (*e.g., Calamites* species) during the Carboniferous period (about 285,000,000 years ago) and appeared even earlier. In its fossil history it is believed to have constituted a larger portion of the flora of the earth than at the present time. Other fossil genera were possibly shrublike (*Calamophyton, Sphenophyllum*). Sphenopsids are characterized by jointed stems and by leaves arranged in whorls. *See* Horsetail; Paleobotany; Tracheophyta.

SPHERE, in Greek geometry, the solid generated by the revolution of a semicircle about its diameter as an axis. In modern elementary geometry it is more generally considered as the spherical surface so generated, or as the locus of points at a constant distance from a fixed point called the *centre.* The constant distance is called a *radius,* and any line segment through the centre and limited at both ends by the sphere is called a *diameter.* In the geometry of surfaces it is defined as the quadric surface passing through the circle at infinity. Any section made by a plane cutting a sphere is a circle, being a *great circle* if the plane passes through the centre, and otherwise a *small circle.* The solid cut off by the plane of a great circle, is a *hemisphere;* that cut off by the plane of a small circle, a *segment.* In analytical geometry, the equation of a sphere in rectangular Cartesian co-ordinates is $x^2+y^2+z^2=a^2$, and in polar co-ordinates it is $r=a$. If the centre is (α, β, γ), the Cartesian equation becomes

$$(x-\alpha)^2+(y-\beta)^2+(z-\gamma)^2=a^2.$$

The surface of a sphere is found by the formula $S=4\pi r^2$, the area being four times that of a great circle. The volume is $\frac{4}{3}\pi r^3$. Analogous to the propositions of plane geometry and trigonometry is a series of propositions relating to the sphere. (*See* Analytical Geometry; Astronomy; Geodesy; Projective Geometry; Solids, Geometric; Surface; Trigonometry.)

SPHERES OF INFLUENCE. "Spheres of influence" is a diplomatic term which signifies the claim on the part of a state to exclusive control within a foreign territory. This control can be economic, military or political, or it may concern the overall government of the territory. The territory which is made the object of a sphere of influence may be either politically unorganized, as were some of the African territories which were transformed into spheres of influence toward the end of the 19th century, or it may be under the sovereignty of a weak government, as were, for instance, Morocco, Ethiopia and Persia when they were divided into spheres of influence at the beginning of the 20th century. The term may refer to a political claim to exclusive control, which other nations may or may not recognize as a matter of fact, or it may refer to a legal agreement by which another state or states pledge themselves to refrain from interference within the sphere of influence.

It is in the latter, legal significance that the term first gained currency in the 1880s. Its coming into usage was closely connected with the consummation of the colonial expansion of the European powers in Africa and Asia. The last stage of that expansion was characterized by the endeavour of all major colonial powers to carry on the mutual competition for colonies peacefully through agreed-upon procedures and the appeal to common standards of action rather than adoption of violent means. Agreements on spheres of influence served this purpose. Thus, the agreement

between Great Britain and Germany in May 1885, the first to make use of the term, provided for "a separation and definition of their respective spheres of influence in the territories of the Gulf of Guinea." In it "Great Britain engages not to make acquisitions of territory, accept Protectorates, or interfere with the extension of Guinea which lies east of" a defined line. "Germany engages not to make acquisitions, accept Protectorates, or interfere with the extension of British influence in that part of the coast and interior of Guinea which lies between the line as above described . . . and the British Colony of Lagos. Both Powers agree to withdraw all Protectorates which they have already established within the limits hereafter assigned to the other party . . . Germany declares herself ready . . . to refrain from making acquisitions of territory or establishing Protectorates on the coast between the Colony of Natal and Delagoa Bay."

This agreement was followed by many of a similar nature, of which article vii of the agreement between Great Britain and Germany of July 1, 1890, concerning East Africa may be regarded as typical. Its text is as follows: "The two Powers engage that neither will interfere with any sphere of influence assigned to the other by Articles I to IV. One Power will not in the sphere of the other make acquisitions, conclude Treaties, accept sovereign rights or Protectorates, nor hinder the extension of influence of the other. It is understood that no Companies nor individuals subject to one Power can exercise sovereign rights in a sphere assigned to the other, except with the assent of the latter."

Agreements of this type concerned territories politically unorganized and therefore subject to occupation by whatever nation complied with the rules of international law relative to the acquisition of unoccupied territory. Agreements of another type, often called spheres of interest, refer to territories which are already politically organized and are to be transformed into a sphere of influence of one of the contracting parties. This type is exemplified by article ii of the Anglo-French agreement of April 8, 1904, concerning Morocco. It provides: "The Government of the French Republic declare that they have no intention of altering the political status of Morocco. His Britannic Majesty's Government, for their part, recognize that it appertains to France, more particularly as a Power whose dominions are conterminous for a great distance with those of Morocco, to preserve order in that country, and to provide assistance for the purpose of all administrative, economic, financial, and military reforms which it may require. They declare that they will not obstruct the action taken by France for this purpose. . . ."

Of the legal effects of such agreements, three are worthy of note. First, agreements establishing spheres of influence create rights and obligations between the contracting parties only. States which are not parties to the agreement are legally free to acquire within the sphere of influence whatever rights they can; they have, however, for political reasons generally refrained from interference and acted as though they, too, were bound by the agreement. Second, such agreements do not affect by themselves the actual control over the territories concerned. So long as the state in whose favour a sphere of influence has been established chooses not to exercise its exclusive rights within that sphere, the actual control over it remains in the hands of those who exercised it when the agreement was concluded. Third, once the state in whose favour a sphere of influence has been established makes use of its rights within that sphere, all states, whether or not they are parties to the agreement, are excluded from the territory to the extent that the former state exercises actual control.

When the period of colonial expansion came to a close with the outbreak of World War I, spheres of influence in the legal sense lost much of their importance. From then on, the term was used more and more for the unilateral political claims of great powers, as old as political history itself, to a predominant position in the territories of other nations, especially those adjacent to their own. Thus, it can be said that the Monroe Doctrine, especially in its application after the time of Theodore Roosevelt, established the western hemisphere as a sphere of influence for the United States. In the aftermath of World War II, the Soviet Union created a sphere of influence as a political fact in the territories of the nations of eastern Europe and of the Balkans, with the exception of Greece (and later Yugoslavia). The political division of the Balkans into spheres of influence in the political sense was foreshadowed in the agreements concluded between Great Britain and the Soviet Union in 1944 and on the face of them intended only to facilitate military operations and not to outlast the war. These agreements were opposed by U.S. Secretary of State Cordell Hull, who was not "a believer in the idea of balance of power or spheres of influence as a means of keeping the peace" (from Cordell Hull, *The Memoirs of Cordell Hull*, vol. ii, p. 1452, copyright 1948 by Cordell Hull, the Macmillan Company, 1948). Regardless of their evaluation, it seems that spheres of influence in the political sense are an inevitable concomitant of the struggle for power of sovereign nations. (*See* INTERNATIONAL LAW; MONROE DOCTRINE.)

BIBLIOGRAPHY.—M. F. Lindley, *The Acquisition and Government of Backward Territory in International Law* (London, 1926); Frederick L. Schuman, "Regionalism and Spheres of Influence," in Hans J. Morgenthau (ed.), *Peace, Security and the United Nations* (1946); W. W. Willoughby, *Foreign Rights and Interests in China*, 2nd ed., vol. i, pp. 130–168 (1927). (H. J. MU.)

SPHERICAL HARMONICS are certain functions which occur in the theories of gravitation, electricity, hydrodynamics, and many other branches of mathematical physics.

If a number of particles of masses $M_1, M_2, \cdots, M_m$, attracting according to the inverse-square law, are placed at points $A_1, A_2, \cdots, A_m$, the *potential* of these masses at any point P, *i.e.*, the work done by the gravitational attractions in bringing a unit mass from an infinite distance to P, is

$$\frac{M_1}{A_1P}+\frac{M_2}{A_2P}+\cdots+\frac{M_m}{A_mP}.$$

If now the points $A_1, A_2, \cdots, A_m$, P are referred to a set of rectangular co-ordinate axes, and the co-ordinates of the points are $(a_1, b_1, c_1), (a_2, b_2, c_2), \cdots, (a_m, b_m, c_m), (x, y, z)$ so that

$$A_1P=\sqrt{\{(x-a_1)^2+(y-b_1)^2+(z-c_1)^2\}}$$

with similar expressions for $A_2P, A_3P, \cdots, A_mP$, and if V denotes the potential, it may be verified by partial differentiations that

$$\frac{\partial^2V}{\partial x^2}+\frac{\partial^2V}{\partial y^2}+\frac{\partial^2V}{\partial z^2}=0.$$

This equation does not involve the masses or the points at which they are placed, and so it is satisfied by the potential produced at a point (x, y, z), in free space, by any gravitating system; it is known as *Laplace's equation*, and it is frequently written in the form

$$\nabla^2V=0,$$

∇^2 denoting the differential operator $\dfrac{\partial^2}{\partial x^2}+\dfrac{\partial^2}{\partial y^2}+\dfrac{\partial^2}{\partial z^2}$ for brevity.

Laplace's equation occurs in problems of Gravitation, Electrostatics and Hydrodynamics; the related equation $\nabla^2V=k\partial V/\partial t$ occurs in problems of conduction of heat, V being then a function of the co-ordinates x, y, z, and the time t, and k being a constant; and the equation $c^2\nabla^2V=\partial^2V/\partial t^2$ occurs in problems of Vibrations of air and ether, c being the velocity of sound or of light, according to the problem considered.

In order to discuss generally problems concerning the potential, it is found more effective to examine solutions of Laplace's equation than to regard the potential as the sum of a number of terms of the type M/AP, each term representing the potential produced by an attracting particle. If $V_1, V_2, V_3, \cdots$ are a number of solutions of Laplace's equation, and if

$$V=V_1+V_2+V_3+\cdots,$$

it is easily verified that

$$\nabla^2V=\nabla^2V_1+\nabla^2V_2+\nabla^2V_3+\cdots=0,$$

and so the sum of any number of solutions of Laplace's equation

is itself a solution of the equation. It is thus possible to build up a solution of Laplace's equation by adding together simpler solutions.

Now let V be any potential function, and suppose that it can be expanded by Taylor's theorem as a series of positive integral powers of x, y, z, thus

$$V = a_0+a_1x+b_1y+c_1z+a_2x^2+b_2y^2 \\ +c_2z^2+d_2yz+e_2zx+f_2xy+a_3x^3+\cdots.$$

Group together all terms which are of the same degree in x, y, z, and denote the groups of terms which are of degrees $0, 1, 2, \cdots, n, \cdots$ in the expansion by

$$V_0(x, y, z),\quad V_1(x, y, z),\quad V_2(x, y, z), \cdots,\quad V_n(x, y, z), \cdots,$$

or, for brevity, by $V_0, V_1, V_2, \cdots, V_n, \cdots$, so that

$$V = V_0+V_1+V_2+\cdots+V_n+\cdots.$$

The expression $V_n(x, y, z)$ is called *a spherical harmonic of degree n*. It may be verified that

$$\nabla^2V_2+\nabla^2V_3+\cdots+\nabla^2V_n+\cdots = \nabla^2V = 0$$

so that the expression on the left vanishes identically; and since the terms written down are of different degrees, $0, 1, \cdots, n-2\cdots$, each must vanish separately, so that spherical harmonics are themselves solutions of Laplace's equation.

If the polar co-ordinates of (x, y, z) are (r, θ, ϕ), so that

$$x = r\sin\theta\cos\phi,\quad y = r\sin\theta\sin\phi,\quad z = r\cos\theta,$$

then $V_n(x, y, z)$ may be written in the form $r^nf_n(\theta, \phi)$, where $f_n(\theta, \phi)$ is independent of r; and $f_n(\theta, \phi)$ is called a *surface harmonic*.

It can be proved that $V_n(x, y, z)/r^{2n+1}$ is also a solution of Laplace's equation, and it is called a spherical harmonic of degree $-(n+1)$; this function and $V_n(x, y, z)$ are called *solid* harmonics.

The term spherical harmonics (German, *Kugelfunktionen;* French, *Fonctions sphériques*) is due to Lord Kelvin.

When referred to polar co-ordinates, Laplace's equation becomes

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2V}{\partial\phi^2} = 0;$$

this was the form of the equation originally given by P. S. de Laplace in 1784, in a memoir on Saturn's rings.

Determination of Harmonics of Given Degree.—Let $a_{r,s,t}x^ry^sz^t$ be an arbitrary expression of positive integral degree n in x, y, z where r, s, t are positive integers (zero included) such that $r+s+t=n$, and the summation extends over all possible sets of values of r, s, t. It is required to find conditions that this expression may be a harmonic.

It is readily proved that the number of terms in the expression is $\frac{1}{2}(n+2)(n+1)$; if the expression is operated on by Laplace's operator, the result is an expression of degree $n-2$ in x, y, z, in which the $\frac{1}{2}n(n-1)$ coefficients are linear combinations of the $a_{r,s,t}$. If the last expression is equated to zero, $\frac{1}{2}n(n-1)$ equations connecting the $a_{r,s,t}$ are obtained, which have to be satisfied if the former expression is to be a harmonic. It follows that the number of independent coefficients $a_{r,s,t}$ is

$$\tfrac{1}{2}(n+2)(n+1)-\tfrac{1}{2}n(n-1),$$

i.e., $2n+1$, and the remaining coefficients are expressible linearly in terms of them.

Consequently there are $2n+1$ independent harmonics of degree n, and any other harmonic of degree n is expressible as a linear combination of these harmonics.

Take $n=2$ as an illustration; the expression

$$a_2x^2+b_2y^2+c_2z^2+2d_2yz+2e_2zx+2f_2xy$$

is a harmonic if $a_2+b_2+c_2=0$, and 5 independent harmonics of degree 2, in terms of which any harmonic of degree 2 may be expressed, may be taken to be

$$yz,\quad zx,\quad xy,\quad 2z^2-x^2-y^2,\quad x^2-y^2.$$

The next step is to construct $2n+1$ harmonics of degree n. This may be effected by considering the condition that

$$(ax+by+cz)^n$$

may be a harmonic, where a, b, c, are constants, and deducing the harmonics from it. By actual differentiation it follows that the condition that $(ax+by+cz)^n$ is a harmonic is

$$a^2+b^2+c^2 = 0;$$

and this condition is satisfied by taking

$$a = i\cos u,\quad b = i\sin u,\quad c = 1,$$

where u is any angle and i stands for $\sqrt{(-1)}$; and so

$$(z+ix\cos u+iy\sin u)^n$$

is always a harmonic.

Now suppose this expression to be expanded by the multinomial theorem, and replace $\cos u$ and $\sin u$ by their exponential values $(e^{iu}+e^{-iu})/2$ and $(e^{iu}-e^{-iu})/(2i)$; when the products of powers of $\cos u$ and $\sin u$ are multiplied out in this way, replace expressions such as e^{miu}, e^{-miu} by $\cos mu\pm\sin mu$.

It is thus possible to express $(z+ix\cos u+iy\sin u)^n$ in the form

$$(z+ix\cos u+iy\sin u)^n = \tfrac{1}{2}g_0(x, y, z) \\ +\sum_{m=1}^{n}[g_m(x, y, z)\cos mu+h_m(x, y, z)\sin mu],$$

where $g_m(x, y, z)$ and $h_m(x, y, z)$ are polynomials of degree n in x, y, z, and are independent of u; and since

$$\tfrac{1}{2}\nabla^2g_0(x, y, z)+\sum_{m=1}^{n}[\nabla^2g_m(x, y, z)\cos mu+\nabla^2h_m(x, y, z)\sin mu] \equiv 0,$$

for all values of u, it is evident that $g_m(x, y, z)$ and $h_m(x, y, z)$ must be harmonics. As an illustration,

$$(z+ix\cos u+iy\sin u)^2 = z^2-\tfrac{1}{2}x^2-\tfrac{1}{2}y^2+2ixz\cos u \\ +2iyz\sin u+\tfrac{1}{2}(y^2-x^2)\cos 2u-xy\sin 2u.$$

To obtain $g_m(x, y, z)$ and $h_m(x, y, z)$ explicitly, use Fourier's rule for determining the coefficients in a trigonometrical series, *i.e.*, multiply the expansion of $(z+ix\cos u+iy\sin u)^n$ by $\cos mu$ and by $\sin mu$, and integrate with respect to u from $-\pi$ to π. It follows that

$$\pi g_m(x, y, z) = \int_{-\pi}^{\pi}(z+ix\cos u+iy\sin u)^n\cos mu\ du, \\ (m = 0, 1, 2, \cdots, n),$$

$$h_m(x, y, z) = \int_{-\pi}^{\pi}(z+ix\cos u+iy\sin u)^n\sin mu\ du \\ (m = 1, 2, 3, \cdots, n).$$

It will be proved later that these functions $g_m(x, y, z)$ and $h_m(x, y, z)$ are linearly independent, and so they form a set of $2n+1$ independent harmonics.

The General Solution of Laplace's Equation.—It has just been seen that any harmonic of degree n may be expressed in the form

$$\sum_{m=0}^{n}a_mg_m(x, y, z)+\sum_{m=1}^{n}b_mh_m(x, y, z),$$

where a_m and b_m are constants. If now, for brevity,

$$\sum_{m=0}^{n}a_m\cos mu+\sum_{m=1}^{n}b_m\sin mu$$

be called $\pi f_n(u)$, it follows that *any* harmonic of degree n is expressible in the form

$$\int_{-\pi}^{\pi}(z+ix\cos u+iy\sin u)^nf_n(u)du,$$

where $f_n(u)$ is a function of u; and, conversely, if $f_n(u)$ is an *arbitrary* function of u, it may be verified by substituting the expression in Laplace's equation that it is a harmonic of degree n.

Next take *any* solution V of Laplace's equation and suppose it to be expansible in a convergent series of positive integral powers of x, y, z; group together all the terms of the same degree and replace each group of terms of the same degree n by an expression of the form

$$\int_{-\pi}^{\pi}(z+ix\cos u+iy\sin u)^n f_n(u)du;$$

it follows that

$$V=\sum_{n=0}^{\infty}\int_{-\pi}^{\pi}(z+ix\cos u+iy\sin u)^n f_n(u)\,du;$$

and this value of V is expressible in the form

$$V=\int_{-\pi}^{\pi}f(z+ix\cos u+iy\sin u,\ u)du,$$

where f denotes a function of the two arguments $z+ix\cos u+iy\sin u$ and u; and conversely if f denotes an arbitrary function of its two arguments, the expression on the right may be shown to satisfy Laplace's equation.

It is possible to remove the restriction concerning the possibility of expanding V as a series of positive integral powers of x,y,z; and hence *the general solution of Laplace's equation may be written in the form*

$$V=\int_{-\pi}^{\pi}f(z+ix\cos u+iy\sin u,\ u)du.$$

This solution was obtained by E. T. Whittaker in 1902; it may be compared with the solution of the equation

$$\frac{\partial^2 W}{\partial x^2}+\frac{\partial^2 W}{\partial y^2}=0,$$

(known as Laplace's two-dimensional equation) namely

$$W=f(x+iy)+g(x-iy)$$

where f and g denote arbitrary functions.

Zonal, Tesseral and Sectorial Harmonics.—The $2n+1$ harmonics of degree n which have been constructed take more simple forms when referred to polar co-ordinates; it is, in fact, possible to establish the expansion:

$$(z+ix\cos u+iy\sin u)^n = r^nP_n(\cos\theta)+2\sum_{m=1}^{n}\frac{n!\,i^m}{(n+m)!}r^nP_n^m(\cos\theta)(\cos m\phi\cos mu+\sin m\phi\sin mu)$$

where $P_n(\cos\theta)$ depends on θ, but not on ϕ, and $P_n^m(\cos\theta)=\sin^m\theta\dfrac{d^mP_n(\cos\theta)}{d(\cos\theta)^m}$.

From the form of the above expansion, it is evident that $r^nP_n(\cos\theta)$ is identical with $\frac{1}{2}g_0(x,y,z)$ so that

$$P_n(\cos\theta)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\{\cos\theta+i\sin\theta\cos(\phi-u)\}^n du,$$

it follows from the statement that $P_n(\cos\theta)$ is independent of ϕ that it is permissible to put $\phi=0$ in the last formula without affecting the result; and so

$$P_n(\cos\theta)=\frac{1}{2\pi}\int_{-\pi}^{\pi}(\cos\theta+i\sin\theta\cos u)^n du,$$

a formula known as *Laplace's integral*.

From the result stated, it follows that $2n+1$ harmonics of degree n are

$$r^nP_n(\cos\theta),\ r^nP_n^m(\cos\theta)\cos m\phi,\ r^nP_n^m(\cos\theta)\sin m\phi,\quad (m=1,2,\cdots,n),$$

and, from the manner in which ϕ occurs in them, they are obviously linearly independent.

An alternative definition of $P_n(\cos\theta)$ is derived from the fact that the potential produced at (x,y,z) by a unit mass at $(0,0,a)$ is

$$1/\sqrt{\{x^2+y^2+(a-z)^2\}}=1/\sqrt{(a^2-2a\mu+r^2)},$$

in polar co-ordinates, where μ denotes $\cos\theta$; when this potential function is expanded in ascending powers of r, the result is of the form

$$\frac{1}{a}+\frac{rP_1(\mu)}{a^2}+\cdots+\frac{r^nP_n(\mu)}{a^{n+1}}+\cdots,$$

where $P_n(\mu)$ depends on μ and n only; and it may be seen from the expression in Cartesian co-ordinates that $r^nP_n(\mu)$ is a harmonic of degree n; it is possible to identify $P_n(\mu)$, defined in this manner, with Laplace's integral; and it can be proved that

$$P_1(\mu)=\mu,\quad P_2(\mu)=\frac{3\mu^2-1}{2},\quad P_3(\mu)=\frac{5\mu^3-3\mu}{2},\cdots,$$

$$P_n(\mu)=\frac{(2n)!}{2^n\cdot(n!)^2}\left[\mu^n-\frac{n(n-1)}{2(2n-1)}\mu^{n-2}+\frac{n(n-1)(n-2)(n-3)}{2\cdot4(2n-1)(2n-3)}\mu^{n-4}-\cdots\right].$$

If a sphere is drawn with centre at the origin, the locus of points on the sphere at which $P_n(\cos\theta)$ vanishes consists of the n parallels of latitude on which the cosine of the co-latitude is a root of the equation $P_n(\mu)=0$; the locus thus divides the sphere into $n+1$ zones, and for this reason $r^nP_n(\cos\theta)$ is called a *zonal* harmonic.

In like manner when $n>m\geqq1$, $P_n^m(\cos\theta)$ vanishes on $n-m$ parallels of latitude, while $\cos m\phi$ vanishes on $2m$ meridians, and $\sin m\phi$ vanishes on $2m$ other meridians; the sphere is thus divided into $2m(n-m-1)$ rectangular regions, with $2m$ triangular regions round each pole, and for this reason $r^nP_n^m(\cos\theta)\cos m\phi$ and $r^nP_n^m(\cos\theta)\sin m\phi$ are called *tesseral* harmonics ($\tau\acute{\epsilon}\sigma\sigma\epsilon\rho\alpha=$ a square).

For similar reasons $r^nP_n^n(\cos\theta)\cos n\phi$ and $r^nP_n^n(\cos\theta)\sin n\phi$ are called *sectorial* harmonics.

The function $P_n(\cos\theta)$ is called a *Legendre coefficient*, or a *Legendre function*, while $P_n^m(\cos\theta)$ is called an *associated Legendre function*.

Normal Solutions of Laplace's Equation.—The zonal, tesseral and sectorial harmonics are of the form $R\cdot\Theta\cdot\Phi$ where R,Θ,Φ respectively involve the co-ordinates, r,θ,ϕ, only; a solution consisting of the product of three functions each involving only one co-ordinate is called a *normal solution*.

Thus, if $X\cdot Y\cdot Z$ is a normal solution of Laplace's equation in Cartesian co-ordinates, it follows, on substitution and division by XYZ, that

$$\frac{1}{X}\frac{\partial^2X}{\partial x^2}+\frac{1}{Y}\frac{\partial^2Y}{\partial y^2}+\frac{1}{Z}\frac{\partial^2Z}{\partial z^2}=0,$$

and so

$$\frac{1}{X}\frac{\partial^2X}{\partial x^2},\quad \frac{1}{Y}\frac{\partial^2Y}{\partial y^2},\quad \frac{1}{Z}\frac{\partial^2Z}{\partial z^2}$$

must be respectively independent of x,y,z; hence they are constants. If these constants are called a^2, b^2, c^2, where $a^2+b^2+c^2=0$, we obtain the solutions

$$V=e^{\pm ax\pm by\pm cz}.$$

Similarly, if $R\Theta\Phi$ is a normal solution in polar co-ordinates,

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{1}{\Theta\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\frac{1}{\Phi\sin^2\theta}\frac{d^2\Phi}{d\phi^2}=0,$$

and so $\dfrac{1}{R}\dfrac{d}{dr}\left(r^2\dfrac{dR}{dr}\right)$ and $\dfrac{1}{\Phi}\dfrac{d^2\Phi}{d\phi^2}$ must be constants, say $n(n+1)$ and $-m^2$; and so Θ satisfies the equation

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\left(n[n+1]-\frac{m^2}{\sin^2\theta}\right)\Theta=0,$$

while $\dfrac{d}{dr}\left(r^2\dfrac{dR}{dr}\right)-n(n+1)R=0$, $\dfrac{d^2\Phi}{d\phi^2}+m^2\Phi=0$.

Possible values of R are r^n and r^{-n-1}; and possible values of Φ are $\cos m\phi$ and $\sin m\phi$. But now n and m are not necessarily integers, and so the question arises of defining $P_n^m(\cos\theta)$ in such circumstances.

This problem is dealt with by means of the hypergeometric function of Gauss defined by the equation

$$F(\alpha,\beta;\gamma;z)=1+\frac{\alpha\beta}{\gamma\cdot1}z+\frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)\cdot1\cdot}z^2+\cdots.$$

A solution of the equation which it is convenient to adopt as the definition of $P_n^m(\cos\theta)$ is

$$P_n^m(\cos\theta) = \frac{1}{\Gamma(1-m)}\left(\frac{\mu+1}{\mu-1}\right)^{\frac{1}{2}m} F(-n, n+1; 1-m; \tfrac{1}{2}-\tfrac{1}{2}\mu),$$

where $\Gamma(1-m)$ denotes the gamma function of Euler which reduces to $(-m)!$ when m is a negative integer (when m is a positive integer, this function differs from the associated Legendre function, as previously defined, by a constant factor). The theory of generalized Legendre functions is consequently merged into the wider theory of hypergeometric functions.

Cylindrical and Ellipsoidal Harmonics.—The cylindrical co-ordinates (ρ,ϕ,z) of a point are defined by the equations

$$x=\rho\cos\phi, \quad y=\rho\sin\phi, \quad z=z.$$

When the solution $e^{k(x+iz\cos u+iy\sin u)}$ of Laplace's equation is expressed in terms of such co-ordinates it becomes:

$$e^{kz}e^{ik\rho\cos(\phi-u)}$$

and this can be expanded in the form

$$e^{kz}\Big[J_0(k\rho)+2\sum_{m=1}^{\infty} i^m J_m(k\rho)\,(\cos m\phi\cos mu+\sin m\phi\sin mu)\Big]$$

where

$$J_m(k\rho) = \sum_{n=0}^{\infty}\frac{(-1)^n(\frac{1}{2}k\rho)^{m+2n}}{n!(m+n)!}$$

This leads to the consideration of the *cylindrical harmonics*

$$e^{kz}J_0(k\rho), \quad e^{kz}J_m(k\rho)\cos m\phi, \quad e^{kz}J_m(k\rho)\sin m\phi,$$

which may also be derived as normal solutions of Laplace's equation in cylindrical co-ordinates, viz.

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial V}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2 V}{\partial\phi^2}+\frac{\partial^2 V}{\partial z^2}=0.$$

The functions $J_m(k\rho)$ are known as *Bessel functions* (*see* Functions, Special). They occur in numerous physical problems of diverse types including Bessel's investigations of planetary motion, Daniel Bernoulli's problem of a vibrating hanging chain and Euler's problem of a vibrating circular membrane.

The functions which occur in a similar manner when Laplace's equation is referred to a system of confocal conicoids as surfaces of reference are known as *Lamé's functions*; these are of a more complicated type than the functions described hitherto, because they cannot be expressed as series with simple coefficients. But when the confocal conicoids are spheroids, the substitution

$$x=(r^2-c^2)^{\frac{1}{2}}\sin\theta\cos\phi, \quad y=(r^2-c^2)^{\frac{1}{2}}\sin\theta\sin\phi, \quad z=r\cos\theta$$

gives a form of Laplace's equation with normal solutions of the type

$$P_n^m(r/c)P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix} m\phi,$$

the spheroids being prolate; with oblate spheroids, c must be replaced by ic.

Later Researches.—Some of the main lines along which progress has been made include methods of transforming the hypergeometric series, obtaining approximate expressions for $P_n^m(\cos\theta)$ when n or m (or both) is large, expanding arbitrary functions of two variables as series of surface harmonics, expanding arbitrary functions $f(\mu)$ in the form $\sum a_n P_n(\mu)$, obtaining approximate formulas for $J_n(z)$ when z or n (or both) is large, obtaining formulas by which the roots of the equation $J_n(z)=0$ may be calculated, expanding arbitrary functions as series of Bessel functions, obtaining integral equations satisfied by ellipsoidal harmonics, and by the harmonic functions associated with confocal elliptic cylinders. These investigations all arose in the first instance out of some problem of mathematical physics, though many of them require the most profound methods of pure mathematics for their solution.

Bibliography.—E. W. Hobson, "A type of Spherical Harmonics," *Phil. Trans. of the Royal Soc.* 187A (1896); E. W. Barnes, "A New Development of the Theory of the Hypergeometric Functions," *Proc. London Math. Soc.* (2) 6 (1908); E. W. Barnes, "On Generalised Legendre Functions," *Quarterly Journal of Math.* 39 (1908); E. W. Hobson, "A General Convergence Theorem," *Proc. London Math. Soc.* (2) 6 (1908); A. Gray and G. B. Mathews, *A Treatise on Bessel Functions* (1922); G. N. Watson, *The theory of Bessel Functions* (1922); W. D. Niven, "On Ellipsoidal Harmonics," *Phil. Trans. of the Royal Soc.* 182A (1891); G. H. Darwin, "On Ellipsoidal Harmonics," *Phil. Trans. of the Royal Soc.* 197A (1901); "On the Pear-shaped Figure of Equilibrium of a Rotating Mass of Liquid," *Phil. Trans. of the Royal Soc.* 198A (1901); T. M. MacRobert, *Functions of a Complex Variable* (1917), *Spherical Harmonics* (1927); E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (1927). (G. N. W.)

SPHEROID, literally a sphere-like body. In geometry the term is applied to figures generated by an ellipse (q.v.) revolving about either its minor axis (giving an *oblate spheroid*) or its major axis (giving a *prolate spheroid*). If the equation of the generating ellipse is $b^2x^2+a^2y^2=a^2b^2$, and the ellipse revolves about the major axis (the x axis), the volume is $\frac{4}{3}\pi ab^2$; if about the minor axis, it is $\frac{4}{3}\pi a^2b$. The equation of the prolate spheroid is $b^2x^2+a^2(y^2+z^2)=a^2b^2$, and of the oblate spheroid it is $b^2(x^2+z^2)+a^2y^2=a^2b^2$. (*See* Ellipsoid.)

SPHERULITE, in petrology, is a small rounded body which commonly occurs in vitreous igneous rocks. Spherulites are often visible in specimens of obsidian, pitchstone and rhyolite as globules about the size of millet seed, with a duller lustre than the surrounding glassy base of the rock, and when they are examined with a lens they prove to have a radiate fibrous structure.

Spherulites are commonest in acid glassy rocks but occur also in basic glasses such as tachylyte. Sometimes they compose the whole mass; more usually they are surrounded by a glassy or felsitic base. Occasionally spherulites have zones of different colours, and while most frequently spherical they may be irregular in outline. In some New Zealand rhyolites the spherulites send branching "cervicorn" processes (like stag horns) outward through the surrounding glass of the rock. Long, elliptical or bandlike spherulites are called axiolites.

Large cavernous spherulites have been called lithophysae; they are found in obsidians at Lipari, Italy, Yellowstone National park, etc. The characteristic radiate fibrous structure is usually conspicuous, but the fibres are interrupted by cavities which are often so arranged as to give the spherulite a resemblance to a rosebud with folded petals separated by arching interspaces. Some of these lithophysae are one inch or more in diameter. Spherulites with typical radiate fibrous structure may be produced in artificial glasses, salt solutions and melts of organic substances. Their study reveals the existence of an equatorial constriction perpendicular to a line along which the growth of the original sheaf of fibres began. New fibres grow at an angle to previous fibres so that the sheaf opens up at both ends, and fibres finally meet along the equator. A cavity is left, shaped like a flattened doughnut or torus, but it is filled by later crystallizations.

(J. D. H. D.)

SPHINX, a composite creature with lion's body and human head. The name is Greek and the sphinx is famous in Greek legend. According to Apollodorus she was a winged, human-headed lion, offspring of two giants, who lived in the vicinity of Boeotian Thebes. She terrorized the people by demanding the answer to a riddle taught her by the Muses—what is it that has one voice and yet becomes four-footed and two-footed and three-footed?—and devouring a man each time the riddle was answered incorrectly. Eventually Oedipus gave the proper answer: man, who crawls on all fours in infancy, walks on two feet when grown and leans on a staff in old age; the sphinx thereupon killed herself. From this tale apparently grew the legend that the sphinx was omniscient, and even today the wisdom of the sphinx is proverbial. The word *sphinx* was derived by Greek grammarians from the verb *sphingein* (to bind or squeeze), but the etymology is not related to the legend and therefore dubious. The earliest form was apparently *phix*, for which no etymology has been suggested; probably the word was not originally Greek.

In ancient near eastern art the monster is well known. The earliest example is one of the most famous monuments of all antiquity, the colossal recumbent sphinx at Giza, in Egypt, dating from the reign of King Khafre (Chephren) of the 4th dynasty, *c.* 2550 B.C. This is known to be a portrait statue of the king, and

the sphinx continues as a royal portrait type through most of Egyptian history, symbolizing both the mighty strength and the protective power of Egypt's ruler. Through Egyptian influence the sphinx became known in Asia, especially in the Levant in the 2nd millennium B.C., where it became a popular subject on cylinder seals and other forms of decorative art. Its meaning in Asia is uncertain; it is no longer a royal portrait but may share with its Egyptian counterpart a certain protective aspect. Asian art always showed a fondness for monstrous beings, and the Egyptian sphinx may have been added to the repertory merely as another kind of monster. Although Mesopotamia was the place of origin of many monsters, the sphinx does not occur there until around 1500 B.C., when it was clearly imported from the Levant. In appearance the Asian sphinx differs from its Egyptian model most noticeably in the addition of wings to the leonine body, a feature that continues through its subsequent history in Asia and the Greek world. Another innovation is the female sphinx, which appears in the 14th century B.C. A few are found in Egypt in the 18th dynasty but it is common in the Levant, which is probably its homeland. What meaning the change of sex may have had is uncertain, but both males and females are found in similar poses and seem to play the same role in art. On seals, ivories and metalwork they sit on their haunches, often with one paw raised, and are frequently paired with a lion, a griffin or another sphinx.

Around 1600 B.C. the sphinx first appeared in the Greek world. Seal impressions and other objects from Crete at the end of the Middle Minoan period and from the shaft graves at Mycenae at the beginning of the Late Helladic Age show the sphinx, characteristically winged, and in the later part of the Late Helladic Age (14th and 13th centuries) the creature appears frequently on ivories, metal plaques and painted pottery. Although clearly derived from the Asian sphinx, the Greek examples are not identical in appearance; they customarily wear a flat cap with a flamelike projection on top. Nothing in their context connects them with the later legend and their meaning at this time is unknown.

Around 1200 B.C. came the cataclysm in both Greek and Asian lands that marks the division between the Bronze and Iron Ages; it is traditionally known in Greece as the Dorian invasion and in Asia as the invasion of the Sea Peoples. For about 400 years no sphinxes appeared in Greece, but in Asia they continued in forms and poses similar to those of the Bronze Age. By the end of the 8th century the sphinx reappeared in Greek art and was common through the orientalizing and Archaic periods (down to the end of the 6th century). It is often associated with griffins, lions and other oriental motifs and is clearly derived from the same source; from its appearance it cannot be a direct descendant of the Bronze Age Greek sphinx. The new Greek sphinx is almost always female and usually wears the long-tiered wig known on contemporary sculptures of the Daedalic style; the body becomes graceful and elegant in line and the wings develop a beautiful curving form unknown in Asia. Sphinxes decorate vases, ivories and metal works and in the late Archaic period occur as ornaments on temples. As in the Bronze Age, the context is usually insufficient to enable their meaning to be judged, but their appearance on temples (especially the large terra-cotta acroteria on a 6th-century Artemis temple at Calydon) suggests a protective function. The handsome stone sphinx atop a votive column set up by the Naxians in the Apollo precinct at Delphi corroborates a sacred meaning of some sort, as does the use of the sphinx on sepulchral monuments. By the 5th century clear illustrations of the Oedipus legend appear on vase paintings, usually with the sphinx perched on a column (*e.g.*, the red-figured Nolan amphora in Boston), but other monuments of classical age show Oedipus in armed combat with the sphinx and suggest an earlier stage of the legend in which the contest is physical instead of mental. Of such a stage the literature gives no hint, but battles of men and monsters are common in Asian art from prehistoric times down to the Achaemenid Persians, and Greek art may have adopted from the near east a pictorial theme that Greek literature did not share.

BIBLIOGRAPHY.—A. von Pauly and G. Wissowa, *Realenzyklopädie der classischen Altertumswissenschaft*, vol. ii (6) (1929); W. Roscher, *Ausführliches Lexikon der griechischen und römischen Mythologie*, vol. iv (1909–15); A. Dessenne, *Le Sphinx, Étude iconographique* (1957). (A. Ps.)

SPHRANTZES, GEORGE (1401–end of the 15th century), Byzantine historian and diplomat, wrote a chronicle covering the years 1413–77. He rose to high office in the service of Manuel II and the later Palaeologan rulers, both in Constantinople and in the Peloponnese. In 1451 Sphrantzes was great logothete (chancellor) in Constantinople and on its capture by the Ottomans he fled to Mistra and then had to retreat to Corfu, where he and his wife finally entered monastic life (1468). His chronicle is probably what is known as the *Chronicon Minus* and not the *Chronicon Maius*. The *Chronicon Maius*, considered by some to be Sphrantzes' expansion of the *Minus*, was probably compiled later by Macarius Melissenus. The chronicle was written for the Corfiotes and deals with the last years of the Palaeologi in Constantinople and the Peloponnese, and it shows marked aversion to the Ottomans and the Latins.

The *Chronicon* was edited by I. Bekker in *Corpus scriptorum historiae Byzantinae*, xxiii (1838), and in J. P. Migne, *Patrologia Graeca*, vol. 156 (1866). A critical edition of books i–ii was made by J. B. Papadopulos (1935).

See G. Moravcsik, *Byzantinoturcica*, vol. i, 2nd ed., pp. 282–288 (1958); G. Ostrogorsky, *History of the Byzantine State*, pp. 416–418 (1956). (J. M. Hx.)

J. HORACE MCFARLAND CO.

SPICEBUSH (LINDERA BENZOIN) FOLIAGE BEARING FRUIT

SPICEBUSH (*Lindera benzoin*), a North American deciduous shrub of the laurel family (Lauraceae), allied to the sassafras and Oregon myrtle (*qq.v.*), native to moist woods from Maine to Ontario and south to Florida and Texas. It grows from 4 to 20 ft. high, with smooth aromatic twigs and foliage, and bears bright yellow, fragrant flowers, which unfold before the leaves, and scarlet fruit (drupe), maturing in late summer. The similar hairy spicebush (*L. melisaefolium*) is found in swamps in the southeastern states. The bark of *L. benzoin* was formerly used medicinally in reducing fever.

SPIDER. The spiders, or Araneae, form a natural order of the class Arachnida (*q.v.*). They are distributed all over the world, from sea level to Mt. Everest, on the ground and on vegetation, in forests and in meadows, in swamps and in deserts, in human homes and in subterranean caves, on the water and, in the case of at least one species, in air-filled tents under water. All spiders are easily distinguished from insects by the subdivision of their bodies into two portions only: the cephalothorax and the abdomen. The former is the result of a fusion of the head with the thorax. The fusion is so complete that the outlines of the head may be recognized only from the shallow depressions which separate its dorsal surface from that of the thorax. The abdomen is attached to the cephalothorax by a very slender pedicel, a feature which spiders share with some other Arachnida. But by far the most important character of spiders, separating them not only from other Arachnida but from all other animals, is the presence of spinnerets on the abdomen.

1. Spinning Apparatus.—The spinnerets are modified appendages of the fourth and fifth abdominal embryonic segments and are supplied with silk-producing glands situated in the abdomen. The ducts of these glands open at the end of specially modified hairs, the so-called spinning tubes situated on the spinnerets. The silk is produced as a fluid which is pressed out of the spinning tubes and solidifies immediately on contact with the air. Among the Arachnida the Pseudoscorpiones are also able to produce silk, but their spinnerets are at the end of the movable

finger of the chelicerae; *i.e.*, of the first pair of appendages situated in front of the mouth, and the silk glands are not in the abdomen but in the cephalothorax. Some Acari are also endowed with the ability to spin silk. In their case the silk glands are really modified salivary glands, as is also the case in such insects as moths. So the spiders are the only possessors of their type of spinning apparatus.

Besides common spinning tubes with very fine ducts, spiders possess also a few larger spinning tubes, so-called spigots, present in small, but fixed numbers among the hundreds of regular tubes. In some species the total number of spinning tubes was estimated to be as high as 1,000, but in others it is less than 100. A spider's

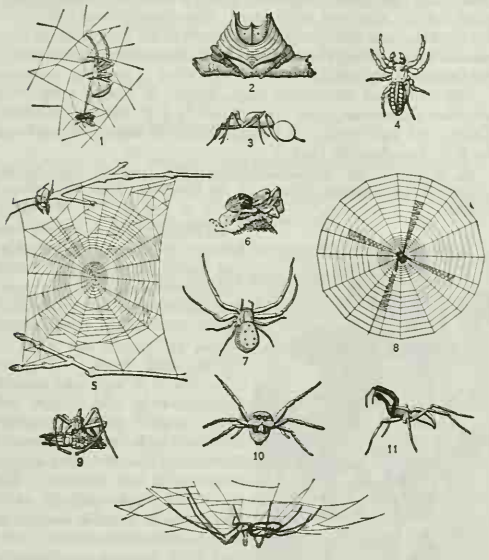


FIG. 1.—TYPES OF SPIDERS, VENOMOUS AND NONVENOMOUS

(1) Theridiid spider, wrapping a fly in sticky threads; (2) argiopid spider, shown from behind (after Vinson); (3) female lycosid spider, carrying egg cocoon; (4) common attid spider (after Blackwall); (5) epeirid spider and web (after Lulham); (6) female *Lycosa*, with egg sac attached to abdomen; (7) thomasid spider (after Blackwall); (8) argiopid spider (after Hingston); (9) grass spider (*Agalena*), pairing (after McCook); (10) male attid spider (after Bristowe); (11) male *Lycosa*, in courting attitude (after Bristowe); (12) linyphid spider, on under side of her sheet web (after Blackwall)

silk thread, as seen with the naked eye, is in reality a cable composed of many fine, individual threads. The diameter of such an individual thread is no more than three ten-thousandths of a millimetre, and in some cases, as in the inner lining of the cocoon of *Agalena naevia*, is less than one-tenth of a micromillimetre. Spider silk is finer, lighter and stronger than commercial silk produced by the silkworm; *i.e.*, by the caterpillar of the silk moth. Naturally, attempts have been made in the past to utilize spider silk, but without success. The reason for the failure lies in the cost of labour involved in the upkeep of a spider colony sufficient to produce large quantities of silk. In captivity each spider has to be kept separately and supplied with food and water, while under natural conditions the size of a garden adequate to accommodate numerous spiders, without danger of loss from overcrowding, is prohibitive from the point of view of reasonable profit. The most promising attempt at keeping spider gardens was made in Madagascar in the 19th century. The species used was the great orb weaver, *Nephila madagascariensis*. But even in Madagascar the cost of labour proved to be too high. At present spider silk is used only in the construction of optical instruments, because it is the finest thread available. For this purpose cocoons of northern species, such as the American *Argiope aurantia*, seem to be satisfactory. The length of the thread which *Nephila* uses in the construction of an egg cocoon varies from 1,500 to 4,000 m.

All spiders have spinnerets, but it must not be imagined that all make webs or snares. On the contrary, the majority of species use their silk for purposes other than the capture of insects: they spin egg cocoons, sperm webs, molting sheets, gossamer threads, attachment discs with cables, lining of burrows, hinges for trap doors, etc., while relatively few species build snares of various types for the capture of prey. In using silk for different purposes, spiders do not use the same glands. Each spider has in its abdomen numerous glands belonging to several different types. Even males and females of the same species differ in this respect, because only females have glands producing silk for egg cocoons. Only spiders which make snares to capture insects produce silk with adhesive droplets. Since spinning of silk is so intimately connected with the activities of all spiders, it is clear that the spinning apparatus is one of the most important features of their anatomy and one of the chief factors which determined the course of their evolution.

2. Webs.—There are several types of webs built by spiders, more or less characteristic of whole families. Sheet webs, funnel webs, tube webs, reticular webs, geometric orb webs, etc., can be distinguished. In the early days of arachnology spiders were even classified in accordance with the type of web they build. Attempts were made later to derive one type of web from another and thus to establish trends in their evolution. Unfortunately all such attempts disregard other evidence of spider relationship and of changes involving the anatomical structure of various organs. In many cases comparatively close relatives belonging to the same family build webs of a very different type, while spiders belonging to widely separated families may build webs of the same type. Thus so-called trap-door nests are characteristic of spiders belonging to the family Ctenizidae among the Mygalomorphae, yet some species of the Dipneumonomorph family Lycosidae, notably the European *Lycosa* (*Tarentula*) *opifex*, make trap-door nests of the same type. *Theridiosoma nechodomae* makes a little geometric spiral web. *Epeirotypus gloriae*, a triangular web and *Wendilgarda theridionina* spins a single fine thread across brooks; yet all three belong to the same subfamily of orb-weaving spiders and occur in the same locality on the island of Puerto Rico. *Linyphia marginata*, a species common in Europe and America, builds a dome web, its close relative *Linyphia communis* makes an inverted dome or bowl web and another relative, *Linyphia pusilla*, weaves flat webs.

The most complex snares are undoubtedly those of the typical geometric orb weavers of the family Argiopidae. A finished web consists usually of three or four outer foundation lines, several inner foundation lines, many radii, a central hub, an intermediate zone and a viscid spiral. Besides this many species build a tent of silk or of a leaf rolled up and held together by threads of silk. In this tent the spider hides during the day, holding in its claws the guy line which is connected with the centre of the web and which transmits the vibrations caused by entangled insects. Before the adhesive spiral is made, the spider makes a corresponding spiral of common silk. When this is finished, the adhesive spiral is attached to the radii in the space between the turns of the common spiral. The latter is clipped off, rolled up by the chelicerae and dropped to the ground as having no further use. Species of *Argiope* weave additional X-shaped bands called stabilimentum, although they have nothing to do with stabilizing the web. Some weave a disc of lacework in the middle of the orb. *Cyclosa conica*, a little orb weaver common both in Europe and America, suspends in the vertical diameter of her web a series of silk pellets to camouflage her egg cocoon which occupies a place in the same row.

A very unusual web is made by *Hypochilus thorelli*, a peculiar relict species found in the Great Smokies and adjoining mountain ranges in the United States and having close relatives only in Tasmania and China. Its web has the shape of a truncated cone suspended by the smaller, apical end from the underside of horizontal ledges, while the wider, open base of the cone, forming the

entrance to the inside of the web, hangs freely down and is held in place by only a few threads. The long-legged spider sits inside the web, flattened against a flat foundation sheet, which is attached directly to the surface of the rock. Another very interesting snare is made by *Dinopis longipes* of Central America. The spider is long, with very long legs, the span of which measures as much as 150 mm., and when the spider hangs fully relaxed with legs and body stretched out in a single line, it looks more like a slightly curved twig than like a spider. Suspended from a horizontal line by a single vertical thread, the spider holds onto it with its spinnerets and its four posterior legs. With its four anterior legs, the spider holds the corners of a rectangular net made of special silk produced by so-called cribellar glands. When an insects falls on top of this trap, the spider immediately displays tremendous activity by swinging the trap in a spiral motion with great speed, until the victim is completely enswathed in silk. A new trap has to be built after each capture. *Hyptiotes cavatus* of America and *Hyptiotes paradoxus* of Europe, both belonging to the family Uloboridae, make triangular snares which they swing with the aid of a guy line held in their legs. Their nearest relatives of the genus *Uloborus* construct orb webs of the same type as the Argiopidae, but use cribellar silk for the spiral.

The construction of tents was already mentioned in connection with the building of orb webs. Many spiders of other families build tents which they use as retreats or hiding places in which they find shelter from rain, cold, light and enemies, or from which they can pounce upon unsuspecting victims. The European and Asiatic aquatic spider, *Argyroneta aquatica*, builds her tent under water in the shape of a bell and fills it with air. Special tents are often used for molting, during which process all spiders, even the most formidable ones, are quite helpless. Some females make complete envelops of tough silk for themselves and their brood and stay inside until the spiderlings become capable of taking care of themselves. Some species winter in special tents made for that purpose under loose bark of trees.

Egg cocoons present a great variety of structure. Some are circular or oval and perfectly flat. Others are globular and smooth. The egg-shaped, pointed cocoons of *Latrodectus geometricus* are covered with numerous little knobs. Some cocoons are white, others are coloured yellow, brown, pink or green. The cocoon of *Argiope argentata* is a flat, elongated hexagon, often grass green on one side and golden yellow on the other. The cocoon of *Hypochilus thorelli* is white when made, but the female then camouflages it with pieces of moss which she collects from the surface of the rock or from the ground. Most cocoons are made with at least two walls, an inner and an outer one. First, the spider spins a sheet, deposits the eggs in it, covers them with another sheet and binds the two together, neatly clipping the edge. Then the outer sheath is made, often of much tougher silk. In *Argiope aurantia* the outer sheath or envelope is almost globular, made of brown, loosely woven silk. The inner sheath is cylindrical, with a round bottom and a strongly constricted neck which protrudes through the top of the outer sheath. The cocoons of *Agalena naevia*, the common American funnel-web spider, are fastened under bark to which their flat base is attached. They are gray, unattractive and dirty looking. The female guards them until she dies.

Attachment, discs or anchors are made by all spiders whenever they want to be safe from injury by falling from an elevation or when they jump from one place to another. An anchor is made instantaneously by pressing the spinnerets against the surface of the object on which the spider happens to be. The pattern of an anchor is characteristic of each species and consists of numerous loops of fine threads brought together to a median line which is then continued as a single cable. On reaching the object of its destination, the spider either breaks off the end of the cable or, if it intends to return, fastens the end and climbs the cable to the point of its anchorage.

Silk is used for enswathing insects captured in a snare. This makes their escape impossible and serves to store them as food for use at a later time. In case the insect is a dangerous adversary, such as a wasp or hornet, silk may be squirted as far as an inch from the spinnerets, a distance sufficient to make the combat safe.

The maximum number of spinnerets which a spider may possess is four pairs. This number, however, is found only in the Liphistiomorphae. In some spiders the anterior median pair is transformed into a flat, perforated plate, the cribellum. In most spiders only six or four spinnerets are present. A few species are known with a single pair of spinnerets, the minimum possible, because the complete loss of spinnerets would require such a readjustment of habits that the very existence of such creatures would become impossible. The size and shape of spinnerets, on the other hand, may vary considerably and apparently have very little to do with the type and complexity of the web. Spiders belonging to the Mygalomorph family Dipluridae have very long and slender spinnerets. They make large, irregular sheet webs or reticular webs which are usually provided with a funnellike retreat, whereas spiders of the Dipneumonomorph family Hersiliidae, also with very long and slender spinnerets, make no webs of any kind.

3. Reproductive Organs.—Another character which spiders alone possess and by which they differ from all other animals is of a sexual nature: it consists in a complete separation of the male copulatory apparatus from the reproductive system proper. The latter is situated in the abdomen in both sexes. The female has two ovaries, two oviducts and a single, very short uterus which opens in the median line between the organs of respiration belonging to the second embryonic segment. Another pair of openings is situated in front or at the sides of the genital opening. These openings lead into sperm receptacles. In more highly developed spiders the latter are connected with the uterus by special fertilization ducts. Various chitinous accessory structures, known as the epigynum, are present in many species. These are more or less intimately connected with the genital opening, are characteristic of each species which possesses them and in some way, not yet clearly understood, have to do with the act of copulation. The reproductive organs of the male are also in the abdomen. They consist of a pair of sperm glands or testes, a pair of deferent ducts and a short, common ejaculatory duct. The male genital opening is in the identical position described for the female genital opening. But the organs of copulation are in no way connected with the reproductive organs and are instead situated on the terminal joint of the pedipalpi. This is what makes them unique in the entire animal kingdom and what necessitated the development of special instincts, of strictly inherited behaviour of male spiders, peculiar to sexually mature male spiders and only to them. Because the female has a pair of seminal receptacles and the male has a pair of pedipalpi, the males have a pair of copulatory organs. In its simplest form the apparatus consists of a bulb ending in a more or less curved spout or embolus. Inside the bulb a blind receptacle for sperm is situated, with a single duct which opens at the end of the embolus. The receptacles have to be filled with sperm before copulation can take place. This requires the construction of a special sperm web, the ejection of the sperm into the web and its transfer into the palpi through the opening of the embolus. How such an arrangement could have originated remains a mystery. Something similar exists in the case of the Ricinulei, another order of Arachnida, in which the metatarsus and tarsus of the third pair of legs in the male are modified as an organ of copulation. Something similar, but without any structural modification, is also present in the Solifugae, also an order of Arachnida, in which the male transfers the sperm into the genital opening of the female by means of his chelicerae. It is not known in which geological age the pedipalpi of male spiders became modified as an organ of copulation. Oligocene spiders of the Baltic amber and Tertiary spiders of the Florissant shale possess fully developed copulatory organs of the same type and as complicated as Recent spiders. The Mesozoic era left a single record of a fossil spider and that so imperfect that nothing can be learned from it. The few fossil spiders from the Upper Carboniferous of Europe and Pennsylvanian of North America, as well as the still fewer fossils found in the Devonian Red Sandstone of Scotland, are not sufficient to determine whether at that distant age male spiders had as yet

unmodified pedipalpi, or whether all available fossil specimens belong to the female sex or at most represent immature individuals. But it is clear that whatever the geological age may be when the pedipalpi became modified into organs of copulation, the change must have involved a corresponding change in sexual habits and this, in its turn, required a change in other habits. It is quite possible and even probable that previous to the development of organs of copulation on the pedipalpi, male spiders used a method of fertilization comparable to that of pseudoscorpions. The males of these small Arachnida attach a spermatophore to a stalk on the ground, grasp with the pincers of their pedipalpi the corresponding pincers of the female and drag her until her genital opening comes in contact with the spermatophore. Since spiders have no pincers on their pedipalpi, or for that matter on any other appendages, males could never grasp a female in the manner of the pseudoscorpions. But they could perhaps have used their pedipalpi to hold a drop of sperm by simple adhesion and transferred it into the genital opening of the female in the best way possible, or may have dropped sperm onto a web and pushed the female until her genital opening came in contact with it. With the development of a special copulatory apparatus in the male, the development of sperm receptacles in the female must have gone apace until the bewildering diversity and complexity of structure and habits has been reached, by which the many thousands of species of spiders are now better distinguished than by other structural characters.

4. Poison.—The third character which is found only in spiders among Arachnida is the presence of poison glands either in the basal joint of the chelicerae or in the cephalothorax and opening by means of a duct near the tip of the fang. With few exceptions the structure of the glands is the same in all spiders. It is an elongated, cylindrical sac, the walls of which have a single layer of glandular epithelial cells and an outer muscular covering composed of spirally wound, striated fibres. A structureless membrane separates the two layers and another thin membrane formed by highly flattened cells protects the outside of the gland. The flow of

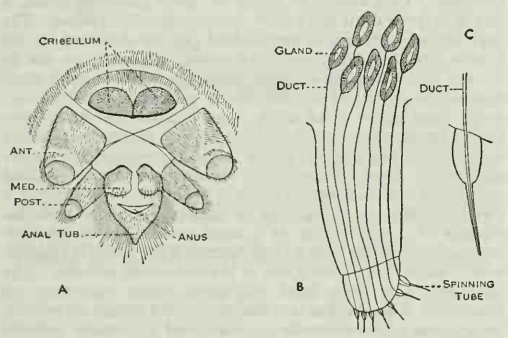


FIG. 2.—SPINNING ORGANS OF SPIDERS

(A) Spinnerets of *Eresus latifasciatus*, ventral view; (B) one of the spinnerets of *Trechalea magnifica*, showing silk glands and ducts; (C) greatly enlarged spinning tube of *Trechalea magnifica*

poison is under control of the spider. In fact, spiders which ensnare their prey do not use their poison, while those who hunt for their prey or hide on flowers and capture insects by grasping them with their fangs when the insects come within their reach kill their victims with poison. All spiders, however, use their poison in self defense, when escape is impossible. Spiders of two families, Uloboridae and Heptathelidae, have no poison glands.

The poison of spiders attracted early attention and was studied many times. Unfortunately, the results of most of such studies are vitiated by the application of wrong and faulty methods. Observations on the effects of bites on man and animals are of very little value even when the species of spider is known and at best may serve only as confirmatory evidence. But experiments conducted in the laboratory have also led to wrong conclusions until ways were found to get the poison in a pure state, without any admixture of other substances such as invariably happens when an entire cephalothorax is used for extraction of poison in a fluid. As now practised, poison is obtained either by dissecting the cephalothorax and removing the glands, or still better by applying an electric current to the immobilized, live spider and collecting the pure poison as it appears in droplets at the ends of the fangs. Poison obtained in this manner may be dried, weighed and used in measured quantities and known concentration. Few spiders are poisonous to man. Among these the black widow alone is dangerous in the United States, as are other closely related species in Europe, Africa, Asia and Australia. All these species belong to the genus *Latrodectus* of the family Theridiidae and their poison is neurotoxic. The only other spiders known to have a neurotoxic poison belong to the genus *Ctenus* of the family Ctenidae. A Brazilian species, *Lycosa raptoria*, causes a creeping ulceration of the skin. Some observations point to the possibility that other species belonging to various families may be poisonous to man, but the majority—even large tarantulas—are no more dangerous to man than wasps or hornets and a great many spiders do not attempt to bite even when taken in one's hand.

5. Respiratory System.—A fourth feature peculiar to spiders among other Arachnida is the presence of two types of respiratory organs in the same individual: book lungs and tracheal tubes. This peculiarity is restricted to a single suborder among the spiders, the so-called Dipneumonomorphae, but the vast majority of spiders belong to it. It seems certain that originally spiders had two pairs of book lungs and no tracheal tubes; the second pair was later replaced by tracheal tubes and finally the first pair was also replaced by tracheal tubes. Thus there exists in spiders a condition unique among the Arachnida. Of late, Tasmanian spiders have been discovered in which only one pair of tracheal tubes remains.

6. Digestive System.—Several features, while not unique among Arachnida, nevertheless are of great importance in the life of spiders. The first of these has to do with the structure of their digestive system. The fore-gut, consisting of a pharynx, esophagus and gizzard, is concerned only with the mechanics of ingestion. The pharynx is so structured that it acts both as a pump and a strainer, preventing coarse particles from passing into the esophagus. The latter is merely a tube connecting the pharynx with the gizzard, but the gizzard is a second, powerful pump. The mid-gut is provided with several branched diverticles, the finest branches ending each in a little spherical ampulla and the whole lot of them held together by interstitial tissue. The mid-gut serves two purposes: it produces digestive enzymes and absorbs the fluid substances into which food is converted after the action of enzymes on its constituents. A pair of maxillary glands situated in the maxillary lobes of the pedipalpi also produce digestive enzymes. The mid-gut forms a large stercoral pouch in which excrements are stored before they are excreted through the hind-gut and anus. Now this arrangement of the digestive system, found also in other Arachnida, makes ingestion of only fluid food possible and necessitates predigestion of food outside the body. It is a mistake to think that spiders suck blood. That is the one thing they never do. Instead, they inflict a wound with their fangs, inject digestive enzymes into the wound and, as the tissues of their prey become liquefied, pump the insect dry, leaving nothing but an empty, chitinous shell behind. For some reason which is not quite clear, spiders have never learned to make use of vegetal food or even to feed on dead bodies of animals. They leave in peace even live insects when the latter do not move. Spiders may be fed in captivity by little pieces of raw meat, but the latter must be attached to a thin thread and dragged or dangled in front of the spider to induce the latter to grasp it.

7. Sense Organs.—The second feature which naturally has a great influence on the entire behaviour of spiders is the structure and function of their sense organs. Spiders belonging to different families differ greatly in this respect. Some have well-developed eyes with keen vision. Such spiders have less perfectly developed tactile organs. Other spiders have eyes with much poorer vision and depend more on the perfection of their sense of touch. The latter is the best developed sense even in spiders with good eyes, while in species with poor eyes it is almost unbelievably fine and

diversified. The body and legs, the pedipalpi and chelicerae and even the spinnerets of spiders are covered with numerous hairs of various kinds, some stiff and stout, others slender and pliable, some smooth, others provided with serrations. Many spiders have plumelike soft hair. Some hairs lie flat on the surface of the skin, others stand at an angle of 30° or more, some are erect. Not all are sensory hairs, but each sensory hair has its own nerve fibre which conveys impulses to the central nervous system. Especially fine, erect hairs arising from the centre of little drumlike membranes are sensitive to slow currents of air. Such hairs are called trichobothria and their distribution has systematic value. Once the trichobothria were suspected of being capable of transmitting sound vibrations, but this proved to be incorrect. Many spiders possess sound-producing organs, but organs of hearing have not yet been discovered, nor is it known how acute hearing is in spiders. It is probable that the range of sounds which they can perceive is quite different from the range of the human ear and presumably much higher.

Except for a few blind species, all spiders possess eyes which are on the carapace. The maximum and most common number is four pairs; but spiders with three, two, or a single pair of eyes are known. Some families have only six eyes, as for example the Dysderidae; but usually the reduction in the number of eyes occurs in different genera of the same family. The so-called anterior median eyes have always a direct retina; *i.e.*, one in which the rods face the dioptric apparatus, consisting of a glass body and a corneal lens. The other eyes have always an inverted retina and are often supplied with a tapetum of pigment cells, reflecting light and giving the eyes a pearly appearance. The disposition of the eyes is characteristic of whole families and is responsible for the diverging directions of the axes of the eyes. Convergent eyes do not occur in spiders. The respective fields of vision are determined by the shape and curvature of the lens, the acuity of vision by their diameter, the size of the image and the distance at which the image of two points falls on a single rod. Each pair of eyes forms images of a size different from that formed by the other eyes and the fields of vision are mostly quite different. Thus a fairly large portion of the horizon is covered, but only small portions are ever covered by overlapping fields of four eyes, still smaller ones by six eyes and scarcely ever by all eight eyes. The best eyes have lenses with spherical surfaces, but the acuity of vision with even the best eyes is far below that of the human eyes.

The olfactory sense is very poorly developed and apparently restricted to what one may call sexual odours. The males of some species are able to follow the invisible spoors left by a female on a smooth surface, such as a sheet of paper. But other scents are not perceived unless the substance is of a highly irritating nature.

8. Behaviour.—Spiders have a highly developed nervous system, possess memory, and their reflexes may be conditioned by methods used in experiments. But since none of the spiders developed a social state comparable to that of insects and only few species at best may be said to be gregarious, the behaviour of spiders is entirely of the type usually spoken of as instinctive. Some of these patterns of inherited behaviour appear early and persist throughout life. Others appear at the time of sexual maturity and disappear rapidly. Generally speaking, the life cycle of a spider may be pictured as follows: A spiderling, on emergence from the egg, is incapable of taking care of itself, because its digestive system is not yet fully developed. The mid-gut is filled with embryonic yolk and ingestion of food is impossible. Consequently spiderlings either stay in a web built by their mother for that purpose before she died, are guarded by their mother inside a brood cocoon or are carried by the mother on her back, as the case may be. Their life is restricted to slow digestion of the embryonic yolk. The completion of its digestion coincides with the completion of the digestive system itself and is accompanied by the shedding of the skin and some changes in the proportions of the body. With the mid-gut emptied, hunger sets in and the spiderlings would attack and destroy each other but for the simultaneous emergence of an inherited desire to escape. Dispersal of spiderlings may be accomplished by different methods. In some species, spiderlings simply run away in all directions. But in the majority of species aerial dispersal takes place. This so-called ballooning coincides in temperate zones with the sunny days of autumn or of spring, if the eggs remain quiescent through the winter. Each spiderling climbs to the end of a blade of grass, a twig, a pole or any protruding object, raises its abdomen and spins a thin thread of silk. As this gossamer thread is caught in the breeze, the spiderling lets go of the support on which it sat, climbs to the middle of the thread and is carried by air currents to a considerable height upward and to great distances horizontally. Gossamer threads have been encountered by airplanes five miles above sea level and one observer reported that he saw spider threads in the stratosphere. Spiderlings have also been caught in the rigging of ships several hundred miles from shore.

When other currents of air finally bring them down to the ground, the spiderlings enter upon the new phase of their life, which will last until maturity. Each spiderling follows the pattern of behaviour characteristic of the species. Some build a web, orb web, sheet web, reticular web, etc., as the case may be; others make burrows or run about on the ground among leaves, remaining vagabonds without a home. From time to time they molt, shedding the entire skin, including the corneal lens of the eyes and the lining of the fore-gut and the organs of respiration. During this period the spiders are quite helpless and cannot eat nor bite. The number of molts varies in different species and even in the same species, but the usual number is 7 or 8 and up to 13. Males of all species and most females never molt after maturity, only some Lycosidae and Theraphosidae do so. The usual behaviour of females is not appreciably changed when they become mature, except that sexual excitation becomes manifest and they respond to the advances of a male in a manner different from that which governs their attitude toward their prey or unwelcome intruders. Otherwise they still lead the same life, build the same type of snares, etc.

In the case of males, the entire behaviour pattern changes as soon as they mature. The filling of their palpi with sperm is now their chief concern. The males of some species use the old web for the purpose, merely adding a few threads of silk to which they attach the drop of sperm as it comes out of their genital opening. Other species weave special sperm webs, even if they have never built a web of any kind before. The pumping of the sperm takes from 45 minutes to two hours time. When the palpi are charged with sperm the male goes in search of a female. This is a rather hazardous undertaking, considering the fact that no female has ever seen a male in her life and that each female is familiar only with animals suitable for food or those that must be handled as enemies. Courtship is the means by which a male overcomes the hostile attitude of a female. In jumping spiders endowed with a good sense of sight, courtship consists of a love dance in the presence of a watching female, but at some distance from her. In spiders with a less keen sense of sight, actions involving touch or smell or a combination of these two senses take precedence. Some orb weavers show a discrepancy in the time of maturation, the males maturing considerably earlier than the females. In such cases one or more males get into the web of a female and live in it until she becomes mature. In some species the male builds a tent in which he stays with the immature female, sharing food with her until she molts for the last time. Some males offer a captured insect to the female to appease her hunger. Copulation lasts but a few minutes. The palpi are used one at a time and courtship is sometimes repeated in the interim. When both receptacles of the female are filled, the relationship between her and the male reverts to that between any two spiders of the same or of opposite sex. The male, being the smaller of the two and the more agile, seeks safety in escape. The female, with the sexual desire appeased and the feeling of hunger returning, treats the male as an intruder if he persists in remaining in her web or stays too close to her on the ground. But the often repeated stories that female spiders always devour their mate after copulation are based on misinterpretation of observed instances when a male was prevented from escaping and later fell victim to the hungry female. The majority of males die naturally soon after copulation, because the supply of their energy is completely spent. In captivity a male may linger for several weeks if kept alone, but he usually refuses food and dies with the abdomen show-

ing considerable shrinkage of all tissues.

Females live longer. In some species they die soon after egg laying, but in other species they may live several years and copulate annually. To the former group belong the majority of spiders in temperate zones, to the latter the larger Lycosidae and the Theraphosidae. Some large tarantulas are known to have lived in captivity as long as 15 years.

9. Enemies.—Spiders have many enemies among vertebrates and invertebrates. Many lizards, some toads and various birds feeding on insects also eat spiders. The large *Bufo maritima* was observed greedily feeding on hairy tarantulas. Solitary wasps are perhaps the greatest enemies of spiders. Many of them depend entirely on spiders as food for their larvae. Collectors know that the easiest way of getting certain species of spiders is by examining the nests of mud daubers. Species of the genus *Pepsis* among solitary digger wasps paralyze Mygalomorph spiders, bury them and lay a single egg on the abdomen of each victim. Some Hymenoptera and Diptera spend their larval life as external parasites of spiders. Nematodes of the family Mermithidae are not uncommon parasites in the body cavity of spiders. Protozoan and helminth intestinal parasites have not as yet been found, perhaps because infection is impossible on account of predigestion and filtering of food.

More than 50,000 species of spiders are known, distributed among more than 2,500 genera, and many new species are discovered every year. The spider fauna of Europe is known fairly completely, that of the United States much less so.

10. Paleontology.—Fossil spiders have already been found in the Red Sandstone of the Devonian. Many species have been described from the Carboniferous of Europe and America. A single species has been reported from the Mesozoic, but the Tertiary of Europe and America and particularly the Oligocene Amber of the Baltic are rich in fossil species, the great majority of which are now extinct.

Bibliography.—The following short list of books may be of help to the uninitiated. More or less complete bibliography published yearly will be found in the *Zoological Record*. L. Berland, "Les Arachnides," *Encycl. Ent.*, xvi (1932); W. S. Bristowe, *The Comity of Spiders*, 2 vol. (1939–41); J. H. Comstock, *The Spider Book*, rev. by W. J. Gertsch (1940); Sukh Dyal, "Spiders of Lahore," *Bull. Dep. Zool. Panjab Univ.* (1935); U. Gerhardt and A. Kästner, "Araneae" in *Kükenthal's Handbuch der Zoologie*, vol. 3 (1937); H. C. McCook, *American Spiders and Their Spinningwork*, 3 vol. (1889–93); K. C. McKeown, *Spider Wonders of Australia* (1936); E. Nielsen, *The Biology of Spiders*, 2 vol. (1932); A. Petrunkevitch, "Systema Aranearum," *Trans. Conn. Acad. Arts Sci.*, vol. 29 (1928); C. F. Roewer, "Araneae," *Tierwelt Mitteleur.* (1928); T. H. Savory, *The Biology of Spiders* (1928); E. Simon, *Les Arachnides de France*, 7 vol. (1874–1937); *Histoire Naturelle des Araignées*, 2nd ed., 2 vol. (1892–1903). (A. Ph.)

SPIDER MONKEY, a group of tropical American monkeys known to zoologists as *Ateles*. They take their English name from the slimness of the body, the elongated limbs and the long tail, the undersurface of the prehensile extremity of which is naked. The thumb is rudimentary or wanting. Their long coarse hair and the broader partition between the nostrils distinguishes them from the woolly spider monkeys (*Brachyteles*). The species are arboreal.

See Primates.

SPIELHAGEN, FRIEDRICH VON (1829–1911), German novelist, was born at Magdeburg on Feb. 24, 1829. He was brought up at Stralsund, studying also at Berlin, Bonn and Greifswald.

He became a master in a *Gymnasium* at Leipzig, but upon his father's death in 1854, devoted himself entirely to writing. After publishing *Klara Vere* (1857) and *Auf der Düne* (1858), he obtained a striking success with *Problematische Naturen* (1860–61). His powerful *Sturmflut* appeared in 1876. He died Feb. 25, 1911.

Spielhagen's *Sämtliche Werke* were published in 1871 in 16 vol., in 1878 in 14 vol.; his *Sämtliche Romane* in 1898 (22 vol.), and these were followed by a new series in 1902. His autobiography, *Finder und Erfinder*, 2 vol., was published in 1890.

SPIKENARD (Nard), a celebrated perfume which seems to have formed one of the most durable aromatic ingredients in the costly unguents used by the Romans and eastern nations. The ointment prepared from it is mentioned in the New Testament (Mark xiv, 3–5; John xii, 3–5) as being "very costly." Pliny mentions that nard spikes reached as much as 100 denarii per pound. The source of the true or Indian nard is now identified as *Nardostachys jatamansi*, a plant of the family Valerianaceae, the fibrous rootstocks or spikes of which are still collected in Bhutan and Nepal. The great lavender (the early *Lavandula spica*) was also once called spikenard.

The great spikenard of the United States is *Aralia racemosa* (*see* Aralia), and another species of the same genus, *A. nudicaulis*, or wild sarsaparilla, is known as wild spikenard.

In the West Indies *Hyptis suaveolens* is called spikenard, and in Great Britain the name plowman's spikenard is given to *Inula conyza*.

SPILITE, a fine-grained or dense igneous rock, usually free of visible crystals and commonly greenish or grayish-green in colour. Spilites are commonly vesicular (blistered) or variolitic (pocked) and show wonderfully preserved pillow structure, a feature in most cases indicative of a submarine origin of the lavas from which they were formed.

The individual pillows are filled with concentric zones of vesicles filled with chlorite and calcite. The formation of pillow structure in volcanic rocks was directly observed by Tempest Anderson in the case of the lava poured into the sea from the volcano Matavanu in Samoa. On the surface of the lava, buds form and expand, giving rise to "pillows" with a glassy crust. The interspaces between the pillows are frequently filled with sediment representing the soft ooze of the sea floor. Not all pillow lavas are spilitic.

The spilites are of basaltic character, but possessing in place of the normal labradorite, a feldspar of the composition of albite. The ferromagnesian mineral is an augite of pale brown colour; spilites are, however, usually very completely decomposed, augite being represented by chlorite and calcite.

Some spilites showing pillow structure are not strictly lavas but shallow intrusions into unconsolidated submarine ooze. Excellent examples of such intrusive spilites are provided by the Devonian spilites of Nundle, New South Wales. Though the term spilite was first used by Alexandre Brongniart (1770–1847) for altered basic lavas free from phenocrysts and possessing well-marked vesicular textures, early in the 20th century a new significance was given to the term by British geologists Dewey and Flett, who urged that a special group of rocks—the spilitic suite—should be recognized.

According to these authors, a large suite of igneous rocks is genetically associated with the spilites. Their composition varies widely, but they all have some characters in common. They possess a high percentage of soda and are usually extensively altered. These rocks are albite-dolerites, minverites, picrites, keratophyres, soda-felsites and soda-granites. The spilitic suite was originally given a status comparable with that of the alkaline and calc-alkaline suites. The spilitic series is regarded as peculiar to districts that have undergone a long continued and gentle subsidence with few or slight upward movements and no important faulting.

The lavas are poured out on the margins of geosynclines while their common sedimentary associates are black shales, limestones and radiolarian cherts. As geosynclines are the centres of subsequent fold movements many ancient spilites have become extensively folded and metamorphosed with production of new minerals, including albite of metamorphic origin. Eruptions having the spilitic facies have occurred repeatedly over a wide area and on a large scale. (C. E. T.; X.)

SPINA, an ancient city of Gallia Cispadana, Italy, between Ravenna and the mouths of the Po. In Etruscan times it was the principal port of Felsina. (*See* Bologna.) Near Comacchio, in the draining of the lagoon of Valle Trebba, a very extensive necropolis was found and about 450 tombs were excavated, ranging from the 6th to the 3rd century B.C. They contained fine Greek, Italic and Etruscan vases.

SPINACH (*Spinacia oleracea*), a hardy, leafy, annual of the family Chenopodiaceae, used as a potherb. First a rosette of large, petioled leaves develops; leaves are triangular-ovate to arrow-shaped in outline and are nearly flat to much crumpled. Later the branched seed stalk grows to one to three feet. A plant

may bear male flowers only, female flowers only, both male and female flowers, or perfect flowers along with flowers of one sex. The seed is a small, one-seeded, dry fruit called a utricle; it is smooth or armed with sturdy, sharp spines. When seed is sown in autumn the resulting plants behave as biennials. Plants may survive temperatures as low as 0° F. High temperature and long days induce seed stalk formation.

PHOTOGRAPH, J. HORACE MCFARLAND CO.

SPINACH, LONG SEASON TYPE

Spinach requires cool weather for best growth, and rich, well-limed, well-watered soil. Seed is sown in rows 8 to 18 in. apart and plants should stand 3 to 4 in. apart in the row. In large fields seed is sown thinly to avoid need for hand thinning of plants. About 1950, F_1 hybrid varieties resistant to the blue mold or downy mildew disease became available.

Spinach culture in the United States received an enormous impetus about 1920 when attention was called to its relatively high content of calcium and iron, and further impetus when its relatively high content of vitamins A and C were emphasized. The later discovery of large quantities of oxalates in the plant cast some doubt upon its true dietary value. Oxalic acid of the plant reacts with calcium therein and possibly some calcium from other dietary sources to form calcium oxalate. The calcium of calcium oxalate is believed to be unavailable nutritionally. Spinach, nevertheless, continues to be a most popular potherb.

New Zealand Spinach (*Tetragonia expansa*) (family Aizoaceae), a tender annual, native to New Zealand, produces greens similar to spinach when cooked. It is a much-branched spreading plant, one to two feet high, with fleshy, triangular leaves and inconspicuous, yellow-green, axillary flowers. The seed is a nutlike fruit containing one to two seeds that often germinate very slowly. The plant is frost-tender and heat-tolerant, furnishing greens at a season when other potherbs do not thrive in warm regions. It is sown in rows two to three feet apart, with the plants about one foot apart in the rows. The leafy tips of the branches are harvested repeatedly as new growth develops.

Mountain Spinach or Orach (*Atriplex hortensis*) (family Chenopodiaceae), of Asiatic origin, is a coarse, hardy annual growing to a height of six feet. The leaves are large, triangular-ovate or arrow-shaped, with toothed or wavy margins and colour varying from pale yellow-green to dark red. The leaves are used as a potherb, but the plant is of little interest to European or American gardeners except in regions not suited to the culture of the commoner leafy vegetables, as the Great Plains of the United States.

(V. R. B.)

SPINAL COLUMN, SURGERY OF. **Neurologic Diseases.**—The spinal column is of importance in the surgery of neurologic diseases because it encloses both the spinal cord (*q.v.*) and the extension of the cord which is known as the cauda equina, inasmuch as it resembles a horse's tail.

In the development of the human embryo the body normally fuses in the mid-line of the back. Sometimes developmental defects of varying severity occur and there is failure of the spinal segments to meet in the mid-line, a condition called spina bifida. Through the bifid spine may emerge the coverings and even the cord or cauda equina elements themselves and these defects may require surgical repair.

The spine, of course, is subject to both direct and indirect injury. Forces, for example, may be transmitted indirectly to the cord with resultant impairment of its activity. Usually surgery has nothing to offer in such cases. If direct force, as in severe fracture or dislocation of the spine, tears the cord, there is no possibility of surgical repair because the fibres in the cord have no capacity for regeneration. However, occasionally the surgical lifting of bone fragments so that they do not press on the cord may be of benefit.

Generally speaking, blood-vessel disturbances of the cord or of the cauda equina are not subject to surgical correction, though tumours which compress the cord, and which are sometimes made up of blood vessels, usually require removal.

Occasionally infection may localize in the jackets surrounding the cord or the cauda equina. In such instances removal of the posterior portion of the spinal column and drainage of the pus may, combined with use of antibiotic drugs, restore cord activity. Tuberculous infection of the vertebral column may cause pressure on the cord or on the cauda and necessitate operative "decompression."

The bony segments of the spinal column, the vertebrae, are separated from one another by resilient cushions, called intervertebral discs. Relaxation of the circular ligaments surrounding the disc allows disc substance to herniate and to press on the cord itself, or on nerve roots supplying the arms or legs, with resultant pain and weakness. In case of persistent pressure on these structures, removal of the disc may become necessary.

In cases of unrelievable pain arising in diseased structures elsewhere in the body, the interruption of the ascending pain pathways within the spinal cord itself may be used as an important method of providing relief from pain.

See Charles Frazier and A. R. Allen, *Surgery of the Spine and Spinal Cord* (New York, 1918); George C. Prather and Frank H. Mayfield (eds.), *Injuries of the Spinal Cord* (Springfield, Ill., Oxford, 1953); F. Keith Bradford and R. Glen Spurling, *The Intervertebral Disc* (Springfield, Ill., 1941). (J. P. Es.)

Orthopaedic Aspects.—Surgical affections of the spine run the gamut of diseases and may be classified as congenital defects, developmental and degenerative disorders, metabolic diseases, infections, tumours and traumas. Each of these classifications may be further subdivided, some cases requiring surgical and others medical management. Often, as in spinal tuberculosis, a combination of surgical fusion and administration of specific drugs provides the best chance for cure. The successful management of a spinal disorder depends on careful consideration of its natural course plus the relative influence of therapy at various stages of advancement.

Operative surgery of the vertebral column is, at times, most difficult because of the inaccessibility of portions of the vertebrae and the proximity to vital structures such as the spinal cord. When spinal affections arise from or directly involve the cord or its investments or the spinal nerve roots, operative treatment is best carried out by the neurological surgeon. The conservative management of some of these cases may occasionally fall to the orthopaedist, however. In such cases, braces, special medical techniques or therapeutic agents take the place of open operation.

World Wars I and II and subsequent conflicts gave a new emphasis to the need for the active treatment and rehabilitation of paralytic patients. The principles developed from wartime experiences were applied to many diseases far removed from the spinal trauma of military combat.

The most common orthopaedic spinal operation is the fusion. First employed by Russell A. Hibbs in 1911, the spinal fusion consists of making a segment of the vertebral column permanently rigid by one of various means. The standard method utilizes pieces of bone taken from the leg (or pelvis) and placed in contact with the hindmost processes of the vertebrae which are to be joined. By a natural process extending over a period of months, new bone arises from the vertebrae, penetrates and replaces the original bone grafts, leaving the entire segment linked by a solid bridge of bone.

Spinal fusion is employed in cases where (1) permanent arrest of motion is necessary to control a disease process (*e.g.*, tuberculous spondylitis); (2) pain is arising from instability and deformity resulting from a congenital defect (spondylolisthesis); (3) severe curvature is developing during the growth process (scoliosis); and (4) pain and instability have resulted from injury. Fusion may be simulated conservatively, when the occasion demands and when perfect immobilization is not needed, by a steel-framed back brace.

The least accessible portions of the vertebrae may be reached by introduction of a long needle, carefully avoiding the spinal cord, nerve roots and major blood vessels. This technique is useful only to secure small pieces of tissue for diagnostic purposes or, occasionally, to drain abscess pockets. It is possible to approach certain of the vertebral bodies by radical means in the rare cases where that is necessary.

The manner of treatment of spinal fractures and dislocations depends on the extent of cord or nerve-root damage. The majority of such cases do not feature serious neurologic signs and are best managed conservatively. (J. Dy.)

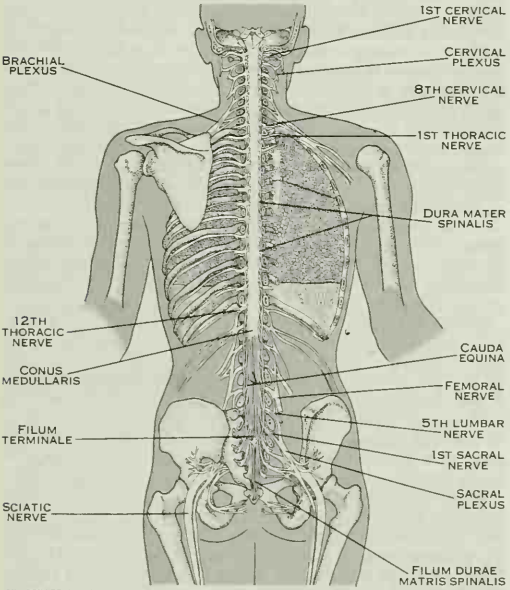
SPINAL CORD in man is the elongated cylindrical portion of the central nervous system located in the upper two-thirds of the vertebral canal of the vertebral column.

ANATOMY

The 18-in.-long spinal cord extends from just above the first cervical vertebra (atlas) at the base of the skull to the level of the first lumbar vertebra, where it is known as the conus medullaris. Small groups of nerve fibres that are attached to the spinal cord unite to form the segmental series of dorsal roots and ventral roots (fig. 1 and 2). (*See* also NERVE: *Spinal Nerves.*) One dorsal root and one ventral root combine in the region of an opening, the intervertebral foramen, between the two vertebrae to form one of the 31 pairs of spinal nerves. The dorsal roots contain the nerve fibres that transmit sensory impulses from the body to the spinal cord, while the ventral roots are composed of the motor fibres that transmit nerve impulses from the cord to the muscles and glands of the body.

The spinal cord does not occupy the entire length of the vertebral canal because the vertebral column continues to grow after the cord has reached its full length. Since the column is longer than the cord, the nerve filaments of each dorsal root leave the cord above the site where the root enters the intervertebral foramen (fig. 1).

Each thoracic, lumbar, sacral and coccygeal spinal nerve is named from the vertebra that is just above its exit from the vertebral column. As there are 12 thoracic vertebrae, 5 lumbar vertebrae, 5 fused sacral vertebrae and 1 (to 3) vestigial coccygeal vertebra, there are 12 thoracic, 5 lumbar, 5 sacral and 1 coccygeal spinal nerves (fig. 1).



BY COURTESY OF CIBA CLINICAL SYMPOSIA, VOL. 1, NO. 6, AND THE CIBA COLLECTION OF MEDICAL ILLUSTRATIONS

FIG. 1.—THE SPINAL CORD

There are eight cervical nerves, although only seven cervical vertebrae, because in addition to a nerve passing below each cervical vertebra one exists above the first cervical vertebra. The long roots of the lower third of the vertebral canal below the conus medullaris form a structure called the cauda equina ("horse's tail").

Meninges.—The spinal cord is enveloped by three coverings called meninges (figs. 1 and 2). (*See* MENINGES AND CEREBROSPINAL FLUID). From within outward they are the pia mater, arachnoid mater and dura mater. These meninges are continuous with comparable layers surrounding the brain. The pia mater is closely attached to the spinal cord and contains the many blood vessels that supply the cord. The pia mater continues as a strand in the middle of the cauda equina known as the filum terminale, which extends to the coccyx (fig. 1).

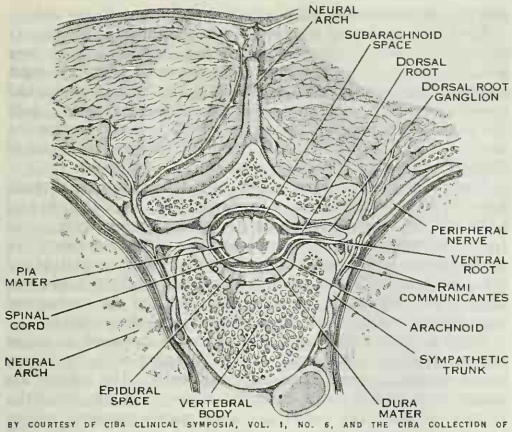
The second meninx is the arachnoid layer, which is connected to the pia mater by delicate strands. The region between the two layers is the subarachnoid space, which is filled with the cerebrospinal fluid. Formed mainly in the brain, the fluid cushions the spinal cord. In the region caudal to (that is, below) the conus medullaris the arachnoid forms a cul-de-sac (fig. 1)—containing the cauda equina and cerebrospinal fluid—into which lumbar puncture needles are inserted to withdraw cerebrospinal fluid for diagnostic examination and to inject spinal anesthetics. This procedure avoids damage to the spinal cord, while injury to the spinal roots is minor.

The outer layer is the dura mater, a tough fibrous tissue surrounding the cord and continuous with the sheaths of the spinal roots and spinal nerves. It extends caudally to the lower sacral region where it is reduced to a strand, the filum durae matris spinalis, attached to the coccyx (fig. 1). A thin subdural space is located between the dura mater and the arachnoid. Between the dura mater and the vertebral column is the epidural space, which contains many blood vessels, especially veins, and fat. A hypodermic needle may be inserted into this space in the lower sacral region for the injection of certain anesthetics (caudal anesthesia).

Cervical and Lumbosacral Enlargements.—The spinal cord has two enlargements: the cervical enlargement and the lumbosacral enlargement. These regions are large because these segments are involved in the innervation of the upper and lower extremities respectively. The fifth through eighth cervical segments and the first thoracic segment comprise the cervical enlargement. The spinal nerves arising from the cervical enlargement form the brachial plexus (fig. 1).

Cross Section.—In cross section the spinal cord consists of an H-shaped gray matter surrounded by white matter (fig. 2 and 3). The gray matter is made up of cell bodies of nerve cells (the part of the cell containing the nucleus and other constituents) and nerve fibres oriented in the cross-sectional plane because the fibres enter and leave the gray matter from the white matter and spinal roots. The white matter consists of nerve fibres that are oriented parallel to the long axis of the spinal cord. These fibres are ascending to and descending from the brain and other levels of the spinal cord. The whiteness is due to the fatty sheath (myelin sheath) that surrounds many of the fibres. Some nonneural cells, called neuroglia, are scattered in both the gray matter and the white matter. The arms of the H that extend dorsally (that is, toward the back of the body) are called dorsal (sensory) horns, while the ventrally (toward the front) directed arms are called ventral (motor) horns. The crossbar of the H is the gray commissure and the intermediate gray. The gray matter is largest in the regions of the cervical and lumbosacral enlargements because of their relation to the innervation of the limbs. In the middle of the gray commissure is the small central canal, which is continuous with the ventricles of the brain. The white matter is subdivided into four columns or funiculi: the dorsal column (between the dorsal horns), the ventral column (between the ventral horns) and the lateral columns (lateral to the gray matter).

Anatomical and Functional Organization of the Spinal Cord.—This may be visualized by describing the orientation of the nerve cells of the spinal reflex arcs (see *Physiology,* below). For example, the withdrawal of the hand upon touching a hot stove before it is felt is a spinal reflex. The heat stimulates the terminal endings of the thermal neurons in the hand. The resulting im-



BY COURTESY OF CIBA CLINICAL SYMPOSIA, VOL. 1, NO. 6, AND THE CIBA COLLECTION OF MEDICAL ILLUSTRATIONS

FIG. 2.—CROSS SECTION THROUGH THORACIC VERTEBRA ILLUSTRATING THE SPINAL CORD, VERTEBRA, MENINGES AND ASSOCIATED STRUCTURES

pulses are transmitted by the fibres of these neurons through the peripheral nerves and dorsal roots into the dorsal horn. Each sensory nerve cell has its cell body in a dorsal root ganglion. Thus each nerve cell is a unit structure extending from its peripheral ending in the hand to its termination in the gray matter of the spinal cord.

In the dorsal horn the nerve fibres synapse (point of contact) with one or more other neurons. The second nerve cells in the arc are called intercalated or internuncial neurons, each with a cell body in the gray matter (fig. 3). Some internuncial neurons send axons into the ventral horn and synapse with the motor cells whose cell bodies are in this horn. The axons of the motor cells reach the voluntary muscles via the ventral roots and peripheral nerves. In short, the stimulus transmitted by a nerve goes to the spinal cord via a dorsal root into the gray matter, then to an intercalated cell and finally to the motor cell and the voluntary muscle via the ventral root and the peripheral nerve. Since three series of cells are involved, this is known as a three-neuron reflex. If the internuncial neuron is not involved, only two cells are in the series (two-neuron reflex).

In the knee jerk reflex the sensory cells synapse directly with the motor neurons in the gray matter, and it is called a two-neuron reflex. The internuncial neuron may cross over from the dorsal horn of one side to the ventral horn of the opposite side and synapse with motor cells. This is one means of co-ordinating one side with the other. An intersegmental reflex is a mechanism relaying impulses to higher and lower spinal cord levels to co-ordinate large areas of the body. The fibres of such internuncial neurons leave the gray matter and go up, down or both (fibres may divide into two or more branches) to other spinal levels in that part of the white matter called the fasciculus proprius and comma tract (fig. 4) and re-enter the gray matter to synapse with motor cells.

The white matter of the spinal cord contains nerve fibres that transmit most of the nerve impulses from the body (below the head) to the brain and from the brain to the body. Microscopic examination of the white matter does not give visible evidence of functional subdivisions.

On the basis of numerous experimental data and examinations of spinal cords from persons with spinal injuries, the white matter can be partitioned into functional regions called tracts or pathways. A tract or fasciculus is a collection of nerve fibres having the same origin, function and termination. Those fibres transmitting impulses toward the brain belong to sensory or ascending tracts, while those fibres transmitting impulses from the brain, form motor or descending tracts.

Pathways of Sensation.—The sensations called proprioceptive sensations are transmitted by the fibres in the dorsal columns. These include position sense, pressure sense, vibratory sense, sense of movement and stereognosis. Position sense is the quality that enables a person to know where the parts of his body are in space when his eyes are closed. Pressure sense is the quality of sensation that results from a deformation of the skin; hence it may be called deep touch. When the stem of a vibrating tuning fork is placed in contact with a bone the vibratory sense is appreciated. Stereognosis, or two-point discrimination, is the faculty of perceiving the weight, form and shape of objects and of distinguishing the distance between two touch stimuli.

These senses are transmitted by fibres that pass from the peripheral nerves and dorsal roots into the dorsal horn of the spinal cord. These fibres with their cell bodies in the dorsal root ganglia have several branches in the gray matter. The shorter branches (collateral fibres) are utilized in the spinal reflex arcs. The main axon continues into the dorsal column to ascend in the column to the lower medulla of the brain. Two points of note are: (1) a fibre in the dorsal column is a part of the same neuron that was stimulated in the body, and (2) this fibre ascends on the same side of the spinal cord as it enters (does not cross the mid-line of the cord).

As a consequence, the sense of position of the toes, for example, is carried by nerve cells that extend from the toes to the medulla without a synapse. In addition, injury to the dorsal column on one side of the cord will result in the impairment of the proprioceptive senses on the same side of the body below the level of injury. Each half of the dorsal column is divided into a fasciculus gracilis and a fasciculus cuneatus (fig. 4). The proprioceptive impulses from the lower half of the body, including the lower extremities, are carried in the fasciculus gracilis, while those from the upper half of the body, including the upper extremities but excluding the head, are carried in the fasciculus cuneatus.

Some light touch sensations (gentle brushing of skin by cotton) are transmitted by fibres in the dorsal column. In the medulla the fibres of each half of the dorsal column synapse with neurons in the nucleus gracilis and nucleus cuneatus. The neurons of these nuclei cross over to the opposite side of the medulla and pass upward to the thalamus of the brain, from which they relay to the cerebral cortex.

Pain and temperature sensations are transmitted by another pathway. The fibres from the periphery, in common with other sensory fibres, enter the dorsal horn of the spinal cord where their terminal branches establish reflex connections with other neurons. At the spinal level at which the fibres enter the cord they synapse with dorsal horn neurons. Some of these cells are relays in the reflex arcs. Other neurons from the dorsal horn send axons across the mid-line to the opposite side through the ventral white commissure (ventral to the gray commissure) to join the lateral spinothalamic tract in the lateral columns (fig. 4). This tract contains

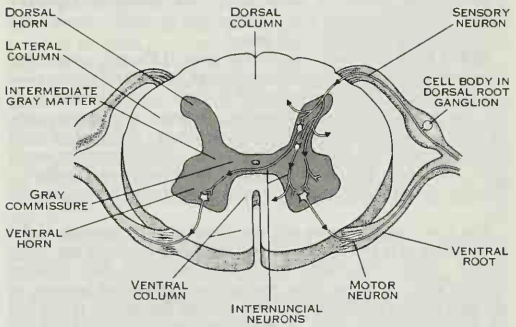
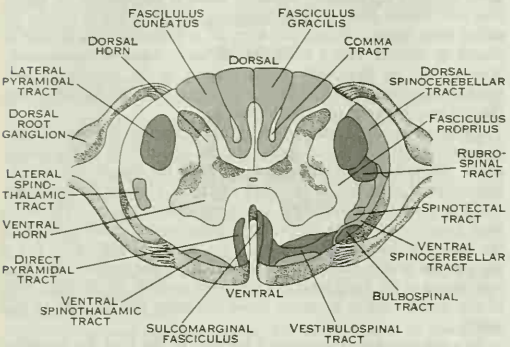


FIG. 3.—DIAGRAM OF CROSS SECTION OF THE SPINAL CORD ILLUSTRATING THE RELATIONSHIP OF THE SENSORY NEURON, INTERNUNCIAL NEURONS AND THE MOTOR NEURON

Arrows on the neurons indicate the direction in which impulses travel

more fibres as it ascends in the spinal cord, for new fibres are added from each spinal level. The tract's name is based on the fact that the fibres of the pathway originate in the dorsal horn of the spinal cord and terminate in the thalamus on the opposite side of the brain. Crude pain and awareness of touch may be brought to consciousness in the thalamus. Impulses from the thalamus are relayed to the cerebral cortex.



BY COURTESY OF CIBA CLINICAL SYMPOSIA, VOL. 1, NO. 5, AND THE CIBA COLLECTION OF MEDICAL ILLUSTRATIONS

FIG. 4.—SCHEMATIC SECTION THROUGH THE SPINAL CORD SHOWING THE LOCATION OF IMPORTANT TRACTS

Dark shaded areas indicate tracts from brain to cord; lighter areas, from cord to brain. Dorsal and ventral horns constitute gray matter

The symptoms resulting from injury to the spinal cord are a consequence of the course of these pathways. For example, if the right half of the spinal cord is cut at a midthoracic level, conscious proprioception would be lost on the right side of the body below the level of the lesion (right abdomen and the right lower extremity). Pain and temperature sensations would be lost on the left side of the body below the level of the lesion. Because of the dual course of their fibres, light touch would be diminished, but not necessarily lost, below the level of the lesion. Perception of sensations from regions above the cut would be unimpaired.

Some unconscious proprioceptive impulses relay from the gray matter of the cord via the spinocerebellar tracts of the same and opposite sides and terminate in the cerebellum (fig. 4). This part of the brain is known as an unconscious proprioceptive organ.

The precise course of the pathways for sensations from the viscera (heart, intestinal tract, kidney and so forth) through the white matter of the cord is not known.

Motor Pathways.—In the ventral horn of the gray matter are large motor cells whose axons pass through the ventral roots to the spinal nerves. These neurons, which innervate the voluntary muscles, are called lower motor neurons. If these cells become nonfunctional, as in poliomyelitis, the muscles are deprived of their nervous stimulation and do not contract. These muscles waste away (atrophy) and do not respond to stimulation (areflexia). This paralytic condition is known as a lower motor neuron paralysis.

The lower motor neuron cells are innervated by neurons of the spinal reflex arcs and by the descending motor tracts of the white matter. These descending tracts from the brain may synapse directly with the motor cells, but usually intercalated neurons of the gray matter are interposed between the descending tracts and the lower motor neurons. Collectively the motor tracts of the white matter are known as upper motor neurons. Their cell bodies are in the brain. Two major subdivisions of these motor tracts are recognized: the pyramidal tracts, which are phylogenetically new, and the extrapyramidal tracts, which are phylogenetically old.

The lateral pyramidal tract or lateral corticospinal tract (fig. 4) arises in the cerebral cortex of the opposite side and crosses over just as its fibres enter the spinal cord. The direct pyramidal tract or ventral corticospinal tract originates in the cerebral cortex and crosses over in the spinal cord. The extrapyramidal tracts include the motor tracts that arise in regions of the brain other than the cerebral cortex. However, these tracts do receive nerve impulses from the cerebral cortex as well as other parts of the brain.

The extrapyramidal tracts include: (1) the reticulospinal tracts from the reticular formation of the brain stem (medulla, pons and midbrain); (2) the vestibulospinal tract, which transmits impulses from the vestibular (balance) mechanisms of the inner ear; (3) the bulbospinal tract, which arises from the medulla or bulb of the lower brain stem; (4) the sulcomarginal tract, a collection of fibres from several regions of the brain stem; and (5) the rubrospinal tract (minor in man) from the nucleus ruber in the midbrain.

When a number of these motor tracts are damaged, an upper neuron paralysis may result. This contrasts with the lower motor neuron paralysis previously mentioned. The reflexes are exaggerated, there is loss of voluntary motion, and the extremities become spastic. The condition, called spastic paralysis, results because the muscles are under control of a spinal cord that still mediates spinal reflex arcs but is deprived of its normal innervation from the brain. Spastic paralysis resulting from an injury of the spinal cord is observed on the same side of the body as the lesion.

The complexity of the nervous system is illustrated, for example, by spinal reflex arcs innervating the voluntary muscles. The motor pathways of the central nervous system discharge on small motor cells in the ventral horn of the spinal cord (fig. 4). These nerve cells whose axonal processes leave the spinal cord through the ventral roots stimulate small muscle fibres in certain sensory endings (neuromuscular spindles) to contract. These activated spindles in the voluntary muscles send impulses via sensory fibres through the dorsal root back to the spinal cord to synapse with the motor cells of the ventral horn. These motor cells stimulate the voluntary muscles to contract. When the voluntary muscles contract sufficiently the sensory endings in the tendons are stimulated to send impulses via sensory fibres through the dorsal root back to the spinal cord to inhibit the motor cells of the ventral horn. In brief, the sensory endings in the muscle and tendon have a critical role in muscle activity.

Autonomic Nervous System.—The autonomic nervous system, which is concerned with the motor innervation of glands, blood vessels and the viscera, has neurons in the gray matter. The nerve cells of the sympathetic division of this system have their cell bodies in the lateral part of the intermediate gray, known as the intermediolateral cell column, of the thoracic and upper lumbar levels. Their axons leave the spinal cord through the ventral roots. The nerve cells of the spinal part of the parasympathetic division of this system have their cell bodies scattered in the gray matter of the midsacral levels. Their axons leave the spinal cord through the ventral roots. The sympathetic nervous system assists the body in meeting emergencies and stress. This includes increasing the rate and force of the heartbeat, the dilatation of the bronchial tubes of the lungs and the inhibition of the motility and the secretory activity of the digestive system. The parasympathetic system contributes to the conservation of bodily energy. This includes decreasing the rate and force of the heartbeat, the constriction of the bronchial tubes and the stimulation of the motility and secretory activity of the digestive system.

EMBRYOLOGY

The nervous system is one of the first organ systems to develop. It differentiates as a long narrow strip of cells, the neural plate, on the dorsal surface of the early embryo. Each side of the plate grows rapidly, folds over and fuses with its counterpart of the other side at the mid-line to form the neural tube with a cavity, the central canal.

During this process the cord sinks below the surface to become covered by embryonic skin. The cells, ependymal cells, lining the

central canal differentiate into the nerve cells of the spinal cord and of the ganglia of the peripheral nervous system and, in addition, some neuroglia cells. In fetal life, the gray matter and the white matter are recognizable. If by some chance the fusion of the neural plate to form the neural tube is incomplete, malformations known as spina bifida may occur.

COMPARATIVE ANATOMY

In all the vertebrates the spinal cord is a tube with a central canal. In the jawless fishes, cyclostomes, the cord is flattened and the neurons are arranged as a pair of thin wings surrounded by unmyelinated fibres. In the sharks (elasmobranchs) the nerve fibres are myelinated, and hence the distinction between white matter and gray matter exists as in the higher vertebrates. In the amphibians the amount of gray matter increases to form the H configuration present in reptiles and mammals. In the reptiles and mammals the increase in the number of cells in the gray matter and of fibres in the white matter is significant. The basic patterns and topographical relations of the gray matter and white matter are similar in these forms. The quantitative increase is related to the complexity of the neural integrative mechanisms in the higher vertebrates.

As compared with the brain, the spinal cord is a conservative structure in vertebrate evolution. (C. R. Nk.)

PHYSIOLOGY

The spinal cord is only a part of the nervous organ that throughout the length of the body forms the meeting place of the nerve paths arriving from and issuing to all regions with which nerve fibres communicate.

To gain a true point of view for understanding the working of the spinal cord it is necessary to refer to the general function of the nervous system in the bodily economy.

Relation to General Nervous System.—An animal of microscopic size may continue throughout its life to be constituted entirely by one single cell. Animals of larger bulk, although each begins its existence as a single cell, attain their development by the multiplication of the original single cell, so that from it there comes to be formed a coherent mass of cells very many millions in number. In these multicellular animals each of the constituent cells is a minute self-centred organism, individually born, leading its own life and destined for individual death. The corporate power of the complex animal is the sum of the powers of those manifold individual existences, its cells.

Of the agencies which integrate the complex animal, one of the most potent is nervous action. A certain number of the unit cells composing the animal are specially differentiated from the rest to bind the whole together by nervous action. These specially differentiated cells are called "neurons." They constitute living threads along which waves of physicochemical disturbance are transmitted to act as releasing forces for the energy in distant cells.

It is characteristic of this nervous system, the system of neurons, that, although ramifying far and wide through the body, it is a continuum from end to end. The peripheral nerves are formed of bundles of neurons lying side by side, but these, although packed close together, are strictly isolated from one another as conductors and remain isolated throughout the whole length of the nerve.

The points of functional nexus of the neurons with one another are confined to one region only of the whole system. All their conductive connections with one another take place solely in the central nervous mass which constitutes the so-called central nervous system, a nervous organ extending axially along the length of the body midway between the body's lateral halves. Thither the neurons converge in vast numbers, those of each body segment converging to that fraction of the central organ which belongs to their body segment. The central nervous organ thus receiving these neurons is called, where it lies in the head, the brain; the rest of it is called in vertebrates the spinal cord, in worms and arthropods the nerve cord. The central organ not only receives neurons that converge to it from outside, but many of its own neurons thrust out their conductive arms from it as nerve fibres carrying nervous influence outward to regulate the activity of glands and muscles.

In the vertebrates the ingoing neurons for each segment, and similarly the outgoing neuron fibres, are collected into a segmental nerve. These are each attached to the spinal cord by two roots, one dorsal, consisting of the afferent fibres, the other ventral, consisting of the efferent fibres.

The Reflex.—The simplest complete reaction of the system is a reflex. There are many reflexes which are extremely complex, being built up of a number of simpler reflexes combined. A reflex is a reaction started by the environment acting as a stimulus upon some nerve which communicates the excitement to other nerves by connections with them in the central nervous organ. The excitement so generated and transmitted finally travels outward from the central organ by one or more of the efferent nerves and through these reaches muscles or glands, producing in them its final effect. The muscles and glands are from this point of view termed effector organs. The reaction is therefore "reflected" from the central organ. The nerve structures which include its trajectory are spoken of as a nervous arc.

The whole purpose of the central nervous organ is to bring afferent neurons into touch with efferent neurons. The whole purpose of reflex arcs is to bind one part of the organism to another part in such a way that what the environment is doing to the organism at one place may appropriately call forth or restrain movement or secretion in the muscles or glands wherever situated in the organism.

Receptor Cells.—There is one condition for the due performance of these reactions which is not provided by the nervous system itself. The afferent neurons are not in most cases so constituted as to be excitable themselves directly by the environment—for instance, they cannot be stimulated by light. Their amenability to the environment, their sensitization to environmental agencies, is effected by special cells adjunct to their peripheral ends. These cells form organs called receptors. They are delicately adapted to be stimulated by this or that particular agent and are classifiable into various species, so that each species is easily excited by a particular agent which is "adequate" for it, and is inexcitable or excitable only with difficulty by agencies of other kinds. Thus in the skin some receptors are adapted for mechanical stimuli (touch) and not for thermal stimuli, while others (cold spots, warm spots) are adapted for thermal stimuli and not for mechanical. As far as it is known, each afferent neuron is connected with receptors of one species only. The receptors thus confer upon the reflex arcs selective excitability. Each arc is thus tuned to respond to certain stimuli, while other arcs not having that kind of receptor do not respond. The receptors, therefore, while increasing the responsiveness of the organism to the environment, prevent confusion of reactions (in-co-ordination) by limiting to particular stimuli a particular reaction.

Proprioceptors.—The system of neurons is thus made accessible to the play of the external world acting on the body. And in addition to those receptors, which are stimulated directly by the external world, there are others lying within the mass of the organism itself, which are excitable by actions occurring in the organism. These are called proprioceptors. They are distributed preponderantly in the muscles and structures functionally adjunct to muscle, such as joints, ligaments, fasciae, etc. The reactions, both passive and active, induced in such motor structures reflexly by environmental stimuli tend therefore secondarily to be followed and accompanied by reflex reactions initiated from proprioceptors.

Conduction.—The process by which the excitement generated in the afferent neuron travels along the reflex arc is known as conduction. Conduction along afferent and efferent nerves differs in some important respects from that obtaining in the nerve centre; *i.e.*, in the piece of the central nervous system connecting the afferent nerve with the efferent nerve. In a nerve trunk the excited state set up in it by a stimulus travels along its fibres as wavelike disturbance at a speed of about 100 m. per second, and does not alter in intensity or speed in its travel. A nerve trunk when ex-

cited at some point along its length transmits the "impulse" (*i.e.*, the wavelike excited state) in both directions (*i.e.*, both up and down each fibre) from the point stimulated. This is true whether the fibre is afferent or efferent. The speed of travel of the nervous impulse along the nerve trunk is the same whether the nervous impulse is weak or intense. The nerve trunk shows practically no delay in its response to an effective, even though weak, stimulus, and its response ceases practically at once on cessation of the exciting stimulus. When excited by repeated brief stimuli the rhythm of the response corresponds closely with that of the stimuli, even when the frequency of the latter is as high as 500 per second. With momentary stimuli a response even so brief as 20 can be given by the nerve trunk. Finally, nerve-trunk conduction is singularly resistant to fatigue, to impoverished blood supply and to many drugs which powerfully affect reflex actions.

Through the central nervous organ the travel of the impulse exhibits departure from these features. Its intensity is liable to be altered in transit. Its time of transit is much longer than for a similar length of nerve trunk. Its direction of transmission becomes polarized; that is, confined to one direction along the nervous path. To a strong stimulus the central reaction, instead of being as brief as an impulse, may endure for a whole second or more. A stimulus not capable of evoking a response from a centre when applied once may by simple repetition become effective (temporal summation).

It is in the gray matter that conduction differs in these respects from conduction in nerve trunks. In the gray matter each afferent fibre breaks up into branching threadlets which ramify in various directions and terminate in close apposition with other neurons. The point of nexus of one neuron with another is termed the synapse. A minute gap of about 200 angstrom units (Å) is the distance between two neurons at the synapse. The absence of cytoplasmic continuity between the neurons is the anatomical basis that accounts (1) for the slight delay in the speed of the impulse at the synapse, and (2) for the polarization of the direction of the impulse.

Reflex Reactions.—When the spinal cord is severed at any point the reflex arcs of the portion of the body behind the transection are quite cut off from the rest of the nervous system in front, including the brain. The reflex reactions elicited from the thus isolated region cannot therefore be modified by the action of the higher nervous centres. In the head the local centres are overlaid by higher centres which cannot by any simple severance be separated from them. By studying, therefore, the powers of the cord behind a complete spinal transection, information can be obtained as to the powers of the purely local or segmental reflex mechanisms.

The so-called "flexion reflex" of the limb is one of the most accessible of the local reflex reactions which can thus be studied with an isolated portion of the spinal cord as its centre.

Let it be supposed that the limb observed is the hind limb. The three main joints of the limb are the hip, the knee and the ankle. Each of these joints is provided with muscles which flex or bend it, and others which extend or straighten it. It is found that the reflex throws into contraction the flexor muscles of each of these joints. It matters little which of all the various afferent nerves of the limb is stimulated, whichever of these the afferent nerve may be, the centrifugal discharge goes to practically the same muscles, namely, always to the flexors of the joints.

Not only does the reflex action not discharge motor impulses into the nerves of the extensor muscles, but if the spinal cord happens to be discharging impulses into these nerves when the reflex is evoked this discharge is suppressed or diminished (inhibited). In this way the latter muscles are prevented from impeding the action of the contracting flexors. This inhibition prevents other reflexes from upsetting for the time being the due action of the flexion reflex, for it renders the muscles opposing that reflex less accessible to motor discharge through the spinal cord, whatever the quarter from which incitation to that discharge may come.

A feature of this reflex is its graded intensity. A weak stimulus evokes in the flexor muscles a contraction which is weak and in the extensor muscles a relaxation which is slight. In the case of the muscle fibre, any strength of stimulus, if sufficient to excite it at all, excites the fibre totally and evokes its full contraction, and this latter remains the same in amount so long as the mechanical and nutritive conditions of the fibre remain constant. With each muscle fibre reacting on this "all-or-none" principle and each nerve fibre similarly responding on that same principle, different amounts of contraction yielded by the muscle under different strengths of stimulation of its nerve signify differences simply in the number of muscle fibres the stimulus excites. In other words the weaker stimuli excite fewer of the nerve fibres innervating the muscle.

When the tension developed by the muscle (*e.g.*, against a stiff spring) reaches the maximal obtainable from it, that fact tells us that all the muscle fibres are then in contraction; when the tension development is less than maximal, some of the muscle fibres are not in action. A stimulus to the nerve which evokes maximal tension from the muscle is one which succeeds in exciting all the nerve fibres innervating the muscle, and any stimulus which evokes less than maximal tension is leaving some of the nerve fibres idle.

The tension developed by a muscle under reflex excitation falls short of its maximal tension, even when the afferent nerve is stimulated maximally. The reflex activates a proportion only of the total aggregate of the muscle's fibres, and therefore a proportion only of the spinal motor neurons innervating the muscle. But within the range of the limit imposed by the maximal number of muscle fibres which it can activate, the grading of the number activated in accordance with the grading of the strength of stimulation of the afferent nerve is very delicate, much more so than in the case of the motor nerve itself. The actual fractional proportion of the muscle's total aggregate of motor units which an afferent nerve, maximally stimulated, can activate, differs from one afferent nerve to another, and for the same afferent nerve in different experiments. A subconvulsive dose of strychnine raises the size of the fraction. The threshold value of stimulus for the reflex is usually higher than for the contraction of the muscle from the motor nerve.

Spinal reflex action no more regards "muscles" than does a motor centre of the cerebral cortex. Both ignore muscles as entities and have in view movements purely. A weak flexion reflex activates a few fibres of each of the several flexor muscles of the limb. The reflex treats a complex composed from some (*i.e.*, those of similar threshold for the particular afferent nerve) motor units of hip flexor, of knee flexor and of ankle flexor as functionally more homogeneous than is the total group of fibres composing any one of these muscles alone.

Reflex inhibition has its seat in the gray matter of the spinal centre. It acts at some point in the reflex arc upstream from the all-or-none motor unit (motor nerve fibre with its group of muscle fibres). The inhibitory reflex is, like the excitatory, capable of very delicate grading; its effect in diminishing or precluding this or that amount of an excitatory reflex is effected entirely by regulating the number of the motor units thrown out of action or kept out of action. Intensity of contraction or of relaxation is thus in both cases a question simply of number of motor units reacting. The reflex centre continues discharging impulses for a certain time after its exciting stimulus has ceased. This after-discharge succeeding a strong stimulus may persist even for several seconds.

Refractory Phase.—Besides characters common to all or many spinal reflexes certain spinal reflexes have features peculiar to themselves or exhibited by them in degrees not obvious in other reflexes. One of these features is refractory phase. The scratch reflex exemplifies this. In the dog, cat and many other animals the hind limb often performs a rapid scratching movement, the foot being applied to the skin of the shoulder or neck as if to groom the hairy coat in that region. This movement is in the intact animal under control of the brain, and can be executed or desisted from at will. When certain of the higher centres in the brain have been destroyed, this scratching action occurs very readily and in an uncontrolled way. When the spinal cord has been severed in the neck this scratching movement of the hind limb can be elicited with regularity as a spinal reflex by merely rubbing the skin of the side of the neck or shoulder, or applying there a weak

electric current to the skin.

In this reflex the stimulus excites afferent nerves connected with the hairs in the skin, and these convey impulses to the spinal centres in the neck or shoulder segments; these in turn discharge impulses into nerve fibres entirely intraspinal, passing backward along the cord to reach motor centres in the hind limb region. These motor centres in turn discharge centrifugal impulses into the muscles of the hind limb of the same side of the body as the shoulder that is the seat of irritation. The motor discharge is peculiar in that it causes the muscles of the hind limb to contract rhythmically at a rate of about four contractions per second, and the discharge is peculiar further in that it excites the flexor and extensor muscles of the joints alternately so that at the hip, for instance, the limb is alternately flexed and extended, each single phase of the movement lasting about an eighth of a second.

Now this rhythmic discharge remains the same in rate whether the exciting stimulus applied to the skin is continuous or one of many various rates of repetition. Evidently at some point in the reflex arc there is a mechanism that, after reacting to the impulses reaching it, remains for a certain brief part of a second unresponsive, and then becomes once more for a brief period responsive, and so on. This phasic alternation of excitability and inexcitability repeats itself throughout the continuance of the reflex. The phase of inexcitability is termed the refractory phase. Its seat lies in the spinal centre. A similar element almost certainly forms part of the co-ordinating mechanism for many other cyclic reflexes, including those of the stepping of the limbs, the movement of the jaw in mastication, the action of the eyelids in blinking, and perhaps the respiratory movements of the chest and larynx.

Fatigue.—Nerve trunks do not easily tire out under even the most prolonged stimulation. Reflex actions on the other hand relatively soon tire. Some are more resistant, however, than are others. The flexion reflex may be continued for ten minutes at a time. As a reflex tires, the muscular contraction which it causes tends to become less intense and less steady. The relatively rapid onset of fatigue in reflexes is counterbalanced by speedy recovery in repose. A long flexion reflex, when from fatigue it has become weak, tremulous and irregular, will recommence after 30 seconds' repose with almost the same vigour and steadiness as if it had not recently been tired out.

The natural movements to which the artificially provoked reflexes seem to correspond do not demand prolonged motor activity, or when they do, demand it in rhythmic repetition with intervening pauses which allow repose.

Reflex Postures.—But there are certain reflexes which do persist for long periods at a time. These are reflex postures. The hind limbs of the "spinal" frog assume an attitude which is reflex, for it ceases on severance of the afferent spinal roots. This attitude is one of flexion at hip, knee and ankle, resembling the well-known natural posture of the frog as it squats when quiet in the tank. Similarly in the decerebrate dog or cat certain muscles (*e.g.*, the limb extensors) exhibit a persistent contraction. These tonic reflexes are related to attitudes. In the dog and cat they are exhibited by those muscles whose action antagonizes gravity in postures which are usual in the animal; thus the extensors of the knee and hip and shoulder and elbow are in tonic contraction during standing. The postural contraction is accompanied by electrical "action currents" like those of other contractions, though weak and indicative of self-smothering from the rhythms being asynchronous in the contributory fibre groups. The postural contraction of the extensor muscles in the maintenance of the erect posture is traceable to a reflex, called the stretch reflex. A stretch (*e.g.*, by a pull upon the tendon) lengthening the muscle by even so little as 0.5% excites reflex contraction of some of its fibres. The passive stretch is a mechanical stimulus to some of the proprioceptive organs (perhaps the muscle spindles) in the muscle. The reflex is therefore unobtainable when the afferent nerve fibres of the muscle itself have been severed. The essential centre for the stretch reflex is spinal, but midbrain and cerebellar centres much reinforce it. In the erect posture the head, neck, tail and jaw would droop and the limbs fold up under their own weight and that of superincumbent parts, were not the "antigravity" muscles checked from yielding by the stretch reflex which their own passive yield induces in themselves. Hence the reflexly maintained "standing" is a multiple stretch reflex. Considerable latitude of actual pose is allowed to the individual parts because each muscle involved acts for itself and develops and regulates its own stretch reflex.

Characteristic of reflex postural contraction is its continuance for long spells at a time without obvious fatigue. Attempts to reproduce this feature by electrical stimulation of afferent nerves or of the muscle's motor nerve are uniformly unsuccessful. The suggestion has therefore been made that the postural contraction of muscle is due to a different process in the muscle fibre from that of the ordinary contraction executing movements and so easily and often studied under electrical stimulation in the laboratory. Forbes, however, offered the satisfactory explanation that the postural contraction, while the same process as that of ordinary reflex and other contraction, involves usually relatively few motor units at a time and these, as they fatigue, are replaced by others. The stretch which is the essential stimulus is automatically shifted to fresh motor units as soon as those already in operation begin to yield and give out.

The eyeballs are eminently organs which, the gaze being a posture, employ active postures throughout the waking day. The proprioception of the neck muscles excites from the cervical spinal cord reflex influences acting on the nerves of the eyeball muscles. An instance is: turning of the neck toward one side excites turning of the eyeballs toward the opposite side. Or again, a clockwise partial rotation of the neck round its long axis excites from the neck proprioceptor a compensatory anticlockwise rotation of the eyeballs. These reflex postures of the eyeballs are sometimes called compensatory, because they tend to keep truly vertical the vertical of the retina despite displacements of the head by the movements of the neck.

It may be mentioned here that the spinal and the bulbospinal animal (cat), although its composite stretch reflex enables it when placed in the erect posture to maintain that attitude, even for hours at a time, is yet not capable of assuming that posture when laid, for instance, on its side. But when the midbrain is retained in addition to hindbrain and cord, the animal has that power of righting itself. These righting reflexes are likewise postural and largely evoked by the otolith organs. Although not truly spinal reflexes, their field of nervous operation overruns into the spinal cord.

Interaction Between Reflexes.—At the commencement of every reflex arc is a receptive neuron extending from the receptive surface to the central nervous organ. This neuron forms the sole avenue which impulses generated at its receptive point can use, whatever their destination may be. This neuron is therefore a path exclusive to the impulses generated at its own receptive point, and other receptive points than its own cannot employ it. A single receptive point may play reflexly upon a considerable number of different effector organs. It may be connected through its reflex path with many muscles and glands in many different regions. Yet all its reflex arcs spring from the one single shank or stem; *i.e.*, from the one afferent neuron that conducts from the receptive point at the periphery into the central nervous organ.

But at the termination of every reflex arc is found a final neuron, the ultimate conductive link to an effector organ (muscle or gland). This last link in the chain (*e.g.*, the motor neuron) differs obviously in one important respect from the first link of the chain. It does not subserve exclusively impulses generated at one single receptive source, but receives impulses from many receptive sources situated in many and various regions of the body. It is the sole path which all impulses, no matter whence they come, must travel if they are to act on the muscle fibres to which it leads.

Therefore, while the receptive neuron forms a private path exclusively serving impulses of one source only, the final or efferent neuron is, so to say, a public path common to impulses arising at any of many sources of reception. A receptive field (*e.g.* an area of skin) is analyzable into receptive points. One and the same effector organ stands in reflex connection not only with many

individual points, but even with many various receptive *fields*. Reflexes generated in manifold sense organs can pour their influence into one and the same muscle. Thus a limb muscle is the *terminus ad quem* of many reflex arcs arising in many various parts of the body. Its motor nerve is a path common to all the reflex arcs which reach that muscle.

Reflex arcs show, therefore, the general features that the initial neuron of each is a private path belonging exclusively to a single receptive point (or small group of points); and that finally the arcs discharge into a path leading to an effector organ; and that their final path is common to all receptive points wherever they may lie in the body, so long as they have connection with the effector organ in question. Before finally converging upon the motor neuron the arcs converge to some degree. Their private paths join internuncial paths common in various degrees to groups of private paths. The terminal path may, to distinguish it from internuncial common paths, be called the final common path. The motor nerve to a muscle is a collection of final common paths.

Certain consequences result from this arrangement. One of these is the preclusion of essential qualitative difference between nerve impulses arising in different afferent nerves. If two conductors have a tract in common, there can hardly be essential qualitative difference between their modes of conduction.

A second consequence is that, each receptor being dependent for final communication with its effector organ upon a path not exclusively its own but common to it with certain other receptors, such nexus necessitates successive and not simultaneous use of the common path by various receptors using it to different or opposed effect.

In the simultaneous correlation of reflexes some reflexes combine harmoniously, being reactions that mutually reinforce. These may be termed allied reflexes, and the neutral arcs which they employ may be termed allied arcs. On the other hand, some reflexes, as mentioned above, are antagonistic one to another and incompatible. These do not mutually reinforce, but stand to each other in inhibitory relation. One of them inhibits the other.

Allied Reflexes.—What happens when trains of impulses, traveling by convergent afferent arcs, meet at the same "common path"? With convergent reflexes that are "allied" in the above sense, this case must be one of extremely frequent occurrence and the management of the interaction between the convergent streams of impulses must be a fundamental factor in nervous co-ordination. The afferent path A discharges into the intraspinal motor neurons of the knee flexor and can activate 75% of them. The afferent path B also discharges into the motor neurons of the same muscle and can activate 75% of them. When the muscle is reflexly contracting to A with a tension 75% of that of the muscle's maximal contraction, stimulation of B is added. The contraction tension in result rises to 85% of the maximal. The inference is that for 65% of the total motor units of the muscle the excitation from A acts upon the same motor units as would that from B. In other words 65% of the total aggregate of the motor units of the muscle are common to A and to B. The stream of impulses arriving at these common motor units from B when A's stream is already activating them leaves their activity unchanged, failing either to increase or to diminish it. A's stream, which is engaging them, precludes B's stream from engaging them. This is termed occlusion.

The final common path under the competition of allied arcs is comparable with a telephone line already busy for one subscriber and for that time engaged and nonavailable to other subscribers. The action of the principle of the final common path may be instanced in regard to allied arcs in the scratch reflex as follows. If, while the scratch reflex is being elicited from a skin point at the shoulder, a second point, distant 10 mm. from the other point but also in the receptive field of skin, is stimulated, the stimulation at this second point favours the reaction from the first point. This is well seen when the stimulus at each point is of subminimal intensity. The two stimuli, though each unable separately to invoke the reflex, yet do so when applied both at the same time. The receptive field of a reflex is really the common area of commencement of a number of allied arcs.

Antagonistic Reflexes.—But not all reflexes connected to one and the same common final path stand to one another in the relation of allied reflexes. Suppose during the scratch reflex a stimulus is applied to the foot not of the scratching side but of the opposite side. The left leg, which is executing the scratch reflex in response to stimulation of the left shoulder skin, is cut short in its movement by the stimulation of the right foot, although the stimulus at the shoulder to provoke the scratch movement is maintained unaltered all the time. The stimulus to the right foot will temporarily interrupt a scratch reflex, or will cut it short or will delay its onset; which it does of these depends on the time relations of the stimuli. The inhibition of the scratch reflex occurs sometimes when the contraction of the muscles innervated by the reflex conflicting with it is very slight. There is interference between the two reflexes and the one is inhibited by the other.

The reflex from the right foot evokes at the opposite (left) knee extension; in doing this it causes steady excitation of extensor neurons of that knee and steadily inhibits the flexor neurons. But the scratch reflex causes rhythmic excitation of the flexor neurons. Therefore these flexor neurons in this conflict lie as a final common path under the influence of two antagonistic reflexes, one of which would excite them to rhythmical discharge four times a second, while the other would continually repress all discharge in them.

In all these forms of interference there is a competition, as it were, between the excitatory stimulus used for the one reflex and the excitatory stimulus for the other. Both stimuli are in progress together, and the one in taking effect precludes the other's taking effect as far as the final common path is concerned.

Again, if, while stimulation of the skin of the shoulder is evoking the scratch reflex, the skin of the hind foot of the same side is stimulated, the scratching may be arrested. Stimulation of the skin of the hind foot by any of various stimuli that have the character of threatening the part with damage causes the leg to be flexed, drawing the foot up by steady maintained contraction of the flexors of the ankle, knee and hip. Here, therefore, there is an arc that discharges into a final path, common to it and to the scratch reflex arc; both these arcs employ the same effector organ, namely, the knee flexor, and employ it by the common medium of the final pathway. But though the channels for both reflexes discharge into the same final common path, the excitatory flexor effect specific to each differs strikingly in the two cases. In the scratch reflex the flexor effect is an intermittent effect; in the nociceptive flexion reflex the flexor effect is steady and maintained. The scratch reflex is set aside by that of the nociceptive arc from the homonymous foot.

The stimulation which previously sufficed to provoke the scratch reflex is no longer effective, though it is continued all the time. But when the stimulation of the foot is discontinued the scratch reflex returns. In that respect, although there is no enforced inactivity, there is an interference that is tantamount to, if not the same thing as, inhibition. Though there is no cessation of activity in the motor neuron, one form of activity that was being impressed upon it is cut short and another takes its place. A stimulation of the foot too weak to cause more than a minimal reflex often will suffice completely to interrupt, cut short or prevent onset of the scratch reflex.

The kernel of the interference between the homonymous flexion reflex and the scratch reflex is that both employ the same final common pathway to different effect—just as in the interference between the crossed extension reflex and the scratch reflex. Evidently, the homonymous flexion reflex and the crossed extension reflex both use the same final common pathway. And they use it to different effect. The motor neuron to the flexor of the knee being taken as a representative of the final common path, the homonymous flexion reflex inhibits it from discharging. Hence if, while the direct flexion reflex is in progress, the crossed foot is stimulated, the reflex of the knee flexor is inhibited. The crossed extension reflex therefore inhibits not only the scratch reflex, but also the homonymous flexion reflex.

Further, in all these interferences between reflexes the direction taken by the inhibition is reversible. Thus, the scratch reflex is

not only liable to be inhibited by, but is itself able to inhibit either the homonymous flexion reflex or the crossed extension reflex; the homonymous flexion reflex is not only capable of being inhibited by the crossed extension reflex, but conversely in its turn can inhibit the crossed extension reflex. These interferences are therefore reversible in direction. Certain conditions determine which reflex among two or more competing ones shall obtain mastery over the final common path and thus obtain expression.

As to the intimate nature of the mechanism which thus, by summation or by interference, gives co-ordination where neurons converge upon a common path, it is difficult to surmise. In the central nervous system of vertebrates, afferent neurons A and B in their convergence toward and impingement upon another neuron Z, toward which they conduct, do not make any lateral connection directly one with the other—at least, there seems no clear evidence that they do. It seems, then, that the only structural link between A and B is neuron Z itself. Z itself should therefore be the field of coalition of A and B if they transmit allied reflexes.

Factors Determining the Sequence.—The formation of a common path from tributary converging afferent arcs is important because it gives a co-ordinate mechanism. There the dominant action of one afferent arc, or set of allied arcs in condominium, is subject to supersession by another afferent arc, or set of allied arcs, and the supersession normally occurs without intercurrent confusion.

Whatever may be the nature of the physiological process occurring between the competing reflexes for dominance over the common path, the issue of their competition, namely, the determination of which one of the competing arcs shall for the time being reign over the common path, is largely conditioned by four factors. These are spinal induction, relative fatigue, relative intensity of stimulus and the functional species of the reflex.

Spinal Induction.—Induction occurs in two forms, one of which has been named immediate induction. The stimulus that excites a reflex tends by central spread to facilitate and lower the threshold for reflexes allied to that which it particularly excites. Thus there tends to be formed a constellation of reflexes that reinforce each other, so that a reflex figure results. If the prepotent stimulus shifts, allied arcs are by the induction particularly prepared to be responsive to it or to a similar stimulus.

Immediate induction occurs only between allied reflexes. Its tendency in the competition between afferent arcs is to fortify the reflex just established, or, if transition occurs, to favour transition to an allied reflex. Immediate induction seems to obtain with highest intensity at the outset of a reflex, or at least near its commencement.

The other form of spinal induction is successive induction. It is in several ways the reverse of the preceding. If the crossed-extension reflex of the limb of the spinal dog is elicited at regular intervals, say once a minute, by a carefully adjusted electrical stimulus of defined duration and intensity, the resulting reflex movements are repeated each time with much constancy of character, amplitude and duration. If in one of the intervals a strong prolonged (*e.g.*, 30 seconds) flexion reflex is reduced from the limb yielding the extensor reflex movement, the latter reflex is found intensified after the intercurrent flexion reflex. The intercalated flexion reflex lowers the threshold for the aftercoming extension reflexes, and especially increases their after-discharge. This effect may endure, progressively diminishing, through four or five minutes, as tested by the extensor reflexes at successive intervals. Now, during the flexion reflex the extensor arcs were inhibited; after the flexion reflex these arcs are in this case evidently in a phase of exalted excitability. The phenomenon presents obvious analogy to visual contrast. The exaltation aftereffect may ensue with such intensity that simple discontinuance of the stimulus maintaining one reflex is immediately followed by spontaneous appearance of the antagonistic reflex.

The so-called mark-time reflex of the spinal dog is an alternating stepping movement of the hind limbs which occurs on holding the animal up so that its limbs hang pendent. It can be inhibited by stimulating the skin of the tail. On cessation of that stimulus the stepping movement sets in more vigorously and at a quicker rate than before. This afterincrease might be explicable in either of two ways. It might be due to the mere repose of the reflex centre, the repose so recruiting the centre as to strengthen its subsequent action. But a similar period of repose obtained by simply supporting one limb—which causes cessation of the reflex in both limbs, the stimulus being stretch of the hip flexors under gravity—is not followed by afterincrease of the reflex, or the afterincrease might result from the inhibition being followed by a rebound to superactivity. This latter seems to be the case. The afterincrease occurs even when both hind limbs are passively lifted from below during the whole duration of the inhibitory stimulus applied to the tail. And the reflex inhibition of the knee extensor by stimulation of the central end of its own nerve is followed by marked rebound to superactivity of the extensor itself. Again, the knee jerk, after being inhibited by stimulation of the hamstring nerve, is more brisk than before the inhibition.

By virtue of this spinal contrast, therefore, the extension reflex predisposes to and may actually induce a flexion reflex, and conversely the flexion reflex predisposes to and may actually induce an extension reflex. This process is qualified to play a part in linking reflexes together in alternating sequence.

A reflex movement must generate in its progress a number of further stimuli and throw up a shower of centripetal impulses from the moving muscles and joints into the spinal cord. Squeezing of muscles and stimulation of their afferent nerves and those of joints, etc., elicit reflexes. The primary reflex movement might be expected, therefore, of itself to initiate further reflex movement, and that secondarily to initiate further still, and so on. Yet on cessation of the external stimulus to the foot in the flexion reflex, the whole reflex comes usually at once to an end. The scratch reflex, even when violently provoked, ceases usually within two seconds of the discontinuance of the external stimulus that provoked it. There is as yet no satisfactory explanation of this.

Relative Fatigue.—Another condition influencing the issue of competition between reflexes of different sources for possession of one and the same final common path is fatigue. A spinal reflex under continuous excitation or frequent repetition becomes weaker, and may cease altogether. This decline is progressive, and takes place earlier in some kinds of reflexes than it does in others. In the spinal dog the scratch reflex under ordinary circumstances tires much more rapidly than does the flexion reflex.

A reflex as it tires shows other changes besides decline in amplitude of contraction. Thus in the flexion reflex, the original steadiness of the contraction decreases; it becomes tremulous, and the tremor becomes progressively more marked and more irregular. Finally, an irregular phasic tremor of the muscles is all that remains. It is not the flexor muscles themselves which tire out, for these, when under fatigue of the flexion reflex contract no longer for that reflex but do contract in response to the scratch reflex which also employs them.

Similar results are furnished by the scratch reflex, with certain differences in accord with the peculiar character of its individual charge. One of these latter is the feature that the individual beats of the scratch reflex usually become slower and follow each other at slower frequency. Also the beats, instead of remaining fairly regular in amplitude and frequency, tend to succeed in somewhat regular groups. The beats may disappear altogether for a short time, and then for a short time reappear.

When the scratch reflex elicited from a spot of skin is fatigued, the fatigue holds for that spot, but does not implicate the reflex as obtained from the surrounding skin. The reflex is, when tired out to stimuli at that spot, easily obtainable by stimulation two or more centimetres away.

The local fatigue of a spinal reflex seems to be recovered from with remarkable speed. A few seconds' remission of the stimulus suffices for marked though incomplete restoration of the reaction. Fatigue seems a process elaborated and preserved in the selective evolution of the neural machinery. One obvious use attaching to it is the prevention of the too-prolonged continuous use of a common path by any one receptor. It precludes one receptor from occupying for long periods an effector organ to the exclusion of all

other receptors. It prevents long continuous possession of a common path by any one reflex of considerable intensity. It favours the receptors taking turn about. It helps to ensure serial variety of reaction. The organism, to be successful in a million-sided environment, must in its reaction be many sided. Were it not for such so-called fatigue, an organism might, in regard to its receptivity, develop an eye, or an ear, or a mouth, or a hand or leg, but it would hardly develop the marvelous congeries of all those various sense organs which it is actually found to possess.

The final efferent-root neuron forms the instrument for many different reflex arcs and acts. It is responsive to them in various rhythms and in various grades of intensity. In accordance with this, it seems from experimental evidence to be relatively indefatigable.

Relative Intensity of Stimulus.—In the transition from one reflex to another a final common path changes hands and passes from one master to another. A fresh set of afferent arcs becomes dominant on the supersession of one reflex by the next. Of all the conditions determining which one of competing reflexes shall for the time being reign over a final common path, the intensity of reaction of the afferent arc itself relatively to that of its rivals is probably the most powerful. Other things being equal, an afferent arc that strongly stimulates is more likely to capture the common path than is one excited feebly. A stimulus can establish its reflex and inhibit an opposed one only if it has intensity. This explains why, in order to produce examples of spinal inhibition, recourse has so frequently been made in past times to strong stimuli. A strong stimulus will inhibit a reflex in progress although a weak one will fail. Thus in inhibition of micturition in the spinal dog a forcible squeeze of the tail will do it, but not a weak squeeze. So, likewise, any condition that raises the excitability and responsiveness of a nervous arc will give it power to inhibit other reflexes, just as it would if it were excited by a strong stimulus.

Crossed reflexes are usually less easy to provoke, less reliable of obtainment, and less intense than are direct reflexes. Consequently crossed reflexes usually are more easily inhibited and replaced by direct reflexes than are these latter by those former. Thus the crossed stepping reflex is easily replaced by the scratch reflex, though its stimulus may be continued all the time and though the scratch reflex itself is not a very potent reflex. But the reverse can occur with suitably adjusted intensity of stimuli.

Again, the flexion reflex of the dog's leg, when fully developed, is accompanied by extension in the opposite leg. This crossed extensor movement, though often very vigorous, may be considered as an accessory and weaker part of the whole reflex, of which the prominent part is flexion of the homonymous limb. When the flexion reflex is poorly elicitable—as, for instance, in spinal shock or under fatigue or weak excitation—the crossed extension does not appear. But, where the flexion reflex is well developed, if not merely one foot but both are stimulated simultaneously with stimuli of fairly equal intensity, steady flexion at knee, hip and ankle results in both limbs, and extension occurs in neither limb. The contralateral part of each reflex is inhibited by the homolateral flexion of each reflex. In other words, the more intense part of each reflex obtains possession of the final common paths at the expense of the less intense portion of the reflex. But if the intensity of the stimuli applied to the right and left feet is not balanced closely enough, the crossed extension of the reflex excited by the stronger stimulus is found to exclude even the homonymous flexion that the weaker stimulus should and would otherwise evoke from the leg to which it is applied.

It was pointed out above that in a number of cases the transference of control of the final common pathway from one afferent arc to another is reversible. Other things being equal, the direction of the transference can be easily governed by making the stimulation of this receptor or that receptor the more intense. A factor largely determining whether a reflex succeed another or not is therefore intensity of stimulus.

Functional Species of Reflex.—A fourth main determinant for the issue of the conflict between rival reflexes seems the functional species of the reflexes. Reflexes initiated from a species of receptor apparatus that may be termed nociceptive appear to dominate particularly the majority of the final common paths issuing from the spinal cord. In the simpler sensations experienced from various kinds of stimuli applied to the skin there can be distinguished those of touch, of cold, of warmth and of pain. The pain ending may be regarded as adapted to a whole group of excitants, a group of excitants that has in relation to the organism one feature common to all its components, namely, a nocuous character.

Considering the organism's liability to various kinds of mechanical and other damage, in a world beset with dangers amid which the individual and species have to win their way in the struggle for existence, nocuous stimuli may be regarded as part of a normal state of affairs. The skin has evolved a specific sense of its own injuries. As psychical adjunct to the reactions of that apparatus, a strong displeasurable effective quality is found in the sensations they evoke. This may be a means for branding upon memory, of however rudimentary a kind, a feeling from past events that have been perilously critical for the existence of the individuals of the species. In other words, if it is admitted that damage to such an exposed sentient organ as the skin must in the evolutionary history of animal life have been sufficiently frequent in relation to its importance, then the existence of a specific set of nerves for skin pain seems to offer no genetic difficulty, any more than does the clotting of blood or innate immunity to certain diseases. That these nerve endings constitute a distinct species is argued by their all evoking not only the same species of sensation, but the same species of reflex movement as regards purpose, intensity, resistance to shock, etc. And their evolution may well have been unaccompanied by evolution of any specialized end organ, since the naked free nerve endings would better suit the wide and peculiar range of stimuli, reaction to which is in this case required. A low threshold was not required because the stimuli were all intense, intensity constituting their harmfulness; but response to a wide range of stimuli of different kinds was required, because harm might come in various forms. That responsive range is supplied by the naked nerve itself, and would be cramped by the specialization of an end organ. Hence these nerve endings remained free.

It is those areas, stimulation of which, as judged by analogy, can excite pain most intensely, and it is those stimuli which, as judged by analogy, are most fitted to excite pain, which, as a general rule, excite in the spinal animal—where pain is of course nonexistent—the prepotent reflexes. The nervous arcs of pain nerves, broadly speaking, dominate the spinal centres in peculiar degree. Pain is thus the psychical adjunct of an imperative protective reflex. It is preferable, however, since into the merely spinal and reflex aspect of the reaction of these nerves no sensation of any kind can be shown to enter, to avoid the term "pain nerves." Remembering that the feature common to all this group of stimuli is that they threaten or actually commit damage to the tissue to which they are applied, a convenient term for application to them is nocuous. In that case what from the point of view of sense are cutaneous pain nerves are from the point of view of reflex action conveniently termed nociceptive nerves.

In the competition between reflexes the nociceptive as a rule dominate with peculiar certainty and facility. This explains why such stimuli have been so much used to evoke reflexes in the spinal frog, and why, judging from them, such "fatality" belongs to spinal reflexes.

One and the same skin surface will in the hind limb of the spinal dog evoke one or other of two diametrically different reflexes according as the mechanical stimulus applied be of noxious quality or not, a harmful insult or a harmless touch. A needle prick to the sole of the foot causes invariably the drawing up of the limb—the flexion reflex. A harmless smooth contact, on the other hand, causes extension—the extensor thrust above described. This flexion is therefore a nociceptive reflex. But the scratch reflex—which is so readily evoked by simple light irritation of the skin of the shoulder—is relatively mildly nociceptive. When the scratch reflex and the flexion reflex are in competition for the final neuron common to them, the flexion reflex more easily dispossesses the scratch reflex from the final neuron than does the scratch re-

flex the flexion reflex. If both reflexes are fresh, and the stimuli used are such as, when employed separately, evoke their reflexes respectively with some intensity, in my experience it is the flexion reflex that is usually prepotent. Yet if, while the flexion reflex is being moderately evoked by an appropriate stimulus of weak intensity, a strong stimulus suitable for producing the scratch reflex is applied, the steady flexion due to the flexion reflex is replaced by the rhythmic scratching movement of the scratch reflex, and this occurs though the stimulus for the flexion reflex is maintained unaltered. When the stimulus producing the scratch is discontinued the flexion reflex reappears as before.

In decerebrate rigidity, where a tonic reflex is maintaining contraction in the extensor muscles of the knee, stimulation of the nociceptive arcs of the limb easily breaks down that reflex. The nociceptive reflex dominates the motor neuron previously held in activity by the postural reflex. And nociceptive reflexes are relatively little depressed by "spinal shock."

Besides those receptors attuned to react to direct noxa, the skin has others, concerned likewise with functions of vital importance to the species and colligate with sensations similarly of intense affective quality; for instance, those concerned with sexual functions. In the male frog the sexual clasp is a spinal reflex. The cord may be divided both in front and behind the brachial region without interrupting the reflex. Experiment shows that from the spinal male at the breeding season, and also at other times, this reflex is elicited by any object that stimulates the skin of the sternal and adjacent region. In the intact animal, on the contrary, other objects than the female are, when applied to that region, at once rejected, even though they be wrapped in the fresh skin of the female frog and in other ways made to resemble the female.

It would seem a general rule that *reflexes arising in species of receptors which considered as sense organs provoke strongly affective sensation prevail, other things being equal, over reflexes of other species when in competition with them for the use of the final common path.*

Of all reflexes it is those of ordinary posture that are the most easily interrupted by other reflexes. Even a weak stimulation of the nociceptive arcs arising in the foot often suffices to lower or abolish the knee jerk or the reflex extensor tonus of the elbow or knee. If various species of reflex are arranged, therefore, in their order of potency in regard to power to interrupt one another, the reflexes initiated in receptors which considered as sense organs excite sensations of strong affective quality lie at the upper end of the scale, and the reflexes that are answerable for the postural tonus of skeletal muscles lie at the lower end of the scale.

One great function of the tonic reflexes is to maintain habitual attitudes and postures. They form, therefore, a nervous background of active equilibrium. It is of obvious advantage that this equilibrium should be easily upset, so that the animal may respond agilely to the passing events that break upon it as intercurrent stimuli.

Results.—Intensity of stimulation, fatigue and freshness, spinal induction, functional species of reflex, are all, therefore, physiological factors influencing the result of the interaction of reflex arcs at a common path. It is noticeable that they all resolve themselves ultimately into intensity of reaction. Thus, intensity of stimulus means, as a rule, intensity of reaction. Those species of reflex which are habitually prepotent in interaction with others are those which are habitually intense; those specially impotent in competition are those habitually feeble in intensity—*e.g.*, skeletal muscular tone. The tonic reflexes of attitude are of habitually low intensity, easily interfered with and temporarily suppressed by intercurrent reflexes, these latter having higher intensity.

The high variability of reflex reactions from experiment to experiment, and from observation to observation, is admittedly one of the difficulties that has retarded knowledge of them. Their variability, though often attributed to general conditions of nutrition, or to local blood supply, etc., seems far more often due to changes produced in the central nervous organ by its own functional conductive activity apart from fatigue. This functional activity itself causes from moment to moment the temporary opening of some connections and the closure of others. The chains of neurons, the conductive lines, have been, by the methods of Golgi, Ehrlich, Apathy, Ramón y Cajal and others, richly revealed to the microscope. Anatomical tracing of these may be likened, though more difficult to accomplish, to tracing the distribution of blood vessels after Harvey's discovery had given them meaning, but before the vasomotor mechanism was discovered. The blood vessels of an organ may be turgid at one time, constricted almost to obliteration at another.

With the conductive network of the nervous system the temporal variations are even greater, for they extend to absolute withdrawal of nervous influence. Under reflex inhibition a skeletal muscle is relaxed to its post-mortem length; *i.e.*, there may then be no longer evidence of even a tonic influence on it by its motor neuron. The final common path is handed from some group of a plus class of afferent arcs to some group of a minus class, or of a rhythmic class, and then back to one of the previous groups again, and so on. The conductive web changes its functional pattern with certain limits to and fro. It changes its pattern at the entrances to common paths. The changes in its pattern occur there in virtue of interaction between rival reflexes: occlusion, substitution by equivalence, inhibition, immediate induction, successive induction and fatigue are factors. A new stimulus striking the receptive surfaces causes in the central organ a shift of functional pattern of the linkage. The central organ is a vast network whose lines of conduction follow a certain scheme or pattern, but within that pattern the details of connection are, at the entrance to each common path, mutable. The gray matter may be compared with a telephone exchange, where, from moment to moment, though the end points of the system are fixed, the connections between starting points and terminal points are changed to suit passing requirements. In order to realize the exchange at work, one must add to its purely spatial plan the temporal datum that within certain limits the connections of the lines shift to and fro from minute to minute. An example is the "reciprocal innervation" of antagonistic muscles—when one muscle of the antagonistic couple is thrown into action the other is thrown out of action.

This is only a widely spread case of the general rule that antagonistic reflexes interfere where they discharge into the same final common paths. And that general rule is part of the general principle of the mutual interaction of reflexes that impinge upon the same common path. *Unlike reflexes have successive but not simultaneous use of the common path; like reflexes mutually reinforce each other on their common path.* Expressed teleologically, *the common path, although economically subservient for many and various purposes, is adapted to serve but one purpose at a time. Hence it is a co-ordinating mechanism and prevents confusion by restricting the use of the organ, its minister, to but one action at a time.*

In the case of simple antagonistic muscles, and in the instances of simple spinal reflexes, the shifts of conductive pattern due to interaction at the mouths of common paths are of but small extent. The co-ordination covers, for instance, one limb or a pair of limbs. But the same principle extended to the reaction of the great arcs arising in the proficient receptor organs of the head, *e.g.*, the eye, which deal with wide tracts of musculature as a whole, operates with more multiplex shift and wider ambit. Releasing forces acting on the brain from moment to moment shut out from activity whole regions of the nervous system, as they conversely call vast other regions into play. *The resultant singleness of action from moment to moment is a keystone in the construction of the individual, whose unity it is the specific office of the nervous system to perfect.* The interference of unlike reflexes and the alliance of like reflexes in their action upon their common paths seem to lie at the very root of the great psychical process of "attention."

The spinal cord is not only the seat of reflexes whose "centres" lie wholly within the cord itself; it also supplies conducting paths for nervous reactions initiated by impulses derived from afferent spinal nerves, but involving mechanisms situated altogether headward of the cord in the brain. Many of these reactions affect consciousness, occasioning sensations of various kinds.

Besides the paths followed by headward-running impulses, the spinal cord contains paths for impulses passing along it backward from the brain. These paths lie in the lateral and ventrolateral columns of the cord. The fibres of which they are composed cross but little in the cord. Their sources are various; some of them come from the hindbrain and some from the midbrain, and in the higher mammalia, especially in man and in the anthropoid apes, a large tract of fibres in the lateral column (the crossed pyramidal tract) comes from the cortex of the neopallium of the forebrain. This last tract is the main medium by which impulses initiated by electrical stimulation of the motor cortex reach the motoneurons of the cord and through them influence the activity of the skeletal muscles.

Of the function of the other tracts descending from the brain into the cord little is known except that mediately or immediately they excite or inhibit the spinal motoneurons by various levels. How they harmonize one with another in their action or what their purpose in normal life may be is at present little more than conjecture. Such terms, therefore, as "paths for volition," etc., are too schematic in their bases to warrant their discussion here.

See also BRAIN; NERVE; NERVE CONDUCTION; NERVOUS SYSTEM; and Index references under "Spinal Cord" in the Index volume. (C. S. S.; C. R. NK.)

BIBLIOGRAPHY.—S. Brock (ed.), *Injuries of the Brain and Spinal Cord and Their Coverings*, 3rd ed. (1949); Charles H. Best and Norman B. Taylor, *The Physiological Basis of Medical Practice*, 6th ed. (1955); L. M. Davidoff and E. H. Feiring, *Practical Neurology* (1955); Francis M. Foster, *Medical Progress 1956* (1956); H. Houston Merritt and Daniel Sciarra in T. R. Harrison and others (ed.), *Principles of Internal Medicine*, 2nd ed. (1954); W. R. Brain, *Diseases of the Nervous System*, 4th ed. (1951); Roy R. Grinker and Paul C. Bucy, *Neurology*, 4th ed. rev. (1949); D. Munro, *Treatment of Injuries to the Nervous System* (1952).

SPINDLE TREE (*Euonymus europaeus*), so called because its tough wood formerly was used for spindles, a small tree of the family Celastraceae, native to Europe and sparingly naturalized in eastern North America. The wood furnishes good charcoal. *See* EUONYMUS; WAHOO.

SPINE, DISEASES AND DISABILITIES OF. The spine can be affected by more than 60 different conditions. The commonest are birth malformations, acquired growth deformities, sprains, fractures and dislocations from injuries; infections, arthritis, herniation of the intervertebral discs and cancer. All may produce pain, limited back movement, muscle spasm and deformity. The pain may radiate into the arms and legs when nerves are irritated. Prolonged pressure may cause paralysis of an extremity. The diagnosis of a given condition is established by suitable examination and laboratory survey, including good X-rays.

Lumbago.—A sprain (frequently called lumbago) results from overstretching of tissues, hemorrhage into a muscle, ligamentous tear or excessive movement of a joint with capsule damage. Common causes are lifting efforts, unexpected jarring or a blow to the spine. The "whiplash" injury to the neck, received in automobile accidents when the occupant's head is thrown backward and then forward, is actually a sprain, characterized by pain, muscle spasm and limited motion. Treatment consists of sedation, protective support for neck or back, and occasionally, weights attached to the head or legs to stretch the injured tissues and relieve pressure on nerves. Recovery occurs in several weeks but may be prolonged.

Arthritis.—Rheumatoid arthritis affects spines of young adults, producing progressive destructive changes in small joints leading to fusion and permanent stiffening. In other cases, the large ligaments about the vertebrae ossify and transform the spine into one piece. Pain is outstanding and is minimized by braces, X-ray therapy and cortisone. Symptoms disappear as the spine becomes rigid and the process inactive.

Degenerative arthritis is manifested by narrowing of disc spaces, sclerosis of adjacent vertebrae and formation of spurs at their margins. Other cases show changes in the facets and the joints become painful. Mild symptoms are alleviated by braces and corsets. Severe pain may necessitate bed rest, leg traction and physical therapy, followed with a back support. (*See* ARTHRITIS.)

Infection.—Infections producing pus in the spine are rare, but when present cause severe symptoms. The patient is usually acutely ill. Isolation of the offending bacteria and utilization of the specific antibiotic drug, together with good nursing care, is essential. Drainage of abscesses facilitates recovery.

Chronic infections, such as tuberculosis, may produce destruction of the vertebrae and adjacent discs. Pus and destroyed tissue accumulate in front of and behind the vertebrae. Pressure on the spinal cord may result in paralysis of the body below the level of infection. Treatment with modern drugs, supplemented by surgery in which the involved segment of the back is made solid, hastens recovery.

Disc Disorders.—The intervertebral discs serve as shock absorbers between adjacent vertebrae. The rubbery central nucleus sometimes undergoes degeneration and allows the contiguous vertebrae to rub on each other. The resultant arthritis produces local pain or radiating pain down an extremity. The nucleus, in other cases, escapes through the back wall of the disc causing spinal cord pressure. This is a protruded or herniated disc, and causes neuritic arm or leg pain, depending upon the location of the spinal lesion.

Treatment consists of traction, supports and physical therapy during the initial attack or in mild recurrent episodes. Severe, unremitting pain necessitates surgery to remove extruded nuclear material and relieve pressure on nerves. Spinal fusion is often performed simultaneously to prevent further collapse of the disc and subsequent pain.

Cancer.—Cancer arises in the spine from cells that normally produce bone and cartilage, from cells in the bone marrow and from covering elements of the vertebrae. The diagnosis is established by characteristic X-ray changes and from tissue obtained by needle biopsy or surgery. Cancer originating from the thyroid gland, gastrointestinal tract, kidneys and prostate gland has a tendency to spread to the spine. The prognosis is poor, but symptomatic treatment minimizes suffering.

Noncancerous growths of the spine, on the other hand, carry a good prognosis.

See also FRACTURES AND DISLOCATIONS. (S. W. BA.)

SPINEL, a group of minerals, the typical member of which is a magnesium aluminate, to which the name was originally restricted. The group includes chromite, franklinite and magnetite, important ores of chromium, zinc and iron, respectively. The magnesia spinel, sometimes known as ruby spinel, is one of the 16 minerals that have achieved importance as a gem, both natural and synthetic (*see* GEM: *Synthetic Gems*). The name is from the French *spinelle*, derived from the Latin *spina* ("thorn"), perhaps in allusion to the sharp, needlelike angles of the crystals. Spinels have the general composition $R''O.R_2'''O_3$, where $R'' = Mg$, Fe, Mn, Zn, and $R''' = Al$, Fe‴, Cr‴. All spinels crystallize in the cubic system, typically in octahedra. Twinning is common, the octahedral face being the twin plane, giving a characteristic form known as the spinel twin.

The group of spinel minerals includes the following members:

Magnesia spinel	$MgAl_2O_4$	Magnesioferrite	$MgFe_2O_4$
Pleonaste	$(Mg,Fe)Al_2O_4$	Magnetite	$FeFe_2O_4$
Hercynite	$FeAl_2O_4$	Gahnite	$ZnAl_2O_4$
Picotite	$(Mg,Fe)(Al,Cr)_2O_4$	Franklinite	$(Zn,Mn)Fe_2O_4$
Chromite	$FeCr_2O_4$		

Cleavage is typically absent in these minerals with the exception of the zinc spinel (cleavage on 111), but an octahedral parting is observable in magnetite and franklinite. The hardness is variable; in ordinary spinel it is 7.5–8, the gem varieties being 8, but magnetite is 6 and chromite is 5.5. The specific gravity varies with the composition, from 3.6 (magnesia spinel) to 5.2 (magnetite). The light refraction shows a range from 1.718 (pure magnesia spinel), hercynite 1.80, to picotite 2.05, chromite 2.10 and magnetite 2.42.

Spinels vary much in colour. Natural magnesia spinels are pink, red and blue. The pure magnesia spinel is colourless, hercynite is dark green in thin slices and picotite and chromite are brown in the thinnest sections. Spinels are readily produced in all colours artificially. Magnesia spinel melts at 2,135° C. and

between this compound and $FeAl_2O_4$ (hercynite) there is complete miscibility as revealed in spinel analyses, but solid solution between (Mg,Fe) Al_2O_4 (pleonaste) and magnetite is very limited. The natural spinels which are used in jewelry are found mostly in gem gravels, the chief localities being Ceylon, Thailand (Siam) and upper Burma.

Spinels occur both in igneous and metamorphic rocks. The home of the magnesia spinel is in thermally altered dolomites where it arises by reaction of alumina with dolomite. There it is usually accompanied by calcite and forsterite. The iron-rich member of the group, pleonaste, is common both in ultrabasic rocks such as dunites and in quartzless argillaceous hornfelses and gneisses. In these latter it is almost universally accompanied by cordierite and frequently by sillimanite, andalusite or corundum. Hercynite is characteristic of the granulites of Saxony in association with garnet, sillimanite, etc.

The chromiferous spinel, picotite, is relegated to the ultrabasic rocks, as dunite, lherzolite and the serpentines derived from them. Gahnite, the zinc spinel, occurs in schists associated with zinc ores and in pegmatites in Finland and at Broken Hill, New South Wales, Austr., while franklinite is associated with zinc and manganese minerals in limestone at Franklin Furnace, N.J.

See also CHROMITE; FRANKLINITE; MAGNETITE.

(C. E. T.; X.)

SPINELLO ARETINO (SPINELLO DI LUCA) (*c.* 1346–1410), prolific late Gothic Italian painter, was born at Arezzo in or soon after 1346. Trained in a local workshop, he seems rapidly to have come under the influence of Orcagna and Nardo di Cione, whose style is reflected in his first major work, a fresco cycle in S. Francesco at Arezzo. A facile artist, Spinello's reputation soon spread beyond his native town. In 1384–85 he painted a polyptych for Monteoliveto Maggiore (panels in the Pinacoteca at Siena, the Fogg museum at Cambridge, Mass., and Budapest), and about 1387 executed a fresco cycle in the small church at Antella, near Florence. A cycle of scenes from the life of St. Benedict in the sacristy of S. Miniato al Monte, Florence, dates from the same time.

To the year 1391 belongs a triptych commissioned for S. Andrea at Lucca, now in the Accademia, Florence, which was painted in conjunction with Lorenzo di Niccolò. In the same year he began work on some frescoes in the Campo Santo at Pisa. Further commissions at Arezzo followed, and between 1404 and his death on March 14, 1410, he seems to have divided his time between Arezzo and Siena, where he undertook work for the cathedral (lost) and executed (1407–10) a cycle of fresco scenes from the life of Pope Alexander III in the Sala di Balia of the Palazzo Pubblico. These and the Pisa frescoes constitute his most important works. Spinello was at best a minor artist with a vigorous sense of narrative, and his interest is due in the main to the transitional character of his late style, which anticipates the realistic painting of the early 15th century.

See G. Gombosi, *Spinello Aretino* (1926). (J. W. P.-H.)

SPINET or SPINNET, names given in England to all small keyboard instruments irrespective of shape, having one string to a note, plucked by means of a quill or plectrum of leather. The earliest name recorded for this instrument is clavicymbalum, which occurs in the rules of the Minnesingers (1404). In the 20th century the term spinet was adopted in the U.S. to describe a small upright piano. *See* HARPSICHORD; PIANOFORTE.

SPINNING, the forming of threads by drawing out and twisting various fibres.

The primitive threadmaking implement consisted of a wooden spindle, from 9 to 15 in. long, which was rounded and tapered at both extremities. Near the top there was usually a notch in which the yarn was caught while undergoing the operation of twisting, and lower down a whorl, or wharve, composed of a perforated disk of clay, stone, wood or other material was secured to give momentum and steadiness to a rotating spindle. Long fibres were commonly attached to a distaff of wood, which was held under the left arm of the operator, but short fibres were spun from carded rolls. After attaching some twisted fibres to the spindle, a rotatory motion was given to the latter either by rolling it by hand against one thigh, or by twirling it between the fingers and thumb of the right hand, after which the fibres were drawn out in a uniform strand by both hands and converted into yarn. When the thread was of sufficient strength, the spindle was suspended by it until a full stretch had been drawn and twisted, after which that portion was wound upon the body of the spindle, and the operation continued until the spindle was filled. The spindle as described here was, so far as is known, the sole apparatus with which yarn was spun until comparatively recent times.

Modern changes have had for their object: (1) the providing of mechanical means to rotate the spindle; (2) an automatic method of drawing out the fibres; and (3) devices for working a large group of spindles together rapidly.

The first improvement consisted in cutting a ring groove in the wharve, mounting the spindle horizontally in a frame and passing a band from a large wheel round the wharve. A rotatory motion was then given to the spindle by turning the wheel with the left hand. After attaching the filaments to the spindle they were attenuated with the right hand, and when fully twisted the thread was moved to form a right angle with the spindle and coiled upon it. In 1533 a citizen of Brunswick is said to have cranked the axis of the large wheel and added a treadle, by which the spinner was enabled to rotate her spindle with one foot and have both hands free to manipulate the fibres.

The flyer, which twists yarn before winding it upon a bobbin, is shown in a drawing by Leonardo da Vinci, together with a device for moving the bobbin up and down the spindle so as to effect an even distribution of the yarn. During the 16th century a machine of the foregoing type changed spinning from an intermittent to a continuous operation. The spindle had affixed upon its outer end a wooden flyer, whose forked legs were far enough apart to enclose a double-flanged spool, and at short intervals bent wires, known as the heck, were inserted in each leg for the purpose of guiding the thread evenly upon the spool. This spool was loosely threaded upon the spindle and one of its flanges was grooved to take a driving band from the large wheel, hence the spindle and the spool were separately driven, but the former at a higher speed than the latter. The twisted filaments were drawn through an eye in the flyer, led along one of its legs and made fast to the spool. By operating the treadle the flyer twisted all the fibres about a common axis once for each revolution, and the spool wound up the length thus spun: the thread being slipped from tooth to tooth of the heck at regular intervals to direct it evenly across the spool. The first stage in the evolution of mechanical spinning was effected by the invention of Lewis Paul of Birmingham. The essential features of this invention consisted in passing carded slivers between pairs of parallel rollers, each succeeding pair of which moved faster than the preceding pair, to attenuate the sliver to the required extent. Richard Arkwright, a barber of Preston and Bolton, obtained patents, in 1769 and 1775 for a machine which consisted of Paul's drawing rollers, and the spindle, flyer and spool from the Saxony wheel, but the spindles and rollers were grouped in sets of four.

In 1829 a spinner, Charles Danforth (1797–1876) of the United States, invented a dead spindle, on the top of which he placed a hollow cap to serve as the winding point, and inside the cap he rotated a spool. In 1828 John Thorp, also an American, invented the ring spinning frame, whose principal feature consisted in the substitution for the flyer of a flanged annular ring, and a light C-shaped traveler. By means of the traveler a thread was held in the best position for winding upon a spool, as well as put under the necessary tension. Between 1764 and 1767 James Hargreaves invented the spinning jenny, by the aid of which 16 or more threads could be spun simultaneously by one person. All the spindles were placed vertically and rotated from a drum, but the rovings were mounted in a movable carriage and passed between a clamp which opened and shut like a parallel ruler. After securely clamping the rovings and attaching them to the spindles, the carriage was drawn out slowly by one hand and the spindles revolved by the other. The rovings were thus stretched to the proper degree of tenuity and sufficiently twisted. This was followed by the inward run of the carriage, when the stretch of spun threads was wound upon the

spindles.

Between 1774 and 1779, Samuel Crompton of Bolton combined, in the mule, the drawing rollers of Paul with the stretching of Hargreaves. But his rollers did not fully attenuate the rovings before twisting them, as is the case with continuous spinning, neither was stretching alone relied upon.

From Crompton's time a host of inventors laboured to render all parts of the mule thoroughly automatic; for details *see* COTTON MANUFACTURE. (T. W. F.)

SPINOLA, AMBROSE, MARQUIS DE LOS BALBASES (1569–1630), Spanish general, was born in Genoa, then practically a protected state under the power of Spain. The family of Spinola was of great antiquity, wealth and power, and a rival of the house of Doria for authority within the republic. Unsuccessful in a lawsuit against the Dorias, Spinola decided to advance the fortunes of his house by serving the Spanish monarchy in Flanders. In 1602 he and his brother Frederick entered into a contract with the Spanish government. It was a speculation in which Spinola risked the whole great fortune of his house. Frederick, who was to form a squadron of galleys for service on the coast, was killed in action with the Dutch (May 24, 1603). Spinola's first landing in Flanders resulted in nothing definite and he returned at the end of 1602 for more men. On Sept. 29, 1603, he laid siege to Ostend and took it on Sept. 22, 1604, winning thereby a high reputation among the soldiers of his time. A visit to the Spanish court at Valladolid resulted in his being appointed commander in chief in Flanders. The wars of the Low Countries consisted at that time almost wholly of sieges, and Spinola made himself famous by the number of places he took in spite of the efforts of Maurice of Nassau to save them.

On the outbreak of the Thirty Years' War Spinola conducted a vigorous campaign in the lower Palatinate, and was rewarded by the grade of captain general. The most renowned victory of his career—the capture of Breda after a long siege (Aug. 28, 1624–June 5, 1625)—is the subject of the great picture by Velázquez, known as "Las Lanzas"; the portrait of Spinola is from memory. The taking of Breda was the culmination of Spinola's career. The enmity of Olivares and the want of money paralyzed his efforts. Appointed plenipotentiary and general in the war over the succession to the duchy of Mantua, Spinola landed at Genoa on Sept. 19, 1629. In Italy Olivares caused him to be deprived of his powers as plenipotentiary and he died on Sept. 25, 1630, at the siege of Casale. (D. H.)

SPINOZA, BENEDICTUS DE (1632–1677), author of one of the greatest metaphysical systems in the history of philosophy, was born of Jewish parents in Amsterdam, Holland, on Nov. 24, 1632. The family surname occurs in various forms: Espinoza, d'Espinoza, Despinoza, de Spinoza and simply Spinoza. The philosopher's Hebrew name was Baruch (meaning "blessed"), which, in accordance with the custom of the time, was rendered into its Latin equivalent, *Benedictus*.

Spinoza's grandfather and father were Portuguese crypto-Jews (that is, descendants of Jews whom the Spanish Inquisition had compelled to embrace Christianity but who remained Jews at heart) who, with many others of their persuasion, had sought refuge in Holland after the successful revolt of the Netherlands against Spain and the granting of religious freedom by the union of Utrecht. His mother, who also came from Portugal, died in 1638, when Spinoza was barely six years old.

The Spinozas were prosperous merchants and respected members of the Jewish community in Amsterdam, and it may be assumed that Spinoza attended the school for Jewish boys founded there about 1638. The curriculum of this school comprised all the Hebrew subjects: the Old Testament, the Talmud, the Hebrew codes and the works of Ibn Ezra, Maimonides, Crescas and others. Outside school hours the boys had private lessons in secular subjects. Spinoza naturally learned Portuguese and Spanish from his parents and Dutch from his environment (Spanish was also the language of the Jewish school); he was taught Latin by a German scholar, Jeremiah Felbinger, who may also have taught him German; and he also knew some French and Italian. When he learned mathematics and physics is not known, but there were plenty of Hebrew books dealing with these sciences, and the Jewish school had a good lending library.

In March 1654 Spinoza's father died. There was some litigation over the estate, as Spinoza's only surviving stepsister claimed it all. Spinoza won the lawsuit, but allowed her to retain nearly everything. Henceforth he had to fend for himself.

His studies so far had been mainly Jewish, but he was an independent thinker and had found more than enough in his Jewish studies to wean him from orthodox theology. Ibn Ezra, Maimonides and Gersonides, centuries earlier, had been, to a greater or lesser extent, critical of the orthodox interpretation of Scripture; moreover, the tendency to revolt against mere tradition and authority was very much in the air in the 17th century. But the Jewish religious leaders in Amsterdam were not inclined to tolerance: they were particularly afraid that heresies which were no less anti-Christian than anti-Jewish might give offense in a country of which the Jews were not yet regarded as citizens; and they retained, furthermore, some of the intolerant spirit of the Inquisition, whose victims they had been.

Spinoza soon incurred the disapproval of the synagogue authorities. In conversations with other students he maintained that there is nothing in the Bible to support the views that God has no body, that there really are angels (as distinguished from mere imaginative visions of them) or that the soul is immortal; and he also expressed his belief that the author of the Pentateuch was no wiser in matters pertaining to physics or even theology than they were. These views were reported to the Jewish authorities who, after trying vainly to silence him with bribes and threats, excommunicated him in July 1656. The fact of his excommunication was formally reported to the civil authorities (a gesture intended to absolve the Jewish community from all responsibility for Spinoza's heresies); and Spinoza was banished from Amsterdam for a short period. After his excommunication he spent a short time at Ouwerkerk, a village near Amsterdam, and then returned to his native city. There is no evidence that he had really wanted to break away from the Jewish community, and indeed the scanty knowledge available would suggest the opposite: on Dec. 5, 1655, he had attended the synagogue and made an offering which, in view of his poverty, must have been a rare event for him; and about the time of his excommunication he had addressed a defense of his views to the synagogue. Apparently he was not entirely indifferent to the authorities' opinion of him.

Already before his estrangement from the synagogue Spinoza had become acquainted with a number of Christians. Among these was Francis van den Enden, a former Jesuit, an ardent classical scholar and something of a poet and dramatist, who had opened a school in Amsterdam in 1652. For a time Spinoza stayed with Van den Enden, helping with the teaching of the schoolchildren and receiving help in his own further education. In this way he improved his knowledge of Latin, learned some Greek and was introduced to the neoscholastic philosophy of such writers as Franco Burgersdijck (d. 1632) and Adrianus Heereboord (d. 1659). It may have also been through Van den Enden's school that Spinoza became acquainted with the "new philosophy" of René Descartes; but in any case Spinoza's other Christian acquaintances were mostly Collegiants who were especially interested in Cartesianism.

At the same time he was learning the science and art of making lenses, in which he became a great expert; and he supported himself partly by grinding and polishing lenses for spectacles, telescopes and microscopes and partly by helping various people with their private studies. One of his friends at this period was Simon Joosten de Vries, whose admiration for Spinoza was so great that he at one time pressed him to accept a gift of 2,000 florins and at another wanted to make him his sole heir. This Spinoza refused and, when left an annuity of 500 florins in his friend's will, would only accept 300. A kind of reading and discussion circle for the study of religious and philosophical problems came into being under the guidance of Spinoza, whose ideas at this period were gradually becoming formed. In order to collect his thoughts and reduce them to a system, he withdrew in 1660 to Rijnsburg, a quiet village on the Rhine, near Leyden.

Rijnsburg was the headquarters of the Collegiants, and Spinoza

lodged there with a surgeon named Hermann Homan. In Homan's cottage Spinoza wrote the *Korte Verhandeling van God, de Mensch und deszelbs Welstand* (*i.e., Short Treatise on God, Man and His Well-being*), the *De intellectus emendatione* (*On the Improvement of the Understanding*), the greater part of his geometric version of Descartes's *Principia philosophiae* (with the appendix of metaphysical thoughts) and the first book of his *Ethica*. Part of the *Korte Verhandeling* consisted of notes that he had prepared in connection with the study circle in Amsterdam, which continued to meet after Spinoza's departure and to which he periodically sent the various parts of the *Tractatus* and other works for study and discussion. Both the *Korte Verhandeling* and the *De intellectus emendatione,* which were originally intended to form one work, were ready in or before April 1662. Already in these writings Spinoza's attitude shows his reaction to Cartesianism.

It was in 1661, during his stay at Rijnsburg, that Spinoza met Heinrich Oldenburg. This man, who was already a member of the Royal society in London and was to be appointed one of its two first secretaries in 1663, is of particular interest here because the most important part of Spinoza's correspondence was with him.

While giving instruction in the philosophy of Descartes, Spinoza made a geometrical version of the second part of Descartes's *Principia,* and this he showed to his Cartesian friends when on a visit to Amsterdam in 1663. They then persuaded him to do the same with the first part, which he did in a fortnight. These two parts, together with an appendix called *Cogitata metaphysica* (*i.e., Metaphysical Thoughts*), were accordingly published by his friends under the title *Renati des Cartes principiorum philosophiae pars I et II, more geometrico demonstratae per Benedictum de Spinoza* (Amsterdam, 1663), with an introduction by Lodewijk Meyer explaining that Spinoza did not share the views expressed in the book. This was the only book published in Spinoza's lifetime with his name on the title page. A Dutch translation by Pieter Balling was published in 1664.

By the early part of 1662 Spinoza had drafted an exposition of his own philosophy in a single work that was to combine the matter of the *Korte Verhandeling* with that of the *De intellectus emendatione;* but he was dissatisfied with the dialogue form employed in this draft and so turned to the geometrical method in the manner of Euclid's *Elements*. Thus originated his major work, the *Ethica ordine geometrico demonstrata,* the first part of which, "De Deo," was finished and in the hands of his friends early in 1663. The title of the whole work was at first intended to be "On God, the Rational Soul, and the Highest Happiness of Man"; and the work was planned in three parts instead of the five into which it eventually developed. The publication of his version of Descartes's *Principia* had been intended to prepare the way for the publication of his own philosophy, for he had both to secure the patronage of influential men as a shield against the hostility of the Calvinists and to show the more philosophically minded people that his rejection of Cartesianism was not due to ignorance.

In June 1663 Spinoza moved from Rijnsburg to Voorburg, near The Hague; and by June 1665 it appears that he was nearing the completion of the *Ethica*. During the next few years, however, he was at work on his *Tractatus theologico-politicus,* which was published anonymously at Amsterdam in 1670. This was "to show that not only is perfect liberty to philosophize compatible with devout piety and with the peace of the state, but that to take away such liberty is to destroy the public peace and even piety itself." Spinoza also told Oldenburg that he wanted to clear himself of the charge of atheism which had been brought against him. It would appear that, with the completion of the *Ethica* in sight, Spinoza was considering the question of its publication, but found it impracticable in view of the contemporary state of public opinion; he therefore set himself to show that the Bible, properly understood, gives no support to ecclesiastical intolerance or to clerical interference in civil and political affairs.

In May 1670 Spinoza moved to The Hague, where he remained until his death. He began to compose a Hebrew grammar, *Compendium grammatices linguae hebraeae,* but did not finish it; instead he returned to the *Ethica,* although the prospect of its publication became increasingly remote. His *Tractatus theologico-politicus* had aroused great interest and was to go through five editions in as many years, but there were many denunciations of it as an instrument "forged in hell by a renegade Jew and the devil, and issued with the knowledge of Mr. Jan de Witt." Jan de Witt, it should be said, was among those of the civil authorities of the United Provinces who showed sympathy for Spinoza; and his murder in 1672 provoked the philosopher to attempt a demonstration of protest that might have cost him his life had he not been prevented from making it public. Further, Spinoza's political situation in The Hague was compromised in 1673, when, lured by disingenuous invitations from the French army, he went in May to Utrecht (then under French occupation; *see* DUTCH WARS) with a view to possible negotiations for peace. He was badly received by the suspicious populace of The Hague on his return several weeks later. When the *Ethica* was finally completed in 1675, Spinoza had to abandon the idea of publishing it, although manuscript copies were circulated among his intimate friends. Spinoza then concentrated his attention on political problems and began his *Tractatus politicus,* which he did not live to finish.

In 1675 the scientist E. W. von Tschirnhaus visited Spinoza and brought about a resumption of his correspondence with Oldenburg, which had been interrupted for about ten years for reasons largely connected with Oldenburg's situation in England. In 1676 Spinoza was visited by Leibniz who, having heard of Spinoza as an authority on optics, had sent him an optical tract as early as 1671 and had subsequently received from Spinoza a copy of the *Tractatus theologico-politicus,* which deeply interested him. According to Leibniz' own account, he "conversed with him often and at great length."

Spinoza, however, was now in an advanced stage of a pulmonary disease contracted by inhaling the glass dust from the lenses. On Sunday, Feb. 20, 1677, he died, at about 3 o'clock in the afternoon. Four days later he was buried in the New church on the Spuy.

In accordance with his previous instructions, Spinoza's manuscripts were sent to the publisher Jan Rieuwertsz in Amsterdam, and his friends Jarig Jelles, Lodewijk Meyer and G. H. Schuller prepared them secretly for the press. *B.D.S. Opera posthuma* (of which there appeared also a Dutch version, *De Nagelate Schriften*), published in Amsterdam before the end of 1677, comprised the *Ethica,* the *Tractatus politicus* and the *De intellectus emendatione* as well as letters and the Hebrew grammar. A treatise on the rainbow and a treatise on the calculation of chances (*Stelkonstige reeckening van den regenboog* and *Reeckening van kanssen*) were printed together at The Hague in 1687 but thereafter long forgotten. The *Korte Verhandeling* was lost to the world until E. Boehmer's publication of it, with a Latin translation, as *Tractatus de Deo, de homine ejusque felicitate* (Halle, 1852), after which it was re-edited, together with the treatise on the rainbow, by J. van Vloten (Amsterdam, 1862).

A word must be said here of the attractiveness of Spinoza's personal character. It is discernible in all the records of his life and is recognized even by those who find themselves forced to reject his philosophical system.

PHILOSOPHY

The philosophy of Spinoza may be regarded as a development from, and reaction to, that of Descartes. Although Spinoza was also much influenced by medieval, especially Jewish medieval, philosophy, he seems to have been much more conscious of the Cartesian influence, and his most striking doctrines are most easily understood as solutions of Cartesian difficulties. As has been said above, he had studied Descartes in detail and written an account of his philosophy in geometrical form. He accepted Descartes's physics in general, although he expressed some dissatisfaction with it toward the end of his life (*see* letter lxxxiii in A. Wolf's edition of the *Correspondence*), but he found three very unsatisfactory features in Cartesian metaphysics, namely the transcendence of God, the substantial dualism of mind and body and the ascription of free will as *liberum arbitrium* both to God and to human beings. In Spinoza's eyes, these doctrines made the world unintelligible. It was impossible to explain the relation between God and the

world or between mind and body or to account for those events which are occasioned by free will.

Spinoza assumes without question that it is possible to construct a system of metaphysics, *i.e.*, an account of the world as a whole, which will render it completely intelligible. It is therefore possible in his view to present metaphysics deductively, that is, as a series of theorems derived by necessary steps from self-evident premisses expressed in terms which are either self-explanatory or defined with unquestionable correctness. His chief work, the *Ethica*, is set out in this manner (*ordine geometrico demonstrata*). The desire to secure impersonality was probably his chief motive for adopting this method of presentation. He appreciates, of course, that the use of the method guarantees true conclusions only on condition that the axioms are true and the definitions correct. Spinoza, like his contemporaries, holds that definitions are not arbitrary, but that there is a sense in which they may be correct or incorrect.

The question is discussed at length in his unfinished work, *De intellectus emendatione*. A sound definition, he holds, should make clear the possibility of the existence of the object defined, or, in the case of a necessary existent, the necessity of its existence. The *Ethica* begins with the definition of substance, the necessary existent. Since the concept of substance (*q.v.*) is among the most disputed in metaphysics, it is clear that the system could be rejected by anyone disputing the definition, however cogent the subsequent reasoning. In fact, however, as Leibniz pointed out, the demonstrations do not proceed with mathematical rigour. The system is closely knit, but not mathematically demonstrated, and the method of presentation creates some difficulties for the reader, who will at first find it easier to follow Spinoza in the prefaces to the different books and in the extended scholia where he drops the strict geometrical method. It seems in fact that, since the basic concepts, substance, attribute and mode, are in dispute and cannot, like those of geometry, be given intuitive content, metaphysics is not most suitably treated step by step, in the geometric way. The system must be accepted or rejected as a whole. If the conclusions are found to be sound, this lends support to the initial axioms and definitions.

From Spinoza's point of view a satisfactory metaphysical system would have to fulfill the following requirements: it must be intelligible in the sense explained above; it must be consistent with Cartesian or some similar physics; it must explain the main facts of human life, especially the occurrence of true and false thinking and of moral experience; and it must provide a worthy object of worship. This last point was undoubtedly of great importance for Spinoza: a deep religious feeling was one of the main motives in the construction of his system.

God or Substance.—Spinoza begins with the idea of substance as that which can be conceived as existing in complete independence (*Ethica*, i, definition 3). Since its existence is to be intelligible and it cannot be caused or explained by anything else, it must be its own cause (definition 1); *i.e.*, necessarily existent. There can be only one such substance, and this is quickly identified with God, the traditional necessary existent, and with the whole of Nature. Rejecting the idea of a transcendent God as unintelligible, but still accepting the traditional concept of a necessary being, Spinoza finds this being in the whole of the universe, which must contain all reality. He is able to accept this as a worthy object of worship and thus arrives at the position of pantheism (*q.v.*). Spinoza's account of the nature of God was the most shocking feature of his philosophy to his orthodox contemporaries and successors, both Christian and Jewish. With the transcendence of God, he denies also God's personality in the ordinary sense and God's providence. Spinoza undoubtedly thought that these doctrines were inseparably connected with a crude anthropomorphism. God is a conscious and, indeed, an omniscient being, but has no free will as ordinarily understood and no purposes: He acts according to the necessity of His own nature, which is to act freely in Spinoza's sense (*Ethica*, i, appendix). God cannot, therefore, have any special providential care for man; and the good man, although he loves God, will not expect God to love him in return (*Ethica*, v, 19).

The Attributes.—Substance has, or is constituted by, an infinite number of attributes (*Ethica*, i, definition 6); but only those two attributes of substance which enter into his own nature are known to man; viz., thought and extension. To say that these are attributes of substance is to say that the universe is a spatial whole which is conscious throughout. Spinoza accepts this conclusion and thus arrives at a panpsychist position: "All things . . .," he says ". . . are alive" (*Ethica*, ii, 13, scholium). The theory that there are infinite attributes is very difficult and seems to lead to inconsistencies in the development of the system For, since all attributes qualify substance throughout, it seems to follow that there should be some entity under each attribute which corresponds to the human mind in the same way that the body corresponds to it. We should therefore expect that all the attributes would be known to man. When pressed on this point, Spinoza replies that the modes under thought which correspond to modes under attributes other than extension are not human minds, but minds of a different kind (letter lxvi in Wolf's edition). This solution, however, is inconsistent with complete parallelism, for it gives the attribute of thought a unique position, in that it corresponds to all the attributes in a way in which they do not correspond to each other.

This difficulty has led some commentators to adopt a subjectivist interpretation of the attributes; *i.e.*, to say that the attributes are simply the infinitely many possible ways in which substance could be conceived. This interpretation receives some support from the extremely obscure definition of attribute given by Spinoza: "By attribute I understand that which the intellect perceives of substance as constituting its essence." It seems clear, however, from the letter quoted that Spinoza thinks of the attributes as objective. Probably the doctrine of infinite attributes is a relic of traditionalism in Spinoza which cannot be satisfactorily interpreted in detail.

The Modes.—Besides consisting of infinite attributes, substance is diversified into infinite modes. The interpretation of the word "mode" has been much disputed, but it is best understood as "part" as long as it is borne in mind that such a part cannot be separated from the whole of which it is a part. Human beings and other common-sense objects, which are typically modes, are parts of nature, and Spinoza once refers to the human mind as part of the infinite intellect of God. There is a one-to-one correspondence between modes under the several attributes, and necessarily so, since every attribute must qualify the whole of substance. This is the doctrine of the parallelism of the attributes. So far we have spoken of finite modes only, but Spinoza speaks also of infinite modes, which are, of course, not parts of substance but coextensive with it.

There are both immediate and mediate infinite modes. This part of the theory is difficult to understand, especially as Spinoza does not illustrate it very fully. The immediate infinite mode under the attribute extension is motion and rest. This is probably to be interpreted as the sum total of motion in the universe, so that this part of the theory is an attempt to incorporate the Cartesian doctrine of the conservation of motion. Similarly, the absolutely infinite understanding, which is the immediate infinite mode under thought, would be the sum total of consciousness or knowledge. The universe as a system of attributes is called *natura naturans*, as a system of modes *natura naturata*.

Causality and Determinism.—The whole scheme is completely deterministic: since the nature of God cannot be other than it is, the whole system of modes, which follows from or expresses the nature of God, must necessarily be as it is. This is a corollary of Spinoza's demand for intelligibility; for only on this condition can we conceive that every modal event has a cause or explanation which lies in another modal event. By "cause" Spinoza means roughly what we should mean by "explanation," and he shares with Descartes the general rationalist view that a cause must have properties in common with its effect.

It follows that a mode under one attribute cannot be caused or explained by a mode under another attribute, for they have no property in common. Any modal happening under one attribute has its explanation entirely in terms of modal happenings under

that same attribute. Since the modes are infinite in number, the complete explanation of any event in concrete terms would be infinitely complex and so cannot be attained by human beings. It is possible for us, however, to explain a particular modal event to some extent by referring to the general laws which govern the occurrence of modal events under the appropriate attribute; *e.g.*, by referring to the laws of physics in the case of material events and to those of psychology in the case of mental events.

Application of the Metaphysical Scheme to Human Beings.—The abstract scheme gains meaning as it is applied to human beings. Spinoza develops this application in the *Ethica*, parts ii–v. A man is a mode under the two attributes, extension and thought. His body is a mode under extension and is not a single thing but a complex mode, owing its unity to the maintenance of a constant pattern of relations between changing parts (*see* the axioms, etc., following *Ethica*, ii, 13). The human mind is what corresponds under the attribute thought to the human body. It is therefore likewise a system maintaining the same pattern of relation through change of parts. Spinoza, however, does not develop his theory of the unity of the mind in the same detail as his theory of bodily unity. The mind is called the "idea of the body." Spinoza sometimes speaks of two modes, the human mind and the human body, while at other times he asserts that they are one and the same thing, conceived now under thought and now under extension. These are simply alternative expressions. A human mind is that part of God's mind which is the knowledge or awareness of a given human body. The relation between a mind and its body is, therefore, intrinsic to it; and Spinoza does not require, as Descartes and Leibniz did, to invoke the miraculous intervention of God to explain the connection between a given human mind and its body.

From the general position it follows immediately that, while there is no causal connection between mind and body, there is a mental correlate for every physical event and a physical correlate for every mental event. If "mental correlate for" is here understood as "knowledge of," we reach the paradoxical result that all events in a given human body are known to the corresponding human mind and that no other events are known to that mind. At one point Spinoza appears to accept this result: "The object of the idea which constitutes the human mind is the body or a certain mode of extension actually existing and nothing else" (*Ethica*, ii, 13). The paradox is softened, however, when we realize that the word "idea" and the closely related term "knowledge" (*cognitio*) are used by Spinoza to signify not only a state of clear and accurate knowledge but any state of consciousness whatsoever, however dim and vague. All states of consciousness must have objects, but the objects need not be in the full sense known. Interpreted in this way, the paradoxical statement resolves itself into the following two propositions: (1) every change in the human body is associated with some change of awareness, however dim and vague, in the human mind (*Ethica*, ii, 12); and (2) a human being can know external physical objects only insofar as they affect his own body (*Ethica*, ii, 26).

Theory of Knowledge.—Spinoza distinguishes three grades or kinds of human knowledge (*Ethica*, ii, 40, scholium 2). He distinguishes states of awareness as adequate or inadequate rather than as true or false; and the argument of *Ethica* ii, 19–31, is directed to show that we cannot adequately know particular things, either our own bodies or their parts, or external objects. We are aware of things only in interrelation and cannot know them as they are in themselves. Thus all sense experience which is the mental correlate of the action of external things on our own bodies and all unscientific generalization from sense experience is inadequate. We cannot state the truth about a given object in itself, because it cannot be isolated from the rest of nature; and we cannot state the whole truth about it, since this would involve the whole of nature. Knowledge of this kind is called by Spinoza opinion or imagination. If, however, the consciousness is directed solely toward those properties which all objects, including the human body, have in common, there is no distortion such as occurs when objects with different properties interact. In such cases we have the second kind of knowledge or reason (*ratio*). In this way Spinoza accounts for the possibility of clear a priori knowledge in geometry, general physics and general psychology.

The third kind of knowledge, intuition (*scientia intuitiva*), seems to be the adequate knowledge of individual objects. It is difficult to see how this is possible on Spinoza's presuppositions. For the proof of the possibility he relies on the fact that it is possible to state of any individual thing that it depends on God as its complete, immanent cause (*Ethica*, ii, 47), but it is difficult to see how this general statement could be held to constitute adequate knowledge of the individual thing. It is possible that by the term "intuition" Spinoza refers to mystical experience, the insight—accompanied by strong emotion—into the dependence of all things, including the human being himself, on the whole of nature.

There are other difficulties in Spinoza's theory of knowledge, for it is hard to see how, if the theory be true, we could either be mistaken about the general properties of things (*i.e.*, have a faulty geometry or physics) or be completely right about some particular thing. The distinction between truth and falsehood about particular things occasions Spinoza great difficulty. According to his general account, the difference must be one of degree only, falsity being a matter of greater confusion or less adequacy. His account of the matter is to be found in *Ethica*, ii, 35, and there is also a long discussion of falsity in the *De intellectus emendatione*. What is clear is that our understanding of an individual thing or event will be adequate in proportion as we can explain its features in terms of the general laws, which are the concern of reason; *i.e.*, insofar as we can make correct causal statements about it. The position has important consequences for Spinoza's psychology of the emotions and for his ethics.

Psychology of the Will and the Emotions.—Spinoza followed Descartes in the attempt to construct a rational psychology. This is possible, according to his general theory, because there corresponds to every corporeal mode not only an idea but also an idea of that idea (*Ethica*, ii, 20). Thus introspective or reflective knowledge is possible. This seems to occasion another break in the strict parallelism of the attributes, giving again a certain primacy to the attribute of thought.

Spinoza rejects the concept of will as ordinarily understood. He does not, of course, deny that voluntary action or deliberately chosen action occurs, but this does not involve free will in the ordinary sense (*Ethica*, ii, 48) and is to be understood in the following way. Willing is primarily thinking of the course of action to be followed: "The will and the understanding are one and the same thing" (*Ethica*, ii, 49, corollarium). This mental state is the correlate of a physical state whose normal outcome is the realization of the action contemplated. When there are no hindering factors, the action in question inevitably follows (*Ethica*, iii, 2, scholium). The illusion of undetermined choice arises from our ignorance of the preceding causes of thought and action. Since Spinoza's ethics are, however, largely concerned with the distinction between bondage (*servitus*) and freedom (*libertas*), it is clear that he must provide some alternative account of freedom which does not involve free will as ordinarily understood. This is to be found in the distinction between the active and the passive emotions, which is closely bound up with his metaphysical theory.

The denial of human free will is another feature of Spinoza's system which has been found highly shocking, as involving the denial of moral responsibility and of the appropriateness of blame or punishment. It is true that Spinoza, anticipating later criminology, pleads for an objective and dispassionate treatment of undesirable human behaviour, but he by no means denies the sociological and political necessity of punishment, holding that most men must be governed by fear.

Every individual thing or mode obeys the fundamental law of nature that it should strive to preserve itself and continue in existence. This *conatus quo in suo esse perseverare conatur* ("endeavour to persist in its own being") is the essence of the individual thing. All finite modes, on account of their limited power, in the end perish. Only substance has the power of unlimited self-maintenance. In human beings *conatus* is on the physical side appetite and on the mental side desire (*cupiditas*), one of the primary emotions. The only other primary emotions are joy (*laetitia*) and grief (*tristitia*). Joy is the psychic accompaniment

or consciousness of the transition to a greater degree of power, while grief is the consciousness of a transition to a lesser degree of power. The other emotions, love, hate, etc., are defined in terms of these and their different possible objects. Spinoza's analytical psychology, which has often been justly admired, is partly based on Descartes's *Les Passions de l'âme*, but without the interactionist presuppositions of that work. Insofar as human beings are able to understand themselves and their environment, they can control events rather than be controlled by them. The striving for, and transition to, a greater understanding is accompanied by desire and joy in their active forms. A man is free insofar as his emotions are active rather than passive. Freedom, for Spinoza, is then a relative notion: human beings are free in proportion as they are able to think clearly, to control their environment and to preserve themselves for the maximum time in an active state. In part iv of the *Ethica* ("Of Human Bondage"), Spinoza stresses the extreme difficulty of attaining any degree of freedom on account of the strength of the passive emotions. In part v, however, he shows that liberty may be attained and the passive emotions mastered through understanding their causes. He is the first to emphasize that the complete understanding of emotions and of their causation deprives them of their overmastering quality. In this he may be said to have anticipated psychoanalysis. He also anticipates a later attitude in his insistence on the completely objective study of the emotions, their causes and effects, "as if it were a question of lines, planes and solids" (*Ethica*, iii, preface).

Moral Philosophy.—Spinoza is content with an apparently subjective definition of "good": "The good is that which we know certainly to be useful to us" (*Ethica*, iii, definition 1). He also accepts the fact that the good for different species (*e.g.*, for man and for horse) is different. For God the distinction between good and evil would be meaningless, as it is essentially relative to the purposes of finite creatures; and the human being, insofar as he understands nature scientifically, makes no moral judgments (*Ethica*, i, appendix). But human beings have enough in common for the philosopher to be able to pronounce on what is really good for them—however different their views may be because of the inadequacy and confusion of their ideas. Thus "good" may be said to have an objective significance for human beings. The characteristically human good is the increase in understanding which leads to the longest possible preservation of the individual in physical and mental health. This is summed up in the phrase "knowledge of God" (*Ethica*, iv, 28). In his actual rules Spinoza may seem to overemphasize the importance of intellectual pursuits, but he is not an ascetic; for, according to his parallelistic doctrine, everything that increases physical well-being will also promote mental health.

Since the consciousness of a transition to a state of increased power or health is joy, while sadness in all its forms is the accompaniment of diminishing power, Spinoza condemns such emotions as remorse and humility as harmful and morbid in themselves, while admitting that they may have useful effects (*Ethica*, iv, 53–54).

Spinoza's ethics are essentially egoistic, but the conduct that he commends is that which would be commended by an altruistic utilitarian. Love may be at times an active emotion, but hatred never (*Ethica*, iv, 45); and Spinoza tries to prove that hatred may be overcome by love (iv, 46). His ideal man is in many respects comparable to the Stoic sage, for his aim is the maximum self-sufficiency, but Spinoza gives a more balanced picture in that he has a stronger realization of the essential interdependence of men. Co-operation and friendship between rational men is not only a means to, but an essential part of, the individual's true good (*Ethica*, iv, 60). The preservation of civil society and of peace is a necessary means to individual well-being. This is the basis of Spinoza's political philosophy, which will be considered below.

It is first necessary, however, to touch upon a difficult feature of Spinoza's individual psychology and ethics, the doctrine of the eternity of the mind. This is contained in the latter half of book v of the *Ethica*, a section which has been especially puzzling to commentators. Any theory of survival in the ordinary sense seems plainly inconsistent with Spinoza's parallelistic doctrine. He nevertheless declares that the human mind cannot be absolutely destroyed with the body, but that something remains of it which is eternal (*Ethica*, v, 23). This doctrine has never been satisfactorily explained. Spinoza connects the eternity of the mind with the third kind of knowledge, intuition, which consists essentially in understanding individuals in relation to God, and with the characteristic accompanying emotion, the intellectual love of God, which is the highest degree of acquiescence. This should be compared with the dictum of the *De intellectus emendatione* that blessedness is the knowledge which the mind has of its union with the whole of nature. Spinoza perhaps means that since men must necessarily fall short of complete activity, their highest state is the understanding of, and acquiescence in, their own necessary degree of passivity. Insofar as this is real understanding, it is shared with God and is, in that sense, eternal. But such a doctrine carries no implication of individual survival.

Political Philosophy.—In *Ethica*, iii and iv, Spinoza stresses not only the desirability but also the extreme difficulty of co-operation between men. This is made especially clear by the fact that the good at which human beings aim, insofar as they are affected by passive desires, is necessarily competitive; *i.e.*, if men conceive that their survival and welfare are dependent on the possession of external goods, they will necessarily be in conflict with each other (*Ethica*, iv, 32–34). The true good of human nature, the knowledge of God, is, however, noncompetitive, so that men cannot come into conflict insofar as they are affected by the active emotion of desire for this good. If this emotion were predominant in men, that is, if men were mostly rational, there would be no need for political organization, since free and rational men naturally wish to co-operate. Most men, however, are under the control of their passive emotions for the greater part of their lives and will necessarily strive for their own aggrandizement to the detriment of others. Like Thomas Hobbes, Spinoza finds the only remedy for this situation in absolute government. He maintains that it is only under government that the distinction between right and wrong, in a strict sense, arises: whereas in the state of nature "right" means "power" and man has the "right" to do whatever he thinks necessary to his own preservation, in civil society the rational man must yield full obedience to the government as the necessary bulwark against anarchy, so that what is commanded by the government is right. In relation to each other, sovereign states are in a state of nature and are bound only by the law of self-preservation; *e.g.*, treaties need not be kept when they are no longer advantageous. A wise government, however, like a rational man, will see that the surest way to self-preservation is the maintenance of peace.

Similarly it will recognize that the surest way to secure the adherence of its subjects is just government with the maximum freedom possible. Spinoza, therefore, unlike most absolutists, lays great stress on the desirability of freedom of thought and speech in matters of religion. He argues that it is impossible to coerce thought and that a man will be a more loyal citizen if he is allowed to speak his mind freely. The state has a right, however, to control the outward expression of religion. What Spinoza most fears is ecclesiastical tyranny and sectarian strife, and he regards the control of religion by the state as a much lesser evil. Among possible forms of government he prefers democracy as potentially the most stable. These views are expressed not only in the unfinished *Tractatus politicus*, but also in the *Tractatus theologico-politicus*. In this work, in which he anticipates modern biblical criticism, he bases a plea for freedom of thought and speech in religious matters on a careful examination of biblical doctrine and its relation to philosophy.

Biblical Criticism.—Spinoza is far ahead of his time in advocating the application of the historical method to the interpretation of the biblical narrative. He argues that it is only in respect of their moral and practical doctrine that the prophets of the Old Testament were inspired. Their factual beliefs were those appropriate to their time and must not be regarded as philosophically significant. Complete freedom of scientific and metaphysical speculation is therefore consistent with all that is important in the Bible. The miracles of the Old Testament are explained as natural

events misinterpreted by the prophets and stressed for their moral effect.

The Influence of Spinozism.—Spinoza has an assured place in the intellectual history of the western world, both as the author of one of the most complete metaphysical systems ever conceived and as an early advocate of freedom of thought. His direct influence on technical philosophy has not, however, been great at any time. Throughout the 18th century he was almost universally decried as an atheist (sometimes, perhaps, by writers whose primary intention was to commend atheism by giving it the prestige of his name, while protecting themselves from condemnation by their formal disparagement of the very ideas that they were presenting with detailed care under cover of mere historical narration). The tone had been set by Pierre Bayle, in whose *Dictionnaire* Spinozism is described as "the most monstrous hypothesis imaginable, the most absurd and the most diametrically opposed to our minds' most evident notions"; and even David Hume felt obliged to speak of the "hideous hypothesis" of Spinoza. Throughout this period he was probably, as Voltaire says, "less read than famous," partly because an exaggerated respect for his geometrical method led to the view that his conclusions were inescapable once his premisses were accepted. He was rendered intellectually respectable by the efforts of literary critics, especially Lessing, Goethe and Coleridge, who admired the man and found an austere excitement in his works. The attraction of his work is due largely to the fact that he expresses an intensely religious attitude entirely divorced from dogma, and it has been felt less by technical philosophers than by poets, such as Goethe and Wordsworth, and by scientists of wide intellectual sympathies, as, for example, Einstein. He has nevertheless been much studied by professional philosophers since the beginning of the 19th century. Both absolute idealists and Marxists have read their own doctrines into his work, and positivists and empiricists, while rejecting his metaphysical approach, have developed fruitfully certain detailed suggestions of his theory of knowledge and psychology.

BIBLIOGRAPHY.—The best critical edition of Spinoza's works is by C. Gebhardt, 4 vol. (Heidelberg, 1924); there is also a good edition of the text by J. van Vloten and J. P. N. Land, 3rd ed., 4 vol. (The Hague, 1914). For English versions see *Spinoza's Chief Works*, ed. by R. H. Elwes, 2 vol. (London, 1883–84); *Ethics*, trans. by Hale White and A. H. Stirling, 3rd ed. (London, 1899); *A Short Treatise on God, Man and His Well-Being*, trans. by A. Wolf (London, 1910); and *The Correspondence of Spinoza*, trans. by A. Wolf (London, 1928).

For general works *see* K. Fischer, *Spinoza's Leben, Werke und Lehre*, 5th ed. (Heidelberg, 1909); F. Pollock, *Spinoza: His Life and Philosophy*, 3rd ed. (London, 1912); J. Freudenthal, *Spinoza: Leben und Lehre*, ed. by C. Gebhardt, 2 parts (Heidelberg, 1927). For biography *see* J. Freudenthal, *Die Lebensgeschichte Spinozas in Quellenschriften* (Heidelberg, 1899); E. Dunin-Borkowski, *Der Junge de Spinoza* (Münster, 1910); L. Brunschvicg, *Spinoza et ses contemporains* (Paris, 1923); and A. Wolf (ed. and trans.), *The Oldest Biography of Spinoza* (London, 1927); also A. Wolf, *Spinoza: His Life and Treatise on God and Man* (London, 1910). For studies of his philosophy *see* the bibliography given by J. A. Gunn, *Benedict Spinoza* (Melbourne, 1925); also L. Roth, *Spinoza, Descartes, Maimonides* (London, 1924), *Spinoza* (London, 1929); R. McKeon, *The Philosophy of Spinoza* (London, 1928); L. Robinson, *Kommentar zu Spinozas Ethik* (Leipzig, 1928); H. A. Wolfson, *The Philosophy of Spinoza* (Cambridge, Mass., 1934); H. H. Joachim, *Commentary on the Tractatus de Intellectus emendatione* (Oxford, 1940); S. Hampshire, *Spinoza* (London, 1956). *See* further the periodicals *Chronicon Spinozanum* (The Hague, 1921 *et seq.*) and *Bibliotheca Spinozana* (Amsterdam, 1922 *et seq.*). (M. KE.)

SPINY SQUIRREL, a group of African ground squirrels, characterized by the spiny nature of the fur. They form the genera *Xerus, Euxerus, Atlantoxerus* and *Geosciurus*, being characterized by the small size or absence of the ears and the long, nearly straight claws. Typical spiny squirrels live in clefts or holes of rocks or in burrows. (*See* RODENT.)

SPIRAEA, a much cultivated genus of shrubs of the rose family (Rosaceae, *q.v.*), comprising about 80 species, natives mostly of the northern hemisphere. They are slender shrubs, two to eight feet high, usually with simple leaves and small white, pink or reddish flowers in dense showy clusters. *S. salicifolia* (queen of the meadow) occurs in England and Scotland. *S. latifolia* (meadowsweet), *S. alba* (white meadowsweet) and *S. tomentosa* (hardhack, *q.v.*) are representative species of eastern North America. *S. douglasi* (western meadowsweet) occurs on the Pacific coast. Among the best-known cultivated forms are *S. prunifolia* (bridal wreath) and the hybrid *S. vanhouttei*. There are dozens of named forms, commonly called spiraeas. Some are forced for early spring bloom. A few so-called spiraeas belong to the genus *Astilbe* of the family Saxifragaceae.

SPIRE, in architecture, a steeply pointed termination to a tower (*q.v.*). Towers built in diminishing stages are sometimes called "spires," but in error; strictly speaking, those are steeples. Spires are in essence a kind of pointed roof, originating in the simple, four-sided pyramidal roof capping church towers in the early middle ages. In their mature Gothic development, however, only reminiscences of a roof remained, and spires became independent architectural elements, with both aesthetic and symbolic significance. (*See* also PINNACLE.)

BY COURTESY OF GERMAN TOURIST INFORMATION OFFICE

SPIRE OF ULM CATHEDRAL (COMPLETED LATE 19TH CENTURY), ULM, GERMANY. AN EXAMPLE OF GERMAN GOTHIC STYLE

From their first appearance to the 12th century, spires were generally abrupt and stunted in form; their history is a development toward slimmer, higher forms and a more organic relationship with the tower below. A late 11th-century tower at Le Puy en Velay shows the nature of early experiments in this direction; the successive stages of the tower recede and are broken up by projecting bays on the sides. Furthermore, the pyramidal roof itself is decorated with corner finials and small, round-headed dormers. In Germany the form was varied by ending each face of the square tower in a gable, the plane determined by adjacent gable slopes establishing the slope of the pyramid, as in the square tower of the abbey church of Laach (*c.* 1156). The pleasanter effect of the octagonal spires over octagonal towers, like another of the towers at Laach or the older of the two towers at Mainz (late 12th century), led to attempts to combine an octagonal spire with a square tower. This was accomplished in various ways, one of the simplest being the so-called broach spire, in which small, triangular planes of a less steep pitch than that of the spire proper occur at the bottom of the corner spire faces, with their lower points coinciding with the corners of the towers, as seen in the 12th-century church of St. Columba at Cologne.

A more developed form of broach spire lies in the substitution of a small pyramid at the bottom of the corner faces, instead of the triangular plane. This is a type which became common in English Gothic spires. It may be described geometrically as the intersection of a four-sided pyramid of low slope by an eight-sided pyramid of steep slope, as in the 14th-century spire of St. Mary's at Stamford. A more elaborate method of softening the junction between square and octagon was to carry the corner faces down unbroken to the top of the tower and fill the triangular corners thus left with decorative finials. An early example is the crossing tower of the church of St. Ours, at Loches, France (12th century).

In the southwest tower of Chartres cathedral (end of the 12th century) a further advance is made in connecting tower and spire, by adding high, gabled dormers on the faces of the spire, over the centres of the tower faces, as well as richly developing the corner pinnacles. In the spire at Senlis cathedral (13th century) a vertical, octagonal stage, or lantern, is placed between the square tower and the octagonal spire. The corners of the square tower are occupied by pinnacles and eight slim dormers surround the spire base. The edges of the Chartres and Senlis spires are decorated with projecting roll moldings, and those at Senlis are further enriched by crockets which form an admirable fretting of the silhouette. It is probable that similar crocketed spires were almost

universally intended over the great towers of most of the French cathedrals. The towers at Laon, for instance, were once so crowned and spires at Reims and Amiens were begun.

Gothic Climax.—It was in Normandy, England and Germany that Gothic spire design reached its climax. Particularly famous are the spires of St. Sauveur at Caen (14th-century church destroyed in World War II) and the magnificent group of spires at Coutances cathedral (13th century), in which the spire dormers and the corner pinnacles are treated with the utmost richness and the sense of height and slimness emphasized in every possible manner. The customary lavishness of the French Flamboyant style is magnificently illustrated in its pierced and intricately traceried spires. The two most remarkable examples are the northern spire of Chartres cathedral (1506–13), by Jehan Texier, and the so-called Tour de Beurre (*c.* 1520) of the cathedral at Rouen.

Spires of a simple type crown many of the Italian campaniles, and during the 13th, 14th and 15th centuries, spires and lanterns were frequently added to earlier towers. In neither Italy nor Spain, however, was the highly developed spire a native form, and Italian spires were simple, slate, tile or metal covered, pyramidal or conical roofs of timber. Burgos cathedral is unique in Spain in possessing two western spires of intricate openwork masonry, which date from the 15th century. These, however, are known to have been designed by a German architect.

German luxuriance, which had already appeared in the numerous varieties of timber spires used on the Rhenish Romanesque churches, found equally congenial expression in the stone spires of the Gothic period. The early, simple type is seen in the solid spires of St. Elizabeth at Marburg (not completed until the middle of the 14th century but probably from a 13th-century design). The cathedral at Freiburg has a spire (1270–88) which is of a new and infinitely richer type. The low, square tower carries a high, octagonal lantern, each side of which is capped by a slim gable and filled with delicate tracery. The corners of the square are filled with rich, triangular pinnacles, one of which is continued up as an open, spiral staircase, giving access to the upper portion. The spire rises above the lantern, a mere cage of openwork tracery, with crocketed edges, amazingly light and delicate in effect. Somewhat similar, although even more intricate in their pierced stonework are the western spires of Regensburg cathedral (late 15th century), by Roritzer. These openwork spires became the rule in Germany, as illustrated in the intricate cage of Strasbourg cathedral (spire, 1435); the west towers of Cologne cathedral, built in the 19th century from medieval drawings; the overcomplicated lacework of the cathedral at Vienna (1433) and the spire at Ulm, over 500 ft. high, built in the last years of the 19th century, from 15th-century drawings.

The spire of Oxford cathedral (1220) is a perfect example of the Early English spire type; it is octagonal, has a marked entasis or convex curve, and is decorated with dormers on each face and corner pinnacles. The great spire of Salisbury cathedral (1250), 406 ft. high, shows the tremendous aesthetic advantage of the greater steepness there present, which became the rule in the later English spires. During the Decorated period it became customary to finish the top of the tower with a parapet, either battlemented or of pierced tracery, and to set back the base of the spire behind the face of the wall. With this change the use of broaches disappeared and tall, corner pinnacles, sometimes supported on corner buttresses, filled the tower corners, as seen in the two western spires of Lichfield cathedral. In other examples a little flying buttress connects the corner pinnacles with the face of the spire, as in the exquisite Perpendicular spire of Louth church, Lincolnshire. Many Perpendicular spires are often ornamented with crockets up the angles. (*See* also Gothic Architecture.)

Renaissance.—The spire was a form never thoroughly absorbed by the Renaissance which preferred domed lanterns crowning towers. Nevertheless, during the 17th century in Germany, most effective, fantastic, spirelike forms were developed. These, based partly on Italian baroque prototypes, usually had profiles of broken concave and convex lines, crowned at the top with a sort of onion dome, the whole rising to a considerable height, and, in imaginative quality, far surpassing any of the Italian examples. It is such fantastic spires crowning the simple towers of village churches which give much of their character to the little towns of south Germany and Austria. At the same time, in England, the spire idea was receiving a much simpler and more straightforward, and equally effective, Renaissance expression, through the efforts of Sir Christopher Wren and his followers, particularly in the churches built after the Great Fire of London in 1666, such as St. Martin's, Ludgate hill, St. Mary-le-Bow and St. Bride's, Fleet street (spire and steeple only remain), all by Wren. St. Mary-le-Strand (early 18th century) and St. Martin's-in-the-Fields (1721–26), both by James Gibbs, which simplified and refined upon their Wren prototypes, are very successful.

Even more effective are many colonial spires, which, although originally based upon the work of Wren and Gibbs, achieved a fresh beauty through still further simplification. Particularly noteworthy is the type in which a small, octagonal, arcaded lantern crowns a simple, square tower and carries, usually above an attic, a simple, slim, white spire, as in the Old South meetinghouse, Boston (1729), and the extremely delicate, white, wooden steeple of the church at Farmington, Conn. (*c.* middle 18th century). Such steeples as that of St. Paul's chapel, New York (*c.* 1767), Christ church, Philadelphia (1754), and St. Michael's, Charleston, S.C. (1742, occasionally attributed to James Gibbs himself), although more monumental in treatment and more English in effect, still retain the characteristic American delicacy. This trend toward the more slender and attenuated proportions reached its climax in the exquisitely light spire of Park Street church, Boston (1819), by Peter Banner.

The 19th-century architects made extravagant use of spires for their picturesquely irregular outlines and visually exciting shapes; Anglican Gothic Revival architects of the late 1840s, '50s and '60s used them with particular effectiveness to culminate dramatically massed compositions. Perhaps because spires were so closely associated with picturesque eclecticism, 20th-century architects have tended to limit them to rather elementary geometric shapes.

(T. F. H.; An. G.)

SPIRES: *see* Speyer.

SPIRITS: *see* Alcoholic Beverages, Distilled.

SPIRITUALISM. The word spiritualism has been used with two meanings: first, for a metaphysical theory which asserts the reality of a nonmaterial spiritual world, which is a view common to all religions; second, for a system of beliefs and practices having to do with the object of communicating with the spirits of those who have died, and for the social institutions of those who hold these beliefs and carry out these practices. It is in this second sense that the word is to be understood in the present article. To avoid confusion with the more general metaphysical meaning of spiritualism, the word spiritism has been suggested as more appropriate for the beliefs and practices connected with spirit communication, but this use has not generally been adopted.

Spiritualist practices commonly take place in the presence of a man or woman (the medium) who is regarded as a human intermediary with the spirit world. Although mediums are most commonly women, some highly gifted mediums have been men, as, for example, the very remarkable medium D. D. Home (*q.v.*). The séance at which mediumistic phenomena occur may be in complete darkness, in reduced light or, more rarely, in full daylight. The medium is sometimes in a state of trance, speaking and behaving without subsequent memory of what has been said or done, but some mediums work in a state of normal consciousness.

The main activities of spiritualism are: (1) the phenomena of mental mediumship; (2) the production of materializations and other phenomena of physical mediumship; (3) spirit photography; and (4) spirit healing.

The phenomena of mental mediumship include ostensible messages from spirits of the departed. They may be messages to their own friends and relatives present at the séance, or the communicators may give information about the conditions of life beyond the grave. In some cases the spirits have dictated books of fiction or history. Usually the messages come through a spirit

control who appears to communicate directly with the sitters; more rarely the communicators appear to communicate by direct voice.

Phenomena of physical mediumship include the playing of musical instruments, transportation of physical objects (apports), levitation of the medium or surrounding objects, raps, etc. These have sometimes been considered to be too reminiscent of horseplay to be appropriate activities of discarnate spirits. They are, however, generally explained as of value only as providing evidence for the reality of spiritual activity. The most important physical phenomenon is the appearance of a filmy substance (ectoplasm) which is generally supposed to be a quasi-material substance derived from the body of the medium which may take the form of a face, hand or complete body. Such materializations have been photographed, but there have also been photographs of what purported to be ectoplasm in which the structure suggests some such material as cheesecloth. Materializations and other physical phenomena occur in darkness or in dim red light, as the exposure of ectoplasm to strong visible light is said to be injurious to the medium. Spirit photographs also have been obtained by some operators on films that may not have been exposed in a camera. It is obvious that such films could be produced fraudulently as, for example, by the use of radioactive material, but they are also stated to have been produced under satisfactory test conditions.

An important later development of spiritualism has been in the direction of spirit healing. Unorthodox healings have in the past been associated with sacred places and religious rites, and medical science is inclined to attribute all such healings to the normal process of suggestion working under favourable conditions. But it is also possible that there is a genuine power of paranormal healing found in certain persons, and from the spiritualist point of view these healers are regarded as mediums who act as agents of spirit doctors. Healings are claimed for a variety of conditions, some of which are regarded as incurable by orthodox medicine; *e.g.,* leukemia.

History.—The attempt to communicate with discarnate spirits seems to be one of the forms that religion may take in human societies and to be widely distributed in space and time. Practices very like those of a modern spiritualistic séance have been reported in various parts of the world, as, for example, Haiti and among the North American Indians, and there is no reason for supposing that these are of recent origin. The record of a materialization séance of long ago is preserved in the account in the Old Testament of Saul's visit to the witch (or medium) of Endor, in the course of which a materialization appeared which was regarded by the king as the prophet Samuel (I Sam. xxviii, 7–19). Certain mediumistic phenomena are reported in the witch trials of the middle ages, particularly the appearance of spirits in quasi-material form and the obtaining of knowledge through spirits. It may be supposed that many of those persecuted for the practice of witchcraft were what would now be called mediums, although their mediumship was coloured by the fact that it was organized in a forbidden cult and the spirits with which communication was established were regarded as devils. Some mediumistic phenomena were also found among those regarded in the middle ages as possessed by devils; *e.g.,* speaking in languages unknown to the speaker and levitation or partial levitation.

Although spiritualistic practices seem to be widespread in space and time, they were virtually unknown in modern civilized society until March 1848, when there were odd happenings at the house of a farmer named Fox at a small town in New York state. Previous occupants of the house had been disturbed by unexplained raps at night. After a severe disturbance by raps during Mr. Fox's tenancy, his youngest daughter, Kate, successfully challenged the supposed spirit to repeat the number of times she flipped her fingers. Once communication had apparently been established a code was agreed upon by which the raps given could answer questions, and the spirit identified himself as a man who had been murdered in the house.

The practice of having sittings for communication with spirits spread rapidly from that time. Kate Fox (afterward Mrs. Fox-Jencken) and one of her sisters gave much of their later lives to acting as mediums in the United States and in England. Many other mediums gave similar sittings, and the movement became widespread. The attempt to communicate with spirits by table turning became a popular pastime in Victorian drawing rooms.

Much of this activity was motivated by mere curiosity and the fascination of the supernatural, but it also had a more serious intention. Many inquirers wished to convince themselves as to human survival of bodily death, others suffered from the loss of loved relatives and friends and found consolation in the belief that they were able to communicate with them, others wanted information about the future life. To promote these serious ends, spiritualist associations or churches were formed.

The rise of this new cult was not allowed to take place without opposition. There was not only verbal condemnation with accusations of fraud but also mob violence. This was, no doubt, partly a popular reaction to a novel system of ideas and practices that were suspected of being based either on fraud or evil. The suspicion of evil was perhaps strengthened by a conjectured relationship to the discarded system of witchcraft. Although individual spiritualists were often members (or even ministers) of Christian churches, the general tendency of the established religious bodies was to suspect the movement and its claim to a new revelation that would either supplement or replace the Christian revelation. The spiritualist practices seemed also to some religious bodies to be a part of the forbidden activity of necromancy. A decree of the Holy Office of the Roman Catholic church in 1898 condemned spiritistic practices, although permitting legitimate scientific investigation of mediumistic phenomena.

If there was strong and often irrational opposition to spiritualism from the early days of the movement, there was also enthusiastic and often somewhat uncritical acceptance. For those who had lost their faith in traditional Christianity, there was offered a new religion based not on an ancient tradition but on facts that could be observed by anyone. For those to whom materialistic ways of thinking had closed the possibility of a life after death, there was offered a new hope of immortality. To those suffering from grief after the death of their loved ones, there was offered the possibility of communicating with them. There were strong emotional involvements in both the rejection and the acceptance of spiritualism that have made difficult an impartial appraisal of the evidence.

Assessment.—From the earliest days, one of the charges against spiritualist mediums was that of fraud—*i.e.,* that they produced their results by the conscious exercise of guessing and trickery—but there does not seem to be any good reason for supposing that this is true of all mediums or of any large number of them. It is, however, true that the conditions of mediumship are such that it is possible for fraudulent mediumship to be both easy and profitable. Those who attend séances for comfort and reassurance are liable to be uncritical of the evidence offered to them, and a medium, using a mixture of guesswork, inference and items of information such as are likely to be true of most people, may mislead the credulous into thinking that they have received information that could have come only from the departed. A confederate with luminous paint, masks and cheesecloth may put up a counterfeit series of physical phenomena which also may be accepted by the credulous. Judgment as to the genuineness of phenomena is made more difficult by the fact that a commercial medium may produce a mixture of genuine and counterfeit phenomena. If some mediums are able to produce ectoplasm and ectoplasmic forms, this capacity may nevertheless be too sporadic for reliable use, and the medium dependent for his livelihood on the production of spirit forms may use cheesecloth as a supplementary method of satisfying sitters. If a mental medium transmits genuine spirit messages, these too may be sporadic and supplemented by skillful guessing. This supplementation may indeed not be fraudulent since the medium may not be aware of any experienced difference between transmitting a spirit message, drawing on his own paranormal powers or using his own powers of intelligent guessing.

It is generally agreed that, when fraud is eliminated, there

remains a considerable residuum of unevaluated spiritualistic reports, whatever may be their explanation. It does not necessarily follow that the spiritualist explanation must be accepted. Some investigators (*e.g*, C. Richet) have been strongly convinced of the reality of both mental and physical paranormal phenomena but have considered that they are to be explained without reference to discarnate spirits in terms of scientific principles not yet understood. Richet admitted that what was observed suggested on the surface an explanation in terms of surviving spirits.

Two lines of evidence may be followed to test this type of explanation: first, there is evidence by recognition, when the sitter claims that he has an overwhelming impression of the personality of the communicator (his character, tricks of speech, etc.) as vivid as if he were in his living presence; second, there is evidence by information in which the communicator gives some piece of information which only should have been known to the deceased. The first of these may be convincing to the sitter himself but can be of no value in convincing anyone else. The confirmed skeptic may also be unconvinced by evidence from information, or even may deny that it is the result of paranormal cognitive powers (telepathy or clairvoyance) possessed by the medium. (*See* PARAPSYCHOLOGY.)

In principle, all evidence by information can be explained by postulating sufficient paranormal cognitive capacity in the medium, but to explain claims of evidence for survival in this way would require this kind of explanation to be stretched very far. At least some men have claimed that there is a strong prima facie case for some communication from human spirits that have survived death. If, however, there is a genuine element of communication from the departed in some mediumistic utterances, there is also much evidence that such communications can be diluted to an unknown extent with error supplied by the medium himself or derived from the sitters. To many of those who have thought that they were really in touch with spirits, it has seemed as if they were in touch with them imperfectly and indirectly, as if one were receiving a telephone message not directly but dictated through an operator who understood it imperfectly and added her own glosses as well as transmitting elements of wrong messages. It is possible to hold that mediumistic communications may be of value in providing evidence of life after death without considering that such communications provide reliable information either on the nature of that life or on any other subject.

See also PSYCHICAL RESEARCH.

BIBLIOGRAPHY.—F. W. H. Myers, *Human Personality and Its Survival of Bodily Death*, 2 vol. (1920); A. Conan Doyle, *History of Spiritualism* (1926); F. A. P. W. Schrenk-Notzing, *Phenomena of Materialisation* (1920); K. H. Porter, *Through a Glass Darkly* (1958); P. D. Payne and L. J. Bendit, *Psychic Sense* (1961); N. Blunsdon, *A Popular Dictionary of Spiritualism* (1962). (R. H. T.; X.)

SPITALFIELDS, a district of Stepney (*q.v.*), London, Eng. The land belonged to the resthouse or "spital" of St. Mary's priory (1197), Bishopsgate. Spitalfields market (6½ ac., fruit and flowers), was established in 1682. Boots and furniture are made and until about 1900 silk weaving, established by Huguenots after 1685, was important. The Eastertide Spital sermon, given at the pulpit cross in the hospital churchyard until the 17th century, now takes place in St. Lawrence Jewry.

SPITHEAD, an extensive anchorage between the Isle of Wight and the mainland of England, the Spit bank forming the western side of the channel leading to Portsmouth harbour. This anchorage was much used by the Channel fleet in the wars of the 18th century, and is specially suitable for the great naval reviews of modern times. Queen Victoria reviewed the British fleet there in 1853 just before the Crimean War and led the fleet to sea. There was another great naval gathering at Spithead for her funeral in 1901, when her body was conveyed from the Isle of Wight to London. The review by Queen Elizabeth II on her coronation (June 1953) was attended not only by ships of war of many nations, but by every branch of Britain's seafaring industry. (R. N. BA.)

SPITSBERGEN is a group of arctic islands situated between lat. 76° 26′ and 80° 50′ N. and long. 10° 30′ and 28° 12′ E., and forming, with Björnöya (Bear Island), Hopen (Hope Island), Kong Karls Land (Wiche Islands) and Kvitöya (Gillis Land), the Norwegian archipelago Svalbard (*q.v.*; *see* also NORWAY). Spitsbergen comprises the islands Vestspitsbergen (15,075 sq.mi.); Nordaustlandet (North East Land) (5,610 sq.mi.); Edgeöya (Edge Island) (1,942 sq.mi.); Barentsöya (Barents Land) (514 sq.mi.), Prins Karls Forland (Prince Charles Foreland) and a large number of smaller islands. The total area of the Spitsbergen island group is 23,641 sq.mi. There is no indigenous population. Most of the inhabitants are miners living at the Norwegian and Soviet mining towns, and there are some trappers and wireless operators. In the 1950s the population was 3,753.

PHYSICAL FEATURES

About 58% of Spitsbergen is covered by ice. The mountains along the west coast of Vestspitsbergen are rather wild with sharp ridges and peaks formed by folded and metamorphic pre-Devonian sediments. The eastern part of the island and Edgeöya and Barentsöya have plateau-formed mountains built up of nearly horizontal, younger sediments. In the northern part of Vestspitsbergen and Nordaustlandet there are more rounded mountains built up partly of granites and gneisses. The highest peaks are in Ny Friesland, where Newtontoppen reaches 5,633 ft. In the southern part of Vestspitzbergen Hornsundtind (4,695 ft.) is the highest mountain. Many large fjords penetrate the west and north coasts of Vestspitsbergen. The most important are Hornsund, Bellsund with Van Mijenfjorden and Van Keulenfjorden, Isfjorden with several branches, Kongsfjorden, Krossfjorden on the west coast and Woodfjorden and Wijdefjorden on the north coast. The west and north coasts of Nordaustlandet are also indented by fjords, but the east coast of this island is formed by the front of the inland ice. Many of the glaciers reach the sea, but in Vestspitsbergen there are large ice-free valleys, such as Sassendalen, Adventdalen, Colesdalen and Reindalen. In many parts of the coast there are extensive coastal plains, which were formed by the sea when its level was higher than it is now.

Structure and Geology.—The principal features of the geology are known. Most formations from early Palaeozoic, and perhaps Archaean, to recent, occur. The oldest rocks appear chiefly on the west and north, including Prins Karls Forland and Nordaustlandet. They are the Hecla Hook series of Pre-Cambrian, Cambrian and Ordovician dolomites, limestones, shales and quartzites which form the Caledonian folds and overthrusts of the west. The folds can be traced as far east as the west of Nordaustlandet. Granite and gneiss formerly believed to be Archaean probably belong to the same formation and were involved in the Caledonian foldings. Outside the area of these old rocks Spitsbergen is mainly a plateau of relatively undisturbed strata lying unconformably on a platform of pre-Devonian crystalline rocks. In the northwest there are Devonian rocks. These are unconformably overlain by Lower Carboniferous or Culm sandstones and shales with some coal. Middle Carboniferous is rare and the Culm is generally succeeded by Upper Carboniferous limestones, Permo-Carboniferous cherts and Permian sandstones and shales. Next come Triassic, Jurassic and Cretaceous sandstones, limestones and shales. In the Cretaceous beds there is coal. Unconformably there follow Tertiary sandstones and shales with several coal seams. Tertiary folding is obvious in the west against the Hecla Hook beds in places and in the east, but in central Spitsbergen is noticeable only in gentle undulations. Heavy faulting occurred in Tertiary times. Intrusions of dolerites and basalt were probably of Cretaceous (Neocomian) date. An extinct volcano and several hot springs (temp. 82° F.) in Bock bay are Quaternary. The strand flat is well developed up to 180 ft. and postglacial raised beaches are marked. Glacial and post-glacial debris on low ground generally mask the solid rock and loose screes form fans on the lower slopes.

Climate and Ice Conditions.—The sea around Spitsbergen is shallow and the ice readily accumulates round the shores. Pack ice prevents access to most shores except for a few months in the year. However, the warm North Atlantic drift sends a branch to the western shores of Spitsbergen, moderating its climate and leaving an open passage which permits vessels to approach the western coast during most months of the year. The fjords are frozen from

October or November to May or June.

Due to the warm drift the climate of Spitsbergen is less severe than in the corresponding latitudes of Greenland. February and March means are about 5° F., and July means are about 43° F. Even in the coldest months of the winter a thaw may set in for a few days, but snow sometimes falls in July and August. Spring comes in June; the snow becomes saturated with water and disappears in places and scurvy grass and willow open their buds. By the end of June the thermometer has ceased to sink below the freezing point at night. In September, autumn sets in. The arctic night begins in October and lasts until the end of February. The annual precipitation at Green harbour is 11.6 in.; precipitation is often less in the interior. Winds are generally light except on the west coast. There is mist on the west.

Flora.—The only trees are the polar willow, which does not exceed two inches in height, and the rare dwarf birch. But at the foot of the bird cliffs some loam has been formed notwithstanding the slowness of putrefaction, and there, in contrast with the brownish lichens that cover the rocks, grows a carpet of mosses of the brightest green, variegated with the golden-yellow flowers of the ranunculus, the large-leaved scurvy grass, the cuckoo flower, many saxifrages, foxtail grass, etc.; while on the driest spots yellow poppies, whitlow grasses and mountain avens are found. Even on the higher slopes, 2,500 ft. or more above the sea, the poppy is occasionally found. In all about 130 species of flowering plants have been collected. The distribution is suggestive: most species are found in Europe and 117 are circumpolar; many of the plants common on the west coast are absent in the east because of the greater cold. In the ice-free valleys of Vestspitsbergen there is a comparatively rich vegetation of *Salix polaris*, grasses and arctic flowers; in some of the larger valleys there are moors with cotton grass.

Fauna.—The Greenland whale has disappeared in consequence of the great havoc made by the early whalers. According to William Scoresby, no fewer than 57,590 whales were killed between 1669 and 1775. Reckless extermination of seals also took place. Walruses are now rarely seen in the waters west of Spitsbergen. Birds, also, have rapidly diminished in numbers. The fulmar meets ships approaching Spitsbergen far away from the coasts. It makes colonies on the cliffs, as also does the glaucous gull or the "burgomaster." Little auks, black guillemots, ivory gulls and kittiwake gulls breed on the cliffs, while arctic (Brünnich's) guillemots, puffins, pink-footed brant and barnacle geese, purple sandpipers, red-necked phalarope and other waders and arctic terns frequent the tundra and its pools. The commonest passerine bird is the snow-bunting. The eider duck breeds on the islands, but its numbers have become noticeably reduced. These birds, however, are only summer visitors to Spitsbergen, the snowy owl and ptarmigan being the only species which stay permanently. About 60 species have been recorded from Spitsbergen and Bear Island; about half of them breed there. Of land mammals, besides the polar bear, the reindeer and arctic fox have been greatly reduced. Norway has rigorously restricted the slaughter of reindeer; before 1868 from 1,500 to 2,000 were killed by hunters in the few summer weeks. The musk-ox has been successfully introduced near Advent bay.

HISTORY

Exploration.—It is probable that the Svalbard discovered in 1194, according to the *Icelandic Annals*, was Spitsbergen; modern knowledge of Spitsbergen, however, dates from its discovery by Willem Barents and Jacob van Heemskerck in June 1596. They first discovered Björnöya and later saw the northwest coast of Vestspitsbergen. After visits by Stephen Bennett in 1603, Henry Hudson in 1607 and Thomas Marmaduke in 1609, the Muscovy company sent an expedition in 1610, which resulted in the setting-up of the whaling industry in 1611, in which English and Dutch whalers took part. Later French, Hanseatic, Danish and Norwegian whalers also participated, and the king of Norway-Denmark laid claim to "this part of Greenland" as it was then considered. Quarrels among the whalers resulted in the division of the coast. The English had whaling rights north of lat. 79° N., whereas the Dutch erected their cookeries at Smeerenburg on Amsterdamöya, where more than 1,000 persons might be assembled in summer. The French were driven away from the coast in 1632 and had to boil their blubber on board their ships. After about 1640 the whales began to withdraw from the fjords and had to be caught off the coast. English whaling at Spitsbergen ceased in 1660, and Dutch whaling in about 1800, by which time the Greenland whale had also been exterminated. Between 1669 and 1778, 14,167 Dutch vessels visited the waters west of Spitsbergen, catching 57,590 whales.

English whalers were the first to winter in Spitsbergen, in 1630–31. Probably it was not until about 1715 that the Russians first visited Spitsbergen, wintering there to hunt walrus, seal, bear, fox and reindeer. Many of them were sent out by the Solovetskoi monastery. Russian activity ceased in about 1850, but by the end of the 18th century Norwegian sealers from Tromsö, Hammerfest and other towns had been hunting in the Spitsbergen waters, and from 1822 they also wintered there to trap fur-bearing animals. This wintering was, however, not regular until the last years of the 19th century.

Many expeditions have made Spitsbergen their base for polar exploration, as C. J. Phipps in 1773, D. Buchan and J. Franklin in 1818, D. C. Clavering and E. Sabine in 1823 and W. E. Parry, who in 1827 reached 82° 45′ N. In the same year the Norwegian geologist B. M. Keilhau visited the islands, and in 1837 Swedish exploration began with Sven Loven's expedition. This was followed by many expeditions, including that of Otto Torell in 1858, A. E. Nordenskiöld in 1864, 1868, 1872 and 1873, A. G. Nathorst, Gerard De Geer and many others. The Swedish work was the foundation for later exploration. But many other nations also participated in the exploration of Spitsbergen, among them the German expedition under K. Koldewey to the eastern part in 1868, Sir Martin Conway who crossed Vestspitsbergen in 1896–97 and the Swedish-Russian expeditions of 1898–1902. In 1896 the Norwegians built a hotel on Hotellneset at Adventfjorden and started weekly trips by tourist steamers, and the following year S. A. Andrée made his fatal balloon ascent from Dansköya. In the first half of the 20th century Spitsbergen was visited by numerous scientific and exploring expeditions, only a few of which may be mentioned. In 1906 and 1907 the prince of Monaco sent a party, led by W. S. Bruce, who worked in Prins Karls Forland, and another party, led by Gunnar Isachsen, began mapping the northwestern part of Vestspitsbergen. This led to Isachsen's expeditions of 1909–10, and later to Norwegian expeditions under the leadership of Adolf Hoel, A. Staxrud and others. After 1948 this work was continued by the Norsk Polarinstitutt, Oslo. From the early 1920s many British expeditions, associated particularly with the Oxford university exploration club, visited Spitsbergen, and some wintered there.

Spitsbergen has also been used as a base for polar flight. Roald Amundsen started from Ny-Ålesund on his unsuccessful flight to the pole in 1925, and on his flight in the dirigible "Norge" across the pole in 1926. R. S. Byrd also started from Ny-Ålesund on his transpolar flight in 1926. G. H. Wilkins landed at the entrance of Isfjorden in April 1928 after his flight from Alaska, and in May Umberto Nobile started there in the dirigible "Italia" which met disaster north of Nordaustlandet.

Political History.—Despite diverse interests in and claims to the islands by British, Dutch, Norwegians, Swedes, Danes, Russians and Americans, the question of sovereignty was long unsolved, and Spitsbergen was a *terra nullius*. Norway initiated a conference on the matter in Christiania (Oslo) in 1910, which was followed by others in 1912 and 1914, all without result. In 1919 F. H. H. Wedel Jarlsberg persuaded the Allied supreme council to grant Norway sovereignty over Spitsbergen, Bear Island and all the lands lying between 74° and 81° N. and between 10° and 35° E. This was put into effect by a treaty of Feb. 9, 1920, signed by Great Britain and the British dominions, the United States, France, Italy, Japan, the Netherlands, Denmark, Norway and Sweden; the U.S.S.R. adhered later. Norway took formal possession on Aug. 14, 1925. Norway then declared the region a part of the kingdom of Norway and appointed a governor (*sysselmann*).

The 1920 treaty forbade the erection of any naval base or fortress within Svalbard. Nevertheless, during World War II Spitsbergen was the scene of serious operations. In Aug.–Sept. 1941 an Allied force destroyed the radio stations, power stations, stocks of coal and oil, etc., and evacuated all Russians to the U.S.S.R. and Norwegians to Scotland. In the autumn the Germans erected a meteorological station in Longyearbyen. In May 1942 two vessels, "Isbjörn" and "Selis," were sent from England to prevent the mines from falling into German hands, but on reaching Grönfjorden they were destroyed by German aircraft. A small Norwegian force from England landed in July 1942 and occupied Barentsburg, Longyearbyen and Sveagruva. After a German meteorological station had been destroyed in Krossfjorden in 1943, a large German force consisting of the "Tirpitz," the "Scharnhorst" and about ten destroyers entered Isfjorden on Sept. 8, 1943, and totally demolished the mining towns of Barentsburg, Grumantbyen and Longyearbyen. In 1944 a German submarine set fire to Sveagruva as well. All towns were rebuilt after the war.

ECONOMY

Mining.—Except for coal deposits, ore and mineral deposits are

rather poor and scanty. Deposits of anhydrite with gypsum, marble, iron ore, asbestos, galena, zinc blende and pyrites have been found, but they have seemed too small or too poor to render mining profitable. Coal has been known since 1610, but it was not until the beginning of the 20th century that the coal deposits were thoroughly surveyed. From 1898 to 1920 many areas, most of them containing coal seams, were claimed by companies and persons from various countries. When Norway took over the sovereignty, a Danish commissioner was nominated to decide upon the ownership to these claims. He recognized 40 areas as private property, of which by 1956 36 belonged to Norway and 4 to the U.S.S.R. All other land is the property of the Norwegian government and is not allowed to be transferred into private possession. Coal mining on a commercial scale was started by the Arctic Coal company, Boston, Mass., in Longyeardalen south of Adventfjorden in 1904 and by an English company on the north side of this fjord. In 1916 the property was sold to a Norwegian group which established Store Norske Spitsbergen Kulkompani A/S. Other Norwegian mines have been worked at Ny-Ålesund on the south side of Kongsfjorden and at Hiorthhamn (Moskushamn). A Swedish mine was worked during 1917–25 at Sveagruva on the north side of Braganzavågen. In 1933 the property was sold to Store Norske Spitsbergen Kulkompani. A Dutch company worked the mines at Barentsburg on the east side of Grönfjorden (Green Harbour) till 1926, but sold the property to the Russian company Arctic Ougol *n* 1932. This company also took over the properties at Grumantbyen and Pyramiden. In the 1950s two Norwegian companies were working coal mines at Longyearbyen and Ny-Ålesund, and the Russian Arctic Ougol was working mines at Barentsburg, Grumantbyen and Pyramiden. With the exception of Pyramiden which is working on Carboniferous coal, all the other mines work on Tertiary coal seams.

Communications.—Besides the coal steamer serving the Norwegian and Russian coal mines a steamer makes six trips during July and August from Tromsö to the west coast of Spitsbergen. In summer Spitsbergen is also visited by tourist steamers and sealers. To aid navigation, lights and radio beacons have been erected in Bellsund-Van Mijenfjorden, in Isfjorden and in Kongsfjorden. *See* ARCTIC, THE. *See* also Index references under "Spitsbergen" in vol. 24. (A. K. O.)

SPITTA, JULIUS AUGUST PHILLIP (1841–1894), German writer on music, was born at Wechold, Hanover, on Dec. 27, 1841. He studied philology at Göttingen and held several teaching posts before going to Leipzig, where he helped to found the *Bachverein* in 1874. A notable teacher, in 1875 he became professor of musical history at the University of Berlin and at the Hochschule für Musik, of which he became director in 1882. His *J. S. Bach* is one of the great classic musical biographies.

Spitta was associate editor of the *Vierteljahrsschrift für Musikwissenschaft* from 1885 and played a leading part in publishing *Denkmäler Deutscher Tonkunst*. His other works include an important critical edition of the organ works of Dietrich Buxtehude and an edition of the complete works of Heinrich Schütz.

SPITTELER, CARL (1845–1924), Swiss poet of visionary imagination, and author of pessimistic yet heroic verse.

Spitteler was born in Liestal, April 24, 1845. He studied theology in Basel and in Heidelberg and was private tutor for eight years in Russia and Finland. In 1881–82, after his return, his first great work, the mythical epic *Prometheus und Epimetheus* (Eng. trans., 1931), appeared shortly before Nietzsche's *Zorathustra*, a work similar in outlook and in its elevated, quasi-biblical, rhythmical prose style. Alien to the time and strange, it was met with misunderstanding and Spitteler determined "to come down to earth." During the next few years, when he had to make his living by teaching and journalism, he began to express himself in a wide range of literary forms. In 1892 a legacy enabled him to settle in Lucerne and devote himself to creative work. He died there on Dec. 29, 1924.

Spitteler's second great work (which won him the Nobel prize in 1919) was the poetic epic *Der Olympische Frühling* (1900–05; rev. 1910) in which he found full scope for bold invention and vividly expressive power. The last years of his life were given up to rewriting of his first work. Tauter in composition than the early version and, like *Der Olympische Frühling*, in rhyming couplets, it appeared in 1924 under the title *Prometheus der Dulder*.

The widely varied peripheral works belong to Spitteler's middle period. He produced, in verse, *Extramundana* (1883), seven cosmic myths of his own invention; *Balladen* (1896); *Literarische Gleichnisse* (1892); and two cycles of lyrics, *Schmetterlinge* (1889) and *Gras-und Glockenlieder* (1906). He also wrote two masterly stories—*Die Mädchenfeinde* (1907; Eng. trans., *Two Little Misogynists*, 1922), a childhood idyll derived from his own experience, and *Conrad der Leutnant* (1898), a dramatically finished *Novelle* in which he approached the naturalism he otherwise hated. His novel *Imago* (1906) so sharply reflected his inner conflict between a visionary creative gift and middle-class values that it influenced the development of psychoanalysis. He published a volume of stimulating essays, *Lachende Wahrheiten* (1898; Eng. trans., *Laughing Truths*, 1927), and some biographical works of charm including *Meine frühesten Erlebnisse* (1914). In 1914 he published a politically influential tract, *Unser Schweizer Standpunkt*, directed against a one-sided pro-German view of the war. An English translation of *Selected Poems* appeared in 1928.

BIBLIOGRAPHY.—*Gesammelte Werke*, 11 vol., ed. by G. Bohnenblust, W. Altwegg and R. Faesi (1945–58); *Briefwechsel mit Adolf Frey* (1933). R. Faesi, *Spitteler's Weg und Werk* (1933) and W. Stauffacher, *C. Spittelers Lyrik* (1950) contain full bibliographical information. *See* also *C. Spitteler in der Erinnerung seiner Freunde*, ed. by L. Beriger (1947). (R. Fa.)

SPLEEN, in anatomy, is an integral part of the lymphatic and blood vascular systems. It is a specially designed lymphatic tissue interposed in the blood stream and is an important organ for (1) removing from the blood stream foreign material, including disease-producing organisms and worn-out blood cells; (2) producing lymphocytes and other blood cells; (3) producing antibodies; (4) storing iron and, in some diseases, lipids; and (5) maintaining constant circulating blood volume. In the higher vertebrates the spleen is located in the upper left quadrant of the abdomen just below the diaphragm. It varies in size in different species. Thus, in man, dogs, rats and mice it approximates 0.25% of the total body weight; in the cat and rabbit 0.05%.

Microscopic Anatomy.—The spleen is covered by a contractile capsule composed of dense connective tissue and a few smooth muscle cells. The trabeculae which are dense connective tissue and are continuous with the capsule, penetrate the organ and form a coarse scaffolding (*see* diagram) supporting a finer framework (or reticulum) of fixed reticular cells and reticular fibres. The two types of lymphatic tissue, white pulp and red pulp, differ chiefly in the relative number of cell types occurring within the reticulum. White pulp or dense lymphatic tissue forms a sheath of varying thickness around the arteries and is reticulum closely packed with lymphocytes. In the red pulp or loose lymphatic tissue, the reticulum is pierced in all directions by venous sinuses (*see* diagram). The reticulum forms the splenic or Billroth cords between the sinuses and is filled with lymphocytes, free macro-

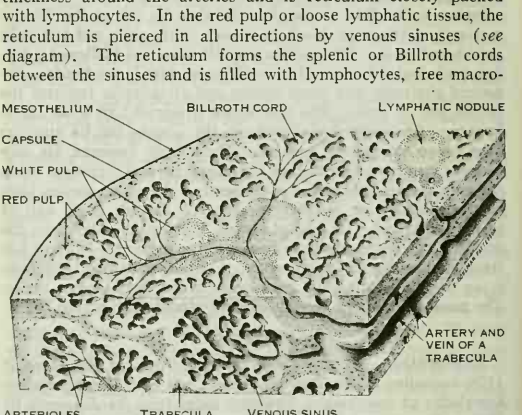


DIAGRAM OF A SPLENIC LOBULE LOCATED JUST UNDER THE CAPSULE OF THE SPLEEN. NOTE THE RELATIONSHIPS OF THE ARTERY AND VEIN TO THE WHITE AND RED PULP

phages (many of which were previously fixed reticular cells) and all circulating blood cells. The venous sinuses contain the same cells but fewer large lymphoid cells and macrophages. They are lined with reticular cells (littoral cells) and not with true endothelial cells, which line ordinary blood vessels. Arterial blood leaving the smallest branches (arterioles) of the trabecular arteries passes through the cords of the red pulp on its way to the sinuses and out through the trabecular veins. Whether the blood has permanent channels throughout the cords (closed circulation) or

simply percolates through temporary passages (open circulation) in the reticulum is not definitely known. Both types of circulation are shown in the figure. In any case, as the blood comes in contact with phagocytes, the phagocytes remove various materials from the blood. The most active phagocytes are the fixed reticular cells and free macrophages of the cords; active to a lesser extent are the reticular cells lining the sinuses and free macrophages in the sinuses.

Functions.—The macrophages of the spleen in the normal animal remove worn-out or fragmented blood cells and digest them. During this process, the iron-containing component of the hemoglobin in the red cells is separated and reutilized by the body. This phagocytic activity may be so hyperactive in some blood diseases and infections as to produce an anemia. The macrophages also remove foreign matter from the blood, including invading parasites. When the foreign substance is an invading organism or a vaccine containing antigens, the spleen forms antibodies against it. Such antibodies make the bacteria or other particulate antigenic material more readily ingested by the macrophages (*see* discussion of the mechanism of immunity to malaria in PARASITOLOGY).

Virulent infections at first generally kill many splenic cells, such as macrophages and especially lymphocytes. Death of lymphocytes results in a shrinkage of the white pulp. If the individual animal survives, lymphocytes and to a lesser extent reticular cells proliferate by mitotic division. Some of the new lymphocytes replace those previously lost, and others develop into monocytes and macrophages and probably into antibody-producing cells. Some authors believe that most of the new macrophages are thus formed from lymphocytes and monocytes and that only a small part arise by the proliferation of reticular cells. The proliferative and developmental processes frequently result in the formation of more splenic tissue, an above-normal number of cells (hyperplasia) and an enlargement of the spleen. This splenomegaly is at times very marked. The extra macrophages in such a spleen are of great functional significance especially during early stages of infection before antibodies appear, which, as mentioned above, increase the effectiveness of individual phagocytes.

Most of the functions of the spleen result from the fact that its macrophages are so oriented that they remove foreign materials from the blood. Other so-called filter organs of the blood are the liver and bone marrow and, to a lesser extent, the adrenal and pituitary glands. The spleen contains the most active macrophages, but the liver, because of its size, can remove as much as 70 times more material. The spleen produces antibody first in response to antigens in the blood. Its output then falls off, and various nonsplenic sources take over and in time surpass it. The spleen and bone marrow among the filter organs are the most important antibody-formers. The liver forms very little antibody although it is the source of most of the other serum proteins. The spleen, in its blood-cell-forming capacity, produces chiefly lymphocytes, monocytes and macrophages. The same is true of the lymph nodes and other lymphatic tissue and the bone marrow.

The spleen contains a sizable fraction of the total blood and acts as one of the reservoirs to maintain a constant volume in the general circulation under various conditions.

Comparative Physiology.—Knowledge of the physiology of the spleen of the lower vertebrates is largely limited to its role in blood-cell formation. H. E. Jordan believes that the spleen is the fundamental blood-cell-forming organ and that it appeared during the evolution of the vertebrates at about the time of hemoglobin-containing red cells. In the hagfish, among the primitive fish, it is a diffuse tissue; in the lamprey it is a more localized mass of tissue projecting into the dorsal wall of the intestine. In the lungfish it is a discrete organ but is embedded in the stomach. In all the rest of the true fishes and higher vertebrates it is a separate organ located in the dorsal mesentery. In its most primitive form, it is the chief site of the formation of red cells, granular leucocytes and lymphoid cells, but among most higher fishes it has largely lost its function to produce granular leucocytes and shares the other functions with other tissues. The function of red-blood-cell formation is largely taken over by the red bone marrow in some reptiles and in all higher vertebrates. Although the spleen retains its function of forming lymphocytes, it shares this function with the bone marrow in the reptiles and birds and with bone marrow and lymph nodes in the mammals. It exhibits many of the foregoing capacities, however, in embryonic development or in disease. For example, it forms red cells during a considerable part of human embryonic development. The normal spleen in adult man never forms red cells, but the red pulp in a disease such as myeloid leukemia produces a large number of them as well as other bone marrow elements. The normal spleen in some animals (*e.g.*, chickens, rats and mice) probably always forms a few red cells and may produce enormous numbers of them in certain infections such as malaria. *See* also BLOOD; LYMPH AND LYMPHATIC SYSTEM.

BIBLIOGRAPHY.—H. E. Jordan, "Comparative Hematology," in Hal Downey (ed.), *Handbook of Hematology*, vol. ii (1938); Alexander A. Maximov and William Bloom, *A Textbook of Histology*, 7th ed. (1957); William H. Taliaferro, "Functions of the Spleen in Immunity," *Am. J. Trop. Med.*, 3:391–410 (1956). (W. H. T.)

SPLEEN, DISEASES OF. The diseases of the spleen (*q.v.*), which are numerous, depend in large measure upon the peculiarities of that organ. The spleen has no outstanding function, yet it has many rather subtle functions. It may be removed with impunity, yet it contributes to the development of immune responses. It is ordinarily useful, yet it may become harmful to the body, particularly when enlarged. In some ways the spleen is a sponge or reservoir containing excess blood; in other respects it is like a large lymph gland situated within the abdominal (portal) circulation; it has many phagocytic cells designed to capture particles and debris.

The spleen may become enlarged (splenomegaly) for a variety of reasons: reaction to certain infections or to immune processes; congestion with excess blood as in polycythemia or in cirrhosis of the liver; stuffing with abnormal cells, as in Gaucher's disease, leukemia or a related disease called myeloid metaplasia. Certain tumours begin in the spleen (lymphosarcoma); peculiarly, ordinary cancers are usually unable to gain a foothold in that organ.

The spleen, being part of the lymphoid apparatus, becomes enlarged in diseases involving this system, as in lymphocytic leukemia and infectious mononucleosis. As part of the reticuloendothelial system, it becomes enlarged when this phagocytic system is overactive as in tuberculosis, rheumatoid arthritis and histoplasmosis. The spleen is easily distensible, so that when the portal circulation is filled with stagnant blood under great pressure due to inability of blood to go through the liver, it takes up extra blood and becomes congested.

In many cases, enlargement of the spleen is accompanied by a reduction in the cells of the blood; thus either the red cells may be affected (anemia), or the white cells (leukopenia) or the platelets (thrombocytopenia). Involvement of all three blood elements is called pancytopenia, and this occurs fairly commonly in hypersplenism, the term given to reductions in blood elements associated with enlargement of the spleen. Sometimes the hypersplenic state is so striking and the blood counts so low that it is desirable to remove the spleen. The spleen also may be removed in certain cases of blood destruction (hemolytic anemia) and blood platelet reduction (thrombocytopenic purpura); however, in these cases all other avenues of treatment, notably the use of cortisone preparations, should be explored first. In one condition, hereditary spherocytosis or congenital hemolytic jaundice, removal of the spleen results in complete cure of the anemia. Here the spleen itself is normal, but the bone marrow produces spherocytes which are actually trapped by the spleen. Removing the spleen removes the trap and cures the condition. In certain tumours and cysts of the spleen it is best to remove the organ, especially since modern surgery allows removal without much reaction. Certain cases of leukemia with large spleens are treated with various chemicals, others by high-voltage X-rays.

(W. DK.)

SPLIT (SPLJET), a city of Croatia, Yugos. (Ital. Spalato), has the finest harbour on the Adriatic coast, with a broad bay affording deep, safe anchorage. This, combined with its central position and good communication with other parts by road, rail and steamer,

made it of great commercial importance, with an extensive trade in cement and marl. Pop. (1961) 99,462, almost entirely Serbo-Croatian. Spalato arose from the palace of Diocletian, who renounced the imperial crown in A.D. 305 and lived there until his death. The palace was then allowed to fall into ruin, but when the incursion of the Avars was over (639) the inhabitants of the ruined city of Salona nearby took refuge there and built up homes within the 9½ ac. covered by the palace, incorporating its walls and pillars. Salona was made a Roman colony in 78 B.C. and was one of the chief ports of the Adriatic. The palace was transformed into an imperial cloth factory. The town was captured several times by the Goths and Huns and in 639 was destroyed by the Avars, but was not entirely deserted until the end of the 12th century. In 659 John of Ravenna, the papal legate, was created bishop of Spalato, as the new city was called. A little later it became an archbishopric and its holders were metropolitans of all Dalmatia until 1033. In 1105 Spalato became a vassal state of Hungary; in 1327 it revolted to Venice, but in 1367 returned to Hungary. It was ruled by the Bosnian king, Tvrtko, from 1390 to 1391 and in 1402 by the famous Bosnian prince, Hrvoje. In 1420 Spalato fell to Venice and ceased to have an independent history. The castle and walls were erected by the Venetians between 1645 and 1670 and dismantled during the French occupation of Dalmatia (1805–13); in 1813 the province came under Austrian rule and remained there until 1918. In that year Italy occupied the town for a short time, but from 1919 to 1941 the town was Yugoslav. It was again occupied by Italy during World War II.

SPLÜGEN PASS, one of the passes across the main chain of the Alps from Switzerland to Italy. The route from Thusis passes first through the celebrated gorge of the Via Mala, then through the Schams basin and past Andeer, beyond which the Rofna gorge gives access to the village of Splügen in the upper reach of the main or Hinter branch of the Rhine (*q.v.*). Leaving to the west the road over the San Bernardino pass, 6,775 ft., the Splügen road (constructed in 1823) mounts south to the pass (6,932 ft.) which forms the political frontier. On the other side the road avoids the old path through the dreaded Cardinello gorge (here passed Alexandre Macdonald's army in Dec. 1800) in order to descend by zigzags to Pianazzo. Thence past Campo Dolcino and Gallivaggio the descent is made to the ancient town of Chiavenna at the junction of the road from the upper Engadine over the Maloja pass, and 17 mi. by rail above Colico, at the northern end of Lake Como.

SPODUMENE, a variety of pyroxene (*q.v.*) consisting of lithium aluminum silicate, is an important ore of lithium and source of ceramic materials. It was named in 1800 from the Greek word meaning "ash coloured," and soon afterward termed by the crystallographer R. J. Haüy, triphane ("appearing threefold") because it exhibited certain characteristics equally in three directions. The hardness is 6.5 to 7 and though generally a dull mineral, some varieties are so highly coloured and transparent as to be valued as gem stones, *e.g.,* the emerald-green hiddenite and the lilac-coloured kunzite, while a yellow or yellowish-green spodumene found as pebbles in Minas Gerais, Braz., resembles, when cut, some kinds of chrysoberyl. The largest natural crystals are found in pegmatite dikes in the United States in Massachusetts, Connecticut and the Black hills of South Dakota. Canada is an important primary producer of spodumene ore and concentrates. Chloride volatilization, alkaline leaching and acid leaching are used to extract lithium from the ore prior to the manufacture of purified lithium chemicals. The extracted lithium is important in commerce in air conditioning, metal alloys, lubricating greases and lubricants, ceramics and welding and brazing. Military applications include H-bombs, nuclear power and high-energy fuels.

Spodumene, $LiAl(SiO_3)_2$ or $Li_2O.Al_2O_3.4SiO_2$, exists in nature as a dense (specific gravity 3.2) α form and changes on heating at 1,000° C. to a much less dense (specific gravity 2.4) β form with a different crystal structure than the natural mineral. The large specific gravity change is used in the benefication procedure known as the "decrepitation" process since the rocklike mineral shatters to small particles when it is heated rapidly to temperatures above 1,000° C., facilitating the separation of accessory mineral impurities. This high temperature form has an extremely low thermal expansion, making it a very important thermal shock-resisting ceramic material. In minor amounts, spodumene is used in glass, enamels and glazes. (F. A. Hl.)

SPOHR, LUDWIG (1784–1859), German composer and violinist, was born at Brunswick on April 24, 1784. He spent his childhood at Seesen, where in 1789 he began to study the violin, and at six years of age was able to take part in chamber music. He learned a great deal from studying the scores of Mozart, and later became a student of Franz Eck.

His first violin concerto was printed in 1803, and in 1805 he became violinist to the duke of Gotha. Spohr failed to have his first two operas (*Die Prüfung* and *Alruna*) produced, but his third, *Der Zweikampf mit der Geliebten,* was successfully performed in Hamburg in 1810, a year after it was written. In 1811 he produced his (first) *Symphony in E flat* and in 1812 composed his first oratorio, *Das jüngste Gericht.* In that year he also became leader of the orchestra at the *Theatre an der Wien,* Vienna, and began his dramatic masterpiece, *Faust,* which he completed in 1813. He resigned his post at Vienna in 1815 and toured Italy, returning to Germany to become conductor of the opera at Frankfurt (1817) and later *Kapellmeister* at Kassel (1822).

In 1822 Mendelssohn, then a boy of 13, visited Kassel; a friendship sprang up between the two, which ceased only with Mendelssohn's death in 1847. Spohr's next three operas, *Der Berggeist* (1825), *Pietro von Abano* (1827) and *Der Alchymist* (1830), attained only temporary success. At the Rhenish musical festival held at Düsseldorf in 1826 his oratorio *Die letzten Dinge* (*The Last Judgment*), the most famous of his sacred compositions, was produced. In 1831 Spohr published his *Violin School.* The year 1834 was saddened by the death of his first wife, Dorette Scheidler, a celebrated harpist. Two years later he married a pianist, Marianne Pfeiffer. During 1833 he worked on an oratorio—*Des Heilands letzte Stunden* (*Calvary* or *The Crucifixion*), which was performed at Kassel on Good Friday 1835 and sung in English at the Norwich festival of 1839 with an effect which he later referred to as the greatest triumph of his life. For the Norwich festival of 1842 he composed *The Fall of Babylon,* which was also a success, though the elector of Hesse-Kassel refused Spohr leave of absence to conduct it. His last opera, *Die Kreuzfahrer,* was produced at Kassel in 1845. Of his nine symphonies the finest, *Die Weihe der Töne,* was produced in 1832.

Spohr's compositions for the violin include concertos, quartets, duets and other concerted pieces and solos, and among these a high place is taken by four double quartets; *i.e.,* octets for two antiphonal string-quartet groups, an art form of his own invention. He died in Kassel on Oct. 16, 1859.

SPOIL FIVE, once known as the national card game of Ireland, is akin to the British napoleon, the American euchre and the French écarté (*qq.v.*). Spoil five and its variant forty-five were carried to the United States and Canada, while the variant auction forty-fives, introduced into Newfoundland in 1912, became the principal club game there.

Spoil five is much like loo (*q.v.*). A pool is formed by equal antes. Each player receives five cards and the next card is turned up for trump. Anyone who wins three tricks must take the pool or jink it (play for all five tricks for the pool plus an extra chip from each player). The hand is spoiled if no one wins the pool, which then is added to the next pool.

The ace of hearts is always the third-best trump, whatever the trump suit. Above it are the five and jack (knave) of trumps; below it, the ace, king, queen. In the red suits the lower cards rank normally: 10 (high), 9, 8 . . . and so on, but in the black suits the rank is reversed: 2 (high), 3, 4 . . . 10 (low). When not trump the spade and club aces rank below the jack, and the diamond ace ranks below the 2.

Forty-Five.—This is a variant in which there is no pool; four or six play, in partnerships of two or three to a side. A side winning three tricks scores 5 points; all five tricks, 10 points. As some play, winning four tricks scores 15 points and all five tricks 25 points. The first side to reach 45 points wins.

In auction forty-fives trump and contract are determined by

bidding. Each trick counts 5 and the highest trump an additional 5; game is 120. (G. MH.)

SPOILS SYSTEM. The practice in the United States under which the political party winning an election rewards its campaign workers and other active supporters by appointment to public office. It involves political activity by public employees in support of their party and the consequence of removal from office if their party loses the election. A change in party control of government necessarily brings new officials to those high positions carrying political responsibility, but the spoils system extends personnel turnover to varying proportions of routine or subordinate governmental positions. In the most extreme form of the spoils system the primary test for appointment is excellence of party service, with almost complete disregard of capacity to perform the duties of the public office concerned.

The term "spoils system" is often applied to the great variety of means by which money may be transferred more or less indirectly from the public treasury to the party coffers, as well as to other abuses of governmental power for partisan advantage. So used, the term includes abuses in the management of public expenditures, irregularities in the granting of franchises and like privileges, partisan favouritism in the administration of criminal law, discrimination in the levy of taxes and other corrupt uses of public power. The term will be limited here to its earlier and narrower usage.

Although patronage existed in colonial times, the spoils system came to full bloom in the period 1800–29. When Jefferson became president in 1801, he found the public offices filled with persons generally hostile to him and to his party. Without espousing the principle of rotation in office, he replaced enough Federalists with Republicans to assure a more even distribution between the two parties of posts in the civil service. Yet in both appointments and removals partisan considerations were not the sole criterion; it remained for President Jackson to adopt the principle of the spoils system in the national government. Spoils practices were instituted earlier in the governments of states and cities. The system first became firmly rooted in New York; it was established almost as soon in Pennsylvania. By 1828 in all the states of the north and west the system existed or strong groups desired to introduce it.

In his annual message in 1829 Jackson presented a reasoned philosophy of the spoils system. "There are," he said, "perhaps few men who can for any great length of time enjoy office without being more or less under the influence of feelings unfavourable to the faithful discharge of their public duties. Their integrity may be proof against improper considerations immediately addressed to themselves, but they are apt to acquire a habit of looking with indifference upon the public interests and of tolerating conduct from which an unpracticed man would revolt." Rotation in office, then, was not an evil but a positive good. "The duties of all public offices are," he declared, "or at least admit of being made, so plain and simple that men of intelligence may readily qualify themselves for their performance; and I cannot but believe that more is lost by the long continuance of men in office than is generally to be gained by their experience."

The common explanation of the rise of the spoils system is that ways had to be found to support the activities of political parties. With the broadening of the suffrage, the use of elections to choose large numbers of officials and other tendencies toward democratization, political parties had to assume heavy burdens in the conduct of campaigns and leadership of debate on public issues. No class existed with the time or inclination to assume these responsibilities and parties had to be maintained by patronage. Practical considerations of party finance were re-enforced by a democratic dislike of the idea of public office as private property and a hostility toward the rise of an officeholding class.

The spoils system is often said to be peculiar to the United States, but similar practices have existed elsewhere from time to time. In fact, the distribution of public offices to strengthen a regime is an ancient practice of statecraft. At the time of the founding of the United States government, the tests for appointment to the civil service were probity, ability and loyalty to the constitution. The test of loyalty aided in buttressing a new government emerging from controversy. The installation of fascist dictatorships in Germany and Italy was accompanied by considerable partisan dilution of the permanent civil service. If the nature of a government changes radically, a degree of turnover in the civil service is likely to occur. The introduction of the spoils system into the national government of the United States coincided with a veritable revolution as new classes came to power in national affairs. Even systems for the selection of public employees on the basis of merit may reserve by effect the higher positions for members of the dominant groups within the community.

The spoils system flourished unchallenged in the United States until after the Civil War. Bills to reform the civil service were introduced in congress in the 1860s but they attracted little attention and less support. In 1871 congress empowered the president to prescribe regulations to govern the selection of civil servants and to employ persons to aid him in determining the fitness of applicants. President Grant appointed a civil service commission, but it was dissolved in 1875 after congress had failed in 1873 and 1874 to provide funds for its support.

The National Civil Service Reform league, established in 1881, led a reform movement which succeeded gradually in restricting the spoils system by the introduction of the merit system for the selection of civil servants. The movement gained impetus from the assassination of President Garfield by a disappointed office seeker and in 1883, the Pendleton act, largely drafted by leaders of the Reform league, was adopted by congress. It provided for the creation of the United States Civil Service commission and for the merit system in the recruitment of some federal civil servants. In the same year New York adopted a civil service law requiring the selection of civil employees through competitive examination. After 1939, all states were compelled to institute merit systems for those employees engaged in security programs aided by federal subsidies.

In the second half of the 20th century about 75% of the U.S. cities had established some type of formal merit system. However, only 25% of the cities covered all employees, and almost half of those reporting cities with a population of less than 25,000 had established no formal civil service system. As the merit system encroached on the spoils system in state and local governments, its application was broadened in the national government by actions of various presidents. A setback to the merit principle occurred after World Wars I and II in the extension to veterans of preference in appointments.

The coverage of civil service legislation, however, does not indicate accurately the extent to which the spoils system has been superseded. In jurisdictions with civil service laws, spoils practices may prevail. In those lacking formal merit systems, ability may govern selection of employees and their tenure may be permanent. *See* CIVIL SERVICE. (V. O. K.; X.)

SPOKANE, a city of eastern Washington, U.S., the county seat of Spokane county, is situated on the high river plain of the Spokane river about 50 mi. from its juncture with the Columbia, at an altitude of 1,898 ft. To the north and east lie the Selkirk and Rocky mountains, generally shielding the valley from the rigours of northeast winds which on occasion cause spells of extreme cold. The population of the city in 1960 was 181,608 and of its standard metropolitan statistical area (Spokane county), 278,333. For comparative population figures *see* table in WASHINGTON: *Population.*

Only five years after the notable Lewis and Clark expedition of 1804–06, the first white habitation of the Pacific northwest, Spokane House, a North-West Fur company post, was built on the Little Spokane river. Ft. Spokane, an Astor establishment, was built across the river from its competitor. When the Hudson's Bay company, successor to the North-West company, moved its operation to Ft. Colville in 1826, that settlement eclipsed the Spokane country. During the missionary period, the Indian disturbances and the early mining boom, Walla Walla held the stage. It was not until 1872 that a settlement was established at Spokane Falls. A village by that name was incorporated in 1881 and the

following year it was chartered and also replaced Cheney as the county seat.

In 1880 the town's population was only 350. With the advent of the Northern Pacific railroad in 1883, however, it mushroomed, and by 1890 it had outstripped Walla Walla and enjoyed third rank in the state. With a revival of mining in the 1880s, this time of the more stable lode-mining type, Spokane became the region's financial and mercantile focal point. In 1886 it had an electric-light plant and a horse streetcar line. An electric trolley system in operation by 1891 served the city until 1935. Additional rail facilities, particularly the Union Pacific (1889) and the Great Northern (1892), and a number of branch feeder lines contributed to Spokane's ultimately becoming the largest railroad centre west of Omaha. From the holocaust of Aug. 4–6, 1889, when 30 blocks in the middle of the city were destroyed by fire, there emerged a more modern city.

The word "Falls" was dropped from the name in 1890. The word Spokane, said to mean "children of the sun," was the name of an Indian tribe which had formerly occupied the valley. It is pronounced to rhyme with "man."

There are three private institutions of collegiate rank: Gonzaga, a coeducational Roman Catholic university founded in 1887; Whitworth, a Presbyterian coeducational college established in 1890; and Holy Names, a Roman Catholic college for women, incorporated in 1907.

The Fox theatre, the Spokane coliseum and the municipal stadium have been the settings for a great variety of entertainment. The city has a symphony and a junior symphony orchestra.

Diversity is the keynote of the city's economy, to which agriculture, mining, lumbering, processing of primary metals, general manufacturing and distribution contribute. An abundance of cheap hydroelectric power has been a major stimulus and was particularly instrumental in the establishment during World War II of the aluminum reduction mill at Mead and the aluminum rolling mill at Trentwood. In 1954 two large oil pipelines reached the city and natural gas was made available in 1956.

The city owns Geiger field serving commercial airlines and Felts field, which serves principally private planes. (H. J. De.)

SPOLETO (anc. Spoletium), a town and archiepiscopal see of the province of Perugia, Italy, 18 mi. N.N.E. of Terni, and 88 mi. N. by E. of Rome. Pop. (1957 est.) 39,728 (commune). It is situated on a hill at the south end of the open valley of the Topino, a tributary of the Tiber. The principal industries are the collection and preparation of truffles and preserved foods, also tanning and the manufacture of earthenware, cotton and wool spinning. Spoleto is also the centre of an agricultural district. There are few towns of Italy which possess so many Roman remains in good preservation under the medieval buildings, and few medieval towns with so picturesque an appearance. There are considerable remains of pre-Roman polygonal walls. There are remains of a Roman theatre, over 370 ft. in diameter, and an amphitheatre 390 by 295 ft. A Roman bridge of three arches, 80 ft. long and 26 ft. high, exists at the lower (north) entrance to the town.

The rock above the town was included within the polygonal walls; but Totila fortified, not this rock, but the amphitheatre, which remained the citadel until 1364, when Cardinal Albornoz destroyed it and erected the present Rocca, which was enlarged by Pope Nicholas V; it is now a prison. The Porta Fuga (the name alludes to the repulse of Hannibal) occupies the site of a Roman gate, but is itself medieval: while the medieval *enceinte* encloses a somewhat wider area than the ancient. The Piazza del Mercato represents the Roman forum; close by is a triumphal arch of Drusus and Germanicus (A.D. 21), and a temple over which is built the church of S. Ansano. The cathedral of S. Maria Assunta occupies the site of a church erected about 602. The present church was consecrated in 1198; the façade with eight rose windows belongs to the middle of the 12th century. Over the main entrance is a large mosaic of Christ enthroned, with the Virgin and St. John, by the artist Solsernus (1207). In the choir and on the half dome of the apse are the finest frescoes of Fra Filippo Lippi (scenes from the life of the Virgin) completed after his death by Fra Diamante.

The church of S. Pietro, outside the town, has a façade richly sculptured with grotesque figures and beasts, of two different dates, about 1000 and about 1200. The little church of SS. Giovanni e Paolo (12th century) contains frescoes of the 12th–14th centuries, notably an early representation of the martyrdom of S. Thomas à Becket and a portrait of S. Francis of Assisi. S. Domenico is a fine example of later Italian Gothic with bands of different coloured stone. Both the church and its crypt contain 14th-century frescoes. The triple-apsed crypt of S. Gregorio probably dates from the 9th century; the upper church was consecrated in 1146 and the Romanesque work covered with stucco in the restoration of 1744. S. Nicolò is a beautiful example of Pointed Gothic. The basilica of S. Salvatore (*il Crocefisso*) at the cemetery belongs to the 4th century A.D. The fine sculptures of the façade, with its beautiful windows, as also the octagonal dome, all belong to this period. It was restored in 1815.

The church of S. Ponziano, which is not far off, belongs to the 13th century.

After the battle of Trasimenus (217 B.C.) Spoletium was attacked by Hannibal, who was repulsed. It suffered greatly during the civil wars. Sulla, after his victory over Crassus, confiscated the territory of Spoletium (82 B.C.). Under the empire it again became a flourishing town. It was situated on a branch of the Via Flaminia, which left the main road at Narnia and rejoined it at Forum Flaminii. An ancient road also ran hence to Nursia. Martial speaks of its wine. Aemilianus, who had been proclaimed emperor by his soldiers in Moesia, was slain by them here on his way to Rome (A.D. 253), after a reign of three or four months. The foundation of the episcopal see dates from the 4th century. Owing to its elevated position it was an important stronghold during the Vandal and Gothic wars. It was beautified by Theodoric (493) and fortified by Belisarius (536) but its walls were dismantled by Totila (546). Under the Lombards Spoleto became the capital of an independent duchy (from 570), and its dukes ruled a considerable part of central Italy. Together with other fiefs, it was bequeathed to Pope Gregory VII by the empress Matilda, but for some time struggled to maintain its independence. In 1155 it was destroyed by Frederick Barbarossa. In 1213 it was definitely occupied by Gregory IX. During the absence of the papal court in Avignon it was a prey to the struggles between Guelphs and Ghibellines, until in 1354 Cardinal Albornoz brought it once more under the authority of the Church. In 1809 it became capital of the French department of Trasimene. In 1860 it was taken by the Italian troops after a gallant defense.

At Monteleone di Spoleto a circular tumulus was found in 1902, with a rectangular chamber in the centre, and in it was the famous *biga* (two horse chariot), the only ancient one found in a complete state, which is now in the Metropolitan Museum at New York. (T. A.)

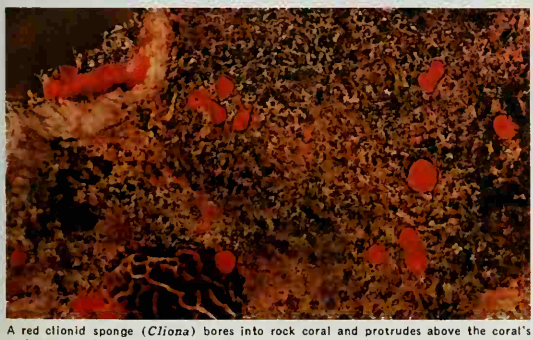
SPONGES, or Porifera (from Latin *porus,* pore, and *fero,* to bear), are the simplest animals of the subkingdom Metazoa, which comprises all the multicellular forms. The body of a sponge consists of a rather loose aggregation of cells supported by a framework of spicules or fibres which form intricate skeletal structure. The differentiation into definite tissues, such as muscular, glandular, nervous, connective, etc., is incomplete and consequently there is but little co-operation and co-ordination among the parts of the body. The majority of sponges are of irregular shape, forming branching or encrusting structures attached to the bottom or to the submerged objects (fig. 1). Some of the sponges have, however, definite form and symmetry.

John Ellis (1765) was the first to demonstrate the animal nature of sponges by observing the water currents produced by a sponge and noticing the contractions on the surface of its body. For a long time sponges were grouped together with the Coelenterates (Linnaeus, Jean Lamarck, Georges Cuvier), until H. M. D. de Blainville (1816) placed them into a special group, Spongiaria, which he considered related to Protozoa. His idea received, however, no recognition from his contemporaries. The present name Porifera was created in 1836 by R. E. Grant, who devoted many years to the study of sponge morphology and anatomy. Separation of sponges from other multicellular forms (Metazoa) was

An angelfish hovers above a yellow vase sponge (*Verongia fistularis*)

Skeletons of several Demospongiae: three bath soonges in the back row and a branching sponge with both siliceous spicules and spongin in the foreground

A commensal sponge (*Suberites*) overgrows shell of crab



A red clionid sponge (*Cliona*) bores into rock coral and protrudes above the coral's surface; coral is broken (upper left) to reveal sponge within

Several common British sponges. Left to right, yellowish *Axinella*, *Raspailia* (with attached grayish *Sycon* "urns") and an irregularly spherical orange *Polymastia*

A VARIETY OF SPONGES

PHOTOGRAPHS, (TOP LEFT) RUSS KINNE—PHOTO RESEARCHERS, INC., (BOTTOM LEFT) W. D. HARTMAN, (TOP RIGHT, CENTRE RIGHT) RALPH BUCHSBAUM, (BOTTOM RIGHT) DOUGLAS P. WILSON F.R.P.S.

SPONGE

Calcareous purse sponge (*Grantia compressa*) encrusts a rock on the Devon coast, England

Underside of stone from shallow water reveals orange *Ophlitaspongia seriata*, brown *Adocia cinerea* and yellow *Halichondria bowerbanki*

Breadcrumb sponge (*Halichondria panicea*) encrusts an overhanging seashore rock. The excurrent openings (oscula) point downward

Squirrelfish and basket sponge in Jamaican waters

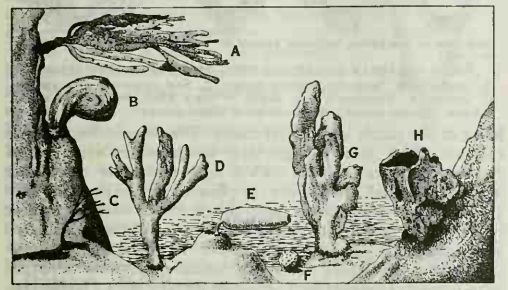
Several *Haliclona* species (blue, turquoise and lavender) grow on mangrove roots; brown area is a *Halichondria* species; in lower right is *Dysidea etheria*

SOME INSHORE SPONGES

PHOTOGRAPHS, (TOP LEFT, TOP RIGHT, CENTRE LEFT) DOUGLAS P. WILSON, F.R.P.S., (BOTTOM LEFT) W. O. HARTMAN, (BOTTOM RIGHT) RUSS KINNE—PHOTO RESEARCHERS, INC.

proposed by T. H. Huxley (1875) and W. J. Sollas (1884), the view still held in 1945. Modern zoology places the sponges into a special branch of Metazoa called, after Sollas, Parazoa.

General Characters.—As the name Porifera implies, the surface of a sponge is covered with a multitude of minute openings or pores through which the current of water is maintained by flagellated cells located inside the sponge body. The water drawn in by the beating of the flagella passes through a complex system of canals and is expelled from several large vents, so-called oscula.



AFTER GIDDER IN "PROCEEDINGS OF THE S.W. NATURALISTS' UNION"

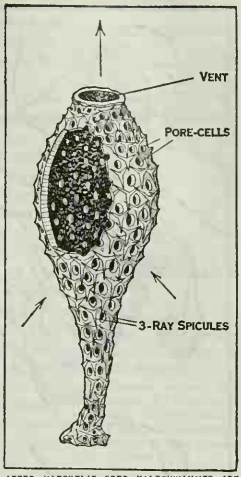
FIG. 1.—COMMON SPONGES OF TEMPERATE SEAS

A. Branched Finger Sponge (*Chalina oculata*); B. Sea Fig (*Suberites ficus*); C. Tow Sponge (*Vibulinus stuposus*); D. *Axinella dissimilis;* E. Common Sponge (*Leucandra fistulosa*); F. Sea Sponge (*Donatia lyncurium*); G. Mermaid's Glove (*Desmacidon fruticosum*); H. Brittle Horny Sponge (*Spongelia fragilis*)

The tissues are kept in place by a skeleton composed of mineral spicules or made of a mesh of elastic fibres (horny sponges). The simplest organization is found in the so-called ascanoid type of sponges represented by a few genera of calcareous sponges, *Leucosolenia* (fig. 2), *Ascute, Ascyssa* and *Dendya*. The latter genus approaches a more complex syconoid type. The vaselike symmetrical body of an ascanoid sponge is made of three layers of tissues forming the walls around a large central cavity (fig. 3). The outside covering consists of epidermis made of a single row of thin and flat cells. The central cavity is lined by a layer of collar cells or choanocytes, recognizable because of the presence of a contractile collar surrounding the base of a single long flagellum (fig. 4). The intermediate layer or mesenchyme (fig. 3, [7]) separating the collar cells from the epidermis consists of gelatinous material with spicules embedded in it and numerous amoeboid cells, the amoebocytes (fig. 3, [8]), freely wandering about. The incurrent pores or ostia are not simple holes in the epidermis but represent intracellular channels passing through special tubular cells, the so-called porocytes (fig. 5). Water is discharged through a large opening known as the osculum (fig. 3, [1]).

In the next more complex type of organization, represented by the genus *Sycon* and called syconoid, the continuous layer of choanocytes is pushed outward at regular intervals to form pouchlike projections or radial canals. The walls of the canals may fuse, leaving between them narrow spaces, the inward-leading canals, lined with the surface epidermis. The spreading of the epidermis and mesenchyme to form a cortex (fig. 6) and the underlying subdermal spaces presents a further advance in the complexity of sponge organization and leads to a final or so-called leuconoid type. The latter type of organization and its numerous variations are represented in a vast majority of sponges. The leuconoid sponge (fig. 6, D) is characterized by clusters of small rounded flagellated chambers formed as a result of the outfolding of a continuous choanocyte layer of the ascanoid sponge. The choanocytes occur only in these chambers. The water is conveyed to the flagellated chambers through the ostia, subdermal spaces, inward-leading channels and pores (prosopyles) and is expelled through the outward-leading pores (apopyles) to the outward-leading canals, larger channels and is finally discharged through the oscula. A large layer of mesenchyme fills up the space between the canals and the flagellated chamber. The large central cavity so prominent in the asconoid type is obliterated and the entire body of the sponge consists of a maze of water channels and flagellated chambers supported by a skeleton made of mineral spicules or spongin fibres. The development of a complex leuconoid type of organization is accompanied by a greater efficiency in producing and maintaining strong currents of water which provide the sponge with food and oxygen for respiration.

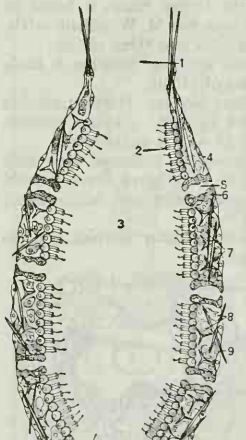
Size, Shape and Colour.—The size of sponges varies from a fraction of an inch, the usual length of small inconspicuous calcareous sponges of temperate seas (*Sycon, Grantia*), to huge masses of irregular or global shape several feet in diameter found in tropical and subtropical waters (*Hircinia* and many other horny sponges). Some of the sponges have definite symmetrical bodies, but the majority grow as irregular encrusting masses or form large branching colonies. Of particular interest are many deep-sea sponges which display greater regularity of growth not found in their relatives inhabiting shallow waters. This secondary symmetry clearly shown in a remarkable species, *Esperiopsis challengeri* (fig. 7), is probably correlated with life in still water at great oceanic depths, which according to the generally accepted view the sponges invaded only in a comparatively recent time.



AFTER HAECKEL'S "DER KALSCHWAMME" (DE GRUYTER & CO.)

FIG. 2.—YOUNG LACE-CHALK (LEUCOSOLENIA PRIMORDALIS) IN "OLYNTHUS" STAGE, MAGNIFIED

Sponges are completely devoid of organs of locomotion and are usually fastened to bottom, rocks, piles or other submerged objects. Those living on muddy bottoms are anchored by roots consisting of bundles of long spicules (fig. 8).



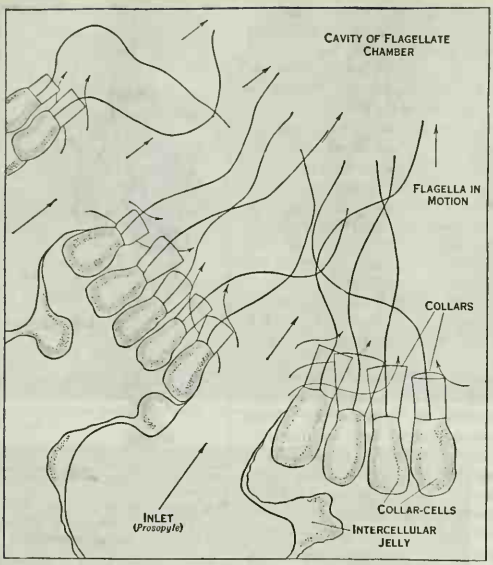
FROM L. H. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

FIG. 3.—DIAGRAM OF THE ASCANOID SPONGE

1, osculum; 2, layer of choanocytes; 3, central cavity; 4, epidermis; 5, pore through porocyte; 6, porocyte; 7, mesenchyme; 8, amoebocyte; 9, spicule

The colour is greatly variable. The deep water sponges are usually drab or brownish, but many of the inhabitants of shallow water lagoons display brilliant yellow, green, red, scarlet and violet colours. Among the most brilliantly coloured sponges are the scarlet sponge, *Tedania ignis*, of the coral reefs in Bermuda, and the red sponge, *Microciona prolifera*, inhabiting the estuaries along the North American coast. Various species of boring sponge, *Cliona*, frequently are found as large bright-yellow masses overgrowing the shell of an oyster or a clam. The greenish colouration of a fresh-water sponge, *Spongilla viridis*, is due to the presence of unicellular green algae which invade the sponge cells and live and propagate inside the sponge body without injuring their host (*see* SYMBIOSIS).

Classifications.—The sponges are classified primarily on the basis of the type of skeleton and shape and size of their spicules.



AFTER BIDDER IN "QUARTERLY JOURNAL OF MICROSCOPICAL SCIENCE"

FIG. 4.—WALL OF FLAGELLATE CHAMBER OF THICK WALLED BANANA-CHALK (LEUCANDRA ASPERA); ARROWS SHOW DIRECTION OF CURRENTS

The chemical composition of the skeletal material serves to separate the calcareous, or chalky sponges, with the spicules composed of calcite, from siliceous and horny sponges in which the spicules are composed of silica or the skeleton consists entirely of horny fibres. Some of the classes and orders of siliceous sponges are not, however, clearly defined and no uniform system of classification has yet been agreed upon by zoologists. The classification commonly used in the United States is based on the works of E. Topsent, H. V. Wilson and M. W. deLaubenfels. The phylum Porifera is usually divided into three classes:

Class I.—Calcarea, the calcareous sponges. Skeleton is made of one-, three-, or four-rayed spicules of calcite.

Class II.—Hexactinellida, the glass sponges. Skeleton consists of a network of separate six-rayed (triaxon) siliceous spicules, finger-shaped flagellated chambers, no surface epithelium.

Class III.—Demospongiae. Sponges of leuconoid type with siliceous spicules, not of triaxon type, or of horny fibres or both. Spicules are differentiated into megascleres and microscleres; small flagellated chambers.

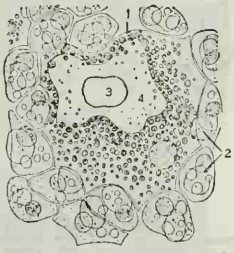
Subclass I.—Tetractinellida—with tetraxon spicules and no spongin. Spicules sometimes absent.

Subclass II.—Monaxonida—large spicules (megascleres) of monaxonal type, spongin may be present or absent.

Subclass III. — Keratosa—the horny sponges. Skeleton consists of a meshwork of spongin fibres. Spicules absent.

For more detailed study of sponge taxonomy, consult monographs and special publications listed by G. C. J. Vosmaer.

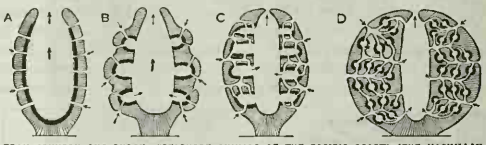
Histology.—The differentiation of cells in sponge tissues is much less pronounced than in the higher forms which possess various organs or systems such as digestive tracts, musculature, kidneys, etc., lacking in Porifera.



FROM L. H. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

FIG. 5.—POROCYTE OF LEUCOSOLENIA WITH SURROUNDING CHOANOCYTES: 1, POROCYTE; 2, CHOANOCYTES; 3, INWARD-LEADING PORE; 4, PORE DIAPHRAGM

In spite of the low degree of organization, cellular elements of the sponge body are specialized to perform definite functions. Several distinct categories of cells can be distinguished in sponge tissues. The first group comprises various cells of the cortex and of the



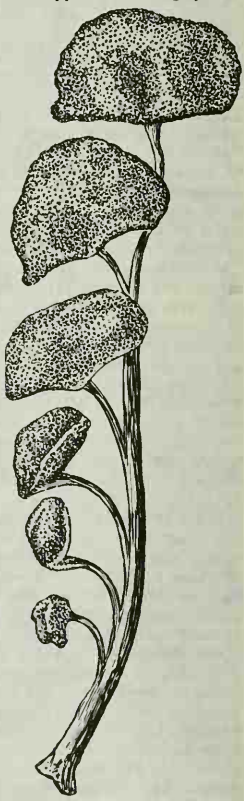
FROM JOHNSON AND SNOOK, "SEASHORE ANIMALS OF THE PACIFIC COAST" (THE MACMILLAN CO., 1927)

FIG. 6.—DIAGRAM OF VARIOUS TYPES OF SPONGE ORGANIZATION
A, ascanoid; B, Syconoid, without cortex; C, final syconoid, with cortex; D, Leuconoid. Choanocyte layer shown in heavy black, mesenchyme cross-hatched. Arrows indicate the direction of water current

lining of the canals and central cavity. The pinacocytes, as these cells are called, form a single layer of epithelium consisting of flattened platelike elements with a large nucleus in the thicker portion of the cell. Sometimes the pinacocytes are provided with the cilia which help in maintaining the current of water. According to Wilson, in some of the sponges the true epithelium is replaced by a continuous "epithelioid membrane" or syncytium with scattered nuclei. The intake of water is regulated by means of special cells, porocytes, pierced by a canal and supplied with a highly contractile diaphragm consisting of a thin sheet of cytoplasm stretched across the opening (fig. 5). The origin of porocytes, which by some of the investigators are considered as degenerate cells, has not yet been determined.

The flagellated cells or choanocytes lining the flagellated chambers are typical structures of sponge organization. They are easily recognized by transparent contractile collars and long flagella (fig. 4) attached to the cell body by means of a basal granule. In appearance and structure the choanocytes resemble the flagellated Protozoa of the family Craspedomonadidae. One representative of this group, a colonial protozoan *Proterospongia* (fig. 16), is regarded as a probable link between the sponges and flagellates.

The inner layer of a sponge body consists of transparent intercellular ground substance of variable consistency which gives support or acts as a scaffolding for widely separated mesenchyme cells. The latter are represented by different kinds of amoebocytes, or wandering cells, varying in size, character of pseudopodia, and type of inclusions. The mesenchyme also contains a large number of undifferentiated cells called archaeocytes, which may develop into more specialized types and apparently play an important role in regeneration. Mesenchyme cells which secrete spicules and spongin fibres are called scleroblasts. According to the type of skeleton they produce (fig. 12), they are designated as calcoblasts (secreting calcite



FROM RAY LANCASTER, "TREATISE ON ZOOLOGY," PART II, "THE PORIFERA" BY E. A. MINCHIN. ADAM & CHARLES BLACK, LONDON (THE MACMILLAN CO., NEW YORK AGENT) 1900

FIG. 7.—ESPERIOPSIS CHALLENGERI, A DEEP-SEA SPONGE WITH A SECONDARY SYMMETRY

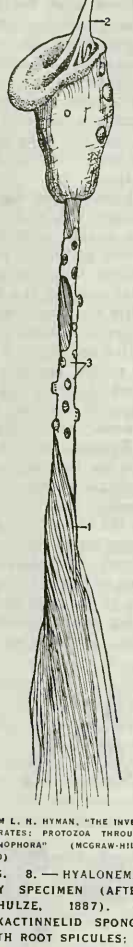
spicules), silocoblasts (producing siliceous spicules) and spongioblasts (forming spongin fibres).

The sex cells, ova and spermatozoa, are apparently derived from choanocytes which lose their flagella and collars and penetrate into mesenchyme. During this stage they look like true amoebocytes. Because of this similarity the old view maintained in the last century by F. E. Schulze, N. Polejaeff and O. Maas that the germ cells arise directly from amoebocytes cannot be discarded without further study.

True muscles in sponges are wanting, but special fusiform contractile cells, the myocytes, somewhat resembling the smooth muscle cells of higher forms have been described. In certain species they form sphincters which regulate the afferent stream of water (fig. 9). Sensory and nerve cells are absent. The contractions of the surface of the body, or of the oscula, noticeable in some of the species, are the result of the direct stimulation of the pinacocytes and myocytes.

Skeleton.—Soft tissues are supported by a complex framework of hard spicules made of calcite ($CaCO_3$) in calcareous sponges or of opal (SiO_2nH_2O) in siliceous sponges. In horny sponges the skeleton consists of a meshwork of elastic fibres made of spongin, an albuminoid of a highly complex chemical structure. The spicules vary in size, shape, and number of axes. Since the taxonomists use them as principal characters for the identification of sponges, elaborate terminology has been developed for designating various skeletal elements (fig. 10). In general there are two principal types: the large skeletal spicules, the megascleres which form the framework of the sponge body; and the small microscleres of variable and irregular shape scattered throughout the tissues.

Various classes of spicules are distinguished by the number of their axes or rays. Thus, monaxon designates the needle-like spicule with one axis; tetraxon has four rays emerging from one point and extending into different planes; three axes crossing at the right angles typical for Hexactennellid sponges produce an element with six rays (fig. 11) or triaxon; several equal rays emerging from a central point characterize the polyaxons known also as asters. The classification and the corresponding terminology become greatly involved as various modifications caused by the loss of rays, bending of axes, etc., are taken into account and each type denoted by special term.



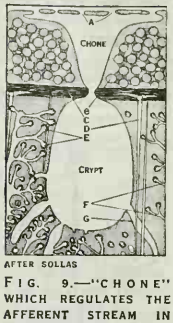
FROM L. H. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

FIG. 8.—HYALONEMA, DRY SPECIMEN (AFTER SCHULZE, 1887). A HEXACTINNELID SPONGE WITH ROOT SPICULES: 1, ROOT SPICULES; 2, GASTRAL CONE; 3, SYMBIOTIC ZOANTHIDS ON ROOT TUFT

Mineral substance of a spicule is deposited around a central axis of organic material. In siliceous sponges the silica is laid in the form of concentric cylinders. Usually two scleroblasts (fig. 12) participate in the formation of a monaxon spicule which is laid down by one of them, the founder, while the other, the thickener, deposits additional mineral material. When the spicule is completed the two cells move apart. Several cells combine in secreting more complex spicules and spongin fibres (fig. 12, K, O).

Chemical Composition.—Mineral constituents of the skeleton comprise a very large portion of sponge body. From 15% to 97% of the ash of siliceous sponges consists of SiO_2, while in calcareous sponges $CaCO_3$ content of ash varies from 84.9% to 86.8%. Other mineral constituents include MgO, P_2O_5, $Fe_2O_3Al_2O_3$, Na_2O, K_2O, SO_3, Si and minute quantities of Mn, Zn, B, Br, Cu and As.

The presence of relatively large quantities of iodine in dried commercial sponges, first demonstrated in 1819 by Andrew Fyfe, is of considerable scientific interest. It is known that iodine present in various species occurs in a combination with an organic compound of the skeleton known as iodospongin. The percentage of iodine in iodospongin varies from 1.5% to 4.86% in bath sponge (*Euspongia sp.*) to 8%–10% in *Luffaria cauliformis*. More detailed studies identified the iodine compound obtained from spongin as di-iodotyrosine, a complex which also contains bromine. Calcareous and siliceous sponges are, as a rule, poor in iodine; but the red sponge *Microciona prolifera* is apparently an exception, for according to the determinations made in 1933 by W. Bergman its iodine content comprises 0.3% of the total dry weight. The physiological role of the iodine in sponge body is not known.

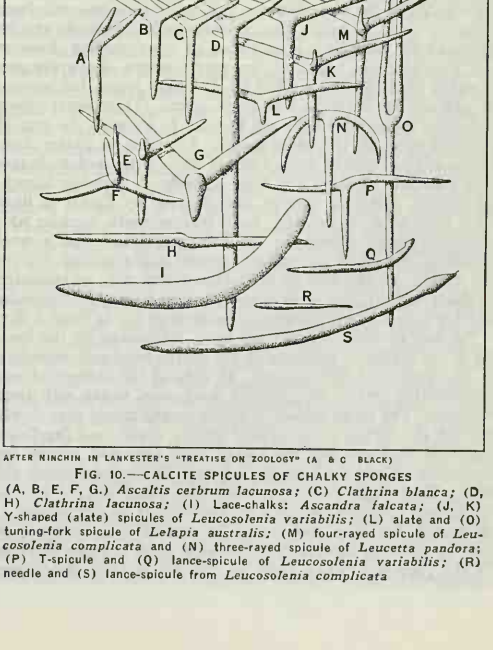


AFTER SOLLAS

FIG. 9.—"CHONE" WHICH REGULATES THE AFFERENT STREAM IN CYDONIUM EOSASTER, A GRAPNEL-SPONGE

(A) Dermal pores; (B) Solid-stars; (C) Contractile cells; (D) Grapnel spirule; (E) Flagellate chambers; (F) Inward-leading canals; (G) Afferent canal

Physiology.—The water pumping mechanism of sponges consists of thousands of choanocytes which line the flagellated chambers. Because of their position inside the sponge body and small size, their action is difficult to observe. Only in calcareous sponges are the choanocytes relatively large and therefore observable with less difficulty. In action the flagellum of a choanocyte makes an undulating spiral movement which starts at its base and proceeds toward the tip. As the flagellum points toward the outward leading pores, apopyles (fig. 4), each of its strokes draws



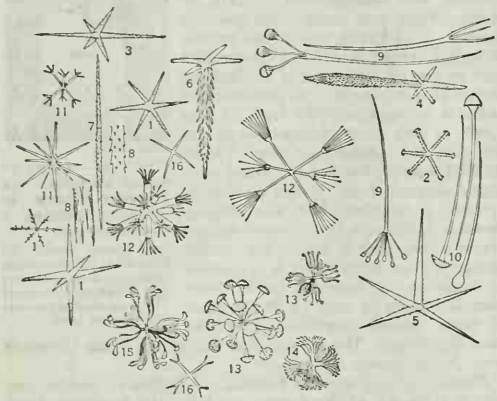
AFTER MINCHIN IN LANKESTER'S "TREATISE ON ZOOLOGY" (A & C BLACK)

FIG. 10.—CALCITE SPICULES OF CHALKY SPONGES

(A, B, E, F, G.) *Ascaltis cerbrum lacunosa;* (C) *Clathrina blanca;* (D, H) *Clathrina lacunosa;* (I) Lace-chalks: *Ascandra falcata;* (J, K) Y-shaped (alate) spicules of *Leucosolenia variabilis;* (L) alate and (O) tuning-fork spicule of *Lelapia australis;* (M) four-rayed spicule of *Leucosolenia complicata* and (N) three-rayed spicule of *Leucetta pandora;* (P) T-spicule and (Q) lance-spicule of *Leucosolenia variabilis;* (R) needle and (S) lance-spicule from *Leucosolenia complicata*

a little water and pushes it forward, gradually filling up the chamber. The pressure so produced stretches the walls of the flagellated chamber and is extended to the outward leading canals and central cavity. Apparently there is no co-ordination of the beating of the flagella, each choanocyte working more or less independently of the others. As a result of the movement of thousands of minute flagella the water is sucked into the flagellated chambers through the dermal pores and inward leading canals and escapes through the outward leading canals and oscula. The velocity of current in each flagellated chamber is rather slow, but as thousands of tiny pumps deliver water into a relatively small number of vents, or oscula, the rate of the current increases. G. P. Bidder showed that in *Leucandra aspera* the current velocity at the osculum was about 8 cm. a second, while the velocity at the same moment past collar cells was only from 2 cm. to 4 cm. an hour. The pressure inside the oscula can be measured by inserting a glass tubing of corresponding diameter and observing the height of the water column above the surrounding level. By this method G. H. Parker found that hydraulic pressure in *Stylotella* was about 4 mm. These observations and other indirect methods can be used to compute the rate of pumping water by the whole sponge. According to Bidder, *Leucandra aspera* requires about 5 gal. of water a day and discharges it at a distance of about 9 in. from its body. A fingerlike branching sponge, *Spinosella sororia*, was observed by Parker to filter 1,575 litres of water per day. F. G. Walton Smith, working with cultivated horny sponges in the Bahamas, found that a wool sponge of 500 cu.cm. in volume pumped approximately 2 litres per minute. At this rate its daily requirement may be estimated at about 600 gal., but observations are still lacking to determine whether such a rate of pumping is maintained for 24 hours a day.

The flow of water through the body, maintained by the beating of the flagella, is regulated by the contractions and expansions of the oscula and dermal pores. In spite of the absence of any co-ordinating centre, the pumping mechanism works with great efficiency. The water sucked in by the sponge moves very slowly through the dermal pores, inward leading canals, and flagellated chambers, and provides the tissues with food and oxygen for respiration. Since the outgoing currents discharged through the oscula have much greater velocity, the water filtered by the sponge and contaminated with the products of its metabolism, is discharged far away and cannot be re-used. The presence of toxic or irritating substances dissolved in water is easily detected by the porocytes, which react by contracting their diaphragms and closing the pores. In this way the sponge protects itself against possible injury. Structure and functioning of the pumping mechanism can be regarded as an efficient adaptation of sponges to the sedentary conditions of life in the ocean.



FROM L. H. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

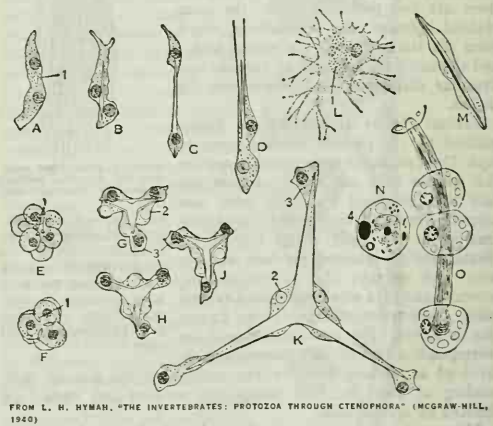
FIG. 11.—SPICULES OF HEXACTINELLID SPONGES

1, hexactines; 2, spiny hexactine; 3, modified hexactine with one long ray; 4, pinulated hexactine; 5, 6, pentacts; 7, uncinate; 8, small part of uncinate greatly magnified; 9, scopules; 10, calvules; 11, oxyhexasters; 12, discoctasters; 13, discohexasters; 14, plumicome; 15, floricome; 16, stauractine

Food and Nutrition.—The food of sponges consists of minute organisms suspended in water. Bacteria probably play an important role in the feeding of horny sponges which frequently inhabit waters containing but small numbers of other microorganisms in suspension. Thus, for instance, principal commercial species of sponges in the Bahamas are found on mud bottoms in waters containing very abundant bacterial flora but poor in plankton. The role of bacteria in the feeding of sponges has not been, however, adequately studied and requires further investigation.

Ingestion and digestion of food can be observed by feeding the sponges with easily identifiable materials such as carmine or carbon powder, milk, starch grains or bacteria, and examining their tissues at regular intervals after feeding. As the flow of water passing from the narrow inward leading channels slows down after entering the flagellated chambers, the suspended particles settle and are caught by collar cells. In calcareous sponges the material entrapped in the collars of choanocytes becomes engulfed and digested by the cytoplasm. Some of the particles may be passed, however, to the amoebocytes surrounding the flagellated chambers (fig. 13). More complete division of functions exists in siliceous sponges. Here food is also caught by the choanocytes but digestion is accomplished only by amoebocytes.

Respiration.—Since special organs of respiration and circulation are lacking, the need for oxygen is supplied through direct exchange between the tissues and the surrounding water. Consequently, the rate of metabolism (O_2 consumption) is controlled by the rate of exchange of water. Hyman found that the O_2 consumption in *Sycon* may decrease by 80% when the oscula are closed. The observed rates per hour in this sponge vary from 0.04 to 0.16 ml. of oxygen per gram of fresh weight. The figures are of the same degree of magnitude as those observed for many other marine invertebrates. Sponges have no organs of excretion. Products of their metabolism and indigestible material, ejected by the amoebocytes, are discharged with the outgoing currents.



FROM L. H. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

FIG. 12.—SECRETION OF MONAXON, A-D, AND TRIRADIATE (E-K) SPICULES (AFTER WOODLAND, 1905); L, CALCOBLAST (AFTER DENDY, 1893); M, SILICEOUS MONAXON OF FRESH-WATER SPONGE (AFTER EVANS, 1901); N, O, SECRETION OF SPONGIN IN RENIERA (AFTER TUZET, 1932); 1, SPICULE; 2, FOUNDER; 3, THICKENER; 4, SPONGIN MASS

Poisonous and Injurious Sponges.—The surface of many sponges is slimy because of the excretion of mucous, which may cause unpleasant but not dangerous irritation of the skin. For instance, the scarlet sponge of the Bermuda islands, *Tedania ignis*, when handled with naked hands, causes painful dermatitis.

The nature of the active substance is not known. Of special interest are the boring sponges of the genus *Cliona*, widely distributed in the coastal water of North America. As the name indicates, the sponge bores through the shells of oysters and other molluscs and in case of severe infestation may destroy them. The method by which the sponge makes its long and branching passages in the shell substance is not known. Secretion of acid has been suggested but not corroborated by observation. Studies conducted by P. S. Galtsoff and V. Pertzoff in 1926 showed, however, that the suspension of cells of *Cliona* behaves as a much stronger acid than that of the nonboring sponge *Microciona*. Mechanical action may also be involved in boring, but the process has not yet been adequately studied.

Regeneration.—The power of regeneration, *i.e.*, the ability to restore lost parts, is very high in sponges. Not only is a piece cut off from a body capable of growing into a complete sponge, but the cells separated by squeezing a piece of sponge through fine silk cloth, and suspended in water, reunite and develop into a new sponge. The phenomenon discovered by Wilson in 1907 is of considerable theoretical interest. It was thought first that the development of a new sponge from reunited masses is to a certain extent similar to the development of an embryo, and that the separated cells undergo dedifferentiations and return to the embryonic state. It was shown by P. S. Galtsoff that certain cells retain their morphological characteristics and the first steps in regeneration are accomplished by the sorting out and rearrangement of the old cellular material.

Under unfavourable conditions sponges collapse, disintegrate and leave only small reduction bodies which consist of a mass of amoebocytes covered by epidermis. With the return of normal conditions these bodies are capable of developing into complete sponges.

Asexual Reproduction.—All fresh-water and some marine sponges reproduce by means of gemmules, small, hard and rounded bodies consisting of a mass of archaeocytes laden with reserve food and surrounded by a columnar layer of flattened cells (fig. 14). The formation of gemmules in fresh-water sponges begins with the gathering of amoebocytes, which are joined by special nurse cells, the trophocytes, carrying food reserves. Some of the amoebocytes arrange themselves in columnar layers and secrete two concentric membranes around the central core. The inner membrane is usually much thicker than the outer. The growing gemmule is invaded by the scleroblasts, which secrete the so-called amphidisks, the spicules having the appearance of ordinary cuff links, and place them between the two membranes of the columnar layer. The amazing part of cell behaviour is that when the gemmule is completed, the columnar cells, nurse cells and scleroblasts depart.

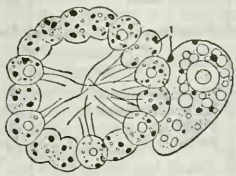
Fresh-water sponges usually disintegrate with the onset of cold weather but gemmules survive and hatch in spring. The content of the gemmule emerges through a small opening, the micropyle (fig. 14, [11]), and develops into a new sponge through the differentiation and rearrangement of the archaeocytes.

Sexual Reproduction.—Hermaphroditism is apparently prevalent among the sponges, although in some of them the sexes are separate. Fertilization of the ovum, studied by J. B. Gatenby in calcareous sponges, is rather unique, for the sperm brought in by current of water first enters the choanocyte which acts as a nurse cell. The fusion of the choanocyte with the nearest egg, which occupies a position close to the flagellated chamber, sets the sperm free. Insemination is then accomplished in a regular manner.

Similar process was described by O. Tuzet in siliceous sponges, with the only difference that sperm enters the amoebocyte, which in turn transfers it to the egg.

Detailed information about the breeding periods of sponges is lacking. In the commercial sponges of the Bahamas (sheepswool and velvet) breeding takes place all the time, but the greatest proportion of specimen with embryos ready for release were found by F. G. W. Smith in March and April.

Development.—The fertilized egg divides into two, four and eight blastomeres which surround a central cavity while they are still arranged in one plane. After the 32-cell stage has been reached the divisions become somewhat irregular. At this time the embryo, called now blastula, has a shape of an oval vesicle, one end of which is flattened and the other rounded. The central cavity of the blastula opens to the exterior at the flattened surface. This opening or pore is surrounded by the cells distinguishable from the others by heavy granular inclusions, while the remaining cells are columnar and have long flagella.



FROM L. H. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

FIG. 13.—AMOEBOCYTE OF A CALCAREOUS SPONGE RECEIVING CARBON PARTICLES FROM CHOANOCYTES (AFTER POURBOIX, 1933); 1, CARBON PARTICLES

Rapid proliferation of granular cells produces a thick mass which bulges into the central cavity (fig. 15A). By this time the embryo, by means of the flagella, forces its way into the nearest flagellated chamber and escapes through the osculum. Apparently through the absorption of water, the layer of invaginated cells is forced out and the embryo acquires an oval shape.

The sponge larva at this stage is called amphiblastula, from Greek, *amphi*, around or on both sides and *blastos*, germ, (fig. 15, B). One half of it consists of columnar and brightly pigmented cells with long flagella, the other half is made of round granular cells. The amphiblastula type of larvae occurs in all families of sponges and is, therefore, typical of the entire group. The larva swims with its ciliated end forward. After a few days of free life it settles on a smooth surface of a stone or shell, with its ciliated portion directed downward, and almost immediately flattens. The ciliated cells become invaginated and the central cavity of the larva, the blastocoele, reduced to a narrow slit. The larva now becomes attached to the substratum. At this stage it consists of two layers. The granular cells form the outer covering and the flagellated cells line the inside cavity. Very soon the flagellated cells develop characteristic collars and are transformed into choanocytes. The appearance of the granular substance between the two layers, the formation of the pores and osculum, and the secretion of spicules complete the development. At this stage the young sponge is usually referred to as "olynthus." Its final transformation into an adult sponge is a matter of continuous slow growth and rearrangement of cellular elements. Although the details of the development may vary in different classes and families the essential steps are similar in all of them.

Individuality of Sponges.—Lack of a central regulatory or co-ordinating organ, great variability in shape or form, and remarkable regenerative power, make it difficult to decide whether sponges are individuals or colonies. This theoretical question, much debated by the zoologists of the last century, has not yet been satisfactorily settled. It is generally considered that the osculum with the contributing part of the canal system is an individual element or a "person," and that several of these units are combined in a colony of individuals grown together and performing common functions. The fact that the osculum controls the flow of water through the adjacent part of the body and is the most conspicuous part of the sponge anatomy, seems to give weight to this view. The boundaries of individual units are, however, indistinct and their individuality is of very vague nature. It is remarkable that, in spite of the absence of central organs, many species of sponges developed very definite shape and symmetry. Their skeletons also show progressive complexity from a rather simple arrangement of spicules to a very complicated symmetrical framework built according to a definite structural pattern.

Relationship.—On the basis of our knowledge of the anatomy and histology of sponges, it is reasonable to suppose that this group evolved from colonial choanoflagellate Protozoa similar to *Proterospongia* (fig. 16). In the subsequent evolutionary process sponges formed a separate branch of Metazoa which failed to develop higher grades of organization characterized by the formation of a head, nervous system, organs of digestion, circulation and respiration present in other multicellular forms. The sponges

retained their ancestral protozoan features in the structure of choanocytes, in the intracellular digestion, and in the mode of production of the skeleton, the elements of which are secreted by single cells or group of freely moving cells. Progressive evolution in sponges consisted chiefly in the perfection of the canal system and in developing a more complex skeleton. Sponges appeared very early in geological history, for their spicules are found in the upper Cambrian deposits.

Ecology.—The distribution of sponges is world-wide. They inhabit all the oceans, seas, freshwater lakes and rivers, and grow on all kinds of bottoms from tidal flats to abyssal depths. Their most favoured environment is, however, that of rocky or hard bottom along the seashores and in coral-reef lagoons. The greatest percentage of more than 2,500 known species of sponges is confined to this habitat. Close association with other invertebrates seems to be one of the most interesting features of sponge ecology. Porous structure and numerous cavities of sponge body provide safe abode for hydroids, worms, shrimps, crabs, molluscs and other invertebrates. Some of these animals invade the sponges accidentally, while others form more permanent association. Thus, thousands of shrimps of the family Alpheidae are found in every loggerhead sponge of the British West Indies. The young shrimp of the genus *Spongicola* enter the cavity of glass sponges (Hexactinellida), where they live in pairs and soon become imprisoned because their grown-up bodies cannot escape through the small openings of the sieve plates guarding the oscula of their host. Sponges of the family Suberitidae are often found overgrowing snail shells occupied by hermit crabs. The relation is obviously advantageous for both parties. The crab is well protected by sponge spicules and its repelling taste, while the sponge is carried about by the crab and gets better opportunities for feeding. Crabs of the genus *Dromia* break pieces of sponges with their legs and place them on their backs where the sponge becomes implanted.

Sponges are apparently distasteful to other animals and are usually avoided. They are eaten only by a few species of nudibranch molluscs.

Diseases.—Widespread mortalities frequently occur among commercial sponges. In the fall of 1938 a devastating disease which rapidly spread from the Bahamas to Cuba, British Honduras, Mexico and the east coast of Florida almost completely destroyed the velvet sponge and killed more than 70% of wool sponges. The disease was associated with an infection by a funguslike micro-organism tentatively identified as *Spongiophaga*. No control measures were suggested to check the infection except the closing of grounds to commercial fishery.

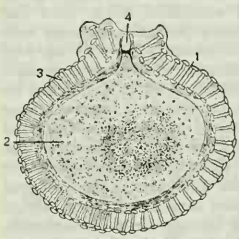
Utilization of Sponges.—Horny sponges (Keratosa) have been in use from the earliest times. In ancient Greece, mothers pacified their crying babies by giving them a piece of sponge soaked in honey. According to Pliny, sponges were employed as paint brushes and mops, and Roman soldiers carried with them a piece of sponge to be used instead of a drinking vessel. This custom makes comprehensible the manner in which vinegar was offered to Christ on the cross—in a soaked sponge (Mark XV, 36). In the 13th century A.D. Arnold of Villa Nova introduced the "burnt sponge" as medicinal preparation to be used internally in scrofula, a term which usually refers to tuberculosis of lymphatic glands, but formerly also covered other diseases. The therapeutic effect, if any, was possibly due to a high content of iodine present in the carbonized sponge in the form of NaI.

TABLE I.—*Common and Scientific Names of Principal Commercial Sponges*

Common Name	Scientific Name	Fishing Areas
Turkey cup; Turkey solid, Fine Levant; Silks	*Euspongia officinalis molissima*	Eastern Mediterranean, Tunis
Turkey toilet	*Euspongia officinalis adriatica*	Adriatic, Tunisia, Eastern Mediterranean
Zimocca	*Euspongia zimocca*	Adriatic, Dardanelles, Greece, Egypt, Tripoli, Tunisia
Elephant's ear	*Euspongia officinalis lamella*	Provence, Dalmatia, Egypt, Tunisia, Algeria
Honeycomb; bath	*Hippiospongia equina elastica*	Mediterranean
Wool; sheepswool	*Hippiospongia lachne*	Florida, West Indies, Honduras
Velvet	*Hippiospongia grossypina*	Bahamas, Honduras, Cuba
Yellow	*Spongia barbara*. Probably also other species	West Indies, Florida, Cuba, Haiti, Caribbean Sea, Australia, New Zealand
Grass	*Spongia graminea*	Florida, Bahamas, Cuba, Mexico, British Honduras
Glove	*Hippiospongia canaliculata* var. *flabellum*	Florida, Bahamas
Reef	*Spongia obliqua*, and probably *S. agaricina* and other species	Bahamas, Cuba, British Honduras, Florida
Hardhead	*Spongia dura*, *S. agaricina corlosia* and probably other species	Bahamas, British Honduras, Haiti

In modern times, besides obvious toilet and household uses, sponges are required for scores of purposes such as the manufacture of special surgical and hygienic preparations, leather dressing, application of glaze to fine pottery, scouring and sponging of cloth. They are extensively employed by tile and bricklayers, painters, decorators, lithographers, jewellers and silversmiths. Probably the largest number of sponges is used for washing railroad cars and automobiles.

Industrial Types.—The trade recognizes several types which correspond to different species listed in Table I. There is, however, no general agreement regarding the specific names of different varieties of commercial sponges.



FROM L. N. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

FIG. 14.—THE GEMMULE OF A FRESH-WATER SPONGE. EPHYDATIA, WITH INNER MEMBRANE (AFTER EVANS, 1901); 1, INNER MEMBRANE; 2, MASS OF ARCHAEOCYTES; 3, A LAYER OF AMPHIDISKS; 4, MICROPYLE

Good commercial sponges occur also in the Philippines and in Australia. Contribution to the world trade of sponges from these countries is, however, insignificant. Several varieties of horny sponges of the Hawaiian Islands are not suitable for the trade because of their extreme softness and lack of durability.

The annual world production of sponges in 1938, the last "normal" year before World War II, was estimated at 2,102,400 lb., valued at $2,556,398. The U.S. contributed 28.8% of the total catch and 41.9% of the total value. The contributions of the principal sponge producing countries for the years 1938 and 1939 are shown in Table II.

In 1943 the production of sponges in the U.S. dropped to 325,822 lb., almost one half of that of 1938, but the value increased to $2,305,070.

The most valuable of the commercial sponges are the Turkey Cup and Turkey Toilet. Best specimens of these varieties sometimes are sold as high as $50 per pound. The total quantity of these high priced sponges offered on the world's market is, however, insignificant in comparison with the wool sponge, which in the western hemisphere constitutes about 80% of the catch.

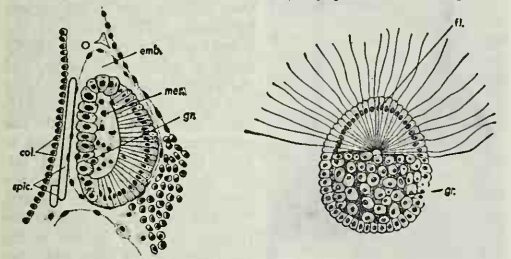
Fishing Methods.—Commercial sponges are taken from the bottom of the sea from tidal level to a depth of about 200 ft. In the shallow inshore waters the sponges are picked by means of a two- or three-pronged hook attached to a long pole. The hooker, as the fisherman is called, operates from a small skiff which he pushes by pole. When the sponge is noticed, sometimes at the depth of 25 or even 30 ft., the pole is lowered and the hook

TABLE II.—*Estimated World Production of Sponges*

Country	1938		1939	
	Quantity (pounds)	(U.S. dollars)	Quantity (pounds)	(U.S. dollars)
United States	606,000	$1,070,848	484,400	$1,161,945
Bahamas	670,000	400,000	470,000	120,000
Cuba	440,000	295,000	330,000	230,000
Greece	87,000	262,680	78,000	190,000
Italy‡	130,000	210,000	*	*
Egypt	†	†	—	—
Turkey	75,000	180,000	55,000	150,000
Libya	70,000	120,000	*	*
Caicos Island§	19,500	11,000	6,850	2,500
Syria-Lebanon	3,000	4,500	*	*
British Honduras	1,100	1,770	1,960	2,660
Philippine Islands	800	600	350	250

*No data available; †100,000 pounds produced in Egypt included in the Greek and Italian production because taken by fishermen of those nationalities; ‡except Libya. §Caicos production sold in Nassau not included.

placed under the sponge, which is then detached from the bottom by a slight pull and is brought to the surface. This primitive and slow method of fishing can be used only in clear and calm inshore waters. Nude diving, without equipment of any kind, is still practised in Greece, along the Libyan coast and in the Philippines. Fishing in the deeper water near Greece and off the Florida coast requires more elaborate diving equipment consisting of a

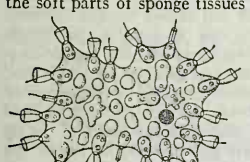


FROM MCBRIDE, "TEXTBOOK OF EMBRYOLOGY" (MACMILLAN CO., 1914)

FIG. 15.—DEVELOPMENT OF GRANTIA LABYRINTHICA (AFTER DENDY)
A, blastula in the embryonic chamber of the mother; B, amphiblastula larva. Col., collared cells lining flagellated chamber; emb., embryonic chamber; fl., flagellated cells; gr., granular cells; mes., mesenchyme cells; spic., maternal spicules. Magnified

regular diving suit, scaphander and boat with mechanically operated air pump. A small dredge, known as "gangave," is used in Libya. The implement is very destructive and its use is prohibited in the western hemisphere. Sponges caught in the Bahamas, Cuba, Turks and Caicos islands are picked exclusively by hookers.

Sponges gathered by divers or hookers are cleaned aboard the boat or placed in "crawls," *i.e.*, small ponds along the shore where the soft parts of sponge tissues rot. The remaining skeletons are taken out, squeezed, rinsed in clear sea water, strung in bunches and dried in air.



FROM L. H. HYMAN, "THE INVERTEBRATES: PROTOZOA THROUGH CTENOPHORA" (MCGRAW-HILL, 1940)

FIG. 16.—PROTEROSPONGIA (AFTER KENT, 1881). A COLONIAL FLAGELLATE CONSIDERED AS A PROBABLE LINK BETWEEN THE FLAGELLATES AND SPONGES

An interesting feature of the sponge industry in the U.S. is a co-operative selling association, the Tarpon Springs Sponge exchange, which auctions all the sponges brought by fishermen. A similar organization, but under direct control of the colonial government, was established in Nassau, Bahamas. The auction is rather unique, for the selling is conducted in complete silence. As the buyers examine the various lots of sponges they write down on a piece of folded paper the highest price they are willing to pay and hand it to the auctioneer. When the auction is over the bids are opened by the representative of the association and the highest bidder is awarded the sponges, provided the price is acceptable to the owner. Before reaching the retail market sponges are sorted, cleaned, bleached and sometimes dyed.

Conservation.—At the end of World War II, no system of conservation or management of natural sponge resources was being exercised by any government. The only protection effective in the U.S., Bahamas, Turk and Caicos islands consisted in the enforcement of the size limit laws which prohibit the taking and selling of sponges less than five inches in diameter.

Sponge Farming or Sponge Culture.—Because of their high regenerative power, sponges can be grown from small cuttings. Experiments in this method of cultivation were made at the end of the last century by H. F. Moore in Florida and were also conducted on a larger scale in the estuarine waters of Tunisia. The practicability of the method was definitely demonstrated by large scale operations conducted by the British government in the West Indies and British Honduras. From 1935 to 1939 more than 140,000 sponges were growing on government grounds in the Bahamas, and more than 700,000 cuttings were planted at Turneffe, British Honduras. Unfortunately, the mortality of 1938–39 was particularly severe among the cultivated sponges and the beds were almost completely destroyed.

The farming method consists in growing sponges from rectangular slices about 2″ x 4″ x ½″, cut with a very sharp knife from a healthy sponge and attached with a string to a piece of rock or cement disk. The cuttings are immediately placed in water. From 16 to 32 pieces can be cut from one adult wool sponge of six to eight in. in diameter. Regeneration begins immediately and the sponge, firmly adhering to its base, grows as a well-shaped round specimen. In the Bahamas, sponges grown from cuttings reach marketable size in about four years.

If correctly conducted, sponge culture may become a profitable occupation, but because of the scarcity of suitable inshore grounds it cannot replace the fishery on natural grounds. The possibility of propagating sponges through sexual reproduction has not yet been explored.

BIBLIOGRAPHY.—E. A. Minchin, "Porifera," Ray Lancaster's *Treatise of Zoology*, v. ii (1900); I. B. J. Sollas, "Porifera," *Cambridge Nat. Hist.*, v. i (1909); L. H. Hyman, *The Invertebrates, Protozoa through Ctenophora*, pp. 284–364 (1940); G. P. Bidder, "The Relation of the Form of a Sponge to its Currents," *Quart. Journ. Micr. Sc.*, v. 67 (1923); P. S. Galtsoff, "Regeneration after Dissociation," Parts I and II, *J. Exp. Zool.*, 42 (1925); P. S. Galtsoff and V. Pertzoff, "Some Physiochemical Properties of Dissociated Sponge Cells," *J. Gen. Phys.*, x (1926); P. S. Galtsoff, H. H. Brown, C. L. Smith and F. G. W. Smith, "Sponge Mortality in the Bahamas," *Nature*, v. 143 (1939); J. B. Gatenby, "Germ Cells, Fertilization and Early Development of Grantia," *Journ. Linn. Soc.*, 34 (London, 1920); further notes, etc., *Quart. Journ. Micr. Sc.*, 71 (1927); A. Hyatt, "Revision of the N. American Porifera," *Mem. Bost. Soc. Nat. Hist.*, v. ii, Part iv (1877); M. W. deLaubenfels, "A Discussion of the Sponge Fauna of the Dry Tortugas," Carn. Inst. of Washington, *Papers from Tortugas Laboratory*, v. 30 (1936); R. V. vonLendenfeld, "A Monograph of the Horny Sponges," *Roy. Soc.* (London, 1889); H. F. Moore, "The Commercial Sponges and the Sponge Fisheries," *Bull.*, U. S. Bur. of Fisheries, v. xxviii (1908) (1910); G. H. Parker, "On the Strength and the Volume of the Water Currents Produced by Sponges," *J. Exp. Zool.*, v. 16 (1914); N. Polejaeff, "Report on the Keratosa," *Challenger Reports, Zoology*, v. xi (1884); G. C. J. Vosmaer, *A Bibliography of Sponges*, 1515–1913 (Cambridge, 1928); H. V. Wilson, "The Sponges," *Mem. Mus. Amp. Zool.* Harv. Univ. 30 (1904) and "On some Phenomena of Coalescence and Regeneration in Sponges," *J. Exp. Zool.*, v. v (1907); H. V. Wilson and J. T. Penney, "Regeneration of Sponges from Dissociated Cells," *J. Exp. Zool.*, v. 56 (1930).

(P. S. G.)

SPONSOR, one who stands surety for another (from Lat. *spondere*, to promise), especially in the rite of Christian baptism, a godfather or godmother. The practice originated not in infant baptism, but in the custom of requiring an adult pagan who offered himself for the rite to be accompanied by a Christian known to the bishop, who could vouch for the applicant and undertake his supervision, thus fulfilling the function performed in the Eleusinian mysteries by the *mystagogus*. The Greek word for the person undertaking this function is ἀνάδοχος, to which the Latin *susceptor* is equivalent. The word "sponsor" in this ecclesiastical sense occurs for the first time, but incidentally only, and as if it were already long familiar, in Tertullian's treatise *De baptismo* (ch. 18), where, arguing that in certain circumstances baptism may conveniently be postponed, especially in the case of little children, he asks, "For why is it necessary that the sponsors like-

wise should be thrust into danger, who both themselves by reason of mortality may fail to fulfil their promises, and may also be disappointed by the development of an evil disposition [in those for whom they become sponsors]?"

The sponsors here alluded to may have been in many cases the actual parents, and even in the 5th century it was not felt to be inappropriate that they should be so; Augustine, indeed, in one passage appears to speak of it as a matter of course that parents should bring their children and answer for them "tanquam fidejussores" (*Epist. . . . ad Bonif.* 98), and the oldest Egyptian ritual bears similar testimony. Elsewhere Augustine contemplates the bringing of the children of slaves by their masters, and of course orphans and foundlings were brought by other benevolent persons.

The comparatively early appearance, however, of such names as *compatres, commatres, propatres, promatres, patrini, matrinae,* is of itself sufficient evidence, not only that the sponsorial relationship had come to be regarded as a very close one, but also that it was not usually assumed by the natural parents. How very close it was held to be is shown by the Justinian prohibition of marriage between godparents and godchildren. On the other hand, the anciently allowable practice of parents becoming sponsors for their own children, though gradually becoming obsolete, seems to have lingered until the 9th century, when it was at last formally prohibited by the council of Mainz (813). For a long time there was no fixed rule as to the necessary or allowable number of sponsors and sometimes the number actually assumed was large. By the council of Trent, however, it was decided that one only, or at most two, these not being of the same sex, should be permitted.

The rubric of the Church of England, that "there shall be for every male child to be baptized two godfathers and one godmother, and for every female one godfather and two godmothers," is not older than 1661; the sponsors are charged with the duty of instructing the child, and in due time presenting it for confirmation. In the Roman Catholic Church the spiritual relationship established between the sponsor and the baptized, and the sponsors and the parents of the baptized, constitutes an impediment to marriage. (*See* MARRIAGE, LAW OF: *Canon Law.*)

SPONTANEOUS GENERATION, in biology, is the theory, now disproved, that living organisms sometimes arise from nonliving matter. It is sometimes referred to as abiogenesis, as opposed to biogenesis, the now-established fact that living organisms arise only from the reproduction of previously existing organisms.

See BIOLOGY: *History: Biogenesis Versus Abiogenesis.*

SPONTINI, GASPARO LUIGI PACIFICO (1774–1851), Italian composer, was born on Nov. 14, 1774, at Majolati (Ancona) in Italy. He was the son of a poor cobbler and was intended for the priesthood. But he obtained lessons from *Kapellmeister* Quintiliani, and in 1791 went to the Conservatorio de' Turchini at Naples, where he studied under G. Paisiello, D. Cimarosa and Fiorivanti. By 1799 he had already written and produced eight operas.

After becoming court composer to King Ferdinand of Naples in this year an intrigue with a princess of the court compelled Spontini to leave Naples in 1800. In 1803 he settled in Paris, where he had no success until the production of *Milton,* a one-act opera, in 1804. Thereafter his powers and ambitions steadily developed, being manifested to the full for the first time in *La Vestale,* produced on Dec. 15, 1807. *Ferdinand Cortez* was received with equal enthusiasm in 1809, but *Olympia* (1819) had less success.

Spontini had been appointed director of the Italian opera in 1810, but his quarrelsome and grasping disposition led to his summary dismissal two years later, and, though reinstated in 1814, he voluntarily resigned his post soon afterward. In 1820 he settled in Berlin by invitation of Frederick William III, commissioned to superintend all music performed at the Prussian court and compose two new grand operas, or three smaller ones, every three years. But he began by at once embroiling himself with the intendant, Count Brühl. *La Vestale, Ferdinand Cortez* and *Olympia*—the last two entirely remodeled—were produced with great success in 1821. But Spontini's fame was entirely eclipsed by the appearance of Carl Weber's *Der Freischütz.* A new opera, *Nourmahal,* founded on Thomas Moore's *Lalla Rookh,* was performed in 1822, and another, entitled *Alcidor,* in 1825; in 1826 Spontini began the composition of *Agnes von Hohenstaufen* (produced in 1829), which undoubtedly will rank permanently as his greatest work.

Dismissed from his court position in 1841, he settled in Paris. In 1850 he retired to his birthplace, Majolati, and died there on Jan. 14, 1851.

SPOON, a table implement, bowl shaped at the end, with a handle varying in length and size. From the derivation of the word the earliest northern European spoon would seem to have been a chip or splinter of wood; the Greek *kochliarion* (Lat. *cochleare*) points to the early and natural use of shells, such as are still used by primitive peoples. Examples are preserved of the various forms of spoons of ivory, flint, slate and wood, many of them carved with the symbols of their religion, used by the ancient Egyptians. The spoons of the Greeks and Romans were chiefly made of bronze and silver, and the handle usually took the form of a spike or pointed stem. Medieval spoons for domestic use were commonly made of horn or wood, but brass, pewter and "latten" spoons appear to have been common about the 15th century. The full descriptions and entries relating to silver spoons in the inventories of the royal and other households point to their special value and rarity. The earliest English reference appears to be in a will of 1259. In the wardrobe accounts of Edward I for the year 1300 some gold and silver spoons marked with the fleur-de-lis, the Paris mark, are mentioned. One of the most interesting medieval spoons is the coronation spoon used in the anointing of the sovereign. The sets of spoons popular as christening presents in Tudor times, the handles of which terminate in heads or busts of the apostles, are a special form to which antiquarian interest attaches. The earlier English spoon handles terminate in an acorn, plain knob or a diamond; at the end of the 16th century the baluster and seal ending became common, the bowl being "fig shaped." At the Restoration the handle became broad and flat, the bowl broad and oval and the termination cut into the shape known as the *pied de biche,* or hind's foot. In the first quarter of the 18th century the bowl became narrow and elliptical, with a tongue or "rat's tail" down the back, and the handle turned up at the end. The modern form, with the tip of the bowl narrower than the base and the rounded end of the handle turned down, appeared about 1760.

BY COURTESY OF VICTORIA AND ALBERT MUSEUM

SILVER APOSTLE SPOONS, MADE IN ENGLAND, 15TH CENTURY

See C. J. Jackson, "The Spoon and its History" in *Archaeologia,* vol. liii (1892); also Cripps, *Old English Plate.*

SPOONBILL, any one of six species of moderately large wading birds constituting the subfamily Plataleinae and, with the ibises (*q.v.*), the family Threskiornithidae (order Ciconiiformes). The head is partly or entirely bare, the plumage white (except in the roseate spoonbill), often with a rosy tinge, some with a nuptial crest or foreneck plumes. They differ from ibises in having a long nearly straight bill, narrow in the middle and broad and flat at the end, which they immerse in feeding, swinging it from side to side and sifting small animals from the mud and water of estuaries, salt-water bayous and lakes. In flight they flap steadily, usually without sailing. They breed in colonies, often with ibises and herons, build a good nest of sticks, sometimes lined with leaves or bark, among reeds or in a low bush or

tree about ten feet up, and lay three to five white eggs blotched with reddish-brown. The beautiful roseate spoonbill (*Ajaia ajaja*), 32 in., head and throat bare, legs red, is deep pink with neck and upper back white, wing coverts and tail coverts deep carmine and tail and sides of breast tawny-buff. It ranges from the Gulf coast of Texas (and a very few in Florida), the West Indies and central Mexico to Argentina and Chile. In many places it was exterminated by plume hunters. The yellow-legged spoonbill (*Platibis flavipes*) of Australia is white with straw-coloured foreneck and a yellow forehead, upper throat and bill. The crested white spoonbill (*Platalea leucorodia*), 24 in., with cinnamon-buff on the foreneck, breeds in marshes of central and southern Europe and Asia, south to Egypt, India and Formosa; the black-billed spoonbill (*P. regia*), white with black wing tips and black legs, from Australia and New Zealand to New Guinea, Celebes, Moluccas and Solomons; the lesser spoonbill (*P. minor*), from Korea and southern Japan to southern China and the Philippines; the African spoonbill (*P. alba*), over Africa south of the Sahara and in Madagascar.

ALLAN D. CRUICKSHANK FROM THE NATIONAL AUDUBON SOCIETY

ROSEATE SPOONBILL (AJAIA AJAJA), FOUND IN TROPICAL AND SUB-TROPICAL AMERICA

(G. F. Ss.)

SPORADES, the scattered islands of the Greek archipelago, distinguished from the Cyclades, which are grouped round Delos, and from the coastal islands of Europe and Asia. The Doric Sporades—Melos, Pholegandros, Sikinos, Thera, Anaphe, Astypalaea and Cos—were by some considered a southern cluster of the Cyclades. In modern times the name Sporades is applied to two groups: the northern Sporades lie northeast of Negropont (Euboea), Skiathos, Skopelos and Ikos being in the department of Magnesia, and Scyros (*q.v.*) in that of Euboea. Scyros has a fine harbour. Skiathos is beautifully wooded; the town overlooks an excellent harbour. Skopelos is also well wooded.

The southern Sporades, off the southwest of Asia Minor, are: Icaria, Patmos (*q.v.*), Leros, Calymnos, Astypalaea (Astropalia or Stampalia), Cos (*q.v.*; Turk. Stanko), Nisyros, Tilos or Episcopi, Syme, Khalki, Rhodes (*q.v.*) and many smaller isles.

SPOROZOA, parasitic Protozoa that produce unicellular spores within a tough envelope or cyst. Since formerly various Protozoa, which had only this character in common and were not truly related to one another, were classed with the Sporozoa, some protozoologists have split up the class Sporozoa into two classes: Amoebosporidia and the true Sporozoa. *See* PROTOZOA.

SPORTS (ARTICLES ON). A universal form of recreation, sports are also one of the great expressions of the aspiration to excel. The victories of athletes were sung by the great poets of classical Greece. Many of *Encyclopædia Britannica's* articles on sports discuss their development in the context of social history. Among the articles in this group are BASEBALL; BOWLING; BOXING; CRICKET; FOOTBALL; GOLF; and WRESTLING.

TRACK AND FIELD SPORTS is supplemented by articles on individual events, such as DISCUS THROWING; POLE VAULTING; SHOT-PUT; etc. WINTER SPORTS summarizes their development and present status, and is supplemented by such individual articles as BOBSLEDDING; CURLING; ICE HOCKEY; and SKIING.

Some of the aquatic sports articles are BOATING; CANOE; MOTORBOAT; ROWING; SKIN DIVING; SWIMMING; WATER POLO; YACHTING.

Among the articles dealing with racing sports are AUTOMOBILE RACING; CYCLING; DRIVING AND COACHING; HORSE RACING AND BREEDING; MOTORCYCLING; RUNNING; and WALKING.

Among the court games to which articles are devoted are BADMINTON; DECK TENNIS; RACKETS; SQUASH RACKETS; and TENNIS. Additional articles on games played with a ball include BASKETBALL; FIVES; HANDBALL; JAI ALAI; LACROSSE; ROUNDERS.

A general view of the sports that have been adapted to the needs of children is given in CHILDREN'S GAMES AND SPORTS.

In some sports of almost universal interest, man's strength and resourcefulness are matched against other forms of life or against forces of nature. Articles dealing with these activities include CAMPING; FISHING; HUNTING; MOUNTAINEERING; and WOODCRAFT. TOURNAMENT deals with the great sport of medieval chivalry. Sports of antiquity that are still practised are treated in ARCHERY; FALCONRY; and FENCING.

GAMES, CLASSICAL describes the athletic festivals that played an important part in the life of ancient Greece and Rome. Articles dealing with modern sports festivals include BRITISH EMPIRE AND COMMONWEALTH GAMES; HIGHLAND GAMES; OLYMPIC GAMES; and PAN-AMERICAN GAMES.

ATHLETICS traces the ancient distinction between the amateur and professional; AMATEUR describes the contemporary rules governing amateur standing. GYMNASTICS discusses the evolution of the art, the types of exercises, and the skills involved. GYMNASIUM deals with types of equipment. The function of gymnastics in educational programs is discussed in PHYSICAL EDUCATION.

For articles on nonathletic games such as cards, board games, etc., *see* GAMES (ARTICLES ON).

SPOTSWOOD, ALEXANDER (1676–1740), American colonial governor, was born in Tangier, Africa, in 1676. He served as aide to Marlborough in the War of the Spanish Succession, and was wounded at Blenheim. He was appointed lieutenant governor of Virginia in 1710, and brought to the colony the privilege of habeas corpus; his term closed in Sept. 1722, but he remained in Virginia, living in Germanna, a settlement of Germans, on the Rapidan in Spotsylvania county; he was deputy postmaster general of the colonies from 1730 to 1739. He was the first representative of the British government in America who fully appreciated the value of the western territory. He recommended the establishment of a Virginia company to trade with the Indians, urged the construction of forts along the frontier, and personally organized and conducted in 1716 an exploring expedition into the Shenandoah valley. He developed the iron industry of Virginia, promoted the religious education of the Indians and tried to advance the interests of education, especially at the College of William and Mary. He died at Annapolis, Md., June 7, 1740.

SPOTSYLVANIA, a county in Virginia, U.S., midway between Washington, D.C., and Richmond, Va.; the scene of four of the major battles of the American Civil War (*q.v.*). At Fredericksburg (Dec. 13, 1862) and Chancellorsville (May 1–3, 1863) Confederate Gen. R. E. Lee inflicted severe defeats on Union armies attempting to move on Richmond. The Wilderness (May 5–6, 1864) and Spotsylvania Courthouse (May 9–20, 1864) were costly but indecisive battles in Union Gen. U. S. Grant's successful campaign of that year. (*See* also CHANCELLORSVILLE; FREDERICKSBURG; WILDERNESS.)

The population of the county in 1960 was 13,819; its area is 409 sq.mi. About half of the area is devoted to agriculture; dairy products are the chief source of farm income. Manufacturing, mainly of lumber and basic lumber products, is the principal employer of the labour force. First settled in the late 17th century, the county was formed in 1720 and named for Alexander Spotswood, then governor of Virginia. Fredericksburg lies on the county's northeastern border. (M. BR.)

SPOTTED FEVER: *see* ROCKY MOUNTAIN SPOTTED FEVER AND OTHER SPOTTED FEVERS.

SPOTTISWOODE (SPOTTISWOOD, SPOTISWOOD or SPOTSWOOD), **JOHN** (1565–1639), archbishop of St. Andrews and historian of Scotland, eldest son of John Spottiswood, minister of Calder, was born in 1565. He was educated at Glasgow university (M.A. 1581), and succeeded his father in the parish of Calder in 1583. In 1601 he attended Ludowick, duke of Lennox, as his chaplain, in an embassy to the court of France, returning in 1603. He followed James to England on his accession, but was the same year nominated to the see of Glasgow, his consecration in London, however, not taking place until Oct. 1610. On May 30, 1605, he became a member of the Scottish privy council. In 1610 he presided as moderator over the assembly in which presbytery was abolished, in 1615 he was made archbishop of St. Andrews and primate of Scotland, and in 1618 procured the sanction of the privy council to the Five Articles of Perth with their ratification by parliament in 1621.

In 1633 he crowned Charles I at Holyrood. In 1635 he was appointed lord chancellor of Scotland, an office which he retained until 1638. He was a spectator at the riot of St. Giles', Edinburgh, on July 23, 1637, endeavoured in vain to avoid disaster by concessions, and on the taking of the Covenant perceived that "now all that we have been doing these thirty years past is thrown down at once." He escaped to Newcastle, was deposed by the assembly on Dec. 4, on trivial charges, and died in London on Nov. 26, 1639, receiving burial in Westminster abbey. His most considerable work was *The History of the Church and State of Scotland* (London, 1655, *seq.*).

See the accounts prefixed to the first edition of Spottiswoode's *History of Scotland* and to that published by the Spottiswoode society in 1851; also David Calderwood's *Hist. of the Kirk of Scotland* (1842–49).

SPRAT, THOMAS (1635–1713), English bishop, was born at Beaminster, Dorsetshire, and educated at Wadham college, Oxford, where he held a fellowship (1657–70). Having taken orders he became a prebendary of Lincoln in 1660. In the preceding year he had gained a reputation by his poem *To the Happie Memory of the most Renowned Prince Oliver, Lord Protector* (London, 1659), and he was afterward well known as a wit, preacher and man of letters. His chief prose works are the *Observations upon Monsieur de Sorbier's Voyage into England* (London, 1665), a satirical reply to the strictures on Englishmen in Samuel de Sorbière's book of that name, and a *History of the Royal Society of London* (London, 1667), which Sprat had helped to found. In 1669 he became canon of Westminster, and in 1670 rector of Uffington, Lincolnshire. He was chaplain to Charles II in 1676, curate and lecturer at St. Margaret's, Westminster, in 1679, canon of Windsor in 1681, dean of Westminster in 1683 and bishop of Rochester in 1684. Although he opposed the motion of 1689 declaring the throne vacant, he assisted at the coronation of William and Mary. As dean of Westminster he directed Sir Christopher Wren's restoration of the abbey. He died on May 20, 1713.

SPRAT, a marine fish (*Clupea sprattus*), one of the smallest species of the genus *Clupea* or herrings, rarely exceeds five inches in length, and occurs in large shoals on the Atlantic coasts of Europe. The sprat spawns in the open sea from February to May. Its eggs are buoyant and pelagic and easily recognized. The sprat is one of the more important food fishes because of the immense numbers in its shoals. In Norway large quantities are tinned in oil, or are eaten fresh, pickled, smoked or prepared as anchovies.

SPRAY GUN AND AIR BRUSH are painting tools that utilize compressed air to atomize a liquid into a controlled pattern. Spraying is the act of breaking up or atomizing a liquid into a multitude of tiny droplets, which are formed when a liquid is drawn into a slender stream or thin film. Since it is not usually practical to form a stream of liquid thin enough to produce a fine spray, it is necessary to utilize a device such as the spray nozzle.

The spray nozzle is the heart of the spray gun and air brush and operates by impinging high-velocity turbulent air on the surface of filaments or films of liquid which when stretched to a thin film collapses to droplets. This action produces a wide range of particle sizes.

An air brush resembles a pencil and is used by photographers and commercial artists for applying colour and shading to drawings, prints and photographs and to accentuate high lights and supply background to films. The more complex spray gun, developed originally from the air brush, is used for less delicate work and requires less skill to operate. Lacquers, paints, varnishes, shellac and other finishes for manufactured articles are applied with spray guns that are designed to be simple to operate, adjust and clean.

Various types and combinations of spray guns are available for different applications. Portable equipment is used for maintenance painting of both new and old construction. Automobiles, furniture and appliances are usually finished under clean, rigidly controlled working conditions with stationary equipment. The finishes may be supplied to the gun by gravity from an overhead reservoir (for low-viscosity lacquers), by suction from a container that is part of the spray gun (for small work and where frequent colour changes occur) or by pressure in a mass-production operation. An adequate supply of dry air is provided by auxiliary equipment, such as compressors, air-reducing valves, oil and water separators and flexible hose. (P. He.)

SPRAYING AND DUSTING MACHINERY. A great variety of equipment is available for applying numerous organic and inorganic substances in solution, suspension or dust form to combat insect pests and plant diseases of field, garden and orchard crops, and to control weeds or other unwanted vegetation. Spraying and dusting are also used for defoliation before harvest, especially for cotton; for applying mineral fertilizers under special conditions; for thinning fruit with blossom-killing chemicals; and for applying various hormone materials to plants to control fruit set, fruit drop, etc. The term sprayer is applied to devices that dispense solutions, and the term duster to those which apply chemicals in the form of dry, finely divided powders or dusts.

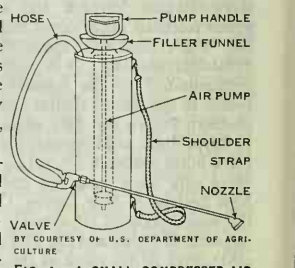


FIG. 1.—A SMALL COMPRESSED-AIR SPRAYER

Spraying Equipment.— Sprayers differ greatly in size and capacity, source of power and other characteristics. They commonly utilize pressure produced by pumps to force the spray solution through specially designed nozzles that break it up into mists, thus covering a maximum percentage of the plant surface with protecting chemicals or reaching a maximum number of insects with contact insecticides. Liquid pumps are commonly used to handle the spray directly, but air pumps that develop pressure in the liquid container, forcing the liquid through the nozzles, are used in some hand sprayers (fig. 1) and in special situations. Cylinders of compressed carbon dioxide or other gases are sometimes used as a source of pressure, and refillable aerosol bombs, similar to but larger than those employed for household sprays, may be used for spraying in greenhouses or gardens. Pressure in these bombs is maintained by the evaporation of a volatile solvent such as Freon.

Nozzles for spraying are commonly of two types. One produces a cone of spray for medium and close applications, and the other produces under high pressure a comparatively solid stream of liquid that breaks up at a considerable distance. Spray guns can be adjusted quickly between these two types. A third and less common type of nozzle produces a flat fan of spray. The solid-stream nozzle is generally used singly for spraying tall trees.

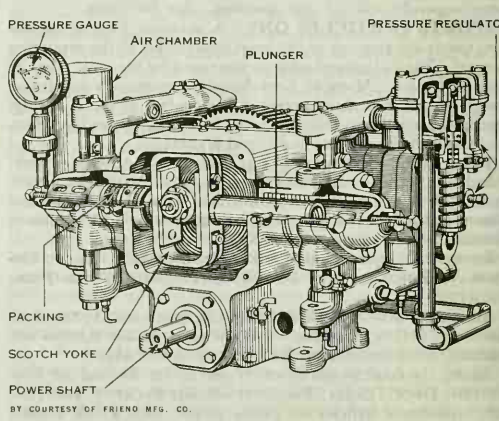
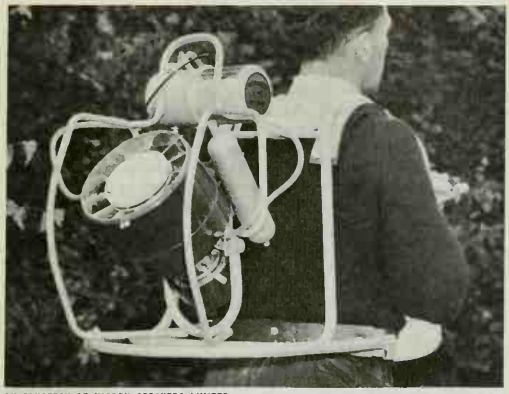


FIG. 2.—SPRAYER PUMP CUT AWAY TO SHOW WORKING PARTS



BY COURTESY OF MICRON SPRAYERS LIMITED

FIG. 3.—BACK-MOUNTED CONCENTRATE OR AIR SPRAYER. A SMALL MOTOR AND FAN DISTRIBUTE THE CONCENTRATED SPRAY

The cone type may be used singly on small sprayers (fig. 1) or in groups, either on a hand-controlled rod with a single cutoff or on a mounted boom. Grape and hop sprayers may use vertical booms with several nozzles on each side of the sprayer. Row-crop sprayers commonly have a horizontal boom with nozzles adjusted to row widths. One nozzle is mounted above the row; frequently there is also one on each side of the row, angled upward and mounted on drop pipes.

Weed-control sprayers may use fan-type nozzles on a straight boom for controlling weeds in small grain, or drop pipes with carefully directed nozzles for spraying under maize or other tall row crops. In the concentrate or air-type sprayer the nozzles are mounted in the air stream from a motor driven fan. Air from the fan carries and distributes the fine spray produced by the nozzles (fig. 3 and 4).

Manual Sprayers.—The simplest form of household or garden sprayer consists of a bicycle-type air pump with a small container attached, and an atomizer nozzle. The knapsack sprayer, as its name implies, is carried on the back of the operator, pressure usually being provided through a diaphragm pump that also actuates an agitator.

In the compressed-air sprayer, pressure is built up with a plunger-type air pump housed in the container with the solution and air. The one shown in fig. 1 has a capacity of from 1½ to 4 gal. Of these same types but of larger capacity are the wheelbarrow and barrel sprayers.

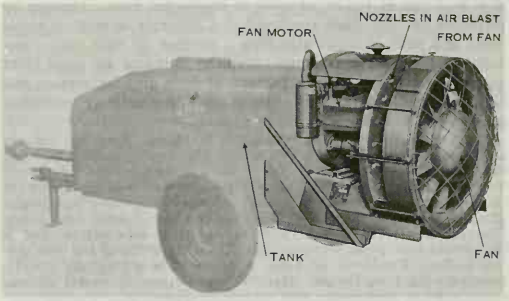
Power Sprayers.—There are three types of power sprayers: those powered with an engine mounted on a frame with the tank, those powered through a power take-off from truck or tractor motor and those deriving their power from a traction drive. With the first two the same type of assembly may be used. This consists essentially of a single-acting, plunger-type, reciprocating pump (usually two or three cylinder), a strainer to remove sediment that might clog the nozzles, a pressure regulator which also acts as a safety valve, a metal or wooden tank containing the liquid and an agitator for stirring it, and such accessories as pressure gauges, cutoff valves and, frequently, air chambers. The pumps obtain their reciprocating motion most commonly from either a crankshaft and connecting rods, eccentric and connecting rods, or a Scotch-yoke assembly (fig. 2), and are available in a wide range of sizes and capacities with pressures up to 1,000 lb. per square inch. Corrosion-resistant valve seats and ball-type valves, usually of hardened stainless steel, are used in the pumps and two types of plunger-cylinder combination—the plunger cup, where the cup is fastened to and moves with the plunger, rubbing against the cylinder wall which is usually coated with porcelain, and the outside-packing type, in which a stainless-steel plunger works back and forth in the stationary packing lining the cylinder wall. The engine-powered sprayer and its engine may be mounted on skids for semistationary work and supplied with portable pipe lines to conduct the solution; it may be fixed as a stationary plant and powered either with an engine or electric motor, in which case the conducting pipes are buried in the ground and lead to risers at strategic points for attachment of the spraying equipment; or it may be mounted on a motor truck or on a two- or four-wheel trailer and driven either by its own engine or by a power take-off connection from the tractor pulling it, or from the transmission case of the truck transporting it.

Occasionally the whole spraying outfit may be mounted on the tractor. Pneumatic tires are generally considered desirable on spray-equipment trailers. Traction-driven sprayers derive their power from the transport wheels and their use is limited largely to row crops. In most instances the power take-off or auxiliary-engine drive is more suitable.

Concentrate or Air Sprayers.—These sprayers, which are also called speed sprayers, are power sprayers with a motor-driven fan which spreads the spray, usually in a concentrated solution at low volume, in a stream of high-velocity air. They vary in size from the carried or wheelbarrow-mounted sprayers with a capacity of one or a few gallons (fig. 3) to giant orchard or row-crop sprayers, operated by one man, that can cover 100 to 200 ac. of crop per day.

The use of an air stream makes it possible to reduce the spray volume to as little as one gallon an acre and thus make great savings in transportation, a factor of considerable importance in rough terrain or where water is scarce. Chemicals are dissolved in water or in especially purified oils of the fuel-oil boiling-point range. (Ordinary fuel oils should not be used since the aromatic compounds they contain are toxic to plants.) The small units are suitable where labour is inexpensive or where the use of large machinery is impractical. The largest units are used by contract sprayers in level areas of extensive culture. Crops sprayed vary from tea and coffee to bananas and orchard crops of all types. Row-crop concentrate sprayers are designed to force a beam of chemical-carrying mist across the field to a distance of 50 to 100 ft., and are adjustable to compensate for wind and uneven land surface.

Dusting Equipment.—Dusters for applying insecticides and fungicides make use of an air blast produced by means of a fan or plunger-type air pump, which blows the dust into a discharge pipe fitted with a straight-tube or fan-shaped nozzle. The many makes and models of dusters are of four types: (1) hand operated, in which the air pressure is built up with a plunger pump or with a crank-operated fan; (2) the traction machine, in which the power for operating is taken from the ground wheels through chains or gears; (3) the power duster (horse or tractor drawn), operated by an auxiliary gasoline engine and suitable for single-spout orchard dusting, or (by changing to a multiple-nozzle unit) useful in row-crop dusting; and (4) the tractor-power duster, mounted on the tractor and driven through the tractor-power take-off. The power and traction types consist essentially of a hopper, agitator,



BY COURTESY OF JOHN BEAN DIVISION

FIG. 4.—MEDIUM-SIZE CONCENTRATE OR AIR SPRAYER. AIR BLAST DISTRIBUTES SPRAY FROM THE NOZZLES

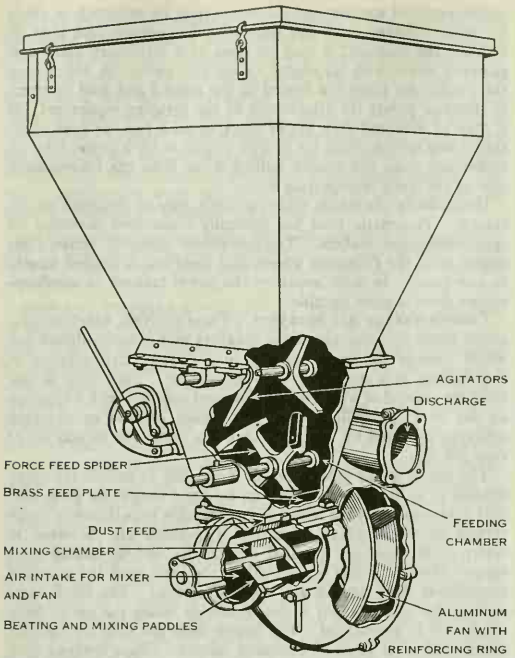


FIG. 5.—TYPICAL DUSTER

feed control, blower fan, mixing chamber, discharge tubes and nozzle or nozzles. Dusters are convenient for gardens or small areas where a few plants must be treated at frequent intervals; concentrate sprayers are generally more suitable for field work.

Granular, dry mixtures were being used experimentally in the late 1950s where complete coverage is unimportant, as in distributing grasshopper bait and some weed-control chemicals. Granulation permits a wider distribution of dry material with minimum drifting.

Airplane Spraying and Dusting.—In the cotton belt of the United States, dusting of the plants by means of airplanes has proved economical. The work is done during daylight, generally in the early morning or late afternoons, when it is considered safest to fly at the necessary low altitudes, and when the moisture on the plants causes the dust particles to adhere better. Commercial firms usually do the dusting on a per-acre basis and commonly dust at a rate of 350 ac. or more of cotton per hour. Dusting by airplanes has also been employed in California to control insect pests and plant diseases attacking a number of fruit, nut, vegetable and small-grain crops. Airplane spraying of weed-killer concentrates has become important in grain-growing regions and brush in rough grazing land often is most economically controlled by spraying from the air.

BIBLIOGRAPHY.—W. S. Hough and A. F. Mason, *Spraying, Dusting and Fumigating of Plants* (1951) has two chapters on modern spraying and dusting machinery; H. P. Smith, *Farm Machinery and Equipment* (1955), contains a good chapter on spraying and dusting equipment as well as a consideration of related machinery; *Sprayer and Duster Manual*, published by the National Sprayer and Duster Association, Chicago, is a small manual with much useful information, and is distributed free to interested persons; Texas Agricultural and Mechanical College, *Handbook of Aerial Application in Agriculture* (1956), covers dusting from the air. Many U.S. and other experiment stations have bulletins on spraying and dusting equipment.

(R. B. G.; W. E. Ls.)

SPRAYS AND DUSTS IN AGRICULTURE. Spraying, dusting and fumigating are the standard methods of applying the pest-control chemicals and some of the other chemical compounds that are of increasing importance in agriculture. In spraying, the chemicals to be applied are dissolved or suspended in water, or less commonly in a petroleum carrier. The mixture is then applied as a fine mist spray to the plants, animals, soils or products to be treated. As an alternative method, dry, finely powdered chemicals may be mixed with an inert carrier and applied with some type of blower. Dry granular materials also have been used instead of dusts for purposes in which adequate coverage can be obtained by this means. In fumigation, gases or the vapours of volatile compounds are held in contact with the materials to be treated—grain in a tight bin, for example.

Sprays and dusts are used to control insect and fungus pests of plants; insects, such as lice and flies on animals; and weeds, by means of chemical weed killers or herbicides. They may also be used for such special purposes as applying mineral fertilizers, increasing or decreasing fruit set or delaying the dropping of nearly mature fruits, and for defoliation to facilitate the harvest of such plants as cotton.

Sprays have advantages over dusts in their ability to adhere to and spread over treated surfaces. Wetting agents or detergents are commonly added to spray mixtures to increase adhesion and wetting of waxy surfaces. Detergents decrease the tendency of water to collect in drops and permit the chemical solution to spread over a leaf in a very thin film, thus bringing the spray chemical into maximum contact with the fungi or insects to be controlled. Dusts can be lighter in weight per unit of active chemical and thus are better adapted for airplane application. They also drift through the foliage and reach leaves that might not be covered by spraying. After about 1950 there was a rapid increase in the use of air sprayers or "wet dusting." The use of a concentrated spray distributed in the air stream from a powerful fan combines many of the advantages of both spraying and dusting.

Fumigation may be used to control insects and some diseases in stored products or to control insects, such as wireworms and grubs, nematodes and sometimes fungi and weeds in soil. The chemical may be applied as a gas or as a volatile liquid. Partial fumigation of the soil may be accomplished by applying the chemical in a spray or granular material and immediately working it into the soil. For more exacting work on small areas like seedbeds the treated soil is covered with a gastight canvas or plastic cover.

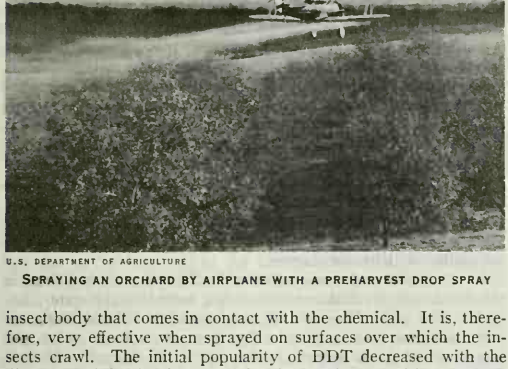
Fungicidal Treatments.—The practice of spraying plants to control plant diseases (*q.v.*) is said to have been discovered accidentally near Bordeaux, France, when a copper-lime mixture applied to grapes near a road to reduce pilfering was observed to control downy mildew, a serious disease of the grape plant. This spray, since known as Bordeaux mixture, remains a standard fungicide, although organic materials are replacing the older copper and sulfur sprays. Among the dozens of modern fungicides, several are especially valuable in general use: captan (trichloromethylthiocyclohexene-dicarboximide), an organic compound sold as a wettable powder, alone or in all-purpose mixtures, and used to treat seeds and bulbs or as a spray or dust on fruits, vegetables and ornamentals for the control of a wide range of fungous diseases; ferbam (ferric dimethyldithiocarbamate), a wettable powder used in sprays and dusts, and as a soil drench to control damping-off diseases of seedling plants; and zineb (zinc ethylenebis-dithiocarbamate), a general purpose fungicide, particularly recommended for blights of tomato and potato and as a soil drench to control crown and root rots of vegetable and flower plants. In general these compounds control more diseases under more conditions and are less toxic to the treated plants than the older copper or sulfur sprays. Streptomycin, an antibiotic, has shown promising results in the control of bacterial plant diseases.

Insecticidal Treatments.—Formerly insecticides were limited primarily to arsenic compounds like Paris green and calcium and lead arsenates for chewing insects, nicotine for soft bodied, sucking insects and oil and lime-sulfur for scale insects. Tobacco powders and infusions were used in Europe before 1800 and pyrethrum powders in Asia Minor at an unknown date. Paris green (copper aceto-arsenite) was used successfully against the potato beetle in Illinois in 1868, and was used against major apple

insects a few years later. Calcium arsenate was the chief insecticide used on cotton from 1920 to 1945.

The compound popularly known as DDT (dichloro-diphenyl-trichloroethane) was among the first of a new group of organic insecticides that includes aldrin (hexachloro-hexahydro-endo, exo-dimethanonaphthalene), lindane (gamma isomer of benzene hexachloride), methoxychlor (trichloro-methoxyphenylethane), parathion (diethylnitrophenyl phosphorothioate) and many others. Most of these are chlorinated hydrocarbons. The organic phosphorus compounds, such as parathion, are among the most effective poisons, but are exceedingly dangerous to operating personnel unless proper precautions are observed. Malathion (ethoxycarbonyl-ethyl-dimethyl phosphorodithioate) shows most of the effectiveness of parathion and is much safer to use. This and other organic phosphorous insecticides showed promise of becoming increasingly important.

Arsenic compounds are toxic only when ingested by the insect. DDT is a nerve poison and paralysis spreads from any part of the



U.S. DEPARTMENT OF AGRICULTURE

SPRAYING AN ORCHARD BY AIRPLANE WITH A PREHARVEST DROP SPRAY

insect body that comes in contact with the chemical. It is, therefore, very effective when sprayed on surfaces over which the insects crawl. The initial popularity of DDT decreased with the discovery of the development of resistance in treated insect populations and with fear of human toxicity from accumulation of the compound as a result of eating contaminated foods. In the late 1950s the tolerance for DDT contamination in milk was set at zero in the United States; that is, no use—such as spraying of dairy animals, their quarters or their feed—that could result in the transmission of the compound to milk was permitted. Directions for control of garden insects specified that DDT should not be applied to foliage to be eaten, or to other edible plant parts that cannot be washed thoroughly or peeled before eating. Malathion, applied not later than three days before harvest, or methoxychlor, not less than seven days before harvest, were the generally recommended insecticides for garden use.

As a rule both the organic fungicides and insecticides are sold as wettable powders for sprays, as dusts or in emulsifiable, concentrated solutions.

Herbicidal Treatments.—Weed-killing chemicals are normally applied as sprays or as dry, granular formulations of compounds that act through the soil. Their use is a comparatively recent development, although ferrous sulfate and other compounds were used to control mustard and similar weeds in grain fields before 1900. Sodium chlorate was used in increasing volume as a nonselective herbicide after 1905, but the simultaneous discovery of 2,4-D (2,4-dichlorophenoxyacetic acid) in the United States and of MCPA (2-methyl-4-chlorophenoxyacetic acid) in England during World War II marked the beginning of the widespread use of herbicidal chemicals. These compounds may be essentially nontoxic to grains and grasses at rates that will kill most broadleaved weeds. An estimate of the herbicides used in North America in 1957 showed 80,000,000 lb. of chlorates, borates, etc., 35,000,000 lb. of 2,4-D and related compounds, 20,000,000 lb. of miscellaneous materials and 25,000,000 gal. of herbicidal oils, with a total value of $65,000,000.

Herbicides may be classed as selective or sterilant. Ferrous sulfate is a selective herbicide for mustard in grain because it adheres to the leaves of the mustard but not to those of the grain. The compound 2,4-D is selective because of its greater toxicity for most broadleaved plants than for grasses, including grains. Trichloroacetic acid (TCA) is selectively toxic to grasses. Sodium chlorate is a general toxin, used to eradicate noxious perennial weeds at the cost of temporary soil sterility or to control all vegetation on railway roadbeds and similar areas. Chlorate is effective primarily as it is absorbed through the roots from the soil, and should be applied so that rainfall or irrigation will carry it into the root zone. Rates of application vary from 1 or 2 to 10 or 12 lb. per square rod.

MCPA and 2,4-D may be translocated downward through the roots of perennial weeds. Rates, in contrast to those for chlorate, are commonly one-half to two pounds an acre, applied in a spray. The compound 2,4,5-T, 2,4-D, with an extra chlorine in the 5-position, is slower in action and more effectively translocated than 2,4-D. It is preferred for controlling brush and some perennials. Great care is necessary when using any of these compounds near susceptible plants, as the slightest drift of the spray, or even the fumes, may cause injury.

The use of a special petroleum fraction known as Stoddard solvent in weeding carrots and other members of the parsley family is a striking example of differential herbicidal action. These plants are resistant while most others are killed by the spray. Other oils are used for various purposes. Moderately heavy oils high in aromatic compounds are useful on such surfaces as ditchbanks and roadbeds.

Miscellaneous Treatments.—Soluble fertilizer materials may be applied in sprays, either to the soil when accurate distribution of small quantities is required, or to the leaves of the plants. Sprays are used extensively to apply iron or manganese to pineapple leaves and less extensively to supply nitrogen—in the form of urea—to apple trees. Defoliant sprays and dusts are used to eliminate leaves from cotton and some other plants before machine harvesting. A cotton defoliant must cause the leaves to drop naturally rather than kill them, since the crumbled dead leaves reduce the grade of the lint. Cyanide dusts and dilute chlorate sprays are among the more successful treatments.

Both toxic compounds like sodium dinitro-*o*-cresylate and hormone type sprays like naphthaleneacetic acid have been used to thin fruits when excessive crops are set. Dinitro kills a percentage of the young fruit and injures the foliage more or less extensively. Hormones cause embryo abortion in young fruits and accentuates natural drop. The same hormone compound may be used later to reduce preharvest drop of nearly mature fruit. Naphthaleneacetic compounds, dilute 2,4-D, or other hormone-type materials may be used to increase fruit set in tomatoes or beans, and to prevent sprouting of potatoes, or leaf and berry drop of holly in Christmas decorations. Wax emulsions are sprayed on nursery stock to reduce water loss during shipping and after planting. New uses for new chemicals, applied as sprays or dusts, are constantly being developed.

See also SPRAYING AND DUSTING MACHINERY.

BIBLIOGRAPHY.—"Insects and Diseases of Vegetables in the Home Garden," *Home and Garden Bulletin No. 46*, U.S. Department of Agriculture (1956) is an excellent brief reference; W. S. Hough and A. F. Mason, *Spraying, Dusting and Fumigating of Plants* (1951) presents a thorough discussion. *See* also L. L. Pyenson, *Elements of Plant Protection* (1951); *Pesticides in Tropical Agriculture*, published by the American Chemical Society (1955); D. E. H. Frear, *Chemistry of the Pesticides*, 3rd ed. (1955); R. L. Metcalf, *Organic Insecticides: Their Chemistry and Mode of Action* (1955); G. H. Ahlgren, G. C. Klingman and D. E. Wolf, *Principles of Weed Control* (1952). Bulletins on spraying and dusting practices are available from most state experiment stations. (W. E. Ls.)

SPRECKELS, a family of U.S. manufacturers, capitalists and financiers who developed large sugar interests in the United States and the Hawaiian Islands. CLAUS SPRECKELS (1828–1908) was born at Lamstedt in Hanover, Ger. Emigrating to the United States in 1846, he operated grocery stores in Charleston, S.C.,

and New York city and a brewery in San Francisco. In 1863, after spending some time in Germany studying sugar manufacturing, he organized the Bay Sugar Refining company in San Francisco with his brother Bernard. He acquired large sugar and shipping interests in California and Hawaii, and came to be known as the "Sugar King." At one time he had a virtual monopoly in the sugar business on the west coast of the United States.

His son RUDOLPH (1872–1958) entered the sugar business at the age of 17. At the age of 22 he and his brother Claus Augustus, in a legal battle with their father and brothers John Diedrich (1853–1921) and Adolph Bernard (1857–1924), won control of the debt-ridden Hawaii Commercial & Sugar company. Rudolph became president of the firm and had it on a paying basis in one year. When he was 26, after making a fortune, Rudolph sold his interests and returned to San Francisco where he became a banker and utility executive. A family reconciliation was effected and in 1905 Rudolph became manager of his father's interests. In 1907 Rudolph contributed $100,000 to finance an investigation of San Francisco city government that resulted in the overthrow of the then current administration. (H. J. SG.)

SPREE, a river of Germany, 247 mi. in length, rising in Saxony, close to the Czechoslovakian frontier, and flowing north past Bautzen, Spremberg and Cottbus, dividing between the first two towns for a time into two arms. Below Cottbus the river splits into a network of channels in the peculiar marshy region known as the Spreewald, then passes Fürstenwalde and Köpenick, threads Berlin in several arms and joins the Havel at Spandau. More than 90 mi. of it are navigable and it is connected with neighbouring rivers by canals.

SPREEWALD, a region of Germany, in the district of Cottbus, a marshy depression of the middle Spree valley, 330 sq.mi. in extent, 37 mi. long and varying from 1 to 7 mi. in width. The Spree river above Lübben splits into more than 200 arms, and in seasons of flood generally overflows portions of this region.

Part of the region is cultivated, part used as pasture land and wooded like a park especially in the lower districts. Fishing, cattle breeding and the growing of vegetables, more particularly small pickling cucumbers, are the chief occupations of the people, some of whom retain their original Wendish language, customs and costumes. Spreewald is a popular summer resort of the Berlin people, and in winter is popular for skating.

SPRINGBUCK (SPRINGBOK; *Antidorcas marsupialis*), a small South African gazelle inhabiting the country south of the Zambezi, but ranging northwestward to Mossamedes. In the more settled parts of Cape Province, the Transvaal and the Orange Free State it now exists only within the enclosures of the large farms. Both sexes carry lyrate horns; the shoulder height of an adult male is about 30 in., and an average pair of horns measures 14 in. along the curve; in the female the horns are more slender. The general colour above is reddish fawn, separated from the white of the underparts by a dark band on the flanks.

SPRINGER, the term given in architecture to the stone from which an arch springs (*see* ARCH AND VAULT).

In some cases this is the stone resting on the impost or capital, the upper surface of which is a plane directed to the centre of the arch.

SPRINGFIELD, the capital city of Illinois, U.S., and the county seat of Sangamon county near the Sangamon river, 99 mi. N.E. of St. Louis, Mo., and 185 mi. S.W. of Chicago, Ill. In the winter of 1818–19 several adventurous settlers located in the area, although John Kelly is acknowledged to be the first to build a cabin (at the northwest corner of Second and Jefferson streets) where the city now stands. Others, generally from North Carolina, Virginia and Kentucky, built near this cabin. In 1821 Sangamon county was created and the little settlement was named temporary county seat; the following year a post office was put into operation. The village was confirmed as county seat in 1825 and incorporated in 1832, and in 1837, with a population of less than 3,000, it was selected as the capital of Illinois.

Those most responsible for the legislative transfer of the capital from Vandalia to Springfield were the Sangamon county representatives, the "long nine" (so called because their aggregate height was 54 ft.). Their leader was young Abraham Lincoln, who came to Springfield from the rude village of New Salem (25 mi. N.W. of Springfield) on April 15, 1837, and lived there until he left for Washington, D.C., on Feb. 11, 1861, to be inaugurated as president of the U.S. Lincoln and the city matured together and the many places identified with his life attract hundreds of thousands of tourists annually.

The old state capitol building, now the Sangamon county courthouse (begun in 1837, completed in 1853), is located in the centre of the business district. There Lincoln served his last term in the legislature, 1840–41, practised before the state supreme court, delivered several of his most notable speeches, including the "House Divided" address, and maintained an office as president elect; and it was there that his body lay in state (May 4, 1865). An unpretentious house at Eighth and Jackson streets, maintained by the state of Illinois as a memorial, is the only home owned by Lincoln. He purchased it for $1,500 in 1844 and in it three of his children were born; the house has been restored and many of the original furnishings are on display.

One and a half miles northwest of the business district is the Lincoln tomb and memorial in Oak Ridge cemetery. The marble burial chamber on the ground floor holds the bodies of Lincoln, his wife Mary and three of their children, Edward Baker, William Wallace and Thomas ("Tad"); their other child, Robert Todd, is buried in Arlington National cemetery. The memorial is 117 ft. tall and is surmounted by a granite shaft. The state of Illinois owns and maintains the structure. The present tomb and memorial is the third erected on this site. The last reconstruction was completed in 1931 with a new interior but without change to the exterior. The First Presbyterian church displays the original pew rented by the Lincoln family. The Wabash railroad passenger station, although converted to freight use, is still standing; there Lincoln delivered his celebrated Farewell address. The locations of Lincoln's three law offices, the Globe tavern, the place where the First Inaugural was written, and other spots of Lincoln interest are marked with bronze plaques.

In the centre and towering over the Illinois state buildings is the statehouse, the fifth capitol building owned by the state. Begun in 1868 and finished in 1888 at a cost of $4,500,000, this massive limestone structure is in the form of a Latin cross and measures 361 ft. in height (to the top of the flagpole, 405 ft.). Extensive interior renovation was done in 1958. The Centennial building, southeast of the statehouse, erected 1918–23 at a cost of $3,000,000, commemorates the 100th anniversary of Illinois statehood; it houses the state historical library, museum, hall of flags, state library and other state offices. In the historical library is a famous collection of Lincoln letters, documents and memorabilia as well as extensive collections concerning the history of Illinois. Immediately west is the state archives building, the third of its type in the U.S.; completed in 1938 at a cost of $820,000, its unique construction provides superior protection to the official, noncurrent records of the state. Immediately northwest of the archives is the state office building, first occupied in 1955 (cost, $11,500,000); within this modern structure is a post office, a cafeteria and the offices of more than 20 state departments. North of the capitol is the state armoury and office building; completed in 1937 and costing more than $1,000,000, it replaced an armoury destroyed by fire in 1934. The supreme court building, adjacent to the Centennial building, is occupied by the two highest Illinois courts and by the attorney general. The governor's mansion, a short distance away, was first occupied by Gov. Joel A. Matteson in 1855 and has been the residence of Illinois governors since.

Springfield is a large wholesale and retail centre for a rich agricultural area; the once significant soft coal industry is no longer important to the city's economy. Principal industrial products include electric meters and electronic equipment, tractors, heavy-duty boilers, castings, automotive equipment, miners' lamps, mattresses and shoes, boxes, paints, livestock feeds, soybean oil and meal and flour and cereal products. City parks total about 1,000 ac. Around Lake Springfield, completed in 1935 to provide water, the city has developed residential and recreational

areas. Pop. (1960) city, 83,271; standard metropolitan statistical area (Sangamon county), 146,539. For comparative population figures *see* table in ILLINOIS: *Population.* (C. C. W.)

SPRINGFIELD, a city of Massachusetts, U.S., on the east bank of the Connecticut river, is the seat of Hampden county and a central city of the Springfield-Chicopee-Holyoke standard metropolitan statistical area (SMSA). This area comprises four cities (Chicopee, Holyoke, Springfield, Westfield) and ten towns (Agawam, East Longmeadow, Hampden, Longmeadow, Ludlow, Monson, Palmer, Southwick, West Springfield, Wilbraham) in Hampden county; one city (Northampton) and four towns (Easthampton, Granby, Hadley, South Hadley) in Hampshire county; one town (Warren) in Worcester county; and one town (Somers) in Tolland county, Conn. Most of these municipalities were part of Springfield when it was incorporated in 1641. Pop. (1960) Springfield, 174,463; Springfield-Chicopee-Holyoke SMSA, 493,999. (For comparative population figures *see* table in MASSACHUSETTS: *Population.*)

Springfield, the financial, commercial and industrial centre of the area, is the third largest city in Massachusetts. Its population has grown at a fairly consistent pace, the 1960 census figure being two and a half times that of 1900. Its area provided ample room for internal population growth so that the surrounding communities did not start growing at a more rapid pace than the core city until the middle of the 20th century.

The metropolitan area covers much of the Connecticut valley lowland that stretches about 30 mi. N. from the Connecticut border and from the Berkshire foothills in Westfield to the eastern uplands in Wilbraham. The Westfield and the Chicopee rivers, two tributaries of the Connecticut, cut the area east and west. The land is fairly flat, although it is broken by a series of traprock ridges which add interest to the terrain.

The region was settled in 1635 by William Pynchon, one of the original patentees of the Massachusetts Bay Colony, and a group from Roxbury, Mass. The patentees moved from what is now Agawam to the east side of the Connecticut river the following year for greater safety from the Indians and they called the community Springfield after Pynchon's home in England. Their leader's autocratic rule ended in 1652 when Pynchon returned to England in the furor caused by his book attacking the Calvinist doctrine of atonement. The settlement was an agricultural and trading community at first; its industrialization was hastened by the building of an arsenal which received congressional endorsement in 1777; by the development of water power, especially in Chicopee and Holyoke, in the early 19th century and by the arrival of the railroad in 1835.

Machinery, electrical equipment and printed goods are the most important products of the highly diversified economy of Springfield. The U.S. arsenal and an armoury, established in 1794, have long been important. More than 800,000 of the famous Springfield muskets were made there during the Civil War. It later became the home of the famous Springfield rifle and the Garand rifle and a principal manufactory of small arms for the U.S. army. The Smith and Wesson revolver factory is located there. G. & C. Merriam company, publishers of Merriam-Webster dictionaries since 1843, has its headquarters in Springfield.

Springfield was incorporated as a city in 1852. Long unique in Massachusetts for its partisan municipal elections and bicameral city council, the city voted in 1959 to change to a strong-mayor, unicameral-council government with nonpartisan elections. Evidence of city planning and of an active cultural life are seen in the so-called municipal group, consisting of the city hall and the municipal auditorium that houses the city's symphony orchestra, and in a nearby quadrangle made up of art and natural history museums, a planetarium and a library. Springfield college, founded in 1885 by the Young Men's Christian association, and American International college, established in Lowell in 1885 as the French-Protestant college and moved to Springfield in 1888, are the two major institutions of higher private education in the city, which has long had a reputation for its excellent schools. The city's extensive park system consists of 2,000 ac. and includes one of the largest zoos in New England. (GE. G.)

SPRINGFIELD, a city of southwestern Missouri, U.S., the county seat of Greene county. Kansas City, 175 mi. N.W., is the nearest larger city. Until the end of World War II Springfield was commercially important chiefly as the centre of a fertile agricultural area, as a distribution point for manufactured products and as the seat of the major shops of the St. Louis-San Francisco railroad. Its rapid growth after that time is attributable mainly to the expansion westward of eastern manufacturing firms. Springfield's population in 1960 was 95,865; that of its standard metropolitan statistical area (Greene county) was 126,276. (For comparative population figures *see* table in MISSOURI: *Population.*)

Springfield was first settled in 1829 by families from Tennessee and Kentucky. The Civil War split friends and families; the merchants and professional men, largely from the southern lowlands, included many slaveowners, while the settlers with rural backgrounds, mainly from the mountain regions, sympathized with the North. Springfield became a strategic point in the war and a battle was fought at Wilson's creek, 10 mi. from the city. In a federal cemetery Confederate and Union graves lie side by side.

There are three liberal arts colleges in the city. Drury (Congregational) was established in 1873, Southwest Missouri State began as a teachers' college in 1905 and Evangel (Assemblies of God) was established in 1953. In addition, there are two college-level religious-training institutions—Central Bible institute (Assemblies of God) and Baptist Bible college.

Besides farm products—chiefly meat, livestock, dairy products, small fruits, eggs and poultry—Springfield produces paper cups, machine belts, portable typewriters, flooring, furniture, truck trailers and men's clothing. Notable institutions are a federal prison hospital and the international headquarters of the Assemblies of God. Springfield is also the seat of a Catholic diocese. Church membership in the city is about 95% Protestant.

Within a 50-mi. radius from Springfield are Ozarks resort areas, with stretches of scenic beauty in between, and the city is a popular stopover for tourists. It is noted for its pleasant tree-lined streets and attractive residential areas. Springfield adopted the council-manager form of government in 1953. Schools were integrated in 1954. (E. R. ST.)

SPRINGFIELD, a city of Ohio, U.S., 45 mi. W. of Columbus and 82 mi. N. of Cincinnati; the county seat of Clark county. Scattered settlements had been made in this region before James Demint built his log cabin on Lagonda (now Buck) creek in 1799. Two years later a townsite was surveyed on the river bottom opposite; the settlement is said to have been named Springfield because of spring water coming down the cliffs bordering the valley. The rapid influx of settlers led the Ohio legislature in 1818 to establish Clark county with Springfield as the county seat. In 1850 it was made a city, its population at that time being 5,108. The population of Springfield in 1960 was 82,723 and that of its standard metropolitan statistical area (Clark county) was 131,440; for comparative population figures *see* table in OHIO: *Population.*

The economic life of the residents rests on the processing of the products of the surrounding farms and the output of over 200 manufacturing establishments. A wide variety of industrial products ensures business stability and well-paid labour.

Educational facilities include Wittenberg university, a Lutheran institution chartered in 1845, Catholic and public schools. The Warder Public library serves all of Clark county. Springfield supports a symphony orchestra. City parks cover over 600 ac.

Ten miles west of Springfield lies the Wright-Patterson air force research and development base, a national centre for defense activity. Many Springfield residents find employment there. (B. H. P.)

SPRING GUN, a device formerly in use against poachers and trespassers. Since 1827 spring guns and all man traps are illegal in England, except as a protection against burglars.

SPRING-RICE, SIR CECIL ARTHUR (1859–1918), British diplomatist, was born in London on Feb. 27, 1859, the second son of the Hon. Charles Spring-Rice (1819–70), and grandson of the 1st Baron Monteagle. Educated at Eton and Balliol college, Oxford, he entered the foreign office in 1882, becoming private secretary to Lord Granville in 1884 and précis

writer to Lord Rosebery in 1885. He went to Washington as third secretary in 1886, and in 1895 proceeded to Berlin. In 1898 he became secretary at Teheran, and in 1901 British commissioner on the Caisse de la Dette in Cairo. In 1903 he went to St. Petersburg (Leningrad), first as secretary and later as councillor of embassy, remaining in Russia till the end of 1905. In 1906 he was sent to Persia as minister, having lately been created K.C.M.G. In 1908 he was created G.C.V.O. and went to Sweden as minister, and in 1912 was appointed ambassador to the United States. He died at Ottawa on his way home to England, Feb. 14, 1918.

See Sir V. Chirol, *Cecil Spring-Rice: In Memoriam* (1919); Stephen Gwynn, ed., *The Letters of Sir Cecil Spring-Rice* (1929).

SPRINGS, a town of the Transvaal, U. of S. Af., lies 28 mi. E. of Johannesburg. Pop. (1960) 135,231, of whom 36,445 were European, 95,911 Bantu, 1,364 Asian and 1,511 Coloured. The rapid growth of the town has precluded monumental building, but the magistrate's court, the municipal theatre (1958) and P. A. M. Brink stadium are noteworthy. The town has been a municipality since 1904 and the seat of a magistracy since 1909. Its life began with the opening of the first coal mine in 1885. In 1908 the mining of gold started at Geduld mine, and the closing in 1913 of the last colliery left gold mining as the sole industry. During the next 25 years Springs became the most highly concentrated gold field in the world. Two of the mines also yield uranium and an extraction plant was officially opened in 1953. Secondary industry has gradually been attracted to the town and factories include those for the making of paper, glass, machine tools, alloy steels, electrical goods and bicycles. Educational training facilities include a branch of the Witwatersrand Technical college and a government miners' training school. Springs has rail and road communication with other important centres including Durban (410 mi.) and Lourenço Marques (393 mi.); Johannesburg is linked by electric railway. Jan Smuts airport, the principal air terminal of the Union, is 18 mi. from Springs by road.

SPRINGS are, in almost all cases, the natural overflow or point of escape from some underground reservoir of water. Their classification may be either according to the geological conditions governing the point of location, or according to the chemical composition of its waters. When considerable chemical impurity is present they are usually termed "mineral springs."

The rocks which constitute the crust of the earth are either permeable or impermeable to water. During most seasons of the year, in temperate climates, a certain amount of rainfall soaks into any crust formed of permeable strata; the part absorbed may be the complete rainfall during winter months, when the air is saturated with moisture, or may fall to near zero during a dry summer, when all the rainfall is returned to the air as evaporation from the surface. That portion of the rainfall which soaks down below the level of plant roots goes to replenish the underground reservoir. The shape of the reservoir varies indefinitely, according to the geological structure of the area; but it is only the shape of its water surface that concerns springs. This surface, known as the water table, divides the fully saturated rocks from those which only hold moisture in their minute pores.

In an area of completely permeable rocks the springs issue at or near the valley bottoms; when, however, the district is made of alternating permeable and impermeable beds, each impermeable bed holds up water on its surface. If the strata are horizontal, small springs may be found all round the outcrops; but if they are tilted or folded the flow of the underground water will be towards the lowest point on the base of the permeable water-bearing bed. At this point the main spring for that local reservoir will be located.

In strata which are, in the main, impermeable, but somewhat brittle the presence of joints and cracks is of prime importance in determining the direction and amount of flow in the underground waters. In jointed rocks the rain water may sink to great depths down one set of joints and rise again along a second, issuing at the surface as a warm or thermal spring; this is the probable cause of the hot springs at Bath. When a permeable bed and an impermeable one are brought into juxtaposition through faulting the flow of water in the permeable one is checked; but since faulting frequently shatters the rock it affords a plane of weakness along which the water will tend to flow. If the water is flowing under pressure due to an overlying impermeable cover, it may reach the surface as an "artesian spring." The water of artesian springs sometimes carries small particles of solid matter in suspension as well as salts in solution. The solid particles are dropped at the point of exit of the spring and may be cemented by the salts deposited from solution. When this takes place a mound is built up, from the summit of which the spring issues. Hence the "mound springs," such as are seen at their best in the great artesian basin of Australia.

Some of the largest springs issue from thick beds of massive limestone. This type of rock is usually well jointed and, being soluble in rain water, the joints and marked bedding planes become enlarged by solution and the whole of the rainfall is absorbed in the mass of the rock and flows underground to issue as large springs. Frequently these springs yield a somewhat hard water of otherwise great purity.

Mineral Springs.—All springs containing noticeable quantities of salts in solution, other than the carbonate and sulphate of lime, are known by this name. The commonest minerals found are common salt, giving rise to "bitter springs"; and iron, sulphur, magnesia, etc., giving "medicinal" waters.

Thermal Springs.—The springs coming under this heading are derived from two sources. Firstly, meteoric waters which have penetrated down to considerable depths and rise again along well-defined fissures issuing as springs at the surface, but with the temperature of the rocks from which they have come; and, secondly, volcanic waters, either in the form of geysers or hot springs (*see* VOLCANO); in these the water may be either meteoric or in part, at any rate, juvenile—that is, water which is issuing at the surface for the first time. Most of these waters contain much mineral substance in solution, which is deposited on cooling and forms marked basins and terraces of sinter, such as the famous pink and white terraces of New Zealand.

(W. B. R. K.)

SPRINGTAIL, any one of a group of small (minute to about $\frac{1}{8}$ in.), primitive, wingless insects of uncertain origin and relationship, so called because many species possess appendages under the abdomen which can be forced down and back, catapulting the insects into the air. They are widely distributed, ranging from the antarctic to the arctic, and are abundant, though often unobserved, in moist plant debris. Certain springtails known as snow fleas are active at near-freezing temperatures and may appear in large numbers on snow surfaces. Although they usually feed on decaying vegetable material or fungi, they occasionally attack living plants, and some species are predatory or feed on animal material. Young, which hatch from spherical eggs, closely resemble the adults.

The lucerne flea causes crop injury in Australia and the garden springtail attacks many cultivated plants in North America. Occasionally springtails are pests of mushrooms, and two species are nuisances in houses. Springtails belong to the order Collembola, which is separated into two suborders, one in which the body is elongate with discrete abdominal segments, the other in which the body is oval or globular, most of the abdominal segments having united with the thorax.

(H. B. M.)

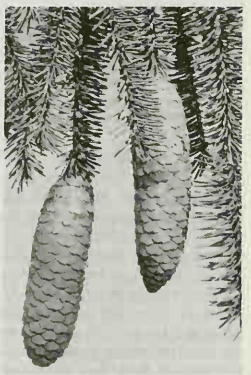
SPRINKLERS, AUTOMATIC. These are a means of combating fire which has the approval of fire insurance companies, who often make a reduction in their premiums when it is employed. The essence of the device is to arrange, under the ceiling of the building to be protected, a series of sprinklers which discharge water as soon as the temperature reaches a certain point. The automatic release of the water is contrived by closing the valves of the sprinklers with fusible metal; the alloy employed being such as to fuse at a temperature of say 150° F. The sprinkler is so devised that the jet spreads the water in a shower, and the distance between the sprinklers is so arranged that every part of the floor is covered. The sprinklers are connected with lines of pipes conveying the water under pressure. In buildings where the water pipes are likely to freeze, the ceiling connections are filled with air under pressure, so that when the valve opens

through the metal fusing, the air escapes and the loss of pressure in the pipes works a lever that opens the valve of the water main. *See* also Fire Protection.

SPRUANCE, RAYMOND AMES (1886–), U.S. naval officer, whose victories in the Pacific theatre during World War II brought him an outstanding reputation as strategist and tactician. He was born in Baltimore, Md., July 3, 1886, and graduated from the United States Naval academy in 1906. Spruance distinguished himself first by his defeat of the Japanese at Midway, June 4–6, 1942. He commanded the 5th fleet in occupation of the Gilbert Islands, invasion of the Marshall Islands, air strikes against Truk, the capture of Saipan, Guam and Tinian, the battle of the Philippine sea and the capture of Iwo Jima and Okinawa. For a short time he served as commander in chief of the Pacific fleet, and then, from 1946 until his retirement, on July 1, 1948, as president of the Naval War college.

He had been promoted to admiral in 1944, and on his retirement congress voted him full pay and allowances for life. Spruance was ambassador to the Republic of the Philippines from 1952 to 1955. (J. B. Hn.)

SPRUCE, an important genus (*Picea*) of evergreen trees in the pine family (Pinaceae), including about 40 species native to the temperate and cold regions of the northern hemisphere. They are pyramidal trees with whorled branches; thin, scaly bark; linear, spirally arranged leaves, each jointed near the stem on a separate woody base; and ovoid to cylindrical pendent cones. The spruces are distinguished from the pines by solitary instead of fascicled leaves, which leave peglike projections on the twig when they fall, and from firs in having pendent, persistent-scaled, instead of upright, deciduous-scaled cones. Spruce trees are subject to severe damage from pests and fire.



ROCHE

NORWAY SPRUCE (PICEA ABIES, SYN. PICEA EXCELSA) BEARING CONES

Seven species of spruce occur in the United States and Canada, 1 in Mexico, 2 in Europe and 29 in Asia. North American spruces have light, soft, compact, straight-grained wood that is strong, easy to work and low in resin content. The wood is resonant and used extensively for sounding boards in pianos and the bodies of violins. It is used also for construction, boats, oars, boxes, food containers, barrels and casks and many other products requiring light, strong, elastic wood. Its chief use has been for pulpwood, 3,000,000,000 to 4,000,000,000 bd.ft. of spruce timber going into pulpwood annually, with cutting exceeding natural replacement. The red spruce (*P. rubens*) is important from Nova Scotia to New York and through the Appalachians to Georgia. Most trees attain heights of 60 to 100 ft. and diameters to 24 in., but specimens 4 ft. in diameter and 120 ft. tall have occurred. The boles are slenderly cylindrical and rise from shallow, widely spreading root systems. Vigorous trees begin to produce seed when 30 to 40 years old, mature in about 100 years and may live 300 to 400 years.

Black spruce (*P. mariana*) and white spruce (*P. glauca*) have almost the same range—from Labrador to New York and the Great Lakes, thence across Canada into Alaska, where they reach the arctic tree line. Both are used for pulp; white spruce produces good lumber, and black spruce is the source of spruce gum. White spruce sometimes reaches a height of 120 ft., but usually is 60 to 70 ft. tall and 18 to 24 in. in diameter. It forms extensive pure stands or is mixed with other conifers, aspen, balsam poplar and birch. The cones of black spruce are purplish, those of white spruce brownish, just prior to maturity.

In the west Engelmann spruce (*P. engelmanni*), the most important spruce lumber tree, attains diameters of 18 to 30 in. (exceptionally to 6 ft.) and heights of 100 to 165 ft. It grows best on loamy soils with high moisture content, but matures slowly. Trees 400 to 600 years old have been reported. It produces large crops of seeds at three- to six-year intervals and the small seeds (average, 135,000 per pound) have a high germination rate, often to 97%. Reproduction by seed is excellent in both humus and mineral soil. Layering sometimes occurs, but trees from such origins usually are stunted. Engelmann spruce ranges from 1,500 to 12,000 ft. in altitude, from British Columbia and western Alberta to Montana, northeastern California and New Mexico. Living timber of Engelmann spruce has been estimated at 63,000,000,000 bd.ft., with half of this reserve in British Columbia.

The blue or Colorado spruce (*P. pungens*) ranges from Yellowstone National park to Idaho, central Utah and southern Arizona, at altitudes slightly below the Engelmann spruce belt. It is smaller than the latter tree and, although of some importance for lumber, the trees grow scattered or in small patches so logging is expensive. Because of the bluish caste of the foliage and its symmetrical growth in open situations it is used extensively as an ornamental. Its deep root system makes it remarkably windfirm and resistant to drought.

The above species have four-angled leaves, but weeping spruce (*P. breweriana*) and Sitka spruce (*P. sitchensis*) have flattened leaves. The weeping spruce, confined to the mountains of northwestern California and adjacent Oregon, has slender, pendent twigs and drooping branches that often reach the ground. The tree has little commercial value. Sitka spruce ranges along the coast from the base of the Alaska peninsula to California and to about 50 mi. inland. It is the largest of the new world representatives, trees 4 ft. in diameter and 200 ft. tall being common; a specimen 16 ft. in diameter and 300 ft. tall has been reported. Most mature trees scale 6,000 to 8,000 bd.ft., but exceptionally large ones have run to 40,000 bd.ft. of merchantable timber. Production of Sitka spruce lumber in Washington and Oregon alone amounted to 106,580,000 bd.ft. in 1948, but after 1950 the cut declined and totaled only 98,150,000 bd.ft. in 1952.

The most important European spruce is the Norway spruce (*P. abies,* syn. *P. excelsa*), native to northern Europe and used extensively in reforestation there and in North America. It approaches Sitka spruce in size, some trees attaining 170 ft. in height and diameters of 5 to 6 ft. The branches have a tendency to droop, the lower ones often touching the ground. Its cones are six to seven inches long, heavy seed producers and confined to the tips of the upper branches. The species ranges from Norway to the Lena river in Russia and from latitude 68° N. to the Pyrenees, the Swiss Alps and the Altai mountains. It furnishes lumber, poles, mining timbers, pulpwood, and limited quantities of material for making stringed instruments. It is planted widely as an ornamental in the old and new worlds. It stands the smoke of cities better than most members of the family although aged trees sometimes are disfigured by blight and insect attack.

Among the old world spruces used ornamentally are the Serbian (*P. omorika*); oriental (*P. orientalis*) of Asia Minor; Himalayan spruce (*P. smithiana*); Alcock spruce (*P. bicolor*) and the tiger's-tail spruce (*P. polita*), both natives of Japan.

Bibliography.—L. H. Bailey, *The Cultivated Conifers in North America* (1933) and *Standard Cyclopedia of Horticulture* (1943); S. J. Record and R. W. Hess, *Timbers of the New World* (1943); W. Dallimore and A. Jackson, *A Handbook of Coniferae* (1948); W. Harlow and E. Harrar, *Textbook of Dendrology* (1958). (I. L. W.)

SPRUE (Psilosis, tropical diarrhoea) is an afebrile, progressively chronic or relapsing disease, characterized by voluminous frothy stools, gastrointestinal distention and discomfort, sore tongue (glossitis), weakness, marked loss in weight and anaemia. It occurs primarily among Caucasians in India, Ceylon, Burma, Indonesia, Indochina, southern China and Korea, the northern coast of South America, Central America and the West Indies, but is rarely diagnosed in Africa. Infrequently it develops in natives of the tropics and in Caucasians in cooler climates with a history of previous residence in the tropics. Sprue is characteristically a disease of middle life and is more common in women than in men.

Aetiology.—Previously regarded by some as of obscure causation and by others as due to a pathogenic yeast in the intestinal

tract, sprue is associated with nutritional deficiency, resulting from (1) inadequate supply of animal proteins, whole cereals and fresh fruits or vegetables containing certain vitamins, iron and calcium salts, together with excess intake of carbohydrates; (2) inability of the intestine to absorb adequate amounts of essential food elements; (3) inability of the stomach to secrete essential components of gastric juice (extrinsic factor of Castle); or (4) a combination of these factors. Previous intestinal disease or some other debilitating condition may be a precipitating or contributory agent.

Symptoms.—Sprue is typically insidious in its onset. The patient develops an inflamed, denuded, extremely sore tongue, copious, frothy, watery stools with associated intestinal tenderness and at times colicky pains, loss of strength and weight. Examination reveals a stool that contains inadequately saponified fats, low blood pressure and at times subnormal temperature, moderate to advanced emaciation, and secondary anaemia (early stage) or pernicious anaemia (later stage). During remissions these symptoms tend to improve, but return with each relapse. The clinical findings as a group distinguish sprue from the usual type of pernicious anaemia, pellagra, chronic colitis, diabetes, tuberculosis and cancer, although any one of these diseases may complicate sprue.

Treatment.—In earlier years treatment consisted of complete rest in bed, frequently on a near-starvation diet, and removal to a more healthful climate. At times cure was effected; all too often, however, after years of invalidism the patient succumbed. An appreciation of the causative factors involved led to administration of liver extract, at times supplemented with iron and purified vitamins A, B-complex and D, and in 1946 to folic acid. The results were frequently dramatic and in most patients adequate treatment over a period of weeks or months resulted in cure, although a daily maintenance dose of 5 mg. of folic acid was advised. Prevention of sprue consists primarily in observance of an adequate, properly balanced dietary and early correction of intestinal disorders.

BIBLIOGRAPHY.—G. G. Lopez, T. D. Spies, J. A. Menendez and R. L. Toca, "Folic Acid in the Rehabilitation of Persons with Sprue," *J.A.M.A.*, 132:906–911 (1946); P. H. Manson-Bahr, "Tropical Sprue and Hill Diarrhoea," in *Manson's Tropical Diseases*, pp. 543–564 (Baltimore, London, 1954); T. D. Spies, "Observations on Treatment of Tropical Sprue with Folic Acid," *J. Lab. and Clin. Med.*, 31:227–241 (1946); R. M. Suarez, "Clinical and Hematological Review of Sprue Based on the Study of 150 Cases," *Ann. Int. Med.*, 12:529–535 (1938).
(E. C. F.)

SPUR, an instrument attached to the heel of a rider's boot for the purpose of goading the horse. (A.S. *spura, spora,* related to *spornan, spurnan,* to kick, spurn.) The earliest form of the horseman's spur armed the heel with a single prick. In England the rowel spur is shown upon the first seal of Henry III, but it did not come into general use until the 14th century. In the 15th century spurs appear with very long shanks, to reach the horse's flank below the outstanding bards. After this time, and until the Restoration, they take many decorative forms, some of which remain in the great spurs worn by Mexican cavaliers. Gilded spurs were reckoned the badge of knighthood, and in the rare cases of ceremonial degradation they were hacked from the knight's heels by the cook's chopper.

In architecture, a spur (Fr. *griffe,* Ger. *Knoll*) is the ornament carved on the angles of the base of early columns; it consists of a projecting claw, which, emerging from the lower torus of the base, rests on the projecting angle of the square plinth; the earliest known example is in Diocletian's palace at Split (Spalato). In Romanesque work the oldest examples are those found on the bases in crypts, where they assumed various conventional forms; being, however, close to the eye, the spur soon developed into an elaborate leaf ornament, which in French 13th-century work and in the early English period is of great beauty.

SPURGE, a name applied to plants known botanically as Euphorbia (*q.v.*). The spurge-flax is *Daphne gnidium,* and the spurge-laurel *Daphne laureola,* both found in Great Britain.

SPURGEON, CHARLES HADDON (1834–1892), English Nonconformist divine, was born at Kelvedon, Essex, on June 19, 1834. In 1852 he became pastor of Waterbeach. His powers as a boy preacher became widely known, and at the close of 1853 he was "called" to New Park Street chapel, Southwark. Exeter hall was used while a new chapel was being erected, but even that could not contain Spurgeon's hearers. The enlarged chapel at once proved too small for the crowds, and a huge tabernacle was projected in Newington causeway. The Tabernacle was opened for service on March 25, 1861. Spurgeon preached at the Tabernacle on Sundays and Thursdays.

His Sunday sermons were sold literally by tons. They are models of Puritan exposition and of appeal through the emotions to the individual conscience, illuminated by frequent flashes of spontaneous and often highly unconventional humour. Collected as *The Tabernacle Pulpit,* the sermons form about 50 volumes. His book of sayings called *John Ploughman's Talks; or, Plain Advice for Plain People* (1869) found many readers. In 1887 his distrust of modern biblical criticism led to his withdrawing from the Baptist union. He died at Mentone on Jan. 31, 1892.

See the *Life* by Schindler (1892) and Spurgeon's *Autobibgraphy,* compiled by his widow and his private secretary from his diary, sermons, records and letters (1897–1900).

SPURN HEAD, foreland of the North sea coast of Yorkshire, Eng., projecting across the mouth of the Humber. From Kilnsea it is 4 mi. long but seldom exceeds 300 yd. wide. Formed of sand and shingle from the rapidly denuding Holderness coast to the north, it is only a few feet above sea level.

SPY, a town near Namur, Belg. Pop. (1955 est. 3,109). There in 1886, in Betche aux Roches cavern, Maximin Lohest and Marcel de Puydt found two nearly perfect skeletons (man and woman) at the depth of 16 ft., with numerous implements of the early Aurignacian period. All the human remains are in the Lohest collection, Liège. The skulls are of the Neanderthal type.

See *L'Homme Contemporain du Mammouth à Spy* (Namur, 1887); G. de Mortillet, *Le Préhistorique* (1900); Sir A. Keith, *Antiquity of Man,* 2nd ed. (1925).

SPY: *see* ESPIONAGE; INTELLIGENCE, MILITARY AND POLITICAL; LAWS OF WAR; WAR.

SPYKMAN, NICHOLAS JOHN (1893–1943), Dutch-U.S. scholar in international politics, was born in Amsterdam, Neth., Oct. 13, 1893. After several years in the near and far east, he came to the United States in 1920, attended the University of California, Berkeley, became professor of international relations at Yale university in 1923 and a U.S. citizen in 1928. Quickly recognized as a teacher of exceptional ability and originality, he approached his subject on the premise that the relations of nations are fundamentally power relations, an approach that aroused controversy at first, but later exerted considerable influence on the teaching of the subject in the United States. He emphasized the impact of geography on politics but rejected the German school of geopolitics (*q.v.*).

He was the organizer and first director (1935–40) of the Yale Institute of International Studies. His most important book, *America's Strategy in World Politics* (1942), exerted some effect upon U.S. defense policy. He died June 26, 1943, in New Haven, Conn.
(F. S. D.)

SPYRI, JOHANNA (née HEUSSER) (1827–1901), Swiss writer whose story for children, *Heidi,* is loved all over the world. Her psychological insight into the child mind, her humour and her ability to enter into childish joys and sorrows give her books their attraction and their lasting value. She was born in Hirzel on Lake Zürich, June 12, 1827; her father was a doctor and her mother a poet. After her marriage in 1852 to Bernhard Spyri, a lawyer engaged in editorial work, she moved to Zürich. Her first book, *Ein Blatt auf Vronys Grab* (1870), was followed by many others. The best known, *Heidi* (2 vol., 1880–81), has been widely translated. Others include *Heimatlos* (1881) and *Gritli* (1882). Johanna Spyri died at Zürich, July 7, 1901. Her love of homeland, feeling for nature, unobtrusive piety and cheerful wisdom gave both her work and her life their unique quality.

See M. Paur-Ulrich, *Johanna Spyri, ein Lebensbild,* 2nd ed. (1940).
(M. P.-U.)

SQUADRON is an English term borrowed from the Italian *squadrone,* which in turn came from the colloquial Latin *exquadra,* an intensive form of the Latin *quadra,* "a square." The *Oxford*

English Dictionary shows the term in English as early as 1562 in the sense of a body of soldiers tactically disposed in the shape of a square. By 1579, however, the term had already come to refer simply to a small compact body of soldiers without regard to a particular formation but understood to suggest a body organized and ready for action; by 1702 it had also obtained a more specific sense, a body of cavalry of from 100 to 200 men. While thus establishing itself as a military term, it was adopted (as early as 1588) in naval usage to designate a certain number of ships that could readily be commanded by a single flag officer. Besides these usages, it served in special contexts where it retained its earliest connotation of something shaped like a square, notably in 17th-century New England as a general term for a district of a town, presumably so used because some of these districts were square in shape. But of all its meanings, the military and naval ones survived the strongest. As a general term it continued to apply to a compact, easily maneuvered body of men, horses or vehicles.

In the U.S. army it refers to a cavalry unit composed of a headquarters and three or four reconnaissance troops—three troops in an infantry division, four in an armoured. In the U.S. navy it applies to one or more divisions of warships or aircraft. In the U.S. air force it may designate a unit at a level of command immediately below, and usually subordinate to, a group, the unit being either a flying or a nonflying unit, or it may refer to a formation of aircraft, usually six or more. In the Royal Air Force a typical fighter squadron consists of from 12 to 24 aircraft, the crews and the maintenance mechanics, all under a squadron leader responsible for ready tactical action, with administrative services provided by the base at which the squadron is stationed. The term is often combined with a designating attributive, as in medical squadron or communications squadron, in each case the squadron being specially organized for the designated purpose.

(Wd. A. H.)

SQUALL AND SQUALL LINE. A squall is a strong wind that sets in suddenly, lasts for several minutes then dies away somewhat less suddenly. Any squall includes a succession of gusts, or irregular fluctuations of wind speed at intervals of a few seconds to a minute; an individual gust may be half again as strong as the average squall wind speed.

Wind speeds in squalls commonly reach 30 to 60 m.p.h., and gusts of 80 to 100 m.p.h. or more occur with the more violent squalls, although gusts of this strength are rare in most localities. Usually, squalls are straightaway winds, although some types of rotary winds are called squalls (*e.g.*, "typhoon squall," a form of waterspout several hundred yards in diameter).

It is common practice to name a squall after the weather phenomenon which accompanies it: rain, snow, hail, thunder, etc. The most spectacular common type is the thundersquall. As a thunderstorm approaches, the wind often dies down, then blows gently toward the storm; finally, as the rain is almost at hand, a violent wind abruptly sets in, blowing outward from the storm and usually lasting for only a short time. (*See* also Thunderstorm.)

A squall of this kind is caused by the spreading out of cold air formed in the rain area of the thunderstorm. A thunderstorm, as a whole, is warmer than the surrounding air at a given level. In the heavy rain region, however, a certain amount of relatively dry air, drawn into the thundercloud from the outside, causes partial evaporation of the rain. This evaporation cools the air, and a cold downdraft forms because the air is then more dense than the surroundings, and sinks in the same way as a heavy liquid will sink when poured into a lighter one. As the downdraft approaches the earth's surface, it is forced to spread out horizontally; the squall occurs near the edge and for a short distance within the boundary of the spreading cold air. If the air below a thundercloud is very dry, all rain may evaporate before reaching the ground, and the squall may be dry, often with dust or sand being picked up from the surface. This is a common cause of desert sandstorms.

Local rain squalls of the same nature are very common in the oceanic doldrums, and several may be visible simultaneously from a ship. The wind squalls are often evident from whitecaps traveling over the sea surface.

A different type of squall occurs with the drainage of cold air down steep slopes. Such squalls derive their force from gravity, and depend on the descending air being colder and thus more dense than the air it replaces. The outstanding example of this type of wind is the bora, an extraordinarily fierce wind (intermittent squalls of 100 to 130 m.p.h.) observed during the winter when cold air originating in Russia spills over the mountains of Yugoslavia to the relatively warm coast of the Adriatic sea. Squally fall winds of the same nature are common on mountainous coasts in high latitudes, where cold air from glaciated highlands spills down fiords (Norway, Greenland) or deep valleys (*e.g.*, "williwaws" of Alaska). The force of the wind is increased where the air is funneled between narrow valley walls.

(*See* also Wind.)

Squalls of a less local character occur with the passage of cold fronts (*see* Meteorology), where large masses of cold air replace warm air. Although the squall is likely to be most pronounced just when the front passes, high winds with gusts are likely to be sustained for several hours. Special names for this type of wind include "norther" (Great Plains, North America); "pampero" (Argentina, Uruguay); and "southerly burster" (Australia).

Squall Line.—A squall line (sometimes, line squall), as the name suggests, is a line along which a number of places simultaneously experience squalls. At one time, this term was used to designate the cold front of a cyclone, but in modern usage applies only to a line of squalls associated more or less directly with thunderstorms. Such a squall line may be several hundred miles long, but with gaps. The line is usually irregular, consisting of several large clusters of thunderstorms, in some cases separate, but more generally connected by relatively inactive cloud masses or sheets formed from old, decadent thunderstorms.

The first sign of an approaching squall line may be layer clouds, alto-cumulus and dense cirrus types, blown out ahead of the thunderstorms by the strong winds in high levels. The squall line when near at hand may appear as a solid wall of cumulo-nimbus cloud, with low ragged clouds beneath and sometimes an arch or straight line of low cloud with an apparent rolling motion. Most generally, however, the main solid wall of cloud is obscured from a ground observer by general low cloudiness, and the only clear evidence of a squall line approaching may be the line of heavy rain, dark low ragged clouds and lightning.

Passage of a squall line is marked by a sudden onset of strong gusty winds, with a marked change in wind direction and a rapid drop in temperature (perhaps 10° or 15° in as many minutes), followed by heavy rain, thunder and lightning, and often hail. Rain amounts of one to two inches in an hour, with one-half inch in the first ten minutes, are not unusual. On the average, rain falls for about 45 minutes, with a gradual decrease in intensity after the first rain burst. Arrival of the squall is also signaled by a sudden increase in barometric pressure, often amounting to one-tenth inch mercury or more, followed after a few minutes by a more gradual fall of pressure; at the same time, the wind gradually returns to what it was before the squall line passed.

Since there are gaps between active thunderstorms in a squall line, a considerable variation in weather conditions is to be expected at different points. Thus a squall line passing two towns, 10 or 20 mi. apart, may cause only weak squall conditions at one and violent winds, heavy rain and hail at the other.

The cumulo-nimbus clouds with squall lines generally extend to heights of five to ten miles, and present a considerable hazard to aircraft because of the strong and turbulent vertical air currents within them. Increasingly accurate weather forecasting, together with pinpointing of the thunderstorms by radar and radio direction-finding (static) devices, however, enable aircraft either to circumnavigate squall lines or to fly through the inactive parts.

A squall line may be found at a cold front, but ordinarily either forms ahead of or moves out ahead of the front under the influence of strong winds in upper levels, so that on the average squall lines are found 50 to 200 mi. ahead of the cold front, on the southeast side of a cyclone. In middle latitudes, squall lines most commonly approach from the west or northwest (at a speed averaging 30–35 m.p.h.), while in low latitudes, for example in west Africa,

they commonly approach from the east.

Squall lines are most frequent during the spring months, in localities occupied by moist tropical air masses but near the zone of the polar front which separates the tropical from the cooler air masses of higher latitudes. The greatest number of squall lines is observed in the Great Plains and middle western United States, a region open to frequent incursions of both tropical and polar air masses.

Phenomena of the same type are observed in eastern India-Pakistan ("nor'westers") and west Africa ("tornadoes"; not rotary storms as are American tornadoes), which are not associated with air mass fronts in the same way as in higher latitudes.

Although the strong winds occurring with squall lines are responsible for considerable damage, most of the damage reported is due to hail and tornadoes occurring with the squall lines. Most tornadoes are associated with squall-line type thunderstorms, and as many as 20 to 30 or even more have been reported with a single squall line (*e.g.*, May 2 and 3, 1956, United States). Damage from one squall line may run into millions of dollars.

The cause of the squall winds with a squall line is in general the same as discussed above in connection with thundersqualls. The original formation of squall lines, as of thunderstorms in general, is most favoured when the air is unstable, having a strong decrease of temperature with height (*see* METEOROLOGY), and also when the air is very humid in the lowest few thousand feet and relatively dry higher up. These conditions are best satisfied in moist tropical air during mid-afternoon when surface temperatures are warmest, and there is a strong preference for squall lines to form at that time.

Once a squall line has formed, usually by some kind of lifting of the air near a cold front which sets off the first thunderstorms, it seems to carry with it its own self-regenerating mechanism. A mass of rain-cooled air beneath the thunderstorms spreads out and lifts the warm moist air ahead of the squall line, forming new thunderstorms which become incorporated into the squall line, and these repeat the same process. In this way, squall-line thunderstorms are able to keep going for many hours even during the night. Six to 12 hours is a common duration, but some squall lines have been observed to continue for as long as 36 hours.

Because it is blown out ahead of the cold front by the strong winds in upper levels, a squall line after 12 or 24 hours may be found several hundred miles away from a cold front near which it started, and a new squall line may have been initiated near the same cold front.

BIBLIOGRAPHY.—Interesting examples of squalls are given by Sir William Napier Shaw, *Manual of Meteorology*, vol. ii (*see* especially pp. 374–380; 429) (1942). Detailed discussions of the mechanism of thundersquall formation are to be found in *The Thunderstorm*, United States Weather Bureau (1949). For complete discussions of squall lines, *see* Sverre Petterssen, *Weather Analysis and Forecasting*, 2nd ed., vol. ii (1956), and article by J. R. Fulks, "Instability Lines," in *Compendium of Meteorology*, American Meteorological Society (1951).
(C. W. N.)

SQUARE. In geometry a square is a plane rectilinear figure having four equal sides and four equal angles, each angle being a right or 90° angle. Such a figure is a special kind of rectangle (equilateral) and parallelogram (equilateral and equiangular). It has four axes of symmetry, and its two finite diagonals (as in the case of any rectangle) are equal. If the side of a square is s, the area is $s \times s$, or s^2, the name "s squared" being derived from this geometric relation. Hence the algebraic square is the product arising from multiplying any algebraic expression by itself; *e.g.*, $(a + b)^2 = (a + b)(a + b) = a^2 + 2ab + b^2$ (*see* BINOMIAL FORMULA). Latin writers called the side of a square its *latus* (side), and hence the square root of a number was often called the *latus*.

The word also is applied to instruments used for drawing or testing right angles, as in the T square and the familiar carpenter's square.

SQUARE DANCE (QUADRILLE), a definitive term designating the most popular and widely known type of American folk dances to distinguish it from those types classified as contra, or longways, and circle dances. Historians trace its origin to both the Kentucky running set of English derivation and to the cotillion, a stately French dance in square formation, popular at the court of Louis XV but supplanted later by the quadrille.

The americanized quadrille or square dance begins and progresses rapidly in beautifully ordered patterns within the framework of a relatively compact square, sets of four couples forming its four sides. To the traditional accompaniment of accordion, banjo, fiddle and guitar and to prompting, patter and singing calls, couples perform a variety of steps, all based on a smooth, "shuffling" walk. Formerly danced in five main figures, the contemporary square dance is comprised of three. *See* also QUADRILLE.

BIBLIOGRAPHY.—B. A. Botkin, *A Treasury of American Folklore* (1944); Anne Schley Duggan *et al.*, *Folk Dance Library* (1948); Gene Gowing, *The Square Dancers' Guide* (1957); Richard Kraus, *Square Dances of Today* (1950); Grace L. Ryan, *Dances of Our Pioneers* (1926, 1939); Curt Sachs, *World History of the Dance* (1937); Lloyd Shaw, *Cowboy Dances* (1939).
(A. S. D.)

SQUASH is the fruit of *Cucurbita maxima* and certain varieties of *C. pepo* (*see* PUMPKIN). Summer squash is a quick-growing, small-fruited, nontrailing or bush type of *C. pepo*. The plants are upright-spreading, 1½ to 2½ ft. high and produce a remarkable diversity of fruit forms, ranging from flattened through oblong to elongate and crooked. Colours range from white

J. HORACE MCFARLAND CO.

SQUASH (CUCURBITA PEPO) SHOWING BLOSSOMS AND FRUIT

through cream to yellow, also green and variegated. The fruit surfaces or contours may be scalloped, smooth, ridged or warty. The fruits develop very rapidly and must be harvested only a few days after they form, before the seeds and rinds harden. They must be used soon after harvest. The entire fruit is usually cut into pieces, boiled and served with butter. The bush forms are grown about two feet apart in rows three to four feet apart. To obtain early yields, seedlings are often started in pots in frames or greenhouses and transplanted into the field after three or four leaves have developed and after danger of frost is past.

Winter varieties of squash, *C. maxima*, are long-vining, generally large-fruited, long-season kinds that can be stored many months if kept dry and well above freezing. The fruit stems are greatly enlarged next to the fruits and the fruits show a wide range of sizes, shapes and colours. They are higher in nutritive value than summer squash or so-called pumpkins, and are commonly baked or made into pies. They are extensively grown in the United States in home gardens, market gardens and on a large field scale for canning. Squash is native to the new world, where it was widely cultivated by the Indians long before European settlement.
(V. R. B.; X.)

SQUASH RACKETS, a game played on exactly the same principle as rackets (*q.v.*) but in a smaller court and with a ball made of India rubber. The rules are in most particulars similar to those of rackets but the scoring is different. In matches in England, hand-in only can win an ace. The rubber is the best of five

games of 9 up. If, however, the score becomes 8 all, the player who reached 8 first has the option of a set to two. In the United States a game consists of 15 points up, and an ace is scored for each rally irrespective of which player is serving. The rotation of service in England is just as in rackets. In the U.S. game, the winner of one rally serves in the next.

The Court.—Squash rackets can be played in courts of many sizes and descriptions, both covered and uncovered, with three or four walls and made of concrete or wood. The first standardization of a court in England was made by the Tennis and Rackets association in 1911. Shortly after World War I the association issued revised standard dimensions, and the great majority of British courts were built to those dimensions thereafter. The standard court has four walls and can either be covered or uncovered. In length it is 32 ft. and in breadth 21 ft. The front and side walls are 15 ft. in height and the back wall is 7 ft. in height.

ENGLISH, BENT ASH FRAME, SQUASH RACKET

The floor in a closed court should be made of wood; in an open court, of cement. In a closed court the front wall should be of composition. The markings on the floor and front wall of the court are exactly similar to those in a rackets court. On the front wall, the top of the play line is 19 in. from the floor and the service line is 6 ft. from the floor. The short line is 14 ft. from the back wall. The service boxes are 5 ft. 3 in. square. In the space between the play line and the floor is generally placed a strip of tin or some other metal in order to differentiate clearly in sound when the part of the front wall "in play" or "out of play" is hit. The lines on the wall and floor are painted red. Maple is considered a suitable wood for the floor. The courts should be lighted from the roof, and nearly all modern courts are fitted in addition with artificial lights for evening play.

The U.S. standard court is of different size from the English, being considerably narrower, 32 ft. by 18 ft. 6 in. The racket is similar to that used in the game of rackets, but shorter in the handle. The ball is of black rubber. Many different forms of ball have been tried, and formerly, when there were no standard courts, the ball which seemed best suited to the particular conditions was generally used. For the modern standard English court, the association adopted three standard balls. The ball used in the U.S. is larger than that in England.

Competition.—*Great Britain.*—Open competitions at squash rackets were unknown in England before World War I. In the following years, however, several were started, the most important being the London interclub tournament (teams of three) for the Bath Club cup; the amateur championship, the Royal Navy and army championships and the women's championships.

The British professional champion in the late 1930s was J. P. Dear. There was no formal competition in professional and amateur squash rackets during World War II. After the war, one of the great professional champions was Hashim Khan of Pakistan, acclaimed as one of the world's leading players. F. D. Amr Bay of Egypt created a sensation in London in the early 1930s and beat the best amateurs and professionals for several years.

United States.—The first courts constructed in the United States, as far as is known, were at St. Paul's school in Concord, N.H., in the early 1880s. The development of the game was more rapid in the United States than in England, and it achieved great popularity in the late 1920s and 1930s. Many courts were built in the large cities, particularly in New York, Boston and Philadelphia. The game is also played in Canada. British teams visited America after 1924, and U.S. teams also toured in England. Among the leading amateur champions were Stanley W. Pearson, W. Palmer Dixon, Myles P. Baker, Herbert N. Rawlins, Jr., and J. Lawrence Pool in about the second and third decades of the 20th century. Beekman Pool, Neil J. Sullivan, Donald Strachan and Germain Glidden played in the 1930s and A. Willing Patterson and Charles Brinton in the early 1940s. They were followed by Stanley Pearson, Jr., Hunter Lott, Edward Hahn, Harry Conlon, Ernest Howard, Henri Salaun and G. D. Mateer, Jr.

In 1954 the first U.S. open championship was held on the courts of the University club in New York city. Hashim Khan, who was playing on U.S. courts for the first time and who never had been beaten in competitive squash rackets, lost to Henri Salaun of Boston, Mass., in the final round. Later Hashim Khan's cousin Roshan Khan emerged as a fine player. In 1933 an international team match for the Wolfe-Noel cup was started between British and U.S. women.

Other Countries.—The French Squash Rackets association was formed in 1926. Outside Europe and the U.S., courts were found in several parts of the British Commonwealth, especially in India and the far east and in Australia, New Zealand and South Africa.

Squash Tennis.—Another game of squash, known as squash tennis, is played in the United States. It had a bigger following than squash rackets until the rise in popularity of the latter game in the late 1920s, after which, since squash tennis is not played at the colleges, the number of players dwindled. By the 1960s it was largely extinct, being kept alive by only a comparative handful of players at a few university clubs in New York city.

Squash tennis is played in the same court as is squash rackets, a combination court for the two games having been adopted in the 1930s; there are minor differences, including service-line markings. In squash tennis a modified lawn tennis racket is used and the ball is inflated, traveling with great speed around the walls, at times rebounding from the front wall to the rear wall and even to the front wall again.

The foremost amateurs at squash tennis have included Fillmore Hyde, Thomas Coward, Harry Wolfe, Rowland Haines, Philip Moore, Milton Baron and Edward Larigan. Walter Kinsella, Frank Ward and Tom Iannicelli have been the leading professional players. Norman Torrance was president of the National Squash Tennis association for many years.

BIBLIOGRAPHY.—Walter Debany, *Squash Racquets* (1950); D. G. Butcher, *Introducing Squash* (1948); Gerald Pawle, *Squash Rackets* (1951); Susan Noel, *More About Squash Rackets* (1948); Allison Danzig, *The Racquet Game* (1930). (E. B. N.; A. DA.)

SQUIBB, EDWARD ROBINSON (1819–1900), U.S. physician and pharmaceutical manufacturer, was born in Wilmington, Del., July 5, 1819. He got his M.D. at Jefferson Medical college, Philadelphia, in 1845 and, although a Quaker, joined the U.S. navy as assistant surgeon in 1847. Assigned to Brooklyn Naval hospital after four years at sea, he proposed making basic pharmaceuticals to replace inferior shipboard drugs. For the navy he first distilled pure ether with steam, a process still in use. In 1858 he opened civilian laboratories in Brooklyn, N.Y. Through professional societies and revision committees of the *United States Pharmacopoeia*, he waged constant war on quacks, nostrums and adulteration. He wrote the first pure food and drug laws for New York and New Jersey (1879–80) but his fight for a federal law was a losing one until six years after his death in Brooklyn, Oct. 25, 1900. His sons Edward H. and Charles sold control of the firm of E. R. Squibb & Sons in 1905.

See Lawrence G. Blochman, *Doctor Squibb: the Life and Times of a Rugged Idealist* (1958). (L. G. BN.)

SQUID, the popular name given to a large number of cephalopod molluscs which include (*e.g.*) the common squids of the Atlantic (*Loligo, Ommastrephes*), the giant squids (*Architeuthis*) and the fire squids (*Lycoteuthis, Abralia*). They are exclusively marine animals of world-wide distribution and form an important part of the population of marine swimming and floating animals.

A typical squid, *e.g., Loligo vulgaris,* the common squid of the northeast Atlantic and Mediterranean, has an elongate slender body (visceral sac) edged by triangular fins, a short square head provided with well-developed eyes and ten arms. On the undersurface of the latter are arranged rows of suckers which are strengthened by tough horny rings. Two of the arms (tentacles) are longer and more mobile than the other eight (sessile arms) and the suckers are concentrated at the extremity as a "hand."

The long tentacles are used for capturing prey, the sessile arms for transferring the latter to the mouth and grasping it while it is being crunched by the horny jaws which are situated around the mouth in the centre of the circlet of arms.

Situated within the muscular tissues of the "mantle" is a long horny structure, the "pen" or gladius. This is the rudiment of a shell and consists actually of a secondary growth (the proostracum) which has replaced the true shell-rudiment. In the genus *Ommastrephes* there is a vestige of the original shell.

The majority of the squids are active, highly mobile animals of aggressive habits and are probably all carnivorous.

Originally the squids were distributed among the two "tribes," Oigopsida and Myopsida, of the ten-armed Cephalopoda (Decapoda). Recently, however, Naef and other students have proposed a more suitable classification of the Decapoda according to which the squids are all placed in a single suborder, the Teuthoidea.

These forms are divisible into two groups originally called by C. Chun the Oigopsida libera and Oigopsida consuta. The first include the large, active and freely swimming forms (*Architeuthis, Loligo, Ommastrephes,* etc.). The second comprise a smaller assemblage of curious and highly modified forms which are less common than the Libera and are very largely planktonic. The distinction between these two groups is based on the mode of attachment of the ventral edge of the mantle to the head. In the Libera this attachment is effected by a stud and socket junction (*see* CEPHALOPODA) between the head and the edge of the mantle and can be freed from the head when necessary. In the Consuta the ventral edge of the mantle is permanently fixed to the head.

The Libera are represented by about 65 genera. Among them *Architeuthis* is the most striking because of its size. The genus includes the largest living invertebrate animals, *Architeuthis princeps* of the north Atlantic attaining a maximum record length of 52 ft. (including the outstretched tentacles). *Sthenoteuthis,* a group of smaller forms, which attain a size of over seven feet, may be also reckoned as "giant squids."

Another interesting group are the phosphorescent squids, *Nematolampas, Thaumatolampas* (*Lycoteuthis*) and *Abralia,* which bear light-organs in certain regions of the body (on the mantle, arms, inside the mantle-cavity and around the eyes). Certain of these when seen alive are very beautiful. *Ommastrephes bartrami,* the so-called "flying squid," is found throughout tropical and temperate seas. Its flight is, strictly speaking, a series of leaps across the surface of the sea which are often strong enough to land it on the deck of a ship. This feat is no doubt rendered easier in high seas. The Consuta are mainly small animals. Their fins are largely reduced and the mantle is swollen and tunlike in *Cranchia, Bathothauma,* etc. In the latter, and in *Sandalops* and *Corynomma* the arms are much reduced and the eyes are situated on stalks. Unlike the larger swimming squids these animals probably float largely at the mercy of the currents and are thus to be reckoned as "plankton." (*See* CEPHALOPODA.)

See G. Pfeffer, *Die Cephalopoden der Plankton-Expedition* (1912); C. Chun, *Wiss., Ergebn. d. Deutsche Tiefsee Expedn.* Bd. xviii (1910); G. C. Robson, *A Monograph of the Recent Cephalopoda,* vol. 1 (1929) and vol. 2 (1932). (G. C. R.)

SQUILL, the name under which the bulbous root of *Urginea* (*Scilla*) *maritima* (lily family) is used in medicine. Squill has been used as a fluid extract, sirup or vinegar for its digitalislike action on the heart, and also as an expectorant and as a diuretic. Use of these galenical preparations became virtually obsolete as their absorption from the intestine is uncertain. Two glucosides, Scillaren-A and Scillaren-B, have been isolated, the former crystalline, the latter amorphous. Scillaren-B and a mixture of the two, scillaren, are available for clinical use, though not officially recognized. Red squill is frequently used as a rat poison.

W. H. HODGE

RED SQUILL (URGINEA [SCILLA] MARITIMA) SHOWING THE POISONOUS SCALES WHICH ARE USED AS RATICIDE

The medicinal squill is a native of the countries bordering the Mediterranean, and grows from the sea level up to an elevation of 3,000 ft. The bulbs are globular and often weigh more than four pounds. Two varieties are met with, white squill and red squill. They are collected in August, when leafless, the membranous outer scales being removed and the fleshy portion cut transversely into slices and dried in the sun.

An allied species, *Urginea indica,* is used in India in the same manner as the European species. There are 80 species or more of the true squills in temperate regions of the old world. Several species are grown in gardens, notably *S. bifolia* and *S. sibirica,* originating from the Caucasus region and southwest Asia. *S. nonscripta* or English bluebell and *S. hispanica* or Spanish bluebell are also widely cultivated. (F. O. K.)

SQUINCH, in architecture, a general term for several means by which a square or polygonal room has its upper corners filled in to form a support for a dome (*q.v.*): by corbeling out the courses of masonry, each course projecting slightly beyond the one below; by building one or more arches diagonally across the corner; by building in the corner a niche with a half dome at its head; or by filling the corner with a little conical vault which has an arch on its outer diagonal face and its apex in the corner. The arched squinch seems to have been developed almost simultaneously by the Roman builders of the late imperial period and the Sassanians in Persia.

Squinches are often found both in Romanesque architecture (*q.v.*) and Islamic architecture. In Italy the form is either the conical type as in the church of S. Ambrogio at Milan or as a succession of arched rings as in the 13th-century central tower of the abbey church at Chiaravalle Milanese. More complex forms with niches and colonnettes are characteristic of the French Romanesque of Auvergne, as in the cathedral of Le Puy en Velay (late 11th and early 12th century); the allied churches of the southwest coast, such as St. Hilaire at Poitiers, use conical squinches of the Italian type. Islamic architecture, borrowing from the Sassanian precedent, makes great use of squinch forms, particularly in the Syrian, Egyptian and Moorish phases; the stalactite work, which is so marked a feature of later Islamic architecture, is, in essence, merely a decorative development of a combination of niche squinch forms. In Gothic architecture (*q.v.*) squinch arches are frequently used on the insides of square towers to support octagonal spires. *See* BYZANTINE ARCHITECTURE; ISLAMIC ARCHITECTURE; ARCH AND VAULT.

SQUINT is a turning of one eye or the other inward (convergent squint, cross-eyedness); outward (divergent squint, walleyedness); or, more rarely, upward or downward (vertical squint). A squint may always affect the same eye (monocular, unilateral squint) or it may affect one eye, then the other (alternating squint). It may arise at any time of life as a result of paralysis of one or more of the muscles of the eye (paralytic squint). Cross-eyedness, the commonest form of squint, occurring in early childhood or infancy, is generally not of paralytic origin; it tends to be familial, and often results from an undue strain of the focusing mechanism (accommodation) of the eyes required by an optical error (farsightedness). However, a squint may have many other causes, including mechanical ones and faults in the neural mechanisms that control the position of the two eyes relative to each other. It is essential for the normal functioning of the two eyes together that they be properly aligned during the early years of life.

If a squint has established itself, the child will neglect the vision in one eye to avoid the double vision that results from improper alignment of the eyes, and will develop also other faulty visual habits. These habits frequently make it extremely difficult or impossible in later life to restore adequate binocular vision. Deep-seated loss of vision may occur in an eye that is allowed to be turned all the time, though the eye is otherwise normal.

Children do not outgrow a squint. To allow normal development of visual function and to prevent loss of vision, it is essential

that medical advice be sought as soon as a squint is observed. No child is too young to be examined or to have some form of treatment. Such treatment may be by eyeglasses, patching of one eye, exercises, locally applied drugs or surgery, or by a combination of any of these means. (H. M. Bu.)

SQUIRREL, any rodent of the family Sciuridae, but sometimes restricted to the familiar arboreal species. Squirrels are small or medium-sized mammals with strong hind legs and well-developed tails, which in tree-dwelling forms are long and bushy and serve as balancing organs or, wrapped about them, for warmth and protection. There are four toes on each front foot, five on each hind. Each side of the jaw has four or five cheek teeth above and four below; these teeth are rooted and their crowns have low cusps adapted for crushing food—nuts, seeds, grasses, tubers, roots and fruits. Internal cheek pouches provide space for temporary storage of large amounts of food, which may be carried elsewhere or may be stored in a cache for later use. Most squirrels are active in the daytime, the year round. Flying squirrels, however, are nocturnal. Ground squirrels may hibernate in cold climates.

JOHN N. GERARD

GRAY SQUIRREL (SCIURUS CAROLINENSIS)

Squirrels as a group are nearly cosmopolitan, occurring in all major land masses except Australia, the extreme polar regions and such islands as Madagascar, New Zealand, New Guinea and Hawaii. They are particularly numerous in India and the tropics. More than 40 living genera and countless species are recognized; fossils are known from the Miocene of Europe and North America. Throughout their wide range squirrels have become modified in structure and habits for life in many different ecological situations; they are good examples of the evolutionary process of adaptive radiation.

Most familiar are the tree squirrels of forested regions. In North America the gray squirrel (*Sciurus carolinensis*) prefers heavily forested lowlands, while the larger fox squirrel (*S. niger*) favours open forests of the uplands; but both species have adapted well to urban habitats and are familiar tamed residents in city parks. Each species stores small amounts of food, chiefly hardwood nuts and seeds, in widely scattered caches. Winter nests are carefully made weather-proof dens in holes of trees or tightly constructed twig and leaf nests high in trees; summer nests are large, flimsily made balls of leaves and twigs. Mating occurs in midwinter, and the young, usually two or three, are born about 45 days later. A second mating takes place in late spring. The smaller red squirrels, or chickarees (*Tamiasciurus*), which occupy a wide variety of woodland habitats but are especially abundant in coniferous forests, store food in one or two large reserves, usually in a damp place, and in numerous smaller units. Usually five or six young are born about 40 days after mating.

The common old world squirrel (*Sciurus vulgaris*) ranges from Norway to Japan, with numerous colour variants in different parts of its range. In Great Britain the common squirrel has been largely replaced by the gray squirrel, which was introduced from North America.

The pigmy squirrel (*Myosciurus*) of west Africa and the dwarf squirrel (*Nannosciurus*) of southeastern Asia are the size of a mouse, while the black and yellow giant squirrel (*Ratufa*) of the Malaya archipelago reaches the size of a house cat. The striped palm squirrels (*Funambulus*) are the commonest species in India. The oil-palm squirrels (*Protoxerus*) and the isindi (*Heliosciurus*) occur in Africa. These tree squirrels are extremely agile and adept at climbing and leaping; they not only ascend but also descend trees head first with ease.

Closely related are the flying squirrels (*q.v.*), which possess furred membranes connecting the fore- and hind limbs, and the ground squirrels (*q.v.*), in which the tail is usually shorter and less bushy and the forefeet are better developed, with larger claws. Adept at burrowing, ground squirrels nest beneath the ground; many of them can climb but seldom do so.

Both albinistic and melanistic individuals are common in American gray squirrels. At least three well-marked colour phases are recognized in the fox squirrels. The complex of Asiatic squirrels is an intriguing zoological puzzle. The sharp division of range between the tassel-eared Kaibab squirrels (*Sciurus kaibabensis*), confined to the Kaibab plateau on the north rim of the Grand canyon in Arizona, and Abert's squirrel (*S. aberti*), on the south rim, is an interesting ecological problem for the zoogeographer.

Squirrels sometimes damage crops and occasionally feed on eggs of desirable birds. Certain species carry diseases transmissible to man, among them the plague, rabies and tularemia. Larger squirrels are often hunted for their palatable flesh and for their pelts, which are used chiefly for trimming. In American folklore one of the ground squirrels, the ground hog or woodchuck (*q.v.*), is a weather prophet (*see* Ground-Hog Day).

See Rodent; *see* also references under "Squirrel" in the Index. (R. H. Ma.)

SQUIRRELFISH (*Holocentrus*), a genus of small (less than a foot long), tropical, marine fishes, numerous about coral reefs. They have notably large eyes, a strong backwardly directed spine at the corner of the preopercle (on the lower side of the head), a long spinous dorsal fin contiguous with a shorter one of soft rays behind it, a short anal fin with one of its initial spines notably long and strong, a forked caudal (or tail) fin and the body covered with firm scales of moderate size. Their colours are bright—red and yellow—and lengthwise stripes on the body predominate. Those with larger mouths (*Flammeo*) are sometimes separated generically from those with smaller mouths (*Holocentrus*). About 12 species of squirrelfishes inhabit the waters of Indonesia; 9, some of which are the same, those of Hawaii; about 7 (of which *Holocentrus ascensionis* is the commonest), the West Indian region. (J. T. N.)

BY COURTESY OF MIAMI SEAQUARIUM

SQUIRRELFISH (HOLOCENTRUS ASCENSIONIS) COMMON TO THE WEST INDIAN REGION

SQUIRREL MONKEY, a small South American monkey of the genus *Saimiri;* three species are known, whose collective range extends from Costa Rica to Bolivia and Brazil. The squirrel monkeys belong to the Cebinae (*see* Capuchin Monkey), from the other members of which they differ by their practically nonprehensile tails, smaller size, large eyes and the backward prolongation of the hinder part of the head.

See Primates.

SQUIRRELTAIL GRASS (*Hordeum* [*q.v.*] *jubatum*), a weedy wild barley, native to Europe, Asia and western North America and naturalized extensively in the eastern U.S. and Canada, often becoming a troublesome weed. It is a perennial, with slender stems usually from 1 to 2½ ft. high, bearing nodding, soft, brushlike, long-awned flowering spikes, from 2 to 4 in. long and about as thick, whence the common name. It is also known as foxtail barley. It is difficult to eradicate, but spraying of TCA (trichloroacetic acid) gives effective control. (N. Tr.)

SRIKAKULAM (formerly CHICACOLE), a town, with administrative district, in Andhra Pradesh, India. The town is on the left bank of the river Languliya, crossed there by a bridge, 55 mi. N.E. of Visakapatnam and 4 mi. from the sea. Pop. (1961) 35,071. Under Muslim rule it was the capital of one of the Northern Circars, and afterward of a British district.

SRIKAKULAM DISTRICT (area 3,901 sq.mi.) was created in 1950 by the division of Visakhapatnam district. Pop. (1961) 2,340,878. The Machkund hydroelectric project was completed there in 1955–56. (S. GL.)

SRINAGAR, capital of the state of Kashmir, 5,250 ft. above sea level, on both banks of the Jhelum river, which winds through the city with an average width of 80 yd. and is crossed by a number of wooden bridges. The curious grouping of the houses, no two of which are alike, the frail tenements of the poor, the substantial mansions of the wealthier, the curious carving of some, the balconies of others, the irregular embankment and the mountains in the background form a quaint and picturesque spectacle. Pop. (1961) 285,257. The artisans of Srinagar enjoy a high reputation. The historic industry of shawl weaving was replaced by carpet manufacture. Other industries are papermaking, leather and silver and copperware manufacture, and wood carving. Srinagar is the seat of the University of Jammu and Kashmir (1948).

SRIRANGAM, a town in Tiruchirapalli (Trichinopoly) district, Madras, India, 2 mi. N. of Trichinopoly city. Pop. (1961) 41,949. It stands on an island of the same name, formed by the bifurcation of the river Cauvery and by the channel of the Coleroon. The town is celebrated for its great 17th-century temple, dedicated to Vishnu, composed of seven square enclosures, one within another, and 350 ft. distant from each other. Each enclosure has four gates with high towers. The outer wall of the temple is not less than 4 mi. in circumference. A remarkable feature is the porch, which has a colonnade of rearing horses, each nearly 9 ft. high. The sculpture is richly carved in highly polished chlorite stone, deeply undercut.

SSU-MA CH'IEN (145/135-*c.* 90 B.C.), is the first major Chinese historian of whom there is any appreciable knowledge, and one of the greatest. His lifework, the *Shih chi* or *Records of the Historian*, completed around 100 B.C., has been widely read not only in China but in Korea and Japan as well, and has exercised a profound influence upon the literature and historiography of those countries.

Scholars disagree on the date of Ssu-ma Ch'ien's birth, some placing it in 145 B.C., others in 135 B.C.; he appears to have died around 90 B.C. In youth he traveled widely through the Chinese empire, gaining firsthand acquaintance with the geography and historic places. Most of his life was spent at the court of Emperor Wu of the Han dynasty. He inherited the court position of grand historian (sometimes translated as "grand astrologer") from his father, Ssu-ma T'an, whom he credits with the conception of the *Shih chi*. Late in his life, Ssu-ma Ch'ien was condemned to suffer castration for defending a disgraced general. Though it was customary under such circumstances to commit suicide, he chose to bear the disgrace in order to complete his history.

The *Shih chi* is a monumental work in 130 chapters covering about 2,000 years of Chinese history from the earliest recorded times to the lifetime of the historian. The existing text is very nearly complete. Divided into five sections, the work begins with 12 chapters of basic annals dealing with the emperors who ruled China. These are followed by ten chronological tables listing in graph form the important events of the past with their dates, and eight treatises on such subjects as music, ritual, astronomy and economics. Thirty chapters entitled "Hereditary Houses" treat the feudal kingdoms of early China. The work concludes with 70 chapters of biographies or memoirs relating the lives of famous men or describing the peoples of Korea, southeast Asia and central Asia with whom the Chinese had contact. Though histories in simple chronicle form existed earlier, this method of dividing material into annals, biographies, etc., was original with Ssu-ma Ch'ien and, in spite of certain disadvantages, was widely imitated in later Chinese historical works.

Ssu-ma Ch'ien's approach to history was essentially biographical. Throughout his work he was primarily concerned with the lives of the great men of the past and the moral inspiration and guidance they provide. Written in a vivid, concise style and enlivened by dramatic scenes and episodes, his history presents an invaluable picture of life in ancient China and has done much to mold the historical consciousness and literary tastes of generations of Chinese.

See Edouard Chavannes, *Les Mémoires historiques de Se-ma Ts'ien*, 6 vol. (1895–1905); Burton Watson, *Ssu-ma Ch'ien: Grand Historian of China* (1958). (B. D. W.)

STAAL, MARGUERITE JEANNE CORDIER DE LAUNAY, BARONNE DE (1684–1750), French author, was born in Paris on Aug. 30, 1684, the daughter of a painter named Cordier. She adopted her mother's maiden name, De Launay. She was educated at a convent at Evreux, later accompanying Madame de Grieu to the convent of St. Louis where the latter was appointed abbess. Mlle. De Launay held a little salon there until 1710, and after the death of Mme. de Grieu entered the household of the duchesse du Maine at Sceaux. She helped her mistress to draw up the *Mémoire des princes légitimes*, demanding the meeting of the states-general. After a short imprisonment in the Bastille, on account of her implication in the Cellamare conspiracy (1718), she returned to the service of the duchess. She had had many admirers and some serious passions, when, in 1735, being then over 50, the duchess arranged her marriage with the baron de Staal. She remained a member of the duchess's household and became the friend and correspondent of Mme. du Deffand. She died at Gennevilliers on June 15, 1750.

Her *Mémoires* appeared in London in 1755, and have often been reprinted. In these memoirs the humours of the "court of Sceaux" are depicted as hardly any other society of the kind has ever been.

Her *Mémoires* were edited by M. de Lescure (1877; Eng. trans. by S. Bathurst 1877, and by C. H. Bell 1892). Mme. de Staal also left two excellent short comedies, performed at the court of Sceaux, and some letters.

STABAT MATER, famous medieval Latin hymn on the Crucifixion, which has enjoyed unexampled esteem throughout the ages, believed to have been written by Jacopone, a Franciscan monk of the 13th century. Though not originally intended for this purpose, it was early introduced into liturgical use and was adopted as one of the Sequences of the Roman missal in the 18th century. Among the many settings of the poem by eminent composers may be mentioned those of Josquin des Prés, Palestrina, Pergolesi, Haydn, Rossini, Verdi and Dvořák.

STABIAE, an ancient town of Campania, It., on the coast at the east extremity of the Gulf of Naples (modern, Castellammare di Stabia). In 89 B.C. it was taken and destroyed by Sulla, and its territory given to Nuceria as a reward for fidelity to Rome. It was destroyed by the eruption of Vesuvius in A.D. 79.

STABILIZATION AGREEMENTS, INTERNATIONAL. International commodity stabilization agreements represent an attempt through concerted international action to mitigate the adverse effects upon trade in primary commodities arising from such difficulties as a persistent imbalance between production and consumption, burdensome stock accumulations and pronounced fluctuations in prices. After the early 1930s such agreements were concluded for a number of commodities important in world trade, including wheat, sugar, rubber, tea, timber, tin, coffee and olive oil. Customarily such agreements were of five years' duration, although some, notably the wheat agreements of 1953, 1956, 1959 and 1962, covered shorter periods.

The Interwar Period.—The international commodity agreements introduced in the period between World Wars I and II were essentially of a restrictive type; that is, they represented an attempt on the part of producers, acting during general economic depression, to arrest downward trends in prices by restricting supplies flowing to the world market. This was usually achieved through quota mechanisms, each participating country being granted a basic quota subject to periodic adjustment in the light of price trends and the market situation generally. Such agreements depend for their success upon an absence of close substitutes for the commodity concerned and upon participation by all the

main producing countries. Some earlier attempts to regulate the market through less comprehensive agreements, such as that for rubber between Malaya and Ceylon (1922–28), failed because higher prices merely stimulated production in and exports by countries outside the agreement. The agreements concluded in the 1930s did, in fact, embrace all the major producing countries and they further extended control over supplies by measures such as the prohibition of the export of planting material from participants to nonparticipants.

It is difficult to reach any general conclusions about the effectiveness of these earlier agreements. The 1933 wheat agreement was undoubtedly a failure, its operative clauses for all practical purposes being abandoned toward the end of the first year of its operation. The scheme of export regulation broke down and difficulties arose over the implied obligations of the four principal exporting countries to reduce acreage. On the other hand, the tea agreements (1933 and 1936 [revised 1938]) and rubber agreements (1934 and 1938) regulated exports in accordance with the agreed quotas and exerted a stabilizing and strengthening influence on prices. In 1937 the League of Nations Committee for the Study of the Problem of Raw Materials reported that the regulation schemes then in operation had, generally speaking, "been an important factor in the improvement in economic conditions experienced in producing countries during the depression as well as in the development of international trade."

World War II and After.—With the onset of World War II and the consequent disruption of supplies and higher prices, such agreements became largely redundant. However, certain of them, including those for sugar, tea and rubber, were extended by protocol, shorn of any effective quota provisions, largely to ensure the ready availability of machinery to meet expected postwar needs. In the case of wheat the protocol of 1942, signed by Canada, the United States, Argentina, Australia and the United Kingdom, made definite provision for a postwar international wheat conference and established an International Wheat council to review the situation meanwhile. Exceptionally for coffee, the special conditions brought about by the war, particularly the loss of the important European market, themselves gave rise to the first international coffee agreement (1940). Known as the Inter-American Coffee agreement, this was essentially an export quota agreement, but it constituted the first important international commodity scheme in which a consuming country, the United States, participated and played a dominating role by providing through the quota system an organized outlet for exports from Latin America.

In the early postwar years there was no immediate resort to international commodity agreements. Although a series of wheat conferences was held, introduction of the first postwar agreement for this commodity was delayed until 1949. The International Sugar agreement of 1937 became inoperative in 1944, although the International Sugar council remained in existence, and it was not until 1953 that a new agreement was successfully negotiated. Regulation of the rubber market was not resumed after the war because of the need to rehabilitate the far eastern producing areas and because of the rapid development of the synthetic rubber industry. Tea export quotas remained nominally in force until 1955, but in the later years they were in excess of 100% of the high basic quotas established under the 1936 agreement (revised in 1938) and had no real restrictive effect.

The subject of international commodity agreements was given detailed consideration by the preparatory committee (1946–47) of the first UN conference on trade and employment and later at the conference itself in Havana (1948). The charter of the proposed International Trade organization, agreed upon at Havana, recognized the importance of international commodity agreements in international trade relations, defining their purpose as follows:

> to prevent or moderate pronounced fluctuations in the price of a commodity with a view to achieving a reasonable degree of stability on a basis of such prices as are fair to consumers and to provide a reasonable return to producers, having regard to the desirability of securing long-term equilibrium between the forces of supply and demand; to maintain and develop the natural resources of the world and protect them from unnecessary exhaustion; to provide for the expansion of the production of a primary commodity where this can be accomplished with advantage to consumers and producers, including in appropriate cases the distribution of basic foods at special prices; to assure the equitable distribution of a primary commodity in short supply.

The charter also laid down that, in drawing up agreements, use should be made of certain principles which represented a major departure from those of the prewar era. In particular, it called for detailed preliminary studies by all interested countries of the difficulties affecting the commodity in question and for adequate participation in any agreement by importing as well as exporting countries; both groups were to have an equal voice in decisions on substantive matters. Another requirement was that agreements should provide, where practicable, for measures designed to expand world consumption of the commodities concerned.

Although the proposed International Trade organization did not come into existence, the principles set out in its charter provided the basis for all subsequent negotiations on commodity agreements. Each of the agreements introduced after World War II was negotiated at a conference convened under UN auspices following a preliminary report by the Interim Co-ordinating Committee for International Commodity Arrangements, and often after a great deal of further research in study groups and elsewhere. In fact, an outstanding development in international relationships concerning commodities after the war was the support given to such study groups, often on a continuing basis. Through an interchange of information and by other means short of the negotiation of an international agreement, these groups were often able to exert a considerable stabilizing influence on the market.

Types of Postwar Agreements.—In the postwar period, to meet the special situations of individual commodities, three main types of agreements were adopted: export restriction agreements, multilateral contracts and buffer-stock agreements.

Export Restriction Agreements.—The first type, adopted for sugar and coffee, conformed most closely to prewar methods, with the important modification that the participants included importing countries, which also assumed rights and obligations. For example, in return for assurances regarding supplies, such countries undertook to import their requirements from participating exporting countries. Price ranges were negotiated for sugar, but not for coffee; in that case the price objectives were couched in more general terms.

Multilateral Contracts.—This type of agreement, adopted for wheat in 1949 and again in 1953 and 1956, was an undertaking on the part of exporting countries to sell to participating importing countries a guaranteed quantity of wheat each year at prices not higher than a specified maximum. Each participating importing country undertook to buy from participating exporting countries on demand a guaranteed quantity at prices not lower than a specified minimum. Under the 1959 and 1962 wheat agreements the concept of guaranteed quantities was abandoned. Importing countries thereafter agreed to import stated percentages of their total commercial imports of wheat from member countries, while the exporting countries collectively agreed to supply all the commercial requirements of the importing countries. These obligations applied only under certain price conditions.

Buffer-Stock Agreements.—This kind of agreement was introduced only for tin, although it was considered also for other commodities not subject to regulation in the mid-1960s, including rubber and cotton. The successive international tin agreements (including the last prewar one of 1937) aimed at minimizing price fluctuations by empowering the manager of the buffer stock to buy tin, to the limit of his resources, when market prices fell below a certain minimum and to sell it when a ceiling was reached. Buying and selling operations were also permissible within the upper and lower limits of the price range in order to prevent steep rises or falls in price. The resources for these operations were provided by producing countries in cash, or in metal to a maximum of 75%. Consuming countries might also make contributions to the buffer stock, but they had not done so by the mid-1960s.

A Special Type.—The international agreements on olive oil (1956 and 1962) represent an approach to a particular commodity situation which is difficult to classify. They were concerned less

with the physical regulation of the market than with measures to enhance the competitive position of olive oil in world markets. For example, the Olive Oil council, established under the agreements, encouraged uniformity in member countries' legislation classifying and defining olive oil. It also concluded with the International Bureau for the Protection of Industrial Property a working agreement for protection against unfair competition, such as false declaration of origin or misuse of trade names. One of the council's most important activities was the initiation of extensive publicity campaigns to encourage the consumption of olive oil.

Summary.—In all the international commodity agreements concluded in the postwar period member countries looked, in varying degrees, beyond the immediate market situation to the basic problems concerning the commodity in question. For example, each of the agreements in effect in the mid-1960s paid some attention to the need to expand consumption. The International Coffee agreement of 1963 was designed primarily to relieve the surplus stock position which had burdened the industry for a number of years; to this end producing countries had undertaken to adjust their production to effective demand at the direction of the International Coffee council. However, progress along these lines tended to be slow, for it impinged upon the basic social and economic policies of the countries concerned.

It was not until 1959 that the scope of international wheat agreements was broadened to enable the International Wheat council to undertake annual reviews of the world wheat situation, and participating countries still reserved for themselves complete liberty of action in the determination and administration of their internal agricultural and price policies. None of the other agreements incorporated provision for a formal or direct system of production quotas in importing and exporting countries, but the view was gaining ground that the onus of correcting structural imbalances in the world market should be shared by buyers and sellers alike.

BIBLIOGRAPHY.—League of Nations, *Report of the Committee for the Study of the Problem of Raw Materials* (1937); International Labour Office, *Inter-Governmental Commodity Control Agreements* (1943); V. D. Wickizer, *The World Coffee Economy* (1943), *Tea Under International Regulation* (1944); A. McFadyean (ed.), *The History of Rubber Regulation, 1934–1943* (1944); K. E. Knorr, *Tin Under Control* (1945); B. C. Swerling, *International Control of Sugar, 1918–41* (1949); United Nations, *Commodity Trade and Economic Development* (1953); Food and Agriculture Organization of the United Nations, *A Reconsideration of the Economics of the International Wheat Agreement* (1952), *Observations on the Proposed International Sugar Agreement* (1953), *The Stabilization of the Olive Oil Market* (1955); Gerda Blau, "International Commodity Arrangements and Policies," Food and Agriculture Organization of the United Nations, *Monthly Bulletin of Agricultural Economics and Statistics* (Sept. and Dec. 1963); United Nations, Interim Co-Ordinating Committee for International Commodity Arrangements, *Review of International Commodity Problems* (annual). (J. M. CK.)

STADE, a town of northwestern Germany in Lower Saxony, situated on the navigable Schwinge, 3½ mi. above its confluence with the Elbe, 20 mi. W. of Hamburg on the railway to Cuxhaven. Pop. (1961) 30,542. The rise of Harburg has deposed it from its former position as the chief port of Hanover.

STADIUM, a large construction which combines space for games and seats for spectators, such as those erected for football. The term is the Latin form of a Greek word for a standard of length; a *stade* = 600 Greek (Olympian) feet or about 606 modern feet. The course for the foot race at Olympia was exactly a *stade* in length, and the word for measurement became transferred first to the race and then to the place in which it was run. Usually the stadiums were U-shaped, with the curve opposite to the starting point. In some instances, the curve led to a slope on which seats were cut into the land. The Romans built several types of structures for similar purposes: circuses; amphitheatres, also called arenas; and naumachia (*q.v.*) which could be flooded for naval engagements. The most famous of the Roman amphitheatres was the Colosseum at Rome. (*See* AMPHITHEATRE.)

Sports played a minor role in men's lives for many centuries after the fall of Rome, and it was not until modern times that permanent stadiums were again needed. In the interim, temporary stands such as those built from time to time in the Place des Vosges in Paris for jousting sufficed.

During the 19th century, systematic physical exercise was revived. This led to organized athletics, which became as popular as in ancient Greece. This revival of interest in athletic games brought about the building of structures adapted to the various sports and providing for many thousands of spectators. The first important stadium of modern times was the reconstruction, in Athens in 1896, of an ancient marble one built about A.D. 143 by Atticus Herodes on the site of a still earlier one (*see* ATHENS: *The Roman Period*). It seated 66,000 persons and was undertaken for the first of the modern series of Olympic games.

Although a few other stadiums were built before 1900, such as arenas for bullfighting in Spain and Mexico, the great era of modern stadium building was the early 20th century. In the United States alone more than 100 large stadiums had been built by 1930.

During the decade of the 1920s, 18 huge stadiums, each seating more than 50,000 spectators, were erected. Most of these were built for college athletic contests, particularly football. Four of the largest stadiums in the world were built in the U.S. The Los Angeles coliseum, built in 1923, was enlarged in 1932 to accommodate 105,000 persons. The Philadelphia Municipal stadium, built in 1926 for the Sesquicentennial exposition, held 102,000. In Chicago, Soldier field, opened in 1926, had a seating capacity of more than 102,000. The University of Michigan stadium at Ann Arbor was built in 1927 at a cost in excess of $950,000 to seat 79,000 spectators. The figure was raised to 87,000 by addition of temporary wooden bleachers. In 1949 seating capacity was raised to 97,000 and in 1956 to 101,000.

The largest sports arena in the world is reputed to be the Municipal stadium at Rio de Janeiro, Braz. It has a seating capacity of 155,000, measures ½ mi. in circumference and is surrounded by a concrete moat 9 ft. wide and 9 ft. deep. For the final game of the 1950 World Cup soccer matches there, the official crowd figures were given at 199,854. The early 20th century also witnessed the development of a new type of stadium, unknown to the ancients and made possible by modern building techniques, the indoor stadium, which could be used night and day the year around. The first important one was the old Madison Square Garden, built in New York city in 1890. It was replaced in 1925 by another of the same name seating 18,000.

The Chicago (Ill.) stadium was the largest indoor stadium in the world at the beginning of the second half of the 20th century. This type could be used for all court sports including those held on ice, and also for public meetings such as conventions, circuses and even concerts.

The years following World War II brought with them a boom in the construction of large indoor stadiums, due primarily to the increased interest in the sport of basketball. Field houses with seating capacities in excess of 15,000 were constructed on the campuses of most large universities throughout the United States.

Olympic Stadiums.—Nearly every country in which Olympic

Leading Stadiums

Place and Name	Capacity
Rio de Janeiro, Braz. (Municipal stadium)	155,000
Los Angeles, Calif. (Coliseum)	105,000
Moscow, U.S.S.R. (Tzentralnyi)	103,000
Philadelphia, Pa. (Municipal stadium)	102,000
Chicago, Ill. (Soldier field)	102,000
Ann Arbor, Mich. (Univ. of Michigan)	101,000
Rome, Italy (Olympic stadium)	100,000
Pasadena, Calif. (Rose Bowl)	90,000
Stanford, Calif. (Stanford Univ.)	90,000
Berkeley, Calif. (Univ. of California)	81,500
New Orleans, La. (Sugar Bowl)	81,000
Columbus, O. (Ohio stadium)	79,000
East Lansing, Mich. (Macklin field)	76,000
Miami, Fla. (Orange Bowl)	75,700
Dallas, Tex. (Cotton Bowl)	75,500
Cleveland, O. (Municipal stadium)	74,000
Urbana, Ill. (Memorial stadium)	71,000
New Haven, Conn. (Yale Bowl)	71,000
Houston, Tex. (Rice stadium)	70,000
Baton Rouge, La. (Tiger stadium)	68,000
New York, N.Y. (Yankee stadium)	67,000
Minneapolis, Minn. (Memorial stadium)	66,500
Philadelphia, Pa. (Franklin field)	65,000
Austin, Tex. (Memorial stadium)	65,500
Pittsburgh, Pa. (Pitt stadium)	62,000
Norman, Okla. (Owen field)	60,000

games were held erected a great permanent stadium for the event. In most instances, entire new stadiums were built for the Olympic Games' spectacular. However, Rome, host city to the 1960 Olympics, reactivated its Olympic stadium with a seating capacity of 100,000. The first stadium of the modern type to provide for the full range of Olympic events was that built in 1908 in London, at Shepherd's Bush, for the IV Olympiad. The stands, part of which were roofed, accommodated 50,000 seats. Other Olympic stadiums of interest include the one built in Stockholm in 1912; the severely simple one of Colombes, near Paris, built in 1927; the handsome brick one built at Amsterdam in 1928; and, most elaborate of all, that built in Berlin in 1936 for the XI Olympiad and last to be held before war broke out in 1939. The Helsinki stadium, built for the XII Olympiad scheduled for 1940 but cancelled by the outbreak of World War II, was the scene of the XV Olympiad (1952).

Plans.—The shapes of stadiums have varied widely, reflecting the variety of uses for which they were built. Some are rectangular in shape with curved corners, others elliptical or U-shaped. In some, the playing field was beneath the ground level with the spectator stands partly or wholly submerged in the excavation. The Yale Bowl, New Haven, Conn., built in 1914 for football, is a closed ellipse in plan. The Harvard stadium, Cambridge, Mass., built in 1912, is U-shape, open at one end to permit both football and track events. Others were built to permit baseball as well. Those built for Olympic games have generally been rectangular to enclose a large playing field, and with rounded corners for the track. The horseshoe shape has also been used as at Ohio State university, Columbus, in 1922. The most flexible shape was that providing two parallel stands, opposite each other, with the narrow ends open. Some of this type were built with superimposed tiers of seats, others with crescent-shaped stands (Northwestern university, Evanston, Ill., 1926) so that as many seats as possible were opposite the centre of the field. Other types have been used for special purposes, such as Yankee stadium, New York city, 1923, built primarily for baseball.

Structure.—The building of large stadiums in the early 20th century was facilitated by a new method of construction—reinforced concrete. In the hands of skillful engineers and careful workmen, especially in Europe, magnificent stadiums built in reinforced concrete daringly exploited the potentialities of the material; *e.g.*, those at Lyon, built in 1920 by Tony Garnier; at Florence, the Giovanni Berta stadium, built in 1933 with fantastic external stair towers; and at Bordeaux, built in 1939, remarkable for a roof of graceful, arched sections.

After World War II, stadiums continued to be built in regions where they had previously been scarce, such as in South and Central America, and more daring and dramatic applications of reinforced concrete were a conspicuous development. Outstanding examples included the stadiums at Rio de Janeiro, São Paulo, Mexico City and Cartagena; and in the United States at Houston, Tex., and Raleigh, N.C.

Continuous research in both materials and techniques opened new possibilities for stadium development, particularly toward reaching the goal of an all-weather stadium which would be free of obstructing columns. (C. L. V. M.; J. T. Ws.)

STAËL, MADAME DE. Anne Louise Germaine Necker, Baronne de Staël-Holstein (1766–1817), French novelist and miscellaneous writer, was born at Paris on April 22, 1766. Her father was the famous financier Necker, her mother Suzanne Curchod, almost equally famous as the early love of Gibbon, as the wife of Necker himself, and as the mistress of one of the most popular salons of Paris. The future Mme. de Staël was from her earliest years a romp, a coquette and passionately desirous of prominence and attention. She was a plain child and a plainer woman, whose sole attractions were large eyes and a buxom figure. She is said to have written her father a letter on his famous *Compte-Rendu* and other matters when she was not 15, and to have injured her health by excessive study and intellectual excitement. Her father's dismissal, and the consequent removal of the family from the busy life of Paris, were beneficial to her. During part of the next few years they resided at Coppet, her father's estate on Lake Geneva, which she herself made famous. They returned to Paris, or at least to its neighbourhood, in 1785, and Mlle. Necker resumed literary work of a miscellaneous kind, including a novel, *Sophie*, printed anonymously in 1786 and a tragedy, *Jeanne Grey*, published in 1790. She married Eric Magnus, Baron of Staël-Holstein, who was the Swedish minister. For a great heiress and a very ambitious girl the marriage scarcely seemed brilliant, but the king of Sweden promised the ambassadorship for 12 years and a pension in case of its withdrawal, and the marriage took place on Jan. 14, 1786. The husband was 37, the wife 20. Mme. de Staël was accused of extravagance, and after 11 years (1797) an amicable separation of goods was arranged. There was no scandal between them; the baron obtained money and the lady obtained an established position. In 1788 she appeared as an author under her own name with some *Lettres sur J. J. Rousseau*. She stood then for a mixture of Rousseauism and constitutionalism in politics. She visited Coppet once or twice, but mostly in the early days of the revolutionary period she was in Paris taking an interest and, she thought, a part in the councils of the Moderates. At last, the day before the September massacres, she fled from France.

She went to Coppet, and gathered round her a considerable number of friends and fellow refugees. In 1793 she visited England and established herself at Mickleham in Surrey as the centre of the Moderate Liberal emigrants. In the summer she returned to Coppet and wrote a pamphlet on the queen's execution. The next year the fall of Robespierre opened the way back to Paris. M. de Staël (whose mission had been in abeyance and himself in Holland for three years) was accredited to the French republic by the regent of Sweden; his wife reopened her salon. She also published several small works, including *De l'Influence des passions* (1796) and *De la Littérature considérée dans ses rapports avec les institutions sociales* (1800). During these years Mme. de Staël was of chief political importance. Benjamin Constant, whom she first met at Coppet in 1794, had great influence over her, as she had over him. Both personal and political reasons threw her into opposition to Bonaparte. Her own preference for a moderate republic or a constitutional monarchy was quite sincere. Her character and Napoleon's were too much alike in some points to admit of their getting on together. For several years, however, she was able to alternate between Coppet and Paris without difficulty. In 1797, as above mentioned, she separated formally from her husband. In 1799 he was recalled by the king of Sweden, and in 1802 he died, duly attended by her. They had three children, Auguste Louis, Albert and Albertine.

The exact date of the beginning of what Mme. de Staël's admirers call her duel with Napoleon is not easy to determine. From the title of her book *Dix années d'exil*, it should be put at 1804; from the time at which it became pretty clear that the first man in France and she who wished to be the first woman in France were not likely to get on, it might be put several years earlier.

In 1802 she published her first really noteworthy book, the novel *Delphine*. In 1803 she returned to Paris. She was ordered by Napoleon not to reside within 40 leagues of Paris. After considerable delay she set out, in company with Constant, for Weimar, where she wintered and then went to Berlin, where she met August Wilhelm Schlegel, who afterward became one of her intimates at Coppet. Thence she traveled to Vienna, where the news of her father's death reached her.

She returned to Coppet as its wealthy and independent mistress, but her sorrow for her father was deep and certainly sincere. She spent the summer at the château with a brilliant company; in the autumn she went to Italy accompanied by Schlegel and Jean de Sismondi, and there gathered the materials of her most famous work, *Corinne*. She returned in the summer of 1805, and spent nearly a year writing *Corinne;* in 1806 she broke the decree of exile and lived for a time undisturbed near Paris.

In 1807 *Corinne*, the first aesthetic romance not written in German, appeared. The publication was taken as a reminder of her existence, and the police of the empire sent her back to Coppet, where she stayed for the summer, and then set out once more for Germany. She was again at Coppet in the summer of 1808 (in which year Constant broke with her, subsequently marrying a

German lady) and set to work at her book, *De l'Allemagne*, which occupied her for the next two years. She decided to publish the book in Paris. The submission to censorship which this entailed was sufficiently inconsistent and Mme. de Staël wrote to the emperor one of the unfortunate letters, at once undignified and provoking, of which she had the secret. The reply to her letter was the condemnation of the whole edition of her book (10,000 copies) as "not French," and her own exile from France. She retired again to Coppet, where she was not at first interfered with, and she found consolation in a young officer of Swiss origin named Rocca, 23 years her junior, whom she married in 1811.

The operations of the imperial police in regard to Mme. de Staël are rather obscure. She was at first left undisturbed, but by degrees the château itself became taboo, and her visitors found themselves punished heavily. Mathieu de Montmorency and Mme. Jeanne Françoise Récamier were exiled for the crime of seeing her. On May 23 she left Coppet almost secretly, and journeyed to Vienna, where she obtained an Austrian passport to the frontier. After some trouble, receiving a Russian passport in Galicia, she escaped from Napoleonic Europe.

She journeyed slowly through Russia and Finland to Sweden, making some stay at St. Petersburg, spent the winter in Stockholm and then set out for England. There she received a brilliant reception and was much lionized during the season of 1813. She published *De l'Allemagne* in the autumn, was saddened by the death of her second son Albert, undertook her *Considérations sur la révolution française* and, when Louis XVIII had been restored, returned to Paris. She was in Paris when the news of Napoleon's landing arrived and at once fled to Coppet. In October, after Waterloo, she set out for Italy with her second husband, Rocca, who was dying of consumption. She returned to Coppet in June, and Byron frequently visited her there. Despite her increasing ill-health she returned to Paris for the winter of 1816–17, and her salon was much frequented. She died on July 14, 1817.

Baron Auguste de Staël (d. 1827) edited the complete works of his mother in 17 vol. (1820–21), with a notice by Mme. Necker de Saussure, and the edition was afterward republished in a more compact form and supplemented by some *Oeuvres inédites*. The *Considérations* and the *Dix années d'exil* had been published after Mme. de Staël's death. Some *Lettres inédites* to H. Meister were published in 1903. In English there are biographies by A. Stevens (1880) and Lady Blennerhasset (1889).

STAFF, MILITARY, a term used to designate the officers who assist the commander of a military force. The basic meaning of the term is similar in all armies, though staff organizations and their functioning vary considerably. The staff is commonly distinguished from the "line," the purely combatant portion of the army.

The staff assists the commander by performing the detailed duties of administration, planning, supply and co-ordination that are necessary to control and maintain his forces. It works out detailed plans in support of the commander's intentions, translates his decisions into orders and issues the orders to the troops. The staff is the commander's channel for maintaining contact with the subordinate units of his force and for receiving the information and advice he requires. A chief of staff normally heads the organization, which is divided into sections responsible for definite fields of activity, such as personnel, intelligence, plans and operations, training and supply. Staff officers have no command authority in their own right; they act always as agents of the commander, issuing orders in his name.

All modern armies, following a concept developed during the 19th century, employ some form of general staff. While the term has had various uses and often designates the staff of a general commanding any major unit of division size or greater, it most commonly denotes the staff of the central headquarters directing the nation's entire army establishment. Such a staff, usually located in the national capital, assists the officer in supreme command of the entire army in the performance of his functions. That officer is responsible to the government for all aspects of the readiness of the army for war, and usually for direction of the army's operations after hostilities have begun. Accordingly, the general staff in peacetime is concerned both with the training and administrative control of existing forces and with the preparation of plans for any war in which the nation might become involved.

EARLY HISTORY

Ancient and Medieval Armies.—As warfare advanced beyond the type of tribal conflict in which a single chieftain could keep his entire force within sight and lead it with his own battle cry, the commander's need for assistants grew. The need was met in a variety of ways, but the institution of the military staff showed no consistent development prior to the Thirty Years' War (1618–48), when regularization of army organization along relatively modern lines began. Until that time, staffs were usually on an informal basis. The commander merely retained a group of officers as his personal party, calling on them for whatever service the moment required, such as reconnaissance or the delivery of orders.

Even in permanently organized formations such as the Roman legion, staff functions were not clearly defined. In the army of the Roman republic, six tribunes headed each legion. The actual command rotated among them, and the five tribunes not currently in command were available for staff duty. In the last days of the republic the legion came under the permanent command of a legate. The six tribunes remained as his assistants, but, with these positions held by men of less stature and experience than under the former system, the tribunes appear to have been little more than orderlies.

In the warfare of the feudal era, with its short campaigns and limited forces, military staff organization had no place. Staff development began when larger armies, representing the ruling dynasties of the emerging national states of Europe, took the field in the 16th and 17th centuries.

Gustavus to Napoleon.—By the early 1630s Gustavus Adolphus had developed the Swedish army into an exceptionally well-organized force. Composed of companies and regiments of uniform size, with a set hierarchy of officers, it had a supply system that functioned smoothly, its soldiers were regularly paid, and its activities were governed by written regulations upheld by courts-martial. While the assistants Gustavus gathered about him were still largely a group of individual officers rather than an organized staff, most of them were assigned titles and definite responsibilities, such as the provost marshal, the judge advocate and the quartermaster. The Gustavian military organization was widely copied, and, indeed, a substantial part of the Swedish king's army passed intact into the service of France a few years after his death in 1632.

Over the next century and a half in most European armies the duties of the quartermaster (*q.v.*) tended to expand until that officer took on many of the attributes of a modern chief of staff. The quartermaster's responsibility for quartering and feeding the troops centred his thought on what lay ahead—the direction and route of the next day's march, the selection, laying out and fortification of the next camp site, and the daily extension of the line of supply. In time that officer became responsible for reconnaissance, for directing marches and ultimately for drafting operational orders, in addition to his basic supply function.

In the French army developed by Richelieu and inherited by Louis XIV in the mid-17th century the emergence of a chief of staff had already begun. The army organization included several officers bearing the title *maréchal de camp*, who handled the quartermaster functions. In one of these, the *maréchal général de camp*, were centralized some of the duties of a chief of staff, serving the marshal of France who was conducting the campaign. At the same time there was a *maréchal de bataille*, who formed the army in battle array and directed its tactical movements. The latter officer soon passed from the scene, and by the end of the 17th century the chief of staff's position, now called the *maréchal général des logis*, had assumed some responsibility for operations as well as intelligence and supply.

In the Prussian army that Frederick the Great led in his campaigns of the mid-18th century, on the other hand, staff organization developed slowly in the presence of military genius. Frederick was, in effect, his own operations officer and chief of staff. He required few assistants but gave careful attention to their professional training. Recognizing a need for officers to perform

reconnaissance, select and fortify camp sites and guide troops on the march, he personally chose and instructed 12 officers in these duties, adding them to his small quartermaster general's staff.

Frederick's interest in the intellectual preparation of officers for high-level duties led him in 1765 to establish an academy where young noblemen were trained in diplomacy and the military art. It was the forerunner of the *Kriegsakademie* that capped the educational system of the German general staff in the 19th century.

Napoleon.—By the time another military genius appeared in the person of Napoleon Bonaparte, the conditions of warfare had been altered in a way that allowed a larger place to military staff operations. The concept of the nation in arms that arose during the French Revolution produced mass armies several times larger than those commanded by Frederick a quarter of a century earlier, but there was also a striking increase in their mobility and maneuverability. Infantry, made more secure against cavalry attack by the socket bayonet and the increased firepower of the flintlock musket, combined with the first really mobile field artillery to mark the end of the rigidly indivisible army. Napoleon organized his army into autonomous corps and divisions, capable of sustaining themselves when separated by a day's march or more. Dispersion of his units not only made it easier to live off the country but permitted rapid movement by the use of parallel roads.

Command centred in the person of Napoleon himself. He required all intelligence reports to be delivered directly to him, and he sought no staff advice when devising his plans and instructions. But a staff was needed to keep track of the disposition of the corps and divisions and to translate Napoleon's decisions into written orders that would set the subordinate units into co-ordinated motion toward the accomplishment of his design. For these relatively technical staff duties Napoleon had an able practitioner in his chief of staff, Marshal Louis Alexandre Berthier.

A studious soldier and a tireless master of detail, Berthier had reduced to writing his conception of ideal staff organization and functioning as early as 1796. More influential in generating thought and study on the subject was the manual of staff doctrine and procedure published in 1800 by another French staff officer, Paul Thiébault, which was soon reproduced in English, German, Russian and Spanish. Both writers stressed the requirement for speed in staff work and for accuracy and conciseness in the writing of reports.

DEVELOPMENT TO WORLD WAR II

General Staff.—Warfare had developed during the Napoleonic era into a complex activity that engaged the full efforts and resources of nations. The work of the military staff consequently rose from the level of planning the next day's march to a more comprehensive form of activity. The 19th century saw the emergence of the concept of the general staff as an organization that prepared war plans in peacetime and systematically collected information that would be useful in the future conduct of operations. In most countries the first activity of this type was topographical reconnaissance and surveying for the production of military maps, because with greater mobility and dispersion of units, generals found it necessary to conduct war on the map rather than by direct observation of the field. The information gathered soon included all that could be learned about the army and military resources of every potential enemy state. The general staff became concerned as well with the intellectual preparation of officers for command and staff duties, through historical study, war games, field maneuvers and attendance at staff colleges.

German Staff System.—The general staff system developed in the Prussian army was the first and for many years the most successful. Extended as the German general staff when the Prussian king became the emperor of a unified Germany, it profoundly influenced the development of modern military staffs in other armies during the years through World War I.

The Prussian general staff took form around the small quartermaster general's staff used by Frederick the Great, which had gained some additional functions under his successor and in 1796 had been directed to prepare a complete military map of Prussia. Under the influence of an energetic lieutenant quartermaster general, Col. Christian von Massenbach (*q.v.*), the staff after 1801 passed beyond the mere gathering of geographical information and began formulating a somewhat theoretical set of plans for possible operations.

The enduring development of the general staff began with Gerhard von Scharnhorst's (*q.v.*) reorganization of the Prussian army following its defeat at Jena in 1806. In form the staff was little changed from Massenbach's day, but Scharnhorst introduced features that kept staff officers in close touch with the commanders and troops and prevented them from living in a world of theory. Under his system officers were selected from all branches of the army to serve on the general staff, and they returned periodically to regular duty with the troops. In addition to maintaining their practical proficiency as soldiers, this rotation permitted a greater number of officers to gain experience in general staff duties, qualifying them for staff assignments when war increased the size of the army.

With the final defeat of Napoleon in 1815, the main body of the Prussian general staff was concentrated in Berlin as the "great general staff," distinguished from the officers of the "general staff with troops," who served in vital staff positions in corps and division commands. In 1821 the two elements of the general staff were placed under a single chief of the general staff, who was thereafter the supreme professional military head of the Prussian army. At the same time, the Prussian king removed the general staff from the control of the war ministry and made it directly responsible to the sovereign.

Thereafter the general staff developed under the leadership of a succession of able chiefs, most of whom held the position for extended periods. Count Helmuth von Moltke (*q.v.*), serving from 1857 to 1888, was the most distinguished. The master plans for speedy mobilization and strategic deployment of the Prussian forces produced by the general staff under his direction led to the rapid victories over Austria in the Seven Weeks' War of 1866 and over France in 1870. Success owed much to Moltke's quick appreciation of the importance of railways as a new factor in warfare. Full exploitation of their potentialities allowed him to deploy the army for war under a plan whose precision and reliability rivaled that of the railroad timetable itself.

To neighbouring countries the effectiveness and efficiency of the German general staff constituted both a menace and a model. During the last half of the 19th century all major nations on the European continent instituted or matured some form of general staff to make detailed plans for war in time of peace. Italy and Austria-Hungary closely followed the German system, and its influence on Russian staff development was marked. During the same period the rising power in the far east, Japan, adopted a high command and staff system that virtually duplicated the German.

Meanwhile in the German government the general staff attained a unique position of power that lasted through World War I. Fortified by the prestige of its victories and protected from civilian control and effective parliamentary questioning by its confidential relationship to the kaiser, the general staff frequently wielded great influence in the selection of ministers and the formulation of foreign policy.

French Staff System.—Attempting to meet the need for a system that would be effective in the absence of a self-sufficient military genius of Napoleon's type, Marshal Gouvion Saint-Cyr as minister of war made a promising beginning toward the creation of a professional staff corps in 1818. He established the *corps d'état-major*, which was to supply officers to staff the headquarters of all units of the French army, drawing its members from the graduates of a newly founded staff college. Saint-Cyr's wise provision that at successive stages in their careers staff officers should perform regular tours of line duty with the troops was abandoned in 1833, and the staff corps became a closed service. Restricted membership, a rather narrow type of training and permanent assignment to staff duties tended to separate the staff corps from the main life of the army and even generated some distrust and disdain of its services by field commanders. Weakness in staff work and in comprehensive planning for war ranked high among

the reasons for the French defeat in 1870–71.

During their remarkable period of national resurgence following the Franco-Prussian War the French soon came to emulate, if they did not directly copy, the German general staff system that had designed their defeat. By the 1880s they had replaced the former staff corps with a true general staff, headed by a chief titled the *chef d'état-major général de l'armée*. Staff candidates passed through an advanced school, the École Supérieure de Guerre, and alternated thereafter between staff assignments and regimental duty.

By 1914 France had a highly proficient general staff that was employed, like its German counterpart, in keeping a prepared file of campaign and mobilization plans up to date while overseeing all other aspects of the readiness of the army for war. Staffs at all levels in the French army were organized, under their chief, into three bureaus. The first bureau handled both supply and administrative matters such as recruitment, organization of units, casualty reports and the assignment, promotion and discipline of personnel. Enemy information was the concern of the second bureau, while the third prepared orders for military movements and operations.

The continuous combat operations of World War I, differing from the set battles of the past, brought the creation of a fourth bureau that assumed responsibility for the duties of supply, transportation and casualty evacuation formerly held by the first bureau. Thus the French system that matured during World War I displayed a distinct separation of staff functions into categories of administration, intelligence, operations and supply. It continued in this form without substantial change in later years.

Russian Staff System.—Imperial Russia was another country where military staff institutions took form around the expanding responsibilities of the quartermaster. Peter the Great appointed the first quartermaster general, in 1701, and the duties of that office grew steadily during the 18th century, often in imitation of the Prussian model. By 1815, when Tsar Alexander I established the "Headquarters Staff of his Imperial Majesty," the quartermaster general's section was the main staff component, responsible for most aspects of military operations. At the recommendation of Baron Antoine Henri Jomini (*q.v.*), formal education of officers for staff duties began with the founding of the Nicholas academy in 1832.

Reorganizations in which the influence of the German general staff concept was apparent followed both the Crimean War (1853–56) and the Russo-Japanese War (1904–05). The imperial Russian army entered World War I under a general staff immediately subordinate to the minister of war—a directing system that appeared to be as well appointed as any other on the continent. That there had been no proper staff training, however, became apparent when the deficiencies soon revealed in staff planning for the supply of the fighting forces contributed greatly to the disorganization and defeat that prepared the way for the Revolution of 1917 and the dissolution of the imperial army.

The Red army was brought into existence in 1918 to defend the new Bolshevik government against counterrevolutionary forces and foreign invaders. To provide their army with competent leadership the Communist masters found it necessary to enlist many former officers of the imperial army, particularly general staff officers and instructors for the military schools. As a result, staff doctrine and organization tended to remain unchanged, and most of the elements of the old general staff continued to function under the new commissariat of the army and navy. Moreover, the traditional German influence was renewed when Soviet officers began receiving training in Berlin military academies following the treaty of Rapallo in 1922.

One notable innovation was the commissar system, which placed a political officer on the staff of every Red army commander to monitor his activities and insure loyalty to the Communist regime. Since no operational order was valid until countersigned by the commissar, he was virtually the equal of the military commander. After 1934 the commissar served only as supervisor of the morale and political indoctrination of the troops, but his cocommander status was restored for a period after the great purge of the military command in 1937 and following the German invasion of June 1941. The commissar was again reduced to the role of political adviser and propagandist in Oct. 1942, and Red army units came under the single direction of their professional military officers. The Red army staff reached a high standard.

British Staff System.—Since Great Britain had failed to keep step with 19th-century developments on the continent, the British army entered World War I with only a decade of experience in the operation of a general staff system. In a period when military thought had centred largely on defense of the British Isles against invasion and the problems of India and other distant colonies, staff development had been slow, despite the founding of a staff college in 1858.

Interest in army reforms began to quicken following the publication in 1890 of *The Brain of an Army* by an Oxford lecturer, H. Spenser Wilkinson. His book explained the importance of the general staff in the successes of the German army and underscored the need for introduction of a similar system in Britain. Pronounced deficiencies in the British command and staff system revealed during the Boer War completed the argument. In 1904 the report of Lord Esher's committee investigating the war office demanded the formation of a general staff, "as the term is understood in all well-organized armies." The position of chief of the general staff was created in that year, the general staff itself being formally constituted in 1906. With expanded functions, the title was changed in 1909 to chief of the imperial general staff.

British staffs in all major army units during both World Wars I and II were divided into three branches. One, called the general staff, handled intelligence and the planning and conduct of operations. The adjutant general's (A) staff dealt with personnel and administration, while the quartermaster general's (Q) staff managed supply and transportation. Broadly stated, the general staff section was concerned with the use of the army while the other two branches looked after its maintenance. The chief of the general staff section co-ordinated the work of the entire staff, collaborating with the chief administrative staff officer, who supervised both the A and Q branches. Only the headquarters of the largest formations employed separate chiefs of staff to head the whole organization.

U.S. Staff System.—Before 1903 there was no general staff in the United States army. The term was sometimes used, but merely as a collective designation for the army's generals and the officers whose assignment to engineer, ordnance and other duties set them apart from the infantry, cavalry and artillery. An officer titled the commanding general of the army had authority over the troops but not over the quartermaster, adjutant general and chiefs of the other technical and administrative bureaus of the war department, whose activities were vital to the successful operation of his forces. Each bureau chief, supreme authority in his own field, reported only to the civilian secretary of war. How little this unco-ordinated system was geared to actual preparation for war was indicated by the statement of the commissary general that his office was functioning perfectly until the Spanish-American War disrupted and disorganized it.

The striking administrative failures of the Spanish-American War revealed the inadequacy of the bureau system and opened the way for reforms under a new secretary of war, Elihu Root. He founded the Army War college in 1902 and the following year obtained from congress authorization to create a general staff. Root's carefully reasoned explanations of the purpose of a general staff, based partly on his reading of *The Brain of an Army*, overcame the fears of some legislators that the change would result in "Prussianization" of the U.S. army. The legislation of 1903 established a chief of staff immediately under the secretary of war and the president, and gave him both the responsibilities of the former commanding general of the army and supervisory authority over the bureaus. To comprise the new general staff corps, officers were selected from the army at large for duty on the war department general staff or on the staffs of field commanders as members of "the general staff serving with troops." Since its beginning the U.S. army general staff corps has been manned by rotation rather than permanent assignment.

As leader of the American expeditionary forces in France in

1917, Gen. John J. Pershing studied British and French organization before erecting his own staff structure. Its final form closely resembled the French, including a chief of staff and five sections devoted to personnel, intelligence, operations, supply and training. When Pershing became chief of staff of the army in 1921 he reorganized the war department along the lines of his World War I headquarters, retaining the "G" system of designation borrowed from the French. The five divisions, each headed by an assistant chief of staff, were called personnel (G–1), military intelligence (G–2), operations and training (G–3), supply (G–4) and the war plans division. This establishment underwent no basic change until the United States entered World War II.

WORLD WAR II AND AFTER

Joint and Combined Staffs.—World War II, truly global in scope and fought by coalitions of great and lesser powers, presented problems of co-ordination and direction that called forth new, higher levels of command and staff organization. The war of 1939–45 saw the rise of air power to full stature as a third armed force, coequal with sea and ground power. Amphibious operations, in which the efforts of all three forces were joined, proved to be a major feature. Thus a controlling organization was needed to encompass both interservice and inter-Allied co-ordination.

As the supreme military body for the strategic planning and direction of the Allied war effort in all areas except the Soviet front, the United States and Great Britain established the combined chiefs of staff early in 1942. Located in Washington, D.C., the combined chiefs of staff organization was directly responsible to the U.S. president and the British prime minister as a combined executive. The three British services were represented by the first sea lord, the chief of the imperial general staff and the chief of the air staff. Within the British government these three offices had been joined since 1923 as the chiefs of staff committee, bearing a corporate responsibility for the command and strategic direction of United Kingdom forces and for supplying military advice to the cabinet and prime minister.

To provide United States membership on the combined chiefs of staff to match this British representation, a new agency known as the joint chiefs of staff came into existence in the U.S. The title followed an agreed definition of terms whereby "combined" signified collaboration between two or more of the Allied nations while "joint" was used to designate the interservice collaboration of one nation. The original members of the U.S. joint chiefs of staff were the army chief of staff; the naval officer holding the double title of commander in chief, U.S. fleet, and chief of naval operations; and the commanding general of the army air forces. Later in 1942 an officer serving as chief of staff to the president became the fourth member.

Beneath the combined chiefs of staff was a planning organization that operated on the committee principle. Combined committees for strategic planning, logistics, intelligence, munitions assignment and other purposes were manned by army, navy and air force staff officers drawn in equal numbers from the two nations.

In March 1942 a reorganization of the U.S. war department increased the effectiveness with which the general staff supported the army chief of staff in his expanding duties at the joint and combined levels. Redesignated the operations division, the former war plans division became the agency through which the chief of staff exercised command authority over the widely deployed forces of the army. The four "G" sections of the general staff concerned themselves only with the development of broad policy, since responsibility for training, the actual operation of the supply system and such technical services as the ordnance, medical and signal branches had been decentralized to three subordinate commands: the army ground forces, the army air forces and the army service forces.

Anglo-U.S. unity in the higher direction of the war was reflected in the field, where forces of both nations were placed under the single authority of a supreme commander with a combined staff. The principle found its fullest expression in the two staffs commanded by Gen. Dwight D. Eisenhower in his Allied force headquarters (AFHQ) in Africa and his supreme headquarters, Allied expeditionary force (SHAEF), in Europe, where officers of the two nations were completely integrated.

The SHAEF staff followed U.S. doctrine in its basic organization, with assistant chiefs of staff for personnel (G–1), intelligence (G–2), operations (G–3), supply (G–4) and civil affairs (G–5). The G–1, G–3 and G–4 staff chiefs were American, while British officers headed G–2 and G–5, but each chief had a deputy of the opposite nationality. In Gen. Walter Bedell Smith the Americans provided an able chief of staff, who was supported by three British deputies. One deputy handled air matters; the responsibilities of the other two followed the usual British division between operations and intelligence functions on the one hand and the "A" and "Q" administrative functions on the other.

In contrast to the western Allies, Germany was poorly organized to deal with its relatively simpler problems of co-ordination. In theory the *Oberkommando der Wehrmacht* (OKW), or armed forces high command, corresponded to the U.S. joint chiefs of staff or the British chiefs of staff committee as the agency for strategic direction and co-ordination of the three services. In practice OKW was dominated by the army and was not a true joint staff. Its control over the armed forces was slight, since the heads of the air force and navy stood in a favoured relationship to Hitler, while the dictator himself assumed control of the army through its high command staff, the *Oberkommando des Heeres* (OKH). Rather than being the supreme military authority, OKW came to be roughly coequal with OKH, controlling German operations in the west while OKH dealt with the Russian front.

When invaded by Germany in 1941, the Soviet Union quickly redesigned its highest governing agencies for centralized control of the war effort. Stalin assumed personal direction of military operations, gathering under himself a supreme headquarters called the *stavka*. Composed of about 20 top military officers of various arms, the *stavka* developed over-all strategic plans for Stalin's approval. The chief of the general staff was a member, and his organization served directly under the *stavka*, providing information and working out detailed operational plans and orders to implement the approved strategy. Normally these operational orders were transmitted to the front commanders for execution, but the *stavka* occasionally detached a command group from its own membership to conduct a particular campaign.

Trends After World War II.—Continuation and elaboration of the unified command and joint staff practices initiated during World War II marked the postwar period. In the United States, the National Security act of 1947 made the air force a separate service but placed it with the army and navy under a single department of defense. The act also gave legal recognition to the wartime joint chiefs of staff (JCS), provided them with a full-time joint staff, and designated the joint chiefs of staff the principal military advisers to the president and the secretary of defense. Amendments in 1949 made them the military advisers to the National Security council as well. Composed of the army chief of staff, the chief of naval operations and the air force chief of staff, JCS after 1949 had a fourth and senior member called the chairman. In most areas where U.S. forces were stationed, JCS set up a unified command in which a single commander with a joint staff controlled units of all three services.

Great Britain also consolidated its wartime experience by making the ministry of defense a permanent office in 1946, with authority over the three services. At the beginning of 1956 the British chiefs of staff committee, like JCS, received a chairman as its fourth member.

In the Soviet Union during 1946 the wartime *stavka* was disbanded and a single ministry of the armed forces assumed control of military affairs. A division into ministries of war and navy occurred in 1950, but the services were united again in 1953 under the ministry of defense. The general staff served directly under the minister of defense, and its chief was one of his deputies. The general staff was basically the former Red army general staff, elevated to responsibility for military planning and co-ordination applicable to all the Soviet armed forces.

New circumstances brought forth new varieties of international staffs, resembling the Anglo-U.S. combined chiefs of staff of World

War II. The 12 nations signatory to the North Atlantic treaty agreed in 1949 to set up regional military committees in which as many as five countries would be represented. Under the treaty an extensive international staff organization was created at the supreme headquarters, Allied powers in Europe (SHAPE).

In the second decade after the end of World War II all the great powers were giving thought to reorganization of their military forces to meet the requirements of atomic warfare. New forms of command and staff organization were likely to evolve as well. *See* also ARMY; STAFF, NAVAL.

BIBLIOGRAPHY.—Paul Bronsart von Schellendorf, *Duties of the General Staff*, 4th Eng. ed. (London, 1905); Walter Goerlitz, *History of the German General Staff, 1657–1945* (New York, 1953); B. H. Liddell Hart (ed.), *The Red Army* (New York, 1956); H. Spenser Wilkinson, *The Brain of an Army*, rev. ed. (London, 1895); W. R. Robertson, *Soldiers and Statesmen, 1914–1918* (New York, London, 1926); Otto L. Nelson, Jr., *National Security and the General Staff* (Washington, D.C., 1946); J. D. Hittle, *The Military Staff: Its History and Development*, rev. ed. (Harrisburg, Pa., 1949); Forrest C. Pogue, *The Supreme Command* (Washington, D.C., 1954); Ray S. Cline, *Washington Command Post: the Operations Division* (Washington, D.C., 1951); Oliver L. Spaulding, *The United States Army in War and Peace* (New York, 1937); R. Ernest Dupuy and Trevor N. Dupuy, *Military Heritage of America* (New York, London, 1956); Philip C. Jessup, *Elihu Root* (New York, 1938); James G. Harbord, *The American Army in France, 1917–1919* (Boston, 1936); also articles from the British *Army Quarterly* and the U.S. *Infantry Journal, Command and General Staff Review, Journal of Modern History* and *Journal of the American Military Institute*. (V. E. D.)

STAFF, NAVAL. The term "naval staff" has an exact meaning in the Royal Navy, but it is often used in a loose general sense, as in business, to denote a body of assistants to a manager or executive head. The naval staff is the group of officers assigned to study the problems which will confront the Royal Navy in war, and to make preparations and plans for solving those problems; to make plans for the operations determined to be necessary in time of war; to develop a war doctrine and have it widely understood among the officers of the service, so that a minimum of specific direction will suffice in wartime; to control education and training for war so that the essential objectives will be known and understood; and to do the advance planning necessary to enable the admiralty to prescribe the ends to be attained when war occurs, while leaving a fleet commander complete freedom to determine how he is to comply with his broad directive and attain the desired results. Cabinet investigating committees in 1910 and 1911 recommended the establishment of a naval staff, and on Jan. 8, 1912, the immediate formation of a "naval war staff" was announced by the lords commissioners of the admiralty. Additions and changes were made many times, notably in 1916–18 and after World War II.

The first sea lord is chief of the British naval staff. He is assisted by a vice chief of staff and two deputy chiefs of staff, all flag officers. The naval staff is located in the admiralty office and is directly responsible for intelligence, planning, operations and tactical doctrine. In addition, there are staff divisions which advise on the policy concerning training in all its aspects, and on the requirements for new ships, aircraft and weapons.

In the United States the chief of naval operations is the naval member of the joint chiefs of staff organization. He heads what is, in fact, the naval staff of the U.S. navy, although it does not bear that name. The naval operations organization performs the naval staff duties, as well as day-to-day administrative functions, much the same as does the admiralty in London. The office of the chief of naval operations is organized on a functional basis. An admiral serves as vice chief and there are six deputies, five of whom hold the rank of vice admiral. The divisions, under the deputies, are headed by rear admirals or captains.

There is no general staff corps in the Royal Navy or the U.S. navy to resemble the general staff corps in various armies. (*See* STAFF, MILITARY.) Service opinion has always been strongly against such a corps, but both navies have attempted to educate officers for staff duties without designating them as specialists. The U.S. navy established the Naval War college at Newport, R.I., in 1884, the first institution of this kind. The Royal Navy has a naval war college and a naval staff college as components of the Royal Naval college, Greenwich. Through the efforts of men like Adm. Sir Herbert W. Richmond, the attempt was made to foster the scientific study of war, to develop logical and consistent habits of thought in connection with the employment of naval force in war and to emphasize that matériel alone will not win wars. Both the Royal Navy and the U.S. navy have always stressed professional competence in seamanship, but Adm. Stephen B. Luce who founded the U.S. Naval War college, Adm. Alfred Thayer Mahan (*q.v.*), who succeeded Luce as head of the War college, Admiral Richmond and other British and American officers attempted through their writing and teaching to bring naval staff work up to the high standards of seamanship which prevailed already. The officers who do the naval staff work in Whitehall, and on the staffs of British admirals, as well as those who do such work in the navy department, and on the staffs of American admirals, are assigned to the admiralty or to naval operations or to a seagoing staff for a period of not more than three years, after which they take some other assignment. Thus the staff work is done by seagoing officers fresh from a nonstaff billet who expect to return to duties which might involve responsibility for executing basic plans which they have helped to formulate while serving on the naval staff.

Staff Corps.—Although it has never had a general staff corps, the U.S. navy does have a staff corps, which must not be confused with such a naval general staff corps. Medical officers, for example, belong to the medical corps. Other corps are: supply, chaplain, civil engineer, dental, medical service and nurse corps. These are known as staff corps and are distinguished from the line officers who exercise military command. Prior to 1900 there was also an engineer corps, and there was a corps of professors of mathematics developed before the days of the naval academy. The naval construction corps was amalgamated with the line in 1940. Certain specialists in engineering and naval construction are designated for "engineering duty only," but are line officers.

The term staff corps is not used in the Royal Navy. There is, however, a medical branch, a dental branch and an instructor branch, as well as the royal naval nursing service. The "general list" of the Royal Navy, which is the same as the line of the U.S. navy, has four classes of specialists: engineering, electrical, supply and seaman. Both navies have warrant officers, all of whom are specialists.

Admirals' Staffs.—A group of officers, without command authority, appointed to assist a commander in chief or an officer in a lesser command, is known as an admiral's staff; it functions in the same manner as an army staff. The minimum staff for a rear admiral in the Royal Navy is usually three officers: a flag lieutenant, who handles signals and messages and accompanies the admiral as his personal aide on occasions of ceremony; an admiral's secretary, who handles correspondence and reports; and an admiral's staff officer who makes operational plans and assists in the details of operations and administration. In the U.S. navy the arrangement is the same, except that the admiral's secretary is known as a flag secretary.

In both navies the admiral uses the flagship's medical officer and supply officer as well as other ship's officers for advice and assistance in their specialties.

The commander in chief U.S. Atlantic fleet was also the supreme allied commander Atlantic under the NATO organization; he had his headquarters at the naval base, Norfolk, Va., with his staff on shore. He had a staff as fleet commander and a NATO staff made up of allied officers for his Atlantic command. The commander in chief was an admiral and he had a vice-admiral, U.S. navy, as his chief of staff in his post as Supreme Allied Commander Atlantic; two rear admirals, one British and one American, served as deputy chiefs of this staff. A rear admiral was deputy fleet commander and chief of the fleet staff. Each of these staffs had six divisions: plans, operations, intelligence, logistics, communications and administration. The plans division was concerned essentially with war plans and drew on the other divisions for assistance as required in this work. The operations division of the NATO staff was concerned with operational planning for war and with plans for exercises which promoted the

smooth functioning of groups involving vessels of the various navies. The operations division included specialists in gunnery, engineering, torpedoes, antisubmarine warfare, air warfare, amphibious warfare and electronics. The administration division was directly concerned with the day-to-day details of correspondence and routine staff work as distinguished from the war staff planning.

The NATO staff had developed tactical and communication doctrines which had been accepted by the various navies so that command of mixed forces could be exercised efficiently at sea.

The commanders in chief of the British home and Mediterranean fleets and the commander in chief of the U.S. Pacific fleet had large staffs fundamentally similar to that of the Atlantic fleet, although the groupings of duties differed. The U.S. navy's commandants of naval districts had staffs usually comprising five divisions: personnel, operations, naval reserve training, logistics and administration. In each large organization the chief of staff co-ordinated all efforts; genuine efficiency required competent teamwork.

Ideal naval staff work afloat was demonstrated in the results of the battle of the Nile (*q.v.*) and the battle of Trafalgar (*q.v.*). Mahan wrote of Adm. Horatio Nelson (*q.v.*): "No man was ever better served than Nelson by the inspiration of the moment; no man ever counted on it less." Before going to sea for the campaign of Trafalgar, Nelson had explained to friends his proposed plans for fighting the combined French and Spanish fleets. Nelson had carefully thought out what the situation was likely to be and had formulated definite plans for meeting that situation. He had been a student of naval tactics all his life, and he was familiar with the experiences of fleet commanders in earlier battles. When he joined the fleet in Sept. 1805 he explained his ideas in detail to a number of his captains and to Adm. Cuthbert Collingwood, his second in command. On Oct. 9 he prepared a secret memorandum which he submitted to Collingwood. A few days later he made an oral explanation of the plan to all of his captains and furnished them copies of the memorandum. This document is a flawless example of completed staff work because it explained the admiral's whole concept of the general situation and how he proposed to attack the enemy; it recorded that the second in command would "have the entire direction of his Line to make the attack upon the Enemy, and to follow up the blow until they are captured or destroyed." It emphasized that the British fleet would concentrate on and overpower the enemy flagship and the rear of the enemy line before the enemy van could assist. The instructions were explicit enough to make the basic plan perfectly clear and to enable each captain to understand what was intended. Nelson also added a sentence which has been widely quoted: "In case signals can neither be seen or perfectly understood, no captain can do very wrong if he places his ship alongside that of an enemy." Nelson won decisively at the Nile and at Trafalgar because of superb war staff work.

Naval staff work is poor, good or excellent depending upon the results in the ultimate test of war, but an intelligent and farsighted staff policy is frequently handicapped by attention to problems of no lasting importance. (J. B. Hn.)

STAFFA (Norse for staff, column or pillar island), uninhabited island of the Inner Hebrides, Argyllshire, Scot., 33 mi. W. of Oban, about 6 mi. N.E. and S.E. from the nearest points of Mull and 6 mi. N. by E. of Iona. It lies almost due north and south, is $\frac{3}{4}$ mi. long by about $\frac{1}{2}$ mi. wide, and its highest point is 144 ft. above sea level. In the northeast it shelves to a shore, but otherwise the coast is rugged and much indented with numerous caves. During the tourist season it is visited regularly by steamer from Oban. In section the isle has first a basement of tufa, from which rise, secondly, colonnades of basalt in hexagonal pillars forming the faces and walls of the principal caves which in turn are overlaid, thirdly, by a mass of amorphous basalt. On the southeast coast is the Clam Shell or Scallop cave. It is about 130 ft. long, and on one side of it the ridges of basalt stand out like the ribs of a ship. Near this cave is the rock of Buachaille ("The Herdsman," from a supposed likeness to a shepherd's cap), a pile of columns, fully seen only at low water. On the southwest shore are the Boat cave and Mackinnon's or the Cormorants' cave. Fingal's cave, the most famous, situated in the southern face of the isle, is 227 ft. long. On its western side the pillars are 36 ft. high. The cave is the haunt of seals and sea birds.

STAFFORD (Family). This famous English family was founded in England by Robert, a younger brother of Ralf de Tosny (Toeni), of a noble Norman house, who was standard-bearer of the duchy of Normandy. Robert received at the Norman conquest a great fief extending into seven counties and became known as Robert de Stafford from his residence at Stafford castle. On the death of his great-grandson, also Robert (d. 1194), the male line became extinct. This Robert's sister had married Hervey Bagot, one of her brother's knightly tenants, who succeeded to the fief in her right; their descendant Edmund de Stafford (that surname having been assumed) was summoned as a baron in 1299. His son, Ralph, conducted the brilliant defense of Aiguillon against the host of France (1346) and fought at Crécy and in the siege of Calais. Chosen a knight of the garter at the foundation of the order, he was created earl of Stafford in 1351.

His son Hugh, who succeeded as 2nd earl in 1372, served in the French wars. From 1376 he became prominent in politics, probably through marriage to a daughter of the earl of Warwick, being one of the four lords on the committee in the Good parliament, and also serving on the committee that controlled Richard II, whom he accompanied on his Scottish expedition in 1385. He died in 1386, on pilgrimage at Rhodes. His son, Thomas, the 3rd earl, married in 1392 the daughter and heiress of Thomas, duke of Gloucester (son of Edward III). She, after Thomas' death, married his brother Edmund, the 5th earl. (The 4th earl, William, also Thomas' brother, died a minor in 1395.) Edmund's son, Humphrey (1402–1460), created the first Stafford duke of Buckingham in 1444, was placed by descent and possessions in the front rank of the English nobility.

Henry, the 2nd duke, was attainted in 1483, but on the triumph of Henry VII, Edward, the 3rd duke (1478–1521), regained the title and estates, and recovered the ancestral office of lord high constable in 1509. He was accused of treason and, after a nominal trial by his peers, was beheaded on May 17, 1521, a subsequent act (1523) confirming his attainder.

Henry (1501–1563), the son of the last duke, was granted by the crown some of his father's manors for his support, and, espousing the Protestant cause (though married to a daughter of Margaret, countess of Salisbury and sister of Cardinal Pole), was restored in blood on Edward VI's accession and declared Lord Stafford, as a new creation, by act of parliament. His second surviving son, Thomas, became involved in plots against Mary I's Spanish marriage and, maintaining that he was next heir, eventually assumed the royal arms. He sailed from Dieppe with two ships in April 1557, landed at Scarborough, seized the castle, and proclaimed himself protector. He was captured and executed for high treason. His father's new barony, in 1637, passed to a cadet in humble circumstances, who was called on, as a pauper, to surrender it to the king, which he did (illegally, it is now held) in 1639. The king thereupon bestowed it on Mary Stafford (the heir general of the line) and her husband, William Howard, in whose descendants it is now vested. Roger, who had surrendered the title, died in 1640, the last male heir, apparently, of the main line.

STAFFORD, a market town, municipal borough and the county town of Staffordshire, Eng., in the Stafford and Stone parliamentary division, on the river Sow, a tributary of the Trent, 132 mi. N.W. of London by road. Pop. (1951) 40,263. Area 8.0 sq.mi. Stafford contained a royal mint before it fell to William the Conqueror, who built a castle there. He appointed as governor Robert de Toeni who assumed the title of baron de Stafford. During the Civil War the town declared for the king, but fell in 1643 when the castle was destroyed. (A replica was begun by Sir George Jerningham in 1810.) Beyond the castle is Bury Ring, an Iron Age encampment. The church of St. Mary, once collegiate, is said to have been founded by Aethelflaed, daughter of Alfred the Great, around the shrine of St. Bertelin; the present building shows transitional Norman, Early English and Decorated styles. It contains a memorial to Izaak Walton (*q.v.*), a native of Stafford.

In pre-Reformation times Stafford had several priories. Since 1909 it has been the seat of a bishopric suffragan in the diocese of Lichfield. Its first charter was granted by King John in 1206 and 15 further charters were obtained from later monarchs. Richard Brinsley Sheridan represented Stafford in parliament from 1780 to 1806. Largely modernized, the town retains some picturesque half-timbered houses, including the gabled "High House" built in 1555, the ground floor of which is now a row of shops. Sir Martin Noel's almshouses are Elizabethan. There are a county technical college and a county school of arts and crafts. Stafford has manufactured boots and shoes since the 18th century and it has also locomotive, grinding wheel and electrical engineering works. Its salt industry has arisen through the availability of local deposits and the Brine baths are in a half-timbered building.

STAFFORDSHIRE, a midland county of England, bounded north by Cheshire and Derbyshire, east by Derbyshire, Leicestershire and Warwickshire, south by Worcestershire and Warwickshire and west by Cheshire and Shropshire. Its geographical area is 1,154 sq.mi. with a population (1951) of 1,621,034.

Physical Features.—The county falls physically into two regions—the Pennine uplands and lowland England. It may itself be divided into three main regions: a middle agricultural belt, the coalfield plateau and the northern uplands and coalfield. Stafford, the county town, lies in the wide middle agricultural belt of Keuper Marl. This belt is cut in two by a fault in the Triassic Sandstone which lies to the north and south. The fault to the south runs through Cannock Chase, a barren tract of former royal forest which, until comparatively recently, effectively divided the fertile plain from the more arid area to the south. The second region is the coalfield plateau of South Staffordshire, the southern part being broken by Silurian and Igneous rock ridges as the Rowley hills. To the west the plateau is lower and is cut by the Stour and its tributaries. The northern uplands and coalfield form the third region. The Potteries coalfield stretches from the northern Triassic area to the Cheshire border. To the north and east stretch the Millstone Grit moorlands, rising above Leek from 800 ft. to 1,500 ft. They are cut by the rivers Churnet, Manifold and Hamps, with their lower lying valleys. Folding and erosion of softer shales has caused a series of edges, as Ipstones edge. Much of the area is bleak and treeless. On the Derbyshire border an area of Carboniferous Limestone is cut by the Manifold and Dove. The greater part of the county falls within the catchment area of the Trent and its tributaries. Parts of northwest Staffordshire drain into the Weaver, and then into the Mersey. In the south and southwest drainage flows to the Severn through the Stour. The Trent rises high on Biddulph moor and flows through the Potteries into a wide valley. The only considerable natural sheet of water is Aqualate mere, near Newport, on the Shropshire border. Rudyard lake (1793), a source of supply for the Trent and Mersey canal, and the Blythe reservoir (1953) of the South Staffordshire Water board are artificial.

History.—There are remains of early habitation, especially in the northeast and isolated finds of the Bronze Age, indicating a movement along the river valleys. There is an Iron Age camp at Kinver. During the Roman occupation important roads were cut through the forests which covered the country, notably Watling street, running roughly east to west across the southern part of Cannock Chase, and Rycknield street, running south-southwest and intersecting Watling street at the military station, Letocetum (Wall) south of Lichfield. A second station on Watling street was Pennocrucium (near Penkridge). From the 7th to the 9th century the land formed part of the kingdom of Mercia. Tamworth, one of the most important of its towns, was much favoured by the greatest Mercian kings. Lichfield grew in importance after St. Chad, bishop of Mercia, moved his see from Repton. The Danes ravaged the county between 872 and 875 but did not settle permanently. In 910 they were defeated at Tettenhall by Edward the Elder. Three years later, Aethelflaed refortified Tamworth and Stafford. In 1016 Edmund led his army against the Danes into Staffordshire; both sides ravaged the county. This is the first mention (Staeffordscir) of the county in the Anglo-Saxon Chronicle.

Staffordshire remained poor and sparsely populated until after the Norman conquest, when the rising of 1069–70 made worse the waste, as is shown by the Domesday survey. There were five hundreds, Totmanslow, Pirehill, Offlow, Cuttlestone and Seisdon.

In the anarchy of Stephen's reign two of the greatest feudal lords in the shire, Henry de Ferrers of Tutbury (created earl of Derby) and Sir Robert Marmion of Tamworth favoured the king. In the Barons' Wars, many of the barons joined Simon de Montfort. In the Wars of the Roses the county generally favoured the Lancastrian cause; Blore Heath, site of the second battle of the wars (1459), is in Staffordshire. Henry Tudor marched through Stafford, Lichfield and Tamworth on his way to Bosworth Field (1485). Part of the county is still in the duchy of Lancaster.

In the reigns of Mary and Elizabeth I religious persecutions were severe. Mary, queen of Scots, was held captive at Tutbury and Chartley castles. The last stand of the Gunpowder plot was at Holbeach house, near Himley (1605). Catholics were again persecuted at the time of the Popish plot (1679).

During the Civil War, Staffordshire inclined toward the parliamentary side though there was strong loyalist support in some families. The battle of Hopton Heath (March, 1643) was the only important engagement on Staffordshire soil apart from sieges. Lichfield cathedral was damaged in the siege of the close (1643). Stafford, first held by the royalists, became the seat of the parliamentary committee which governed the county. Charles II on his flight after the battle of Worcester (1651) hid at Boscobel, Moseley Old hall and Bentley hall.

Staffordshire's later history is largely that of its industries. As the production of iron and its manufactures in the south and the mining of coal have long continuous histories, so has dairy farming in the central region. William Camden, in the reign of Elizabeth I, was not the first to write of the richness of the grazing lands nor of the pit-coal mines. He also mentioned glass making. Though the mines had long been worked, the output of coal in the north and Cannock coalfields probably increased rapidly during the later 16th century. Then the forest trees on Cannock Chase had largely been cut for ships and charcoal burning, but not until the 17th century did coal replace charcoal for smelting. Local records give a vivid picture of an industrial system among the blacksmiths and nailers in the neighbourhood of Walsall at the beginning of that century, pointing to the Industrial Revolution as an impetus to the development of mines and industries long established. The copper works near Cheadle received similar impetus. The making of pottery, though of earlier origin, became important only in the 18th century through the initiative of the great master potters, including Josiah Wedgwood. They appreciated the importance of cheap transport and supported the cutting of canals.

The University College of North Staffordshire was founded in 1949 at Keele.

Prominent Staffordshire men include Izaak Walton and Dr. Samuel Johnson; Elias Ashmole, the antiquary; Hugh Bourne, founder of Primitive Methodism; Sir Robert Peel, member of parliament for Tamworth and prime minister; and the two great admirals, Lord Anson and Earl St. Vincent. R. B. Sheridan was member for Stafford.

Architecture.—Ecclesiastical and secular buildings of architectural interest include the beautiful, small, red sandstone cathedral at Lichfield, St. Peter's at Wolverhampton, the parish churches of Checkley (11th–17th centuries) with three Saxon crosses, Clifton Campville (13th–14th centuries) with woodwork, Ingestre (by Christopher Wren?), Hoar Cross (1876), the Roman Catholic church at Cheadle (A. W. N. Pugin), Tixall gatehouse (*c.* 1580), Broughton hall (16th century) and the Oak house, West Bromwich (16th–17th century). At Croxden, near Uttoxeter, are remains of the Cistercian abbey. The medieval castles of Tutbury and Chartley are ruins; Tamworth castle is the town museum. In Brewood, Abbots Bromley and other towns and villages are groups of old houses.

Agriculture, Industry and Communications.—About 75% of the land is agricultural, Staffordshire being a milk supplying county. The rough grazing and grass of the moorlands give way to mixed arable and dairy farming along the Cheshire and Shropshire borders south toward Watling street. Cannock Chase di-

vides this area from the mid-east grassland-dairying. Arable land to the south includes market gardens around Lichfield and Tamworth. There, in contrast to the predominance of fodder crops in the rest of the county, such crops are outweighed by wheat, potatoes, sugar beet and vegetables. The Forestry commission grows timber on the Chase. The National trust owned 1,781 ac. in 1955 and protected 1,774 ac.

The Black Country and Potteries are the main industrial areas, but industries are spread through the county. Stafford has big engineering works, shoe factories and salt works. Shoes are also manufactured at Stone, sanitary earthenware at Armitage, silk and rayon at Leek, and paper at Tamworth, while Burton is famed for brewing. The Black Country is made up of a number of separate towns including Walsall, Wednesbury, Bilston and Wolverhampton, each with a particular industrial tradition. The area as a whole is noted for iron ore and coal mining and for the manufacture of hardware and engineering products.

The greater part of the railway network is operated by the Midland region of British railways. The main line, London to Crewe, runs through Tamworth, Lichfield (Trent valley) and Stafford. The Birmingham-Crewe line runs via Wolverhampton and Stafford and the cross line from Bristol to Derby runs through Birmingham, Lichfield and Burton. The Western region Birmingham-Shrewsbury line passes through Wolverhampton.

The canals, of which those forming part of the Birmingham navigation are today the most important, formed an important link in the late 18th and early 19th centuries. James Brindley, designer of the Trent and Mersey, or Grand Trunk, canal (1766–77) is closely connected with the county. This canal is an essential part of the inland navigable waterways system.

Population and Administration.—The area of the administrative county is 1,071 sq.mi. and its population (1961) 985,113. There are six county boroughs and eight municipal boroughs. The county boroughs are Stoke-on-Trent (pop., 1961, 265,506), Wolverhampton (150,385), Walsall (117,836), Smethwick (68,372), Burton upon Trent (50,766) and West Bromwich (95,909); and the municipal boroughs are Lichfield (14,077), Newcastle-under-Lyme (76,433), Stafford, the county town (47,814), Tamworth (13,555), Wednesbury (34,511), Bilston (33,077), Rowley Regis (48,166) and Tipton (38,091). There are 17 urban and 10 rural districts. Old market towns are Abbots Bromley, Brewood, Cheadle and Eccleshall. There are two cities, Lichfield and Stoke-on-Trent. The county is in the Oxford circuit and assizes are held at Stafford. There is one court of quarter sessions and 20 petty sessional divisions. Stafford and Wednesbury have separate commissions of the peace while Lichfield, Newcastle-under-Lyme and the county boroughs also have separate courts of quarter sessions. The county is divided into the archdeaconries of Stafford and Stoke-on-Trent within the diocese of Lichfield. From 1948 the county and parliamentary boroughs returned 19 members.

BIBLIOGRAPHY.—The William Salt Archaeological Society (now the Staffordshire Record Society), *Collections for a History of Staffordshire* (1880 *et seq.*); *Transactions* of the North Staffordshire Field Club (1866 *et seq.*); S. Shaw, *The History and Antiquities of Staffordshire*, 2 vol. (1798, 1801); C. J. B. Masefield, *Staffordshire*, 4th ed. (1930); Staffordshire County Council, *Development Plan* (1951); W. White, *History, Gazetteer and Directory of Staffordshire*, 2nd ed. (1851); W. H. B. Court, *The Rise of the Midland Industries, 1600–1838*, 2nd imp. with corr. (1953); J. Myers, *The Land of Britain, Pt. 61* (1945).

(M. Gz).

STAGE, in architecture, an elevated floor, particularly the various stories of a bell tower, etc. The term is also applied to the plain parts of buttresses between cap and cap where they set back, or where they are divided by horizontal strings and paneling. It is, too, the floor or platform on which plays are acted, whence the term has come to signify both the theatre and the drama (*q.v.*). *See* also STAGE DESIGN, below. From its etymological meaning of a station comes the sense of a place for rest on a journey, the distance between such places, etc. *See* also THEATRES (STRUCTURES).

STAGE DESIGN. Until about 1900 the pictures created on any stage, generally known as scenery, had no relation to the feeling for form, colour, composition or light which art from 1850 to 1900 had made a common possession. The picture of the world within the theatre was invariably dull, without illusion and rigidly painted,—painted, in fact, with the same merciless and metallic precision which led Manet to exclaim before one of Meissonier's paintings of charging cuirassiers, "Everything is of steel except their breast-plates." Trees stood invariably in regimental rows and their foliage hung in separate parallel layers above their trunks. The walls of hovels or palaces reminded the most careless eye that they were merely painted canvas; the sky flapped, an obviously painted sheet. The ugliness of the total result was all the more glaring because, in an age of realistic playwriting, actors were using every artifice they possessed to make an audience feel that they were Russian or Silesian peasants, Norwegian town councillors, or English clergymen and poets. Every word in the theatre was trying to convince its hearers that what they saw was not actors, but human beings in a world as real as the one they had just left to enter the theatre; that people "were like that"; that life had this significance and that meaning. But everything on the stage reminded the audience that, after all, what they were seeing was a play in a thoroughly artificial place, the theatre, that had no relation to life. Even when a poet held the stage the result was equally ludicrous. It was difficult for the singers of Wagner to transport an audience to the beginning of the world, where the gods were deciding the destiny of men, when everything about them seemed an enlarged landscape done in chromolithograph. Mary Garden wandered in a park that had neither magic nor mystery, in the supposed gloom of palace halls as bland as lithographs of English country houses, current in the '50s. Any symbolism that might have lurked in Maeterlinck's fairy tale of the terror and foreboding of young love was dissipated; and Mélisande's cry "I am not happy" seemed less a comment on the state of her soul than on the preposterous stage settings that surrounded her.

The inadequacy of this type of scenic background, so prevalent at the turn of the century, did not lie alone in its artificiality nor in the fact that it was an obvious and flimsy convention. Artificiality and the conventions of formally painted designs, as proved by the work of Inigo Jones, Bérain and the Bibienas, can breed beauty in the theatre just as easily as the same tradition in the 19th century bred ugliness. The reason for the change which makes painted scenery in one epoch enhance the meaning of the play and in another detract from it, is our changed attitude towards the theatre. The creative dramatists of the past 40 or 50 years, whom for convenience sake we may call modern, such as Ibsen, Shaw, Hauptmann, Chekov and Strindberg, Synge and O'Neill, are modern in the sense that one and all they do not regard the theatre as a place of pure entertainment. And they have succeeded so well that though we may still go to the theatre merely to be amused, we also go to watch a play with the same expectation with which we open a novel treating of modern life—in the hope of seeing more clearly into the springs of character, of getting a fresh sense of human destiny.

The Modern Scenic Movement.—The blatant dissonance between modern plays that were attempting to reconceive a world and scenic backgrounds that merely reflected the conventions of a purely theatrical theatre, was so preposterous that it could not endure. For that reason, directors of modern theatres in every country in Europe, beginning with Germany, and finally in the United States, enlisted painters or designers to make the world on the stage the world of the play. This is the genesis of the so-called modern scenic movement. It arose, not because modern art already existed, but because it became essential to the theatre in order to quicken the reactions of the audience until the world the dramatist asked them to believe in had the plastic reality of a world before their eyes or the imaginative quality of a poet's dream. The source of modern scenery has not been dogma. It did not arise because Gordon Craig prophesied that only screens could back a stage; modern scenery would have occurred whether or not Craig had ever existed or penned a line. He, however, was one of those who prophesied that a change must come and that stage settings must be designed to interpret the play they housed. Every manner and method of modern painting, has, with

the aid of some mechanical device, been put to work in the theatre: the plaster dome, in conjunction with the flexibility of incandescent lamps which made the painted sky seem a heaven and filled the stage with all the ambient light of day; the spot-light, manipulated to create the chiaroscuro of Daumier; light mixed with colour in the manner of Monet and Seurat making colour vibrate, and achieving the atmosphere of the impressionists, or enmeshed in gauze, giving the atmosphere of Whistler's nocturnes. Every technique of modern art has been exploited, whether of impressionism, post-impressionism, cubism or futurism. However, from the point of view of pictures constructed within the proscenium frame of the theatre, almost every method at one time or another, in the hands of a particular director and in conjunction with the methods of his actors, has succeeded in convincing an audience and projecting the interpretation of a dramatist's ideas.

The point needs stressing because the history of modern scenery is usually regarded from the point of view of a mere chronology of begetting, and is continued in the strain of prophecy. A designer reading any of the current histories of his art would feel that the entire problem before him was the choice of a style. And the value to him, as a craftsman, of almost all discussions of the subject, is hereby rendered nil. The theatre, it is supposed, is about to be re-created, purged of evil, saved and given its final form if he, the designer, only makes the right choice; screens, a stage without curtains, scenery made of nothing but curtains, a circus theatre where the actor can mingle with the audience, scenery that is only light, scenery that is merely symbolic colour, scenery that is wooden trestles, and so on, depending upon which method at the moment has had the latest critical success in Europe. It is essential for the scene designer to avoid this critical approach if he wishes to analyse the basis of his craft. For whatever his preferences may be as a painter, he will find that the moment he sets out to design stage settings his problem is everywhere the same. He will find the problem of relating a particular play to a particular audience, the problem, as conceived by the director, of how the meaning of that play can be put over the footlights, and dramatized for that audience. His methods will be deduced not from any abstract aesthetics of the theatre but from the exigencies of that particular situation. For not all his convictions as to why imitation and realism was a defunct method prevented George Bernard Shaw from continuing to write plays so provocative that they must be acted in country-house gardens, Salvation Army barracks and physicians' consulting rooms, for which the designer has to build the walls, find the furniture and even evoke the flowers. On the other hand, not all the dogmas in the world as to the value of realism prevented Shaw, or any other playwright, from reconstructing Antony's adventure in Egypt or Adam's adventure in the Garden of Eden, for which there is an equally avid audience; and the designer will have to find a decorative equivalent for the sphinx, or some symbolic equivalent of the tree of knowledge. And whether theatre audiences are those of London, Prague, Berlin or Warsaw, the designer will find that they came to the theatre to share experiences that are not their own, so that in the course of a single season the designer may have to help in making plausible the romantic scruples of a Hungarian schoolboy or a Hungarian servant, the feudal scruples of Japanese noblemen, the political dogmas of Slavonic revolutionists, the moral codes of American revolutionists and Russian peasants.

Staging of "The Faithful."—Some of the problems of scenic design as worked out by the present writer at the Guild theatre of New York, although not final in any sense as examples of theatre art, are typical of the problems which every stage designer must face in order to evolve design, whatever the pattern he may ultimately adopt. *The Faithful,* by John Masefield, tells the story of 47 Ronin, retainers and henchmen of a Japanese nobleman, tricked by a rival into committing an affront to the emperor's person for which the only expiation was suicide; of the voluntary exile of these Ronin as pariahs, and their final triumph over their dead master's enemy. It is a poet's celebration of the nobleness of loyalty. What is the producer's problem? Obviously something more than to demand poetic background for a poetic play, because the tale of the 47 Ronin, as much of a household story in Japan as Washington and the cherry-tree in America, is totally unfamiliar to Western audiences. He can count on no background in the audience's mind. But Japan is the background which he must create decisively five minutes after the curtain is up, for the reason is that the unforgivable act committed was nothing more than a mistake in ritual, a failure of ceremonial observance, not to the emperor himself, but vicariously to the emperor's person as personified by his envoy. The rigid code of Japanese feudalism is a tradition difficult for Westerners to conceive. The danger is that they will find it, if not preposterous, at least not plausible, so that they cannot identify themselves with the hero. The problem here is that to an audience not composed of orientals or members of Japanese societies, the normal reaction would be "Stab oneself because of a mistake in court etiquette? How preposterous! What a trivial people!"

The playwright cannot solve this problem in the many ways that are offered to the novelist; he has no preliminary chapters or incidental comment to prepare the mind of his audience, as his story proceeds, to make them understand the rigid Samurai code of honour and accept the tragic alternatives imposed on its adherents. Aware of this problem, the designer finds, in making his first researches into the field of Japanese costume and architecture, that ritual ruled not only court etiquette but all Japanese life of the period, to the most trivial detail. Everything is prescribed and ordered; the degree to which the head is shaved, and the degree to which the hair is looped forward on the head; the precise way the obi or white silk belt is tied that holds the pleated skirt in place; the length of the sleeves, as well as the cut of the gown. Everything is ordered ceremonial, even the squares of the matting, the placing of the sword in its rack, the one picture niche where the solitary landscape scroll is shown and the way the single group of flowers is arranged under it. There are in the simplest acts, of receiving a friend and taking tea in his presence, countless opportunities of outraging him by breach of etiquette.

In a world of this sort, it begins to be comprehensible that a mistake in court etiquette could involve the death penalty. Hence every effort—and here the producer co-operates—is made to convey the ceremoniousness of the Japanese feudal world, in the gait of the actors, their gestures, their genuflections, the mode of entering or leaving a room. A Japanese is found to train them; he ties every obi, adjusts every costume, sets every hat and wig. The room which is the opening scene of the play is almost archaeological in its accuracy, and the line of every costume is established with realistic nicety and with an attention to correct cut and material that could hardly be outdone by the Moscow Art theatre. In short, realism as a method is employed in order to make a poetic theme plausible.

On the other hand, once the world of the Samurai is realized, and the tragedy is enacted, there is no further need of such realistic devices. The ragged Ronin wander in the snow, but the snow may be the snow gorges of Japan, painted upon a six-fold Japanese screen, without diminishing either the reality or the suffering of these outlaws. Here a deliberately decorative method seems the inevitable choice. Japanese landscape is known to Western audiences through the popularization of Japanese prints. A screen setting evokes this purely decorative pictorial tradition and suggests the countryside of Japan far more effectively than any attempt to reconstruct it realistically. Thus meticulous realism and deliberate decoration can be employed simultaneously in the same play, and for the same purpose. As pictures the snow scenes are far more effective. But in the theatre the palace scene was far the more important of the two. For if that scene succeeded in conveying Japan effectively to the audience, the world of the Japanese could be taken for granted for the rest of the evening. A white sheet and the shadowgraph of a few pine boughs might have done just as well for the winter hills. This instance reveals how superficial it is to codify stage sets according to purely pictorial principles.

Staging of "Liliom."—Molnar's *Liliom* presents an analogous problem of transplantation. not from one century to an-

other but from one country to another at the same period. The producer's problem in Liliom was, in essence, that his story deals with the creed of a gangster and a "tough" who is fundamentally romantic. Liliom falls genuinely in love with the servant, Julie, and evinces a belief in "the eternal feminine" worthy of a latter-day Faust. He is transported by the prospect of fatherhood. When he stabs himself rather than be captured by the police in an attempted hold-up that fails, his last cry is "Julie . . . my little girl . . . my little cricket." When, after facing the judgment of heaven and ten years' probation, he returns to earth as a beggar for an opportunity to win redemption, he steals a star to give to his daughter, and then in a moment of exasperation beats her, as he formerly beat her mother. The child cries "he hit me . . . but I didn't feel anything—as though someone had stroked my hand . . ." "Nothing has happened" her mother assures her, divining who the beggar is, "it is possible, my child, for someone to hit you—so that you don't feel any pain." Liliom is not forgiven in heaven; he is redeemed on earth by the persistence of human compassion. The beauty of the final scene made the play at the Guild theatre an allegory, the vindication of a romantic faith, and dictated the scenery. The play was something more than the story of a foreign underworld; its realism was made an ironic cloak for poetic truth. Thus the amusement park where Liliom flourished as a bully was made as authentically Hungarian as possible, so that the audience might accept the reality of a romantic gangster and not reject him as a maudlin invention. On the other hand, precisely because the play seemed important as an allegory because of its poetic truth, the tawdry squalor of Liliom's surroundings had to be invested with beauty, even the tumbledown shack in which he lived. And the railroad embankment, where he attempted his hold-up and met his death, was composed with some of the dignity and severity appropriate to tragedy.

Scene 2 represents the park where Liliom falls in love. The point is that, bully, seducer and braggart though he is, he does fall genuinely in love. He is transfigured and the moment becomes as beautiful for this Hungarian hooligan and a kitchen maid as it does for all lovers. Therefore, the park where they meet must be made beautiful. The hour is dusk, when, as Whistler reminded us, "the evening mist clothes the riverside with poetry as a veil, and the tall chimneys become campaniles, and the warehouses are palaces in the night and the whole city hangs in the heavens"; the "exquisite hour" as Verlaine calls it. Therefore the problem for a designer was to clothe that park with poetry as with a veil; with tree silhouettes which are hung on transparent gauzes; and putting into those tree silhouettes all the grace and the loveliness appropriate to the trees of a pastoral bower destined to arch over young lovers. Had the scene been one of cynical seduction, this quality of tenderness and brooding would have been irrelevant.

Scene 6 represents heaven. Liliom sees heaven in terms of the police courts with which he is familiar. He can conceive of no other bar of judgment; heaven is merely the last police court presided over by a police-court judge; the attendant angels are to his eyes only detectives and policemen. The stage directions emphasize the fact by stating that their costume is the usual police uniform. Nevertheless, Liliom has died; he is in heaven awaiting eternal judgment. If the play is to be properly interpreted this seat of judgment cannot be the proverbial heaven with a celestial throne. On the other hand, if the allegory of the play is to be made plain it cannot be shown simply as a police court. This dilemma was solved by making the scene as far as Liliom's eye ranged a drab room, with the ordinary judge's bench at one end; but the solid walls ended at the wainscoting and window frames. Beyond that was emptiness; the room sat without walls or roof under a blue sky as illimitable as stage light could make it. By this combination of realism and fantasy, the reality of heaven was established for the audience, and at the same time the fact was made clear that to Liliom its judgments were no different from what he had experienced on earth. These four scenes from two plays entirely different in type demonstrate how contradictions of realism and decoration in stage pictures are determined by problems of dramatic interpretation and not primarily by pictorial considerations. The picture is a result of a conviction as to what the meaning of a play is and how it can be reflected in pictorial form.

The Method of Abstraction.—Settings that eliminate the stage picture entirely are designed for the same reason and by the same process. *The Tidings Brought to Mary,* by Paul Claudel, was played on the same platform for its entire five acts, which according to the stage directions take place in a barn, a farm kitchen, an orchard and a road at the edge of a forest. Fedor Komisarjevsky, the director, chose the unit platform because he felt that the play was mediaeval not only in its story but in its quality. It was a mystery, couched in the mystic faith of a truly mediaeval catholicism. A single square block served as a bench and, when it was covered with a formal cloth, as a supper-table and an altar. The orchard became eight nuns clothed in Byzantine splendour, each holding a branch of silver flowers, who stood in a remote semi-circle. The sky behind them at this moment became green to suggest the foliage of out of doors. For the road-side scene at night, the same backdrop became the inky blue of a wintry night spangled with a few stars. And a cold steel-blue light thrown over the platform gave the illusion of winter snow. For the miracle the sky blazed forth gold, like the background of a 14th century mosaic, and aureoled the actors. The same intention was carried through the costumes: many were mediaeval French, but others—those of the father and mother—were clad in the traditional robes of early religious paintings, so that the mother had a suggestion of the Mother of Christ, the father, of one of Christ's disciples, and in its formal grouping, the supper of farewell suggested an apostolic meal.

In *Man and the Masses,* by Ernst Toller, the formal platforms became almost invisible. The settings were made wholly by the massing and the movement of the actors. The human beings on the stage provided all the scenery that was needed, for the play was conceived as an abstract allegory. One leader who tries to limit the workers to the bloodless methods of a general strike, is called simply The Woman. The agitator who incites them to revolution is called The Man. Though the workers are supposedly in their union hall, their angry indictments of capitalistic society are chanted in unison as a formal chorus. How could this meeting be placed in a realistic setting of an actual union headquarters, without making such formal choruses seem ridiculous? The title literally translated is "mass-man." The force of Toller's conception is his vision of mass-man. The working-men are a single stubborn unit, welded together in resentment and anger, from which only voices of the young and the old emerge to cry, threaten and lament. If a group of 30 or 40 actors on the stage are to give this feeling of mass power, they cannot be scattered, they cannot move about much; they must be welded. Hence a stage setting was devised for the drama that was a literal cross-section of an amphitheatre; the players never moved from the spot where they were wedged; they rose with waving arms and clenched fists to face The Woman, who pleaded with them; they towered over her; she was below them, supplicating. When this compact mass of 30 actors stood they seemed to represent "the masses"; it seemed that they could sweep down on The Woman and overwhelm her. Then The Man sprang out of their midst and slowly backed up step by step to the top of the amphitheatre. As his eloquence grew they swung slowly round, looking up towards him, turning their backs on The Woman as they ignored her pleas. Finally, at the moment when they were won over to a doctrine of violence, they were a solid mass, crouching under him waiting to spring; as he cried "revolution" they took one step up, the only step they took in the entire act, and lunged together like a huge beast, echoing his cry "revolution." Each one had turned completely round during the scene without once moving from the spot on which he or she stood. They remained a mass, and by a single mass movement expressed their change of faith.

Varieties of Interpretation.—The designing of scenery is thus never a purely pictorial problem; it is part and parcel of the act of interpreting a play. Whether a designer chooses abstraction or realism, a picture-stage which imitates the world as we see it

in rooms with the fourth wall knocked out into which we peep, or a formal stage which symbolizes the world, is in itself a relative thing. The method of design chosen is only good even from a pictorial point of view provided it touches the imagination of an audience, and really convinces it of the truth and reality of the play. Setting the play is only part of the process of acting it so as to give it dramatic force; no method, however beautiful it is as a picture, is the right method if it fails to convince an audience; and there are as many types of audience as there are kinds of play. No play has only one meaning; it can mean all things to all men. The imagination of the director must decide and the pictorial imagination of the scene designer must second him. If you can make your audience accept Liliom's story as a modernistic fairy-tale, an allegory from the outset, Pitoev's method of presenting it in clear hard formal outline is the right one. If the play means more as a story in which the allegory is implicit, then the more realistic method of the Theatre Guild is better; similarly with *He Who Gets Slapped.* To compare the two, one would have to sit through the two productions as played by two different casts.

There is no one way of producing any play, even classic masterpieces. In fact every one, whether Euripides or Shakespeare, has to be constantly re-interpreted, in order to become relevant and remain living. If Shakespeare's characters seem important as figures rescued from an actual historic past then the elaborate and exact costuming of the Duke of Saxe-Meiningen or Henry Irving is not only necessary but effective. To one audience, which conceives them primarily as creatures of Shakespeare's time, the bare boards of the Elizabethan stage will make them live more completely. To another audience, particularly sensitive to Shakespeare's word-magic, formal scenery, even a single curtain, is enough. To a German audience, for whom much of the poetry is lost in translation, more pictorial backgrounds are necessary, into which that lost glamour must be projected. But to both German and American audiences for whom Shakespeare's protagonists are imaginative figures living in the world of the imagination, they become most real when part of an unreal world, a forest of dyed strips of tapestry or gauze (Barker's production of *A Midsummer Night's Dream* and Reinhardt's production of the same), or a blood red stairway (Jessner's production of *Richard III.*), or among golden screens (Craig's production of *Hamlet* with Stanislavski at the Moscow Art theatre). And if the tragedy of Hamlet begins to seem too remote and archaic to us, we can make him a human being, for certain audiences, by putting him in modern dress. The art of scene designing is a living art because it can find no final formula. It will live precisely as long as it can discover new meanings in every play of every epoch. Scenery is the background of a play, and as such is part of the continuous adventure of projecting the meaning of a play across the footlights, finding, firing, and then fusing the imagination of an audience. And that adventure will always be a different one in every country and every epoch.

Structural Innovations.—The technical and structural changes which determine the design of modern scenery are two: first, the modern theatre, in contrast to the 18th century theatre, has become an extremely shallow box; second, that black box can be completely flooded with electric light, modulated and controlled. The theatres of the late 17th and 18th century fixed the architectural type of theatres, which with changes of ornament and the depth and arrangement of balconies is the type still prevalent to-day. But there is one vital difference: the proportion of the space occupied by the stage in relation to the space occupied by the audience in the auditorium. The 18th century stage averaged in depth 1 to 1½ times the depth of the auditorium pit, and was rarely less than 75 feet. The modern stage, except in opera houses and European state theatres, is rarely more than ½ the depth of the auditorium, and averages at the best 25 to 30 feet.

Moreover the 18th century stage was carefully calculated to enhance painting in perspective. The stage floor sloped up The side flats receded until the rear of the stage was very near the true vanishing point, so that an alley of trees or an avenue of columns receded very much as they would in nature. But on the modern stage the spectators are looking into a space so shallow that any illusion of depth is difficult to attain. In an average production there is rarely more than 8 to 10 ft. from the garden wall to the distant mountains; often there can be only 4 or 5 ft. from a window to the sky.

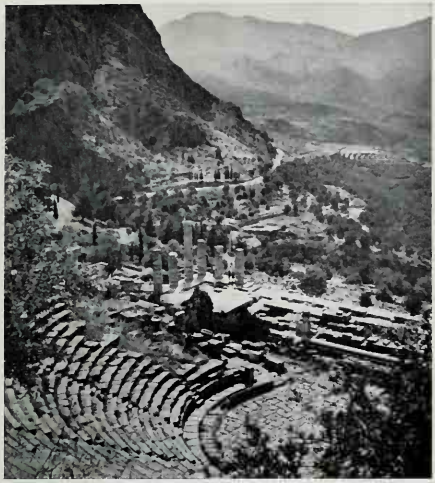
The early technique of shifting scenery was done by having the side flats on grooved slots, set in pairs, tied to a rod which connected with a series of windlasses in the cellar. At a given moment one turn of these windlasses would shoot the rear flats out like a shutter, and withdraw the front flats out of sight, achieving a complete change of scene with great precision. A beamed gridiron under the ceiling, with another series of windlasses, hoisted and lowered the clouds and the platforms on which gods and goddesses descended. Owing to the shallow space of the modern theatre and the suppression of the apron in front of the curtain, the front rows of spectators can see clear to the roof and to the side walls. The modern stage settings therefore have had to become a three-sided box with a lid over the top, when interiors are called for; for exteriors a curved sky, either of canvas or of plaster is used, instead of a sky in flounces as seen in the 18th century theatres. There is so little room on the stage for stacking scenery that most of it has to be broken apart, or unlashed, for each change of scene, and hoisted into the air where it hangs until it is needed again, and as certain sets may be 24 and sometimes 30 ft. high they have to be hoisted high enough to be out of sight when not in use. Hence the gridiron from which scenery is hung—handling is facilitated by having the lines that haul it attached to counterweights grouped at one side of the stage,—is the most important part of standard stage equipment. The opening which the audience sees, the proscenium frame, is in reality a mouse-trap shutter at the bottom of a tall shaft. The average proscenium opening is 24 to 30 feet. The curtain in actual use is rarely drawn higher than 20 ft. for exteriors, and 12 to 14 ft. for interiors; the remaining space is occupied by flounces and draperies. And the gridiron must be a minimum of 65 ft. in height; for hauling scenery out of sight, 80 to 85 ft. is better. At the Century theatre, New York, it was about 100 ft. above the stage floor.

In any case the space behind the stage opening is mainly storage space for hanging scenery in mid-air. The place occupied by the largest scene is approximately one-seventh of the total stage space, and the proportion will not vary greatly in American theatres or most commercial theatres of Europe.

The system outlined above is not by any means the best either mechanically or aesthetically. It is, in method, as archaic as hoisting sails; indeed the workmen on the stage who shift scenery are known as the crew to this day. The revolving stage introduced by Lautenschläger from Japan was used extensively, and did away with the necessity of a gridiron almost entirely. At Dresden, Linnebach at the State theatre and Hasait at the Opera constructed elevator stages worked by hydraulic power, so that heaviest sets could be set in the cellar, hoisted to stage level and slid into place with extraordinary speed. The cellar being two or three storeys deep, the settings for two or three entire productions could be kept intact indefinitely.

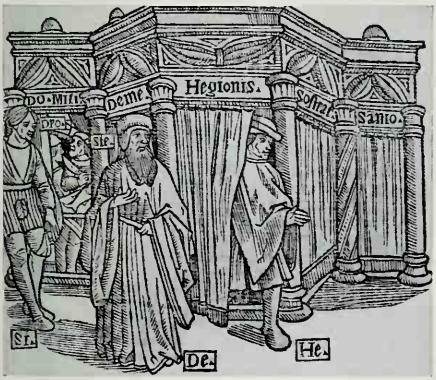
These mechanical marvels have never developed because the economic trend of the time is against them. Land has become too expensive to allow the amount of space and excavation required; rising costs everywhere in Europe make the duplication of any such installation prohibitive. In the United States where a single elevator platform 12 by 30 costs approximately $25,000 to install, an entire stage so equipped would represent a capital investment that only the very wealthiest theatrical producer could carry. Thus theatres tend to become smaller and shallower and the theatre everywhere approximates to American conditions where everything has to be crowded into a standard building site of 100 ft. square.

Important architectural experiments have been made which aim at changing the conventional "peep-show theatre," as its critics have called it, where we look into a stage picture through a more or less gilded frame. Theatres have been projected with the stage as a formal platform, where the actor emphasized in space, picked

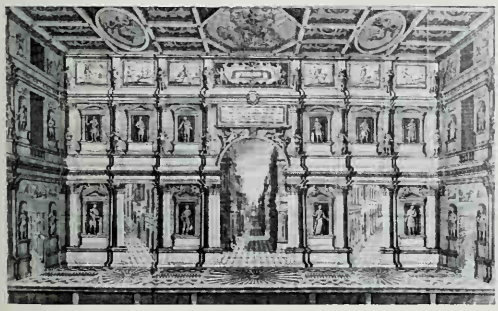


Remains of the Greek theatre at Delphi with the valley of the Pleistus as the natural background to the stage

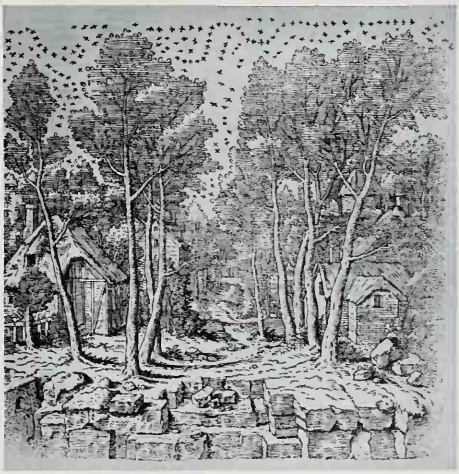
Close-up of the stage and the central portion of the *frons scaenae* (façade of stage building) of the partially restored Roman theatre at Orange, France, c. A.D. 50



Scenery for the *Adelphi* by Terence from an illustration in the edition of Terence's plays published in Lyons, France, in 1483. Victoria and Albert museum



Engraving of the permanent scene for the stage of the Teatro Olimpico, Vicenza, Italy, designed by Andrea Palladio (1508–1580) in 1580 and completed by Vincenzo Scamozzi (1552–1616) in 1585. Victoria and Albert museum



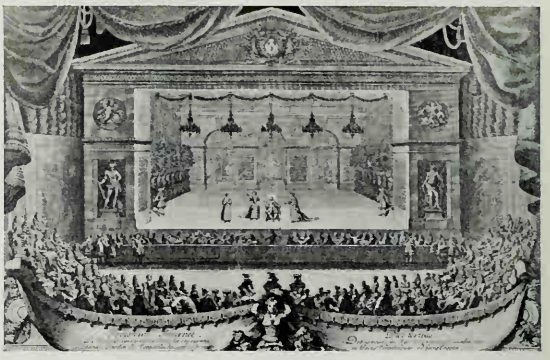
Stage design by Sabastiano Serlio (1475–1554) for a scene of a satiric play from the English edition (1611) of Serlio's book *Architettura* published in 1545. The British museum

GREEK, ROMAN AND RENAISSANCE STAGING

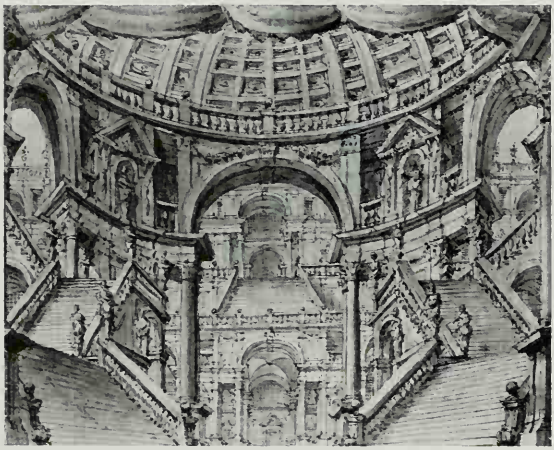
BY COURTESY OF (CENTRE LEFT, BOTTOM LEFT) VICTORIA AND ALBERT MUSEUM, (BOTTOM RIGHT) THE TRUSTEES OF THE BRITISH MUSEUM; PHOTOGRAPHS, (TOP LEFT) THE MANSELL COLLECTION, (TOP RIGHT) THECLA, (BOTTOM RIGHT) JOHN R. FREEMAN & CO.

STAGE DESIGN

Stage designed by Inigo Jones (1573–1652) in the Italian manner for a production of *Florimène* in 1635. Devonshire collection, Chatsworth, England



Setting for a performance of Molière's *Le Malade imaginaire*, shown being given at the palace of Versailles in 1674. Engraving by Le Pautre dated 1676. Victoria and Albert museum



Drawing for an interior setting by Giuseppe Galli Bibiena (1696–1757), a member of a family of Italian scenic artists whose work greatly influenced stage design throughout Europe and particularly that of the operatic stage. Victoria and Albert museum

Back cloth for a prison setting by Philippe J. de Loutherbourg (1740–1812), scenic designer for David Garrick at the Theatre Royal, Drury Lane, London. Victoria and Albert museum

A scene from *Miss Sara Sampson*, by Gotthold Lessing, designed about 1884 by George II, duke of Saxe-Meiningen (1826–1914). The Staatliche Museen, Meiningen, Germany

Detail from the sketch by Victor Simov for the scenery of Anton Chekhov's *The Seagull* staged by Konstantin Stanislavski (1863–1938) at the Moscow Art theatre in 1898

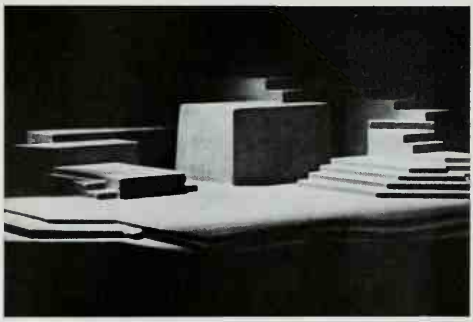
STAGE DESIGN OF THE 17TH, 18TH AND 19TH CENTURIES

BY COURTESY OF (CENTRE LEFT AND RIGHT) VICTORIA AND ALBERT MUSEUM, (BOTTOM LEFT) THE STAATLICHE MUSEEN, MEININGEN, THURINGEN, GERMANY: PHOTOGRAPHS, (TOP LEFT) COPYRIGHT DEVONSHIRE COLLECTIONS, THE TRUSTEES OF THE CHATSWORTH SETTLEMENT, (TOP RIGHT) GIRAUDON, (BOTTOM RIGHT) NOVOSTI PRESS AGENCY, LONDON

Naturalistic setting by David Belasco (1853–1931) for his New York production of *Tiger Rose* in 1917

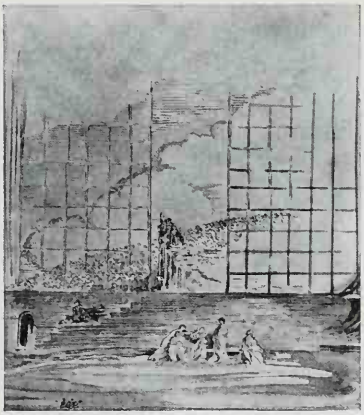
Symbolist setting by Lee Simonson for George Bernard Shaw's *Back to Methuselah*, New York Theatre Guild, 1922

Sketch by Adolphe Appia (1862–1928) for a scene in Wagner's *Die Walküre*, 1892, showing one of his plastic sets, a sculptured or architectonic structure shaped to the movement of the actor



Norman Bel Geddes (1893–1958) stage design for *Hamlet*, 1931, a non-realistic setting that combined many stage levels with special lighting to evoke the varying moods of the play

Stage design by Jo Mielziner for a scene in Arthur Miller's *Death of a Salesman*, 1949, a unit setting in which different parts of a house are visible at the same time



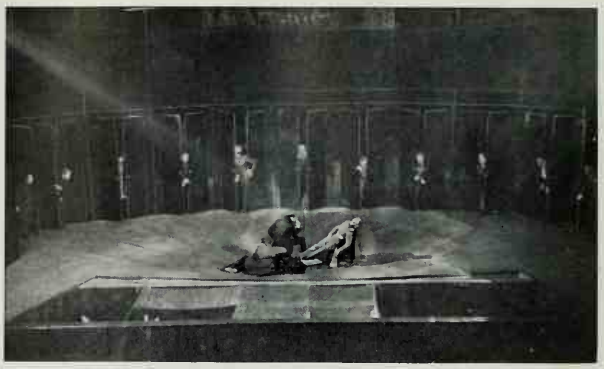
Design for the forum scene in Shakespeare's *Julius Caesar* by Edward Gordon Craig, 1922. Victoria and Albert museum

19TH- AND 20TH-CENTURY STAGE DESIGN

BY COURTESY OF (TOP LEFT AND RIGHT) THEATRE COLLECTION, NEW YORK PUBLIC LIBRARY, (CENTRE LEFT) FONDATION ADOLPHE APPIA, COLLECTION SUISSE DU THEATRE, BERNE, (CENTRE RIGHT) HOBLITZELLE THEATRE ARTS LIBRARY, UNIVERSITY OF TEXAS, (BOTTOM LEFT) JO MIELZINER, (BOTTOM RIGHT) EDWARD GORDON CRAIG, ESQ.; PHOTOGRAPHS, (BOTTOM LEFT) PETER A. JULEY & SON, (BOTTOM RIGHT) VICTORIA AND ALBERT MUSEUM

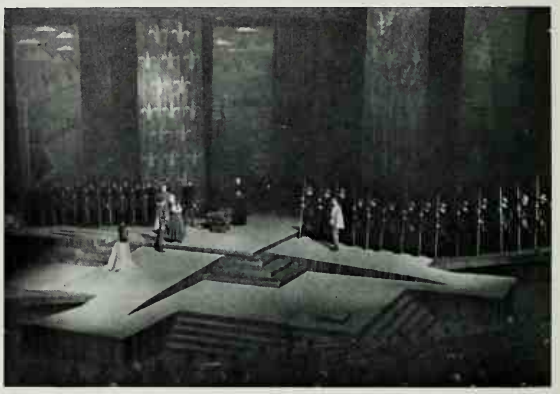
STAGE DESIGN

A simultaneous setting by Robert Edmond Jones (1887–1954) for Eugene O'Neill's *Desire Under the Elms*, Provincetown Playhouse production, 1924



The anti-illusionist stage of Vsevolod Meyerhold (1874–1942) designed for his production of Nikolai Gogol's *Revizor* (*The Government Inspector*), Moscow, 1926, showing the mahogany doors, and movable sloped platform in the style of the Greek exostra

Set for Molière's *L'École des femmes*, Paris, 1936, directed by Louis Jouvet (1887–1951), one of many directors who rejected naturalism and symbolism in favour of artificiality in staging



Scene from Friedrich von Schiller's *Die Jungfrau von Orleans* designed by Paul Walter and directed by Hans Schüler at the National theatre, Mannheim, Ger., 1957–58

Epic setting for a scene in *Mutter Courage und ihre Kinder* staged by the author, Bertolt Brecht (1898–1956), and Erich Engel and produced by the Berliner Ensemble in 1949

Theatricalist setting by Oliver Smith for the Ascot Gavotte scene in *My Fair Lady*, a musical adaptation of Shaw's *Pygmalion* by Alan Jay Lerner and Frederick Loewe, New York, 1956

20TH-CENTURY STAGE DESIGN

BY COURTESY OF (TOP LEFT) THEATRE COLLECTION, NEW YORK PUBLIC LIBRARY, (TOP RIGHT) SOCIETY FOR CULTURAL RELATIONS WITH THE U.S.S.R., (CENTRE LEFT AND RIGHT, BOTTOM LEFT) MORDECAI GORELIK COLLECTION, PHOTOGRAPHS, (CENTRE LEFT) LIPNITZKI, PARIS, (CENTRE RIGHT) ADOLF FALK FOTOGRAFIE, MANNHEIM, GERMANY, (BOTTOM LEFT) THE BERLINER ENSEMBLE, (BOTTOM RIGHT) FRIEDMAN-ABELES

out by light, can mingle freely with the audience and is partially surrounded by them, a sort of electric contact being the goal sought. The most notable theatres of this sort have been the projects of the American Geddes, the Dutchman Vanderwelde and the Austrian Stranad. Jouvet's platform stage designed for Copeau existed in Paris for a number of years; a circus theatre by Poelzig, "the theatre of the 10,000," was actually built in Berlin, but like Copeau's theatre it was abandoned. Quite apart from economic problems, what thwarts the development of these playhouses is the fact that modern playwrights will not write enough plays that can be given in any such manner. O'Neill and Shaw, Toller, or Werfel, revert to realism quite as often as they depart from it and write plays which require the three-sided box room or the illusion of distance in an actual world which can only be achieved, in the theatre, by looking through a frame. Until modern playwriting achieves some unity of tradition, these formal stages must remain isolated experiments or dreams on paper.

Lighting.—The most important effect that the shallow-box stage has had on designing is the fact that perspective planning and painting become preposterous; anything that leads the eye from the front of the stage directly across it to the back wall, makes the eye realize that the picture is only 20 ft. deep. Hence the kind of composition modern scene design employs is not the Renaissance composition of distance into which the eye is led by receding parallel perspective, but distance suggested in the Japanese manner by planes cutting one another like the distant peak of Fujiyama appearing directly over the brink of a hill. The eye, leaping from one to the other, imagines the distance and gets the suggestion of space. Thus a space in modern settings is suggested rather than revealed: the base of one column will suggest an entire temple; one pillar in a gloom pierced by a stained glass window will convey an entire cathedral, and because there is no intermediate form between the pillar and the window, the blackness will seem as deep as the nave of a church. Even the window may not be shown but a shaft of light may be thrown from a point out of the spectator's range of vision in order to increase the scale by the same process of suggestion.

Light is the scene-painter of the modern theatre; light, not paint, draws the scene, and by varying planes of intensity, models it, and makes the corner of a building solid, a hill remote, a sky infinite though it be only two arms' length from the garden wall. If a stage 25 or 30 ft. deep is flooded with even radiance, nothing will persuade the eye that it is anything but a shallow box, however suggestive the arrangement of the setting itself may be. Settings are constructed now as in the Renaissance period, of canvases stretched on wooden frames and hinged or lashed together, but the paint on them is a mere preparation for the light to be thrown on. No setting can be designed until one is certain how one is going to light it; the colour painted is conceived as an undertone of what it will seem when lighted on the stage.

The railway viaduct in *Liliom,* for example, if it is to seem massive and menacing must have the light thrown across it, not at it, otherwise it will become flimsy canvas. Thrown from right to left the single pier casts a sharp shadow which gives a sense of solid masonry. The whole forepart of this scene must have a different quality and intensity of light from the distant silhouette of factory chimneys. With the right balance of light they will recede and seem distant even though they are actually only 8 ft. behind the arch culvert. And both foreground and background must be properly related to the sky if they are not both to fall into it.

So much for the scene; but there are the actors. They must be modelled and emphasized in the areas in which they move; sharply picked out at certain moments, where the action calls for it, at others almost blotted into half shadow, for this scene is one of ambush, of two thieves waiting to rob a cashier, and a rim of shadow into which they can retreat is part of the process of creating the atmosphere of terror and suspense when the crisis comes. Every scene has pools of light waiting for the actors, light which we do not see or know is there until the actor walks into it at a given moment.

Emotionally we react to light far more quickly than to any other medium in the theatre. We react to it even more quickly than to the actor's voice. In fact it sets the key of our emotions, and determines what effect the actor's voice will have on us. Lights not only paint and model the setting, emphasize its planes, and define its form; light also sustains the mood of a scene, and dramatizes its meaning. Crude or blatant lighting can war against the emotions an actor is trying to convey, and often destroy them entirely, all the more so because the lighting of any scene often does not remain static, but is dimmed or heightened, pulses and fluctuates with the action, like an orchestral score.

Change of light alone created every difference essential to the five acts of the *Tidings Brought to Mary.* The park scene of *Liliom* consisted of two gauzes on which the silhouettes of trees were pasted. The entire mood of the scene's poetic dusk was created by a balance of lights. In *Man and the Masses,* the contrast between the mystic dream scenes and the sculpturesque force of the mob scenes is entirely a matter of lighting. In the course of the play there were several hundred subtle shifts of lighting timed to coincide with the action of the play. The tree of knowledge in *Back to Methuselah* was entirely made with light, projected from the rear. The sphinx in *Peer Gynt* became a towering form by the same process. The sacred tree on the rim of the Persian desert in *Marco Millions* was merely a silhouette on a semi-transparent back-drop. But it towered in mysterious majesty because of the way light was thrown upon it, from both front and rear.

The technical basis of this important phase of modern theatre art is the spot-light—in contrast to the flood light—which can be controlled and focussed accurately upon one particular spot. To be effective these spot-lights must be separately controlled so that each can be set at any point of intensity from full to dim and fluctuated back and forth if need be during the progress of the play. Hence the rheostats, or dimmer board, that controls these lights, is to the designer's conception what the central nervous system is to the ideas in the brain. Increased subtlety both in design and control of electric light, and particularly the colour of light, is the key to theatrical design of the future. That must wait not only on the playwright and stage designer, but on the lighting engineer, who, more than anyone, holds the key to the modern theatre's destiny. (*See* also THEATRES [STRUCTURES].)

BIBLIOGRAPHY.—E. Gordon Craig, *The Art of the Theatre* (1905; 3rd ed., 1924); H. K. Moderwell, *The Theatre of To-day* (with bibl. 1915; 3rd ed., 1928); E. Stern and H. Herald, *Max Reinhardt und seine Bühne* (1918; Eng. trans. by M. S. Gudernatsch, ed. O. M. Sayler, 1924); I. Pichel, *On Building a Theatre,* Theatre Arts Mag. Monographs No. 1 (1920) revised and reprinted as *Modern Theatres* (1925); O. M. Sayler, *The Russian Theatre under the Revolution* (1920; 2nd ed., 1923), and *Inside the Moscow Art Theatre* (1925); L. B. Campbell, *Scenes and Machines on the English Stage during the Renaissance* (1923); K. MacGowan, *Continental Stagecraft* (1923); R. E. Jones, "Drawings for the Theatre" and Norman Bel Geddes, "A Project for the Theatrical Presentation of the Divine Comedy of Dante," *Theatre Arts Inc.* (1925); W. R. Fuerst and S. J. Hume, *XX. Century Stage Decoration* (2 vols. 1928); Sheldon Cheney, *Modern Stage Decoration* (1927); A. Nicoll, *The Development of the Theatre* (1927); R. Füloep-Miller, *Das Russische Theater* (1928); S. Cheney, *The Theatre, 3000 Years of Drama, Acting and Stagecraft* (1929); L. Simonson, *The Stage Is Set* (1932); *Settings and Costumes of the Modern Stage* (The Studio Winter Number, 1933); *Theatre Art* (Catalogue of the International Exhibition, Museum of Modern Art, 1934); *Architecture in the New Theatre* (Theatre Arts Inc. 1935). (L. SI.)

STAGE EQUIPMENT

The technical equipment of a modern stage consists of two types of machinery: (1) the mechanical equipment used to facilitate "setting and striking," *i.e.* shifting of scenery; and (2) electrical equipment to control the lighting of the stage.

A house has been described by one modern architect as a machine for living. A stage has always been a machine for giving plays. Even the Greek stage, usually supposed to be indifferent to scenic changes and scenic effects, was equipped, by the 4th century B.C. and probably earlier, with three standard pieces of machinery: (1) the crane used to lower the *deus ex machina* (the god out of the machine) from the roof of the palace façade; (2) the *periaktoi,* triangular prisms placed in the façade of the palace wall, revolving on wooden pins and turned in unison, each face painted

to indicate a different milieu so that three successive indications of place might be achieved, and (3) the *eccyclema;* or rolling platform which emerged from the palace doors after the tragic catastrophe, on which the murdered victim and often the aggressor were shown as a tableau.

The chief mechanical devices of the theatres in the Middle Ages, the Renaissance both in Italy and Elizabethan England, and the European theatres of the 17th and 18th centuries were the trap through which actors appeared and disappeared, often to the accompaniment of effects of smoke and flame; the moving platforms on which Christ was assumpted to the balcony heaven of an open air mediaeval mystery, or the classic gods descended 20 or 30 strong in the *gloire* that concluded so many 17th and 18th century spectacles and opera; and, when mobile scenery was adopted as part of the technique both of playwriting and play production, the windlasses and ropes which hoisted scenery out of the way or dropped it into place.

Until the latter part of the 19th century very little had been added to these basic stage mechanics. The revolving stage of the Japanese popular theatre of the 18th century was copied by Lautenschläger in Munich in 1898; the State and municipal theatres of Germany built in the early 1900s operated the traps mechanically, making them elevator stages which rose from a cellar level as needed. In addition sliding platforms, each carrying a complete setting, known as wagon stages, were utilized and often operated electrically.

In the United States none of these mechanical improvements was generally adopted, the reason being that the conditions of production remained radically different in the American theatre from those that prevailed in Europe. On the Continent theatrical production, particularly in Russia, Germany and the Scandinavian countries, was mainly in the hands of permanent resident companies operating endowed municipal or State theatres which gave a wide repertory of plays ranging from intimate realistic comedies or tragedies to large-scale spectacles and opera. The productions were built on the premises so that workshops as well as a large storehouse were an integral part of the architectural plan, as can be seen in the municipal theatre at Stuttgart with its large theatre at one end, a smaller theatre at the other, and the long façade of the storehouse serving as a connecting link between the two. The complete mechanization of such stages was considered worth the very large investment required in order to facilitate the giving of a different play every night of the week. This type of elaborately mechanized theatre is still standard in Europe, as exemplified by the theatres recently built in Göteborg and Malmö, Sweden.

In the United States, particularly after the decline of the local stock company at the end of the 19th century, the individual company, usually assembled for the occasion, was the producing unit, and much of its income was derived from touring on "the road" through all the principal cities and many of the smaller ones of the entire country. For this reason the mechanical arrangements of the stage remained primitive. Until late in the 19th century the rigging of most stages was exactly like that of a ship. The ropes needed to hoist the scenery up and down were attached to a fly gallery and tied to the belaying-pins on a pin rail, which was an almost exact counterpart of the pin rail of a frigate. Massive or solid scenery was not often built, the entire system of scenic construction consisting of breaking the sets into units not more than 5ft. 9in. wide so that when the settings were dismantled they could be slid through a baggage car door, its standard height being about 6 feet. The American theatre remains very largely an itinerant theatre to this day, the production being assembled usually in New York, then taken for a "tryout" on "the road," then brought back to New York; and, after a New York run, if it is a success, taken on tour again. Instead of being designed and built for the dimensions of a permanent playhouse, it must fit the variable stage dimensions, often exiguous, of any number of theatres.

American theatres have therefore generally been built to rent to travelling productions, either for a percentage of the box-office gross or for a fixed guarantee per week, and it has been the rule to provide as limited mechanical equipment as possible in order to reduce the landlord's overhead expense. The lighting system in these theatres is usually old-fashioned and even primitive, and controls the minimum of apparatus; a few overhead borders and footlights. The company must bring in all its lamps, cables and "dimmer boards," or rheostats, needed to control them. The only mechanical improvement that became common is the substitution of the counterweighted lines operated from the floor, which replace the dead pull of the old rope sets worked from the fly gallery. The houses built in the last century usually had a single trap, known as the Hamlet trap, placed so that Ophelia could be conveniently buried. The trapping of commercial stages is not uniform, and it is under these conditions that even the most modern scenic productions have been made, and to meet these conditions that they have had to be designed and devised. Where revolving stages or sliding stages are needed and used, as is occasionally the case, they also have to be carried on tour and assembled for each opening performance.

Before the deflation of 1929 very few technically adequate theatres had been built, the Century Playhouse, since torn down, a few theatres such as the Martin Beck theatre and the Guild theatre in New York being the exceptions, and a few new opera houses such as those in Chicago and San Francisco. In the main, however, the professional theatres in the United States have been either rudimentary or inadequate in equipment, and except for the houses on "the road" large enough to house musical comedies on tour the stages are too shallow and lack the off-stage space needed to manipulate efficiently and without great waste of time and labour the scenery and properties used during a typical performance.

The technical equipment of American theatres has recently improved due to the fact that they are now being built on college and university campuses for a continued repertory of student productions, and so begin to approximate the rational layout of the repertory houses of Europe. Notable examples are the new theatre at Stanford university, the Wisconsin Union theatre at Madison, Wisconsin, opened in 1939, and the auditorium at the University of Indiana, Bloomington, Indiana, opened in 1940, as well as such small theatres as the recently opened theatre at Amherst and the large playhouse projected for Dartmouth college. In these theatres, not restricted by the size of a conventional city lot, and where the cost of land is not an item of initial expense, stage dimensions are ample, the stage running from 35ft. to 40ft. in depth and 70ft. to 80ft. in width from wall to wall; they are provided with an adequately high gridiron, approximately 70ft., and a complete counterweight system as well as a stage floor that can be trapped over a third of its area. In addition there are workshops where an entire production can be fabricated and painted. At Stanford, at Wisconsin and at Bloomington two auditoriums are provided: a large auditorium with a seating capacity running from 1,300 to 3,000 which can be sub-divided by curtains for smaller audiences, and a smaller theatre for more intimate or experimental productions. At Bloomington and Wisconsin an electrically operated forestage has been provided which can at the touch of a push button be dropped below auditorium level to become an orchestra pit, raised above auditorium level for formal productions, making a forestage for an Elizabethan production if needed, or dropped to auditorium level to hold additional seats.

The funds available for the building of these theatres, running from approximately $500,000 to over $750,000 are, even so, not ample enough to provide for any further mechanization of the stage equipment, such as electrically driven revolving stages or wagon stages; even on a college campus the additional offstage space is rarely available which would be required to house these wagon stages when not in use. The plans for the Dartmouth theatre, however, call for two such reserve wagon stage spaces, and the prize winning design of the William and Mary competition (see *Theatre Arts* for June 1939) also indicate their use.

The more important technical equipment of the modern theatre remains the lighting control system, and the distinguishing feature of a modern production is the subtle variations of light which accompany and dramatize the dramatic action of the play. The

need for these changes, known as "dims," was felt centuries before the mechanical means for their effective use had been invented. The dialogues of Leone di Somi, a Mantuan stage director, written about 1565 (see Appendix B, Allardyce Nicoll's *The Development of the Theatre,* 2nd edition, Harcourt, Brace and Company, 1937) contain the following indications of how variations of light emphasize the mood of the scene:

"Now it has been a custom, both in ancient and modern times, to light bonfires and torches in the streets, on the housetops and on towers, as a sign of joy; and hence arises this theatrical convention—the imitating of such festive occasions. The lights are put there for no other purpose but to imitate, in the very first scene, this mood of gaiety."

"I suppose, then, that these lights would not appear in a tragedy."

"Perhaps they would not be so wholly out of place even in such a play. Quite apart from the fact that there are tragedies with happy endings, we note that nearly all tragedies open in a happy strain; and consequently it will not be unfitting to arouse the mind, so far as we may, to this happiness, although disasters and deaths are to ensue later. I remember once I had to produce a tragedy of this kind. During all the time when the episodes were happy in mood I had the stage brightly illuminated, but so soon as the first unhappy incident occurred—the unexpected death of a queen—while the chorus was engaged in lamenting that the sun could bear to look down on such evil, I contrived (by prearrangement, of course) that at that very instant most of the stage lights not used for the perspective were darkened or extinguished. This created a profound impression of horror among the spectators and won universal praise."

A device for dimming the candlelight of the period can be found in Nicola Sabbattini's well-known treatise, *The Practice of Making Scenes and Machines in Theatres,* printed in 1638.

Di Somi continues:

"It is a natural fact—as no doubt you are aware—that a man who stands in the shade sees much more distinctly an object illuminated from afar; the reason being that the sight proceeds more directly and without any distraction toward this object, or, according to the peripatetic theory, the object impinges itself more directly upon the eye. Wherefore I place only a few lamps in the auditorium, while at the same time I render the stage as bright as I possibly can."

Thus the general scheme and balance of our stage and auditorium lighting was anticipated by nearly 400 years. But the subtleties of modern stage lighting were not finally outlined until Adolphe Appia wrote his epoch making work, *Die Musik und die Inscenierung* (1899), and described in detail a typical light plot for a production, never made, of *Tristan und Isolde,* in the appendix to the volume:

"Picture of the stage at the rise of the curtain: A great torch at stage centre. The somewhat narrow stage space is filled with enough light so that the actors can be distinctly seen, without dimming the brilliance of the torch and above all without dimming the shadows which this source of light casts. The forms which demarcate the stage setting are seen only hazily. The quality of the light veils them in an atmospheric blur. A few barely visible lines in the stage setting indicate the forms of trees. Gradually the eye becomes accustomed to this; above, it becomes aware of the mass of a building in front of which one perceives a terrace.

"During the entire first scene Isolde and Brangaene stay on this terrace and between them and the foreground one senses a declivity, the forms of which one cannot identify clearly. When Isolde extinguishes the torch the entire setting is enveloped in a monotonous half-light in which the eye loses itself without being arrested by a single definite shape. Isolde, as she flies towards Tristan, is enveloped in mysterious shadow which intensifies the impression of death that the right half of the stage has already induced in us. During the first ecstasy of their encounter both remain on the terrace. At its climax (page 112, piano 8°, ff° in orch. *Mein!*) we perceive that they come toward us imperceptibly from the upper terrace and by means of a barely noticeable ramp reach a lower platform further in the foreground. This platform, the ramp leading from the terrace to it and another incline which leads to the forepart of the stage itself, make an uneven terrain on which the glowing dialogue of the scene that follows takes place.

"Then, when their desire is sufficiently appeased, when a single idea possesses them and when we become increasingly aware of the death of time, only then do they reach the foreground of the stage where (pp. 136–37) we perceive for the first time a bench that awaits them. The whole mysterious half-lighted space of the stage becomes still more uniform in tint. The structural forms are blotted out in the shadow of the background and even the different planes of the stage floor are no longer visible.

"Whether as a result of our visual reactions induced by the gleam of the torch and the shadows it casts, or whether because our eye has followed the path that Tristan and Isolde have just traversed, however that may be, we become aware of how tenderly they are enfolded by the world about them. During Brangaene's song the light grows still dimmer. Even the figures of the actors no longer have any clear definition. Finally, when passion awakes once more, grows and threatens by its overwhelming power to annihilate any scenic investiture that might be given it, at this moment (p. 162, first ff° in the orch.) a pale red light strikes at stage right in the background. King Mark and his men-of-arms break in upon the scene. Slowly, cold and colorless, the dawn begins. The eye begins to recognize the general forms of the stage setting and the painted flats that suggest them become clear and hard in quality for the first time. With the greatest effort at self-control Tristan comes back to contact with the world once again, challenging Melot, who betrayed him, to a duel.

"In the setting, now cold in color and hard as bone, only one spot is veiled from the dawning day and remains soft and shadowy: the bench at the foot of the terrace."

These ideas again anticipated the technical means of accomplishing them. It is only in recent years that lighting control systems have been perfected to the point where such fluctuations of atmosphere and light on the stage, to the accompaniment of the changes of dramatic mood and dramatic movement of the script, can be accomplished.

In order to understand this it is important to realize the elaboration of the lighting apparatus needed to light a play and its disposition. The footlights and borders which used to be the principal source of illumination are now unimportant adjuncts of lighting equipment. The focusing lens' spot light with coloured mediums of either glass or gelatin is the essential unit. They are usually assembled on two upright pipe battens on either side of the stage, known as the "light tormentors," and are also hung from one or more pipes overhead, the total often reaching 60 or more spot lights. In addition, in order to illuminate the actors without relying on bright foots which, if used with sufficient strength to light the actors' faces invariably cast distracting shadows on the scenery back of them, spot lights are located in the ceiling beam of the auditorium so that the angle of throw is approximately 60° and hits the actor head-high from the front of the forestage to a point within 6 or 8ft. upstage of the proscenium where the overhead and tormentor lights can illuminate him. These lights are focused in groups to cover various acting areas, and their colour is determined by the nature of the scene: whether daylight or night, whether they are projected to illuminate the actors or give colour and tone to the scenery. They are then successively brought up to a given point of intensity, *i.e.* "dimmed up," or "dimmed down" to a lower point of intensity in overlapping groups, making a light plot which approximates in general scheme and intention Appia's light plot quoted above.

The greatest flexibility of control is therefore essential, and the control board—commonly called the switchboard—remains the nerve centre and the most important single piece of apparatus on the modern stage. Up to recent years the standard rheostat or dimmer consisted of a wound core of resistance wire through which the current was shunted. Thus, when the light was dimmed a major portion of the current was turned into heat. The control was entirely mechanical, each dimmer having a dimmer handle which operated the dimmer from full to out. A certain number of these dimmers could be geared together on a shaft and worked by a master handle. However, as there was a limit to the number that could be geared together without creating too much friction or making the mechanical control extremely difficult to operate, it was usually impossible to select all the lights needed for a particular change and operate them together in a mechanically efficient way. One either needed several operators or often, on a large switchboard, an operator with the reach of an orang-utan. For it must be remembered that in the dimming changes of a modern light plot the light changes start with 50 or more lamps each set at different degrees of intensity—some may be full, some only half bright, others only a quarter up, etc. The change does not take them uniformly either to full or out. Those half-way up may only come down a few points in intensity; others may go to full out; others may come up in the opposite direction from nearly dim to nearly bright. And it is the drawback of the mechanical gearing of dimmer handles on a single shaft that it is impossible to gear them until they are all brought to the same level, so that the task of stopping some when they reach the various predetermined

levels of intensity is an almost impossible one to achieve accurately by hand.

This has been made possible recently by several improvements in the lighting control system. The first is a reactance dimmer in the form of a thyratron tube which directly regulates the voltage of each circuit. The second is a rheostat known as an auto-transformer which automatically transforms the voltage of the lamp to correspond with the desired intensity, the contact being a carbon or copper brush actuated by a motor. Both these dimmers have a variable capacity, the auto-transformer having a wider variable load. The old-fashioned resistance dimmer of 1,000 watts had to be connected with a lamp or lamps of a total 1,000 watts capacity in order to operate smoothly from full to out. The auto-transformer dimmers will take a lamp or circuit of 4,000 watts capacity or a lamp of 10 watts capacity from full to out or to any intermediate point, thus immensely increasing the flexibility of lighting control.

The reactance type of dimming control, though it is in process of being simplified, is much more complicated in wiring layout, more expensive both in initial and installation costs, and occupies more on-stage space than the auto-transformer type. The operating mechanism on the auto-transformer control board consists of a series of potentiometers which actuate, on a 24-volt circuit, the motors attached to the auto-transformers. The control is through a series of small handles over a graduated semi-circular arc which set each circuit at any point between full and out. Without affecting the set-up of the lights then on, a second set of handles controlling these same circuits is pre-set at any number of other points of predetermined intensity from full to out, and the change from the first pre-set to the second is made electrically, each circuit stopping at the point indicated and pre-set. Two master handles are provided which are electrical master controls so that, as on the board at the Wisconsin Union theatre, any of 56 circuits can be connected with one master, the remainder with the other master, and the light change be made on these two sets of selected circuits simultaneously at different speeds either in the same or opposite directions, *i.e.* either from full to out or from out to full, or to any number of pre-selected intermediate points of intensity for each circuit. The system thus gives complete control of any number of lighting units in any number of possible combinations in a way that no mechanical system of dimmer control is able to achieve. This lighting control system in one type or another will undoubtedly become standard in the near future.

It has the further advantage that the control wires can be extended to a point in the auditorium or in the projection booth at the rear of the auditorium, and a portable dimmer board be connected there so that the lights can be as effectively controlled from the auditorium as from backstage if the occasion requires.

An additional feature of modern lighting equipment is the lighting of the cyclorama. The illusion of depth and space in an open air sky is best achieved by a plaster dome, but as this is a heavy and elaborate piece of construction, and requires a great deal of off-stage space if it is not to interfere with the exigencies of scene shifting—more space than can usually be given to it or paid for—a cyclorama made of a textile is the usual substitute. This is a cylinder of cloth or canvas laced to a semicircular pipe above and below, and lighted from overhead and below by lights in parabolic reflectors with glass mediums that are spectroscopically accurate to act as light primaries—red, blue and green. By combinations of these at various intensities in conjunction with the proper painting of the cyclorama itself the colour of the cyclorama can be changed through almost the entire range of the spectrum. As painted canvas is very much affected by changes in humidity and can rarely be made to hang properly, the better practice is to use dyed cotton velour; and in order to facilitate changes of setting during scene intermissions, the up- and down-stage arms are made in the form of travellers, the velour being heavy enough to be drawn back without settling into wrinkles.

The mechanization of the stage for scene shifting seems to offer no further possibilities of practical development differing from what has already been achieved. Future technical developments will undoubtedly be in the electrical field, the development of low voltage high wattage lamps, increased flexibility of focusing lens units, increase in the range of variable load of rheostats and added flexibility of electric master and sub-master control.

BIBLIOGRAPHY.—These devices can be studied in detail in Gustave Cohen's *Histoire de la mise en scène dans le théâtre religieux français du moyen age* (1926) and *Le livre de conduite du régisseur et le compte des dépenses pour le mystère de la Passion joué à Mons en 1501* (1925); W. J. Lawrence's *The Elizabethan Playhouse and Other Studies* (Second Series, 1913); the section of Diderot's *Encyclopaedia* devoted to stage machinery, as well as in G. Bapst's *Essai sur l'Histoire du Théâtre*, and the richly illustrated portfolios of 18th century settings and stage mechanisms *Court Theatres of Drottningholm* (Swedish Architectural Historical Society). (L. St.)

DEVELOPMENT OF STAGE LIGHTING EQUIPMENT

The first authentic description of artificial light as an integral part of a stage presentation is by the Italian Serlio who placed candles and torches behind flasks of coloured water, the light being projected through holes in the scenery. He also used pans and trays as reflectors to intensify the light; but one can imagine that the resulting illumination was never objectionably bright. Even in the early covered theatres the chief light sources were torches or pine knots, open lamps with floating wicks burning vegetable or animal oils, and candles of wax or tallow. "Palladio's theatre" in Italy and "Black Friars Inn" were the first of these indoor theatres and the first indoor productions in England did not take place till 1576.

The removal of the theatre from the open air into buildings was the first active force to bring about better artificial illumination. Inigo Jones brought the Italian theatre influence to England and used the proscenium arch. Gold decorations were applied to the theatre interiors reflecting light from huge clusters of candles. These were hung from the ceiling in a corona above the stage, and supplied all the general illumination. Even with candles, all the writers of the time referred to the lighting as "brilliant."

In 1628 Joseph Furtenbach described the use of oil lamps and candles as "floats" or "footlights," as well as "wing lights" the latter being now known as "borders." In 1775 when David Garrick placed a row of candles below and in front of the stage, masked from the audience by metal screens, they were an innovation in stage lighting. Not till the 18th century did oil lamps generally replace candles. The John Street theatre in New York, erected in 1767, was illuminated by candles for 30 years.

The invention of the Argand lamp in 1783 and the use of camphine, kerosene oil and other higher grade illuminants brought a far superior light to the theatre. The new lamp, which was a simple oil and wick combination provided with a glass chimney, was the first important scientific advance in light sources in 4,000 years. For the first time the illuminant was enclosed rather than being an open flame.

These lamps gradually supplanted the candles. Lamps hung in clusters from the ceilings and projected from walls, balconies and boxes. They were used as footlights, borderlights and wing lights. Despite the improvements, there was still only sufficient illumination to make the actors and scenery recognizable.

Then came gas. In 1781 William Murdock of England claimed a patent for its manufacture, and in 1803 F. A. Winsor gave demonstrations on "the feasibility of gas" in the Lyceum theatre in London. In 1796 it was used in Philadelphia for exterior illumination; and in 1816 a complete gas lighting system was placed in operation in the Chestnut Street Opera house of the same city. Gas was manufactured on the theatre premises for the theatre only. An installation was, therefore, an expensive affair, and gas did not come into general use before 1850.

All of the original gas burners were open flames or flames burning in open glass chimneys. But in 1886 the Welsbach incandescent mantle was invented which increased the efficiency many times. They were used in borders, foots, wings, and bunch lights continuing the old conventions.

Gas for the first time permitted the control of all the lights on the stage from one central location. Elaborate systems were devised with mains sometimes more than 12″ in diameter, and

thousands of feet of feeder pipe, both flexible and rigid. The Lyceum theatre in London and the Boston theatre in America were two of the first to be equipped with these new controls which were called "gas tables." The first electric control boards were fashioned after these gas tables. They were located near the prompter who could reach a master valve controlling the entire layout. In addition to the master there were series of individual dimmer valves, and on some boards flash buttons by means of which spurts of flame could be made to give quick bursts of light.

Chief among the defects of this system was the risk of fire. Many theatre fires of this time were started by the blazing spirit wads on long poles used for lighting gas jets among the scenic pieces. Dancers with their thin costumes were sometimes set afire by the footlights. Because of these dangers, code regulations were established which fixed the position and number of burners in various locations, and made compulsory the use of guards and screens over open flames.

It became the custom at this time to dim the lights in the auditorium during the play, for at last the theatre had discovered that contrast between light and darkness was as important as light alone for visibility and maintaining interest.

Gas lights also left their impression on the scenic design and painting for many years to come. The light source was not small nor bright enough for small individual units; and the old-fashioned borderlights, footlights, and wing lights were retained, altered only to accommodate the burners replacing the oil floats, chimneys and candles.

Henry Irving experimented with effects of realism, using individual lights to achieve a more plastic illumination and devising an arrangement for drawing thin coloured silks or screens in front of others. He also started the use of sectionalized strips and borders—later employing spot lights—and thus increased the flexibility of existing apparatus. At this time also the cyclorama made its appearance, a piece of equipment still most useful on the stage.

In 1816 the calcium light or "lime light" was invented by Thomas Drummond, and by 1860 had come into general use in the theatre. From this unit came the common expression, "in the limelight" which is frequently used today. This light source was extremely white and concentrated and therefore could be used behind a lens or in front of a parabolic reflector as a spot light. The apparatus was used to "follow" characters about the set and in a few instances for general illumination of the stage. It required the constant attention of an operator.

The beginning of electricity in the theatre was the invention in 1808 of the arc by Sir Humphry Davy, who had also experimented with incandescent lamps as early as 1802. However, this apparatus, although invented before the "lime" lights, did not come into use till many years later. Frequent adjustment was necessary as the carbon was consumed, but the arc was a great advance over the calcium light. Two types of units were built around this source. One used a parabolic reflector behind the carbons, the other was equipped merely with hand operated shutters. The first device was designed in 1846 and used by M. J. Duboseq and M. Foncault in the production of *The Prophet* at the Paris Opera. Five years later M. Duboseq also developed a lightning machine of a type still used today, a rainbow projector, and a luminous fountain—all built about the arc as a light source.

In 1878 Paul Jablochkov introduced his Jablochkov candle, two vertical parallel sticks of carbon separated by a layer of insulating compound. In 1879 the Bellecour theatre in Lyons, France, installed 52 of them. But this method never came into general use for it could not compete with the incandescent lamp invented by Edison in 1879.

This invention brought to the theatre for the first time a clear, relatively high powered, easily controlled light source. For the first time there was light which could be used as a plastic medium, coloured, varied in form and intensity, used in ways which had never before been possible. Wagner in his operas had suggested possibilities in the use of light which stimulated the minds of the producers. He created a demand for technical developments which now began to look possible. Qualities of light were available to those who would use them, and theatre lighting as an art had arrived.

But the invention of the incandescent lamp alone did not immediately bring modern stage lighting to the theatre. However, its advantages over the older types of illuminants were recognized almost at once; and the change-over was more rapid than it had been with any other type of lamp. In 1880 and 1881 a series of experiments in the Paris Grand Opera resulted in the first theatrical installation of incandescent lamps. The greatest impetus was received from a theatre at the Electrotechnische Ausstellung at Munich in 1882. Incandescent and arc lights were used for auditorium as well as stage lighting; and units on the latter were equipped with remotely controlled colour screens. The tracker wire mechanical method was used for this colour changing and still is.

The first completely electrified theatres in Germany were the Landestheater in Stuttgart, the Residenztheater in Munich, the Staatsoper in Vienna, and the Staattheater in Brunn—all in 1883. The Savoy theatre in London and Bijou theatre in Boston installed electric lighting in 1882. The first New York theatre to be equipped was the People's theatre on the Bowery, in 1885. In 1882 a congress of theatrical managers assembled and issued a formal report favouring the illumination of halls and theatres by means of electricity. In 1871 this new method was so well established that a law was passed prohibiting the use of gas in all theatres of Prussia.

The original electric incandescent lighting units were footlights, borderlights, and wing lights, following the old conventions. Arcs were still used wherever concentrated light sources were necessary. The only bulbs available were of the carbon filament type which gave a relatively dull glow of a definitely orange colour.

Gradually, however, materials other than carbon were discovered and osmium, tantalum, moulded tungsten and finally drawn tungsten were used in filament constructions. The use of nitrogen and later nitrogen and argon in the bulbs further increased the efficiency and decreased the size of the lamps. Today, with coiled and coiled-coil filaments, gas filled bulbs, and heat resisting glass jackets, 500-watt lamps only 1½in. in diameter are in general spotlight service, and 5,000-watt lamps, although much larger are not uncommon.

With new light sources available, equipment was developed to use them more effectively. Steel MacKaye, a producer and actor, designed the modern effect machine as well as many other mechanical contrivances for the theatre. Mario Fortuny developed the Kuppelhorizont or dome, still the most perfect means of simulating natural or spectacular sky and background effects. He also developed a system of lighting named after him which utilized reflectors of coloured silk placed behind the lamps to tint and diffuse the light. These silk panels were movable so that the colours could be changed and the illumination softened.

David Belasco and his electrician, Louis Hartmann, abolished footlights—a suggestion made years before by Henry Irving—and began the use of individual sources from the auditorium and above the proscenium for acting area lighting. By means of these instruments as well as special reflector borders and baby lens spots they developed a standard of realism now in general practice in theatres throughout the United States. In Germany Lennebach, Hasait and many others made more contributions to the technical development of the theatre. They invented the rolling cyclorama and made up elaborate systems of light control. The Fortuny system was discarded and used only in rare instances.

Today the development of smaller lamps with hard glass jackets and concentrated filaments, and new types of reflecting surfaces have made possible the design of two new types of spotlights. By reducing the diameter of the lamp bulb and using a Fresnel or stepped lens, an instrument having a high efficiency and giving an even field with soft edges, useful for acting area lighting and easy to blend into a general effect, has been developed. The second type of spotlight differs entirely from this, the light from the lamp being gathered and redirected by a reflector rather than a lens.

An ellipsoidal reflector has the valuable characteristic of re-

directing light which is generated at one of its conjugate foci through the other. This is the principle of the ellipsoidal reflector spotlight. A lamp with tubular glass jacket and concentrated filament, generally of the base-up burning type, is placed with its filament near the focal point within a reflector. A gate or aperture is placed between the reflector and the conjugate focal point and "filled" with the reflected light. One or two lenses acting as an objective then project an image of the aperture on the field to be illuminated. In this way fields of even intensity are produced, the beam may be shaped by means of shutters, iris, or other controls at the aperture, hard or soft edges are possible through focusing of the objectives, and efficiencies comparable to those of the Fresnel lens spotlight are obtained.

The two above mentioned types of spotlights have to a large extent replaced the "standard" plano-convex lens "spots."

Besides the spotlights, three other types of lighting equipment are widely used in stage presentations. First are the striplights—really the first form of apparatus as the old wing or borderlights were strips. The latest development in striplights is the individual reflector type in which each lamp is placed in a spun reflector equipped with either a coloured glass roundel or a gelatin, or glass colour frame.

Floodlights are specially shaped reflectors for individual lamps. They are made in narrow beam or wide beam types and accommodate small or large lamps. Wide angle floodlights are used for illuminating back drops and cycloramas, and, in the U.S., are matte finished ellipsoidal reflectors. Banks of them equipped with three or more filters can give effects which simulate any degree of sky colour from midday to midnight including the brilliant hours of sunset and dawn.

Special effects, the fourth class of lighting equipment, often are built about the spotlight as a base; but they may include anything from painted rainhows and moving clouds to water ripples of cut glass strips, a burst of fireworks, or a waving flag. The effects are generally painted on mica discs enclosed in a sheet metal protective bousing and driven by electric or clockwork motors. These "effects" are placed in front of spotlights, an additional lens being added to each to further concentrate the light; and objective lenses project the images of the moving discs.

As opposed to the complicated design of these "projection effects" is the simplicity of the shadow projector. This apparatus is very old in principle but the form now used in stage work is the Linnebach lantern—named after Adolph Linnebach, the designer. It consists of a deep box painted black inside and surrounding a concentrated filament lamp or arc. One side of the box is open, and over this is placed the slide carrying the design to be projected. Because the lamp filament or arc has a definite size, sharp projections such as are obtained with lens units are impossible. But very wide areas can be covered at relatively short projection distances; and, especially for silhouettes of skylines and crude outlines, this method has no equal.

In step with improvements in equipment has been the development of new colour media. Fortuny and other of his contemporaries used coloured silks and reflectors for tinting the light. Lamp dip appeared after the invention of incandescent lamps. Today sbeet gelatin is still probably the most widely used colour medium. It is available in large sizes, a great variety of colours, is inexpensive, and most of the colours are relatively fast. Efforts are constantly being made to increase their permanence.

Gelatin, however, is very unsatisfactory in damp climates as it absorbs water readily and proceeds to "melt." In any climate after being subjected to heat from lamps for several hours it becomes very brittle and tends to shrivel. Acetates, such as celluloid, are much stiffer and do not absorb water, but they are very inflammable and the fire hazard precludes their use. Gelatin may be ignited but supports combustion only with difficulty, and tends to smoulder rather than burst into an open flame.

Cellophane bas been used with a great degree of success in recent years. It is available in 11 colours besides clear, black, and two degrees of clear diffused. It is thinner than gelatin, is not greatly affected by dampness, and supports combustion only with difficulty. Some of the colours are quite fast, but others rather unstable; and the limited colour selection is, of course, a great disadvantage.

From many standpoints glass is the ideal theatrical colour medium despite its weight, initial cost and fragility. It has come into use only recently and more because new equipment required it than because it was developed to a new degree of perfection. "Gelatins" burned out with alarming rapidity when used in front of high powered spotlights, but when the Fresnel lens and ellipsoidal reflector spotlights came into use with their greater light outputs and concentrated beams, new colour filters had to be found at once. Coloured sheet glass in both "flashed" and "pot" forms was available, but none of it was heat resisting. What to do? Well, the glass, once it cracked under heat, was not likely to break again. Therefore, why not break it in a regular pattern before using it? This system was tested in practice and was successful. Coloured sheet glass is now cut into strips, mounted between sheets of asbestos paper and the whole bound in a substantial metal frame. Some of these frames have been tested in front of 5,000 watt Fresnel lens spotlights and have become so hot that solder melted when placed in contact with the glass, yet there are practically no cases of the glass failing under such conditions. Glass is not the perfect colour medium. It is fragile, heavy and expensive. Its colour range is limited, and sheets of glass often vary in colour and transmission by as much as 50% from one end to the other. The field of plastics at present seems to promise a great advance in colour control in the theatre. But as yet there are no transparent materials which are available in heat resisting quality. But one day in the not too distant future we may expect a material which can be used without a stiffening frame to support it, one which is impervious to heat or moisture, available in a multitude of fast colours. It will be light yet sturdy, and less expensive than glass filters. It is not too much to expect.

For the remote control of all the old gas lighting equipment, the gas table had proved very satisfactory. In the first switchboards only intensity control was thought necessary, although colour controls had already been included in many of the previous gas installations. The first electric dimmers were water barrels with two metal plates put in a weak solution of brine. One plate rested at the bottom of the barrel, the other could be raised or lowered, thus varying the resistance of the circuit by varying the amount of brine between the plates through which the current had to flow. These units were extremely unsatisfactory as the water evaporated rapidly; the heat, if the barrels were overloaded, produced a vile smelling vapour; and the settings for any particular lamp intensity were never twice the same.

The next type of intensity control was a coil of German silver resistance wire tapped at a number of points, these taps being connected by wires to the points of a rotary switch. By turning the switch the resistance in the circuit could be varied at will. The resistance coils were generally mounted in the basement while the switches were located on the stage floor. Because the resistance to current flow uses energy which is transformed into heat, the coils became hot, especially when overloaded—and for that reason were dangerous. A large number of wires leading from the control switches to the coils were also necessary, and this was eliminated by the resistance plate dimmers next developed.

In this type the resistance wire was laid out on a plate, either itself an insulator or coated with an insulating and heat resisting compound. The switchpoints were connected with the wire. The entire assembly was coated with more heat resisting material until only the contact surfaces of the switchpoints were visible. The tapping arms were pivoted directly to the resistance block. This type of control unit is still the most widely used on American stages.

The most obvious fault of the resistance control has already been mentioned:—that energy is used in reducing the current flow and is then dissipated in heat. This fault was recognized as early as 1890 and in 1896 a reactance type of dimmer bank was installed in the Earl's Court theatre in London. Here a double reactance with one coil slipping in and out of the second was used to control each circuit, alternating current being used rather than direct. The plan worked well enough but alternating current was not always available.

Other faults of the resistance dimmer are their large size and weight and the fact that they will control only loads of approximately the same capacity as their own rating. But nearly all types of alternating current control units will operate with lamp loads from their maximum capacity down to zero. They are primarily voltage and not current regulators.

The most modern types of alternating current controls are the auto-transformer dimmer, and the electronic tube and reactance circuit. The first is manufactured in sizes up to 5,000 watts for single phase loads and up to 10,000 watts for three phase loads.

The latest development in theatrical intensity control equipment is the electronic tube and reactor circuit, and, still more recently, electronic tubes which themselves carry a considerable load current. By means of these devices switchboard sizes have been tremendously reduced, and a panel the size of a moderately large radio receiver is now sufficient to control the lights in an entire theatre.

Switching and dimming equipment have been improved and reduced in size, weight and cost. Pre-set and fader circuits have been developed which permit the instant switching of a large number of controls by means of a single master lever, or allow an operator gradually to change the lighting from one predetermined setting to another, some units dimming, some brightening, but all in proportion to their individual readings.

As has been mentioned previously, tracker wires and mechanical link movements are still the standard means of remote colour control in Europe. In the U.S., self-synchronous motors and electromagnetic controls perform the same service. The former permit an operator at the central control panel to move colour frames or lighting units themselves in exactly the same way in which he manipulates levers and dials. Only electric circuits connect the control and the lighting equipment—there are no moving cables or pulleys. Electromagnets perform quick changes, such as colour frame shifts, upon being energized by means of a push button or switch, but cannot be made to control slow movements.

BIBLIOGRAPHY.—S. R. McCandless, *Syllabus of Stage Lighting;* Ward Leonard Electric Co., *Theatre Lighting Past and Present;* Friedrich Kranich, *Buhnentechnik de Gegenwart;* Siemens-Schuckert, *Buhnen-Regeltransformatoven;* Franklin Institute, *Light Thru the Ages;* Edison Electric Co. of Boston, *History of Stage and Theater Lighting;* Louis Hartmann, *Theatre Lighting;* Selden and Sellman,

Stage Scenery and Lighting; Westinghouse Lamp Division, *Artificial Light and Its Application;* Theatre Fuchs, *Stage Lighting, Lighting by Century.* (E. Ke.)

STAGE DIRECTING: *see* Acting, Direction and Production.

STAGNELIUS, ERIK JOHAN (1793–1823), one of the outstanding poets of the romantic movement in Sweden. The son of a pastor who became bishop of Kalmar, he was born on the island of Öland, Oct. 14, 1793. He had no regular schooling, but received a good grounding in the classics. After studying law at Uppsala university, he became a civil servant in Stockholm in 1815. Little is known of his later life. He seems to have led a very isolated existence, and it is doubtful whether the ardent love poems to "Amanda" were actually addressed to a real person. His health was undermined by heavy drinking and he died poor and lonely at Stockholm, April 3, 1823.

In his earliest known poetry, Stagnelius adopted a neoclassical style and showed an Epicurean attitude to life. He remained consistent in his aspiration for clear, concise form; his sensuality also persisted in his choice of imagery, although after 1815 Epicureanism gradually gave way to extreme philosophical idealism. Some of the poems from his later theosophic period are difficult to interpret. He was influenced not only by August von Schelling, but by such mystics as Jakob Böhme, L. C. de Saint-Martin and Emanuel Swedenborg. The underlying elements of his philosophy are derived from various neo-Platonic and Gnostic systems. The human soul, or Anima, has left its home in the spiritual world, Pleroma, and is subject to Demiurgos, the ruler over sensual reality. This realm is felt to be a bewitched and enchanted world, a desert or a prison, whence man, like the rest of creation, longs to return to his heavenly home.

Stagnelius' poetic inspiration is wide-ranging and he is master of an unusual variety of forms. Among his notable works are a monumental epic in hexameters on the Christianization of Russia, *Wladimir den store* (1817), and several plays which, though lacking in dramatic intensity, contain passages of brilliant rhetoric and melodious verse. Some are set against a background of early Scandinavian mythology, others (*e.g.,* the blank verse play, *Bacchanterna,* 1822) have a classical setting. But Stagnelius is, above all, a lyrical poet. Of a more metaphysical cast of mind than Keats, of whom he was an almost exact contemporary, he shared with him unusual lyrical intensity and powers of illumination.

Of the few works published in his lifetime was the collection of poems *Liljor i Saron* (1821). The critical edition by F. Böök, 5 vol. (1911–19), has an indispensable commentary.

See F. Böök's monograph on Stagnelius (1919; abbreviated and rev. ed., 1954). (S. B.)

STAHL, GEORG ERNST (1660–1734), German chemist and physician, one of the propounders of the phlogiston theory which held that "phlogiston" was the material of fire, was born on Oct. 21, 1660, at Ansbach. He was graduated in medicine from Jena in 1684, and became court physician to the duke of Saxe-Weimar (1687), professor of medicine at Halle (1694) and physician to the king of Prussia at Berlin (1716), where he died on May 14, 1734.

Known chiefly in connection with his doctrine of phlogiston (*see* Chemistry: *The Phlogiston Theory*), the essentials of which, however, he owed to Johann J. Becher (1635–82), he also propounded a view of fermentation which, in some respects, resembles that supported by Justus von Liebig 150 years later. He was a firm believer in animism. In his youth he believed in alchemy, but later he warned against its frauds. He observed that acids are of different strengths, that the basic component of common salt differs from that of potash, and he recognized the particular nature of the earthy constituent of alum. His views of combustion were first presented in his *Zymotechemia fundamentalis* (1697). His chief medical treatise was *Theoria medica vera* (1707). His principal chemical work was *Fundamenta chymiae* (1723). (R. E. O.; X.)

STAHLBERG, KAARLO JUHO (1865–1952), Finnish statesman, born on Jan. 28, 1865, began his public career by holding alternately university and administrative posts. He became a member of the diet in 1904 and of the Finnish government in 1905, but resigned in 1907 and was appointed a professor of administrative law in 1908. At the outbreak of World War I the government was Finnish in name only, being composed entirely of Russian officials. Stahlberg was then president of the diet and had to act as the chief public representative of Finnish opinion. He firmly withstood the increasingly oppressive tendencies of the imperial Russian government. After the Russian Revolution of 1917, Finland became an independent republic, and in 1919 Stahlberg was elected to the presidency. He withdrew from the candidature for the next presidency after his term of office ended in 1925. He was a candidate in 1931 and again in 1937, but was not re-elected. He died on Sept. 22, 1952, in Helsinki, Fin.

STAINED GLASS, a term that is generally understood to refer only to glass windows that have been coloured by such methods as the fusion of metallic oxides into the glass, the burning of pigment into the surface of white glass, or the joining of white with coloured pieces of glass.

The origin of stained glass is obscure. It probably came from the near east, the home of the glass industry, and mosaic windows of glass set in plaster work, which is known from the 17th century in Egypt and elsewhere, are probably of great antiquity. But it is not likely that the art goes farther back than the 9th century; it is doubtful if before that time glass was made in a sufficient variety of colours. The art would most naturally spread first to Italy, and Venice may have been a centre as early as the 10th century. An Italian panel in the Victoria and Albert museum, tentatively ascribed to the 13th or 14th century, is fully Italian and Romanesque in style and suggests a native tradition.

Actually the earliest reference to stained glass in the accepted sense of the term (that is, windows not merely coloured, but pictorial also) is in a manuscript which records that Adalbéron, bishop of Rheims from 969 to 988, rebuilt the cathedral and redecorated it with windows representing various stories (Richer, vol. ii, lib. 3). An earlier reference, in a 9th-century life of St. Ludger, certainly relates to coloured windows, but it is not clear that they were pictorial. Mâle gives, as the earliest mention of leading, a reference to the *Miracles of Saint Benoit* where it is related that in the last years of the 10th century the church of Fleury-sur-Loire was set on fire and it was feared that the leads of the windows would be melted.

Which are the earliest windows extant is a matter of dispute, but the evidence, both literary and stylistic, seems to favour certain figures of prophets in Augsburg cathedral, which may date from the middle of the 11th century.

TECHNIQUE

A stained-glass window is a translucent mosaic held together by lead; that is the simplest conception of its technique. Actually other considerations come into play in the design and execution of a window. The lead is not merely a connecting medium, but in all good design plays a part of its own; it outlines the main constituents of the design, giving definition and rhythm to the masses of colour.

From the account given by Theophilus, a very accurate notion can be derived of the technique as practised when the art first came to its perfection in the 12th century. The general scheme for the glazing of a church was, of course, the care of the clerics, and it may be assumed that the artists among them, the illuminators, would prepare the first drawings. One such series of drawings, the Guthlac Roll, is preserved in the British Museum (Harley Roll, Y 6). But the actual cartoon for the glass was drawn on one end of a whitewashed board supported on trestles. The vacant space at the other end was used for laying out the glass and for the general business of glazing.

The cartoon was marked to indicate the various colours, and the next step was to shape a piece of glass to the outline required by the design. This was done, firstly and roughly, with a hot iron and then more carefully with a tool known as the "grozing iron," a flat piece of iron with a notch at one end, rather like a modern spanner. This must have been a slow and laborious process, but the grozing iron seems to have continued in use until about 1500

when the modern method of diamond cutting began to be employed. The next step, in a fully developed stained-glass window, was to paint the glass with details which the bare design of glass could not give. At first the use of paint was confined to an opaque brown (grisaille, *q.v.*), used, not as colour, but as a means of outlining the design in further detail, reinforcing the effect of the leads. This pigment consisted of powdered glass mixed with a metallic oxide (probably iron), to which was added sufficient gum to make the mixture adhere. If "high lights" were desired, the whole surface of the particular piece of glass was covered with a thin coating of this paint and the lines scratched through with a pointed stick. Shading effects were produced by stippling with a brush. The glass, having been painted, was next fired in the kiln, to fuse the enamels to the glass. In mediaeval times the glass was fired in a pan, which was filled to the top with alternate layers of glass and whiting. When the glass was fused sufficiently, it was brought out and cooled and then rearranged on the glazing table. The next step was the leading. This was (and still is) done by means of strips of lead, in section like the letter H, but with a thicker cross-bar to represent the "core" of the lead; the upright strokes represent the "tapes." At first the leads were cast, but in the 17th century the lead vice was introduced. This is a kind of mangle, with two toothed wheels like coins with milled edges, between which a strip of lead is squeezed, the soft metal emerging in the form already described. Sometimes the wheels had the glazier's name engraved on them, which thus became impressed on the lead as it passed through the machine, and these names are occasionally found on old leading. When the leads, cut to the required length and shape, had been inserted between the pieces of grozed glass, they were next soldered together at the points of junction, and cement or putty was rubbed into the crevices between glass and lead. The window was then ready to move into place, where it was fixed by means of leaden strips soldered to the leads and attached to iron saddle-bars let into the masonry. Larger windows were made in smaller units and these units fitted into an iron framework or "armature" which itself often formed a geometric design contributing to the general effect of the window.

From the 12th to the 15th century the technique of stained glass remained practically unchanged, though one minor innovation had a great effect on the character of design. This was the invention, early in the 14th century, of a yellow stain derived from a solution of silver. It formed a very thin film on the glass and was therefore very transparent; on clear glass it varied in tone from pale lemon yellow to deep orange; but it could also be applied to blue glass to produce a brilliant green. Its effect on design was far-reaching and will be noticed when we come to the history of the craft. In the 16th century the technique of glass-painting became more scientific; the glass itself was thinner and smoother and more transparent; two layers of different colours could be put together to produce a third colour; if these layers were fused together, one layer could be ground away or "abraded" (and by a later process "eaten" away by fluoric acid) to give delicate effects beyond the reach of a leaded mosaic. Finally, towards the middle of the 16th century, the art of painting in enamel on glass appeared, perhaps as a result of economic causes which made a scarcity of pot-metal. Hitherto glass had been coloured throughout its substance and was known as "pot-metal," though to obtain sufficient translucency in a dark colour like ruby the glass was "flashed," that is to say, clear glass whilst still pliant was dipped into molten coloured glass, and so received a translucent film of the desired colour. But now ground glass was mixed with various metallic oxides (copper for green, cobalt for blue, manganese for purple, and so on) and the design was then painted on to a clear sheet of glass and fused on. For some time this technique was continued with the old technique of leaded pot-metal; but in the 17th and 18th centuries the old technique disappeared, and the art of stained glass was virtually dead.

HISTORY

Eleventh and Twelfth Centuries.—The windows at Augsburg, ascribed by Fischer to the middle of the 11th century, have already been mentioned. To the end of the 11th century belongs a window at Le Mans (France) painted with the Ascension. For the 12th century we have in England a figure from a Tree of Jesse in the minster at York, and at Canterbury there is a magnificent series representing the genealogy of Christ which may be dated about 1180. At Chartres there are more important and extensive examples—in the west window is a Tree of Jesse design which is undoubtedly the best surviving stained glass of this century; also the beautiful figure of the Virgin and Child known as La Belle Verrière. Other considerable specimens of 12th century glass exist at Angers, Saint Denis, Châlons-sur-Marne, Bourges, Strasbourg, Poitiers and Le Champ (Isère). The earliest of these windows, such as those at Augsburg and Chartres, are strongly Romanesque in character. The windows at Canterbury come at the end of the Romanesque tradition. They still preserve the monumentality so characteristic of Romanesque design, but the figures are far removed from the rigorously schematic treatment of the Augsburg prophets; there is the first flush of humanism in their bold features. There is a sense of movement in their limbs; a sense of vigour in the formal draperies. The colouring of this Canterbury glass is still typical of the 12th century; the bare flesh is cut in pink glass, whilst the rest of the design is a quiet harmony of blues, pale green, white, brown and yellow, arranged in fairly broad masses with no effort towards "jewel-like" quality. The architectural settings are rudimentary, in most cases representing a simple column on each side of the figure, with an arch spanning the head.

It is probable that most of the early windows were of the single-figure, monumental character of the Augsburg and Canterbury examples. But about the middle of the 12th century a new type of window came into being which consisted, instead of a single representative figure, of a medallion (or several medallions) painted with pictorial subjects, such as incidents from the life of Christ, or from the life of some saint. This change was accompanied by various subsidiary developments. The necessity of telling a story in a series of medallions led to the decorative arrangement of these medallions within an "armature" or iron framework which was itself a subject for pattern, and very beautiful designs were evolved, showing a skilful counterplay of circles, lozenges, quatrefoils and squares, with their appropriate borders and interspaces. This development in its turn necessitated larger windows and this necessity had a profound effect on the development of Gothic architecture itself. A third development was in the actual colour of the glass. The pictorial treatment called for a more detailed and more varied play of colour, and the technical resources of the glass-maker were equal to this demand. The result was that kaleidoscopic or jewel-like glow of colours with which stained glass is always popularly associated. It is in glass of the early 13th century that this property is supremely evident.

Thirteenth Century.—With these early 13th century medallions we first become fully conscious of the part played by the actual painting in stained glass. It is true that the earlier figures had a broad effective scheme of drawing, but with the 13th century medallions a fine dramatic sense of line was developed and continued in force throughout the century. And with the sense of line came that superb sense of dramatic composition illustrated in so many of the medallions of this period at Canterbury, Lincoln, Chartres, Sens, Bourges, Le Mans, Rouen, Cologne, Marburg and Klosterneuburg (near Vienna). The highly stylicized motives of the decorative borders of the 12th century were no longer in keeping with this dramatic mode, and as the century advances we accordingly get a naturalizing of these motives. The formal acanthus leaves of classical inspiration give place to more natural plant motives, to scrolls of oak and vine; these coil with a more organic and vigorous movement, until eventually, in the 14th century, the artists begin to copy directly from their own observation—from the ivy and the wild strawberry.

Towards the last quarter of the 13th century, a greyer mood seems to have descended on our glass-painters. It may be that they were required to give more light in the churches, or that the supply of ruby glass, always somewhat precarious, gave out. Whatever the reason, there developed in the latter part of the

ENGLISH STAINED GLASS

"The Patriarch Jared." 12th century window in the south transept of Canterbury Cathedral. As in much of the mediaeval glass in England, the bare flesh is cut in pink glass, while the rest of the design is a harmony of blues, pale green, white, brown and yellow, arranged in fairly broad masses.

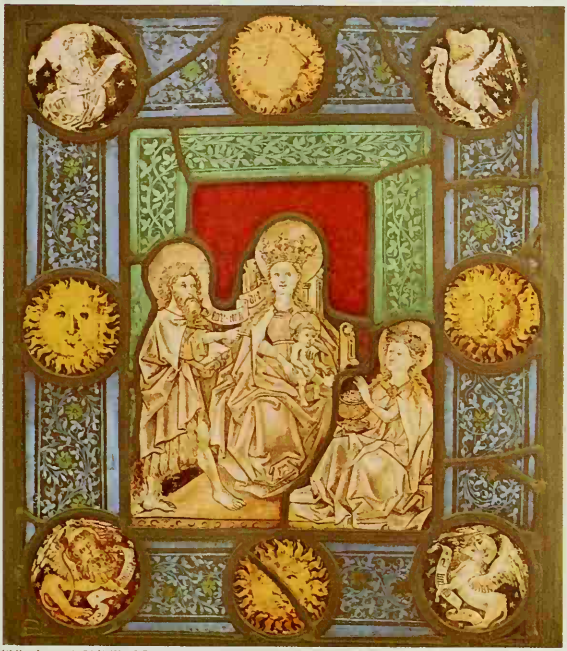
STAINED GLASS

ENGLISH STAINED GLASS

"St. Anne Teaching the Virgin." From the east window of All Saints church, York, dated about 1440. The colouring is simple though rich in quality, a characteristic of English glass of this period. St. Anne wears a ruby cloak and bonnet and a blue gown. The Virgin's dress is white, ornamented with a pale yellow pattern, and the background of figures and canopy is blue covered with a diaper. All other parts are white with some yellow stain

EXAMPLES OF 15TH- AND 19TH-CENTURY STAINED GLASS

A scene from "St. George and the Dragon," designed by D. G. Rossetti and executed by William Morris. 1860. Victoria and Albert museum



"Virgin and Child"; 15th century, German

Philip I of Spain; a window from the Chapel of the Holy Blood, Bruges, Belgium. Late 15th century

PHOTOGRAPHS, TAKEN ESPECIALLY FOR ENCYCLOPÆDIA BRITANNICA, INC., BY C. H. CANNINGS

STAINED GLASS

The Tribe of Benjamin, one of the 12 windows of the Tribes of Israel, created by Marc Chagall for the Synagogue of the Hadassah-Hebrew University Medical Center, Jerusalem, Israel; 1961

MODERN STAINED GLASS

HADASSAH MEDICAL RELIEF ASSOCIATION, INC. 1961. ALL RIGHTS RESERVED BY HADASSAH FOR MARC CHAGALL/ANDRE SAURET

13th and early 14th centuries the technique generally known as *grisaille* painting. It implies the use of large areas of clear quarries painted in black or brown enamel with simple and unobtrusive patterns of foliage. Grisaille work at this period is of two kinds: it may consist of geometrical patterns of quarries, often outlined in coloured bands, on which are painted designs of foliage, often in imitation of a trellised plant, or it may be used in combination with subject medallions. The best instance of this latter method is perhaps seen in the windows of the Chapter House at York, which, though now much decayed and obscured, are supreme examples of grisaille work.

These characteristics of 13th century stained glass are general wherever the main stream of Gothic art (*q.v.*) flowed. Distinctions of nationality need not, and indeed cannot, be made. In point of time the medallion windows at Canterbury perhaps come first; then, closest in time and style, come the similar windows at Sens; Chartres comes next and is very impressive in its quantity and scale; at Bourges and Rouen there is glass very similar to that of Chartres; the glass at Lincoln and Beverley in England must be placed somewhere before the middle of the century. But back again in France we find a more distinctive school developing in Paris during the second half of the century, in connection with the building of the Sainte Chapelle and the extension of Notre Dame. The distinguishing feature of this latter style was the general use of a background of trellis-work in blue and red glass, which, in the mass, has a rather unpleasant purple effect. This style scarcely penetrated to England, though an example may be seen in the east window of the south transept of Christ Church, Oxford.

Fourteenth Century.—Midway in the 14th century comes the Black Death, and this catastrophe marks the end of the pure Gothic tradition, in this as in all other arts. Its most marked effect was perhaps reflected in the break-up of the inter-community of Christian art. No doubt other forces, economic or political, were operative too; but the disorganization caused by the Black Death among artists and craftsmen all over Europe resulted in the formation of more limited and more local schools. Onward from 1350 the styles of England, France, Austria and Germany diverge more and more from each other. And within all these countries the local differentiations between various centres are often very marked. The disorganization not only affected traditions, but even more so technique. It seems that the ability to manufacture some of the brilliantly coloured kinds of glass was lost at this time, never fully to be recovered until the 19th century. There was also a Cistercian ban on pictorial windows, which no doubt had some influence in this direction, during the 14th century. But the glass-painters were not slow to react to the new situation; the invention of yellow stain early in the century had provided them with one new resource; and the growing taste for line design, as witnessed in the grisaille glass already mentioned, led to the full exploitation of clear white glass. When once the use of white glass had been forced on the glass-painters, its aesthetic value was appreciated. It may be also that the resultant effect on the lighting of the churches was appreciated, and that the glass-painters began consciously to aim at a clearer effect. Architectural developments kept pace with, and even aided, this changing colour-scale. The windows grew higher and wider, the tracery more intricate, demanding of the glass-painter an effect that was not too heavy for the expanse to be filled, or too clumsy for its delicacy.

Further developments were due not so much to economic factors as to a change in spirit. The 14th century saw the full emergence of that movement towards humanism which had found its protagonist early in the 13th century in St. Francis of Assisi. The tendency towards naturalism which we have already seen creeping into the decorative motives of stained glass was a far wider phenomenon, and involved all art and literature. The new spirit of Franciscan naturalism only manifested itself in plastic art by slow degrees, but it culminated in the Renaissance. It involved a transformation of Christian iconography, and the glass-painter, along with other artists, now deserted the fixed types of early Gothic art for the living types before him. Subjects are now given a contemporary setting; for his figures, his costumes and his architecture the painter looked around him and found models before his eyes. A playful freedom of line gradually developed; as the brush-stroke grew lighter, it grew easier. The features are rendered with more grace and verve; there is almost an excess of curve and curl in eyes and lips and locks. An individualization of the features begins to creep in; we begin to feel the presence of real people, friends and contemporaries of the glass-painter, no longer abstracted, no longer idealized. The folds of the garments lose their angular restlessness and sweep and sway in graceful curves. We get, too, the entry of fashion. The garments in the 12th and 13th centuries are formal and of classical derivation, except in so far as they are ecclesiastical or military. They continue to be somewhat non-committal during the first half of the 14th century, but with the growth of naturalistic treatment and the revival of figure-subjects we get the unmistakable evidences of contemporary fashions. The representation of costume then becomes a matter of course, and is of great assistance in the precise dating of stained glass.

But the most typical and expressive development of ornament is seen in the canopy. This form of architectural ornament no doubt originated in the 12th century, when an architectural unit—generally an arch—was introduced to indicate a building or interior scene. This simple unit became more elaborate during the 13th century, but remained essentially illustrative in intention. Then we get as an early characteristic of the 14th century, the elaboration of this detail into purely ornamental features. The architecture is given the value of pure decoration. At first this is rendered in one dimension, and slender pillars and fretted pinnacles, interchanged in colour, form a simple and effective setting for the figures within the framework. But as the architecture itself becomes more elaborate, its representation in the glass develops. By the end of the 14th century we find an elaborate use of perspective. In the 15th century the spirit of play breaks out even here, and the niches are peopled with angels and coy minstrels. As the century advances the canopies grow more complex, though more representational.

Fifteenth Century.—With this development of the architectural features of stained glass ornament is perhaps bound up a curious retrogression during the 15th century—the abandonment of natural plant motives for stylicized diapers. Artists forsook the carefully observed ivy, oak and vine, and introduced instead such dreary motives as that usually described as the "seaweed pattern"—an elongated, shapelessly indented leaf, generally painted in reserve on a blue or ruby glass, and seemingly kept as "stock" to be used whenever the design required an indeterminate background. It was a step toward that still more slipshod expedient—the stencilled diaper background of the 16th century. The borders, which had always been an integral part of stained glass design, develop in conformity with the general trend. But their aesthetic function becomes less obvious and they gradually lose decorative value. In the 12th century the border had been the essential framework of the design, enclosing within its definite pattern the otherwise unconfined design. In the 13th century the border existed as a foil for the sequence of separate designs. Then, in the 14th century, when the subject-panel became isolated from its background, the border became a purely decorative feature, contributing a separate unit to the general aesthetic effect; in the 12th century the border framed the design—in the 14th it edged the frame of stonework, its naturalistic form suiting this purpose admirably. Then in the 15th century, with its strange efflorescence of canopy work, the border tended to lose all aesthetic justification; it existed merely as a strip dividing the design from the stonework, and providing, as it were, a transition from one to the other. In this role it naturally tended to become more formal in character. Already in the 14th century the typical border of creeping foliage had been varied with alternations of devices such as crowns, chalices, heraldic badges, grotesque animals, small figures of angels, etc.; an exceptional example is to be seen in the famous Bellfounders' window in the north aisle of the nave, York minster, where the border consists of bells (outer lights) and musical monkeys (centre light). The 15th century saw some recovery of the sense of colour. Coloured glass of a variety of colours, but too even and flat in tone to be quite satisfactory, became available on the continent and was largely imported into England. Already in 1408 the east window at York minster, made by John Thornton of Coventry, is full of a new vitality and gaiety. The covenants for the execution of the glass in the Beauchamp chapel, Warwick, which are dated 1447, specifically provide for glazing the windows "with Glasse beyond the Seas, and with no Glasse of England; and that in the finest wise, with the best, cleanest and strongest glasse of beyond the Sea that may be had in England." This window was carried out by John Prudde, one of the greatest of English glass painters, and the result is versatile in the extreme, the colours, jewelled and counterchanged to the utmost limits of ingenuity, blazing with the greatest richness imaginable.

In Germany also, the development was in consonance with the richness and complexity of late Gothic art in general, and there the develop-

ment may be particularly associated with the name of Hans Wild, whose principal works date from about 1470 to 1480. His most famous window is in the cathedral at Ulm, and shows great boldness and original fantasy in the use of foliage motives. The window becomes a luxuriant trelliswork of flowering plants and pinnacles in the depths of which the human figures tend almost to disappear.

One characteristic of this period is the development of heraldry in glass. We have practically no knowledge or records of the use of heraldry in stained-glass windows previous to the 14th century, though there are three 13th-century shields in the apse windows of Westminster abbey which serve to indicate an early use of heraldic glass. We know from records that these shields were originally placed in the aisle windows, which were of grisaille. But the particular circumstances which brought heraldic glass into general use are hardly consistent with the spirit of early Gothic art. One of the characteristics of the 14th century is the introduction of the personal element into the windows. Donors began to desire their portraits within the design of their window; memorials to the immediate dead began to assume a representational form. At first no attempt at faithful portrayal could be made; the mode and technique of glass painting at that time forbade it. The shield of arms had therefore the very definite use of identification. But it was soon realized that heraldry had an aesthetic value of its own, and its use multiplied. Glass proved to be especially suitable to the rendering of the formal patterns of the shields, and the brilliancy doubly attractive in this personal emblem. The ground of the shield, which was usually filled in with a diaper pattern, gave the glass painter an opportunity to exercise his skill in brushwork and his fertility in the invention of decorative motives. The technique of scratching a design with a point through a dense ground of colour, always a pleasant technique in glass painting, was encouraged by this development. The glass painter kept pace with the herald, and did not hesitate to attempt the utmost elaboration of curved shield and exuberant mantling. With the Renaissance heraldic glass took on a more restrained appearance—at least, in England, where the common use of the garter and the wreath tended to confine the design within sober limits. Elsewhere, especially in Switzerland, heraldic glass became, during the course of the 15th and 16th centuries, the most typical use of glass painting.

The 15th century saw a great growth and intensification of the individualism we have already noted as arising in the 14th century. Such a development naturally gave rise to the dominance of artists of unusual talent, and to the formation of schools dependent on such artists. A distinct style became established as a local tradition and sometimes persisted over two or three generations. Such schools were formed all over Europe, and particularly in the great ecclesiastical cities. In England, for example, definite schools can be associated with York, Coventry, Canterbury, Winchester, Oxford, Gloucester, Wells, Lincoln, Norwich and Westminster.

The different styles are not always easy to distinguish, and the "schools" are rather in the nature of empirical groupings based on the density of distribution.

Toward the end of the 15th century England suffered a general invasion of foreign artists and craftsmen, a growing influx that reached its culmination early in the next century with the arrival of Torrigiano and Holbein. At this time, in the Netherlands, the art of stained glass had reached a development which, in accomplishment and modernity, was far ahead of the insular style in vogue among English glass painters. The wool and cloth trades flourished in spite of the wars. The new merchants, earnest Lollards as they were, were eager to turn their wealth to some good purpose, and thus they made the 15th century an age of church building and charitable foundations. And these merchants, moreover, were men of some culture, whose business had brought them into contact with foreign luxuries—particularly those of the Netherlands. When, therefore, they required glass for the beautifully furnished churches which they began to construct in various parts of the country, they demanded glass in keeping with their luxurious tastes, and for this they turned to the men from the Netherlands. In answer to the demand these men came across and, in accordance with the regulations, established themselves within a recognized sanctuary such as the liberty of St. Thomas in Southwark, and assumed English names. As time went on, they became a menace to the existence of the London company or guild of glaziers, and a continual series of disputes between the guild and the foreigners marks the beginning of the 16th century. But the proper tradition of glass painting was entirely lost. In the middle ages stained glass thrived as an art subordinate to, or in community with, architecture. With the Renaissance, and the shifting of the dominant emphasis of artistic effort from architecture to painting, the glass painter found himself in a dilemma. He had to adapt his art to the new conditions, or suffer from the general neglect of arts subordinate to architecture. He attempted to save himself by adopting the aesthetics of painting, and the history of stained glass henceforth is the history of this false step and of all the disastrous consequences.

Sixteenth and Seventeenth Centuries.—But while this false aesthetic was developing, there grew up in France and the Netherlands a full-blooded Renaissance style of glass painting which makes the first half of the 16th century a distinctive epoch in the history of the art. It is a development of the pictorial treatment of the windows on the largest possible scale; the window is regarded as one large canvas; the intersecting mullions are ignored. On this canvas the coloured glass is spread in generous masses and heightened by all the resources of line and shade. The windows gleam with everything that is rich and ornate in colour, design and subject. The glass is thin, but this fault is counteracted by the breadth of the treatment. The Netherlandish type is well represented in England by the work of imported artists, of whom Barnard Flower is the most important. This artist came from the Netherlands, probably encouraged by Henry VII, who made him king's glazier in 1505. Flower executed the whole of the windows for Henry VII's chapel at Westminster, and was entrusted with the contract for King's College chapel, Cambridge, but died before he had completed more than four of the windows. It is probable that among the work which may be attributed to Flower are the windows of Fairford church in Gloucestershire, perhaps the most complete unit of stained glass remaining intact in England. King's College chapel, begun by Flower after designs probably by the Flemish artist, Dirick Vellert, was completed by a mixed company of English and Flemish glaziers.

In France, Rouen was the great centre of glass painting, and that city still offers an incomparable display of the glass of this period, though many other towns, such as Evreux, Châlons-sur-Marne, Dreux, Beauvais, Auch, Troyes and Montmorency, are famous for their windows. In the Netherlands, the most important windows of this period are found in Antwerp, Hoogstraten, Brussels, Liège and Amsterdam, while the windows at Gouda, painted by the brothers Dirk and Walter Crabeth and their pupils, toward the end of the 16th century, are of extraordinary range and completeness. In Germany important schools existed at Strasbourg, Augsburg, Munich, Freiburg-im-Breisgau, Nuremberg, and above all at Cologne, the general character of all of which was not essentially different from that of the dominant Flemish or French type. In Italy the period is mainly associated with the name of Guglielmo de Marcillat (1467–1529, sometimes erroneously called William of Marseilles). Windows known to have been made by him exist at Arezzo and Rome, and a very typical example from the cathedral of Cortona, dated 1516, is in the Victoria and Albert museum. At Florence, where there are also important earlier windows in the cathedral which are said to have been designed by Ghiberti and Donatello, other windows in the style of Marcillat may be seen in the church of Santa Croce. The building of the cathedral at Milan caused an important school of glass painting to develop there, and the work of Cristoforo de Mottios and Nicolo da Varallo carried out during the second half of the 15th century is of great beauty. The Milan school continued in full activity during the 16th and 17th centuries. In Spain, especially at Seville, León and Avila, magnificent windows of this period exist, but they are in all cases the work of Flemish glass painters.

In the 17th century we find the highest level of glass painting in Switzerland, though actually that level had been surpassed in its own country during the 16th century, when Holbein and Burgkmair were leading designers of the panels typical of that country. But the real significance of Switzerland in the history of glass painting is that there alone was evolved an appropriate technique of enamel painting. For various reasons of a social kind, glass painting in that country had acquired a more domestic character; it actually became part of the furniture of the people, and a panel for the window of a room was as natural, and even more usual, than a picture for the wall. Glass paintings thus became intimate and portable—a kind of *Kleinmalerei* comparable to the development of *Kleinplastik* in sculpture. Within the limits of this littleness, the art of enamelling glass developed an appropriate technique. It was, indeed, the technique of painting—of translucent painting—but then the Swiss panel was in all essentials a painting. To the same category belong the roundels painted in grisaille which were produced in large quantities during the first half of the 16th century. In these types of glass painting the specific problems of distance and architectural harmony did not enter into the question; everything was made for nearness and intimacy. And on this scale stained glass was inappropriate, ungainly and without effect.

The same observations might be made of the Dutch and Flemish glass paintings of the 16th and 17th centuries, which illustrate the enamel technique in both its good and bad aspects, and it was through these, rather than the Swiss, that the new style penetrated to England. Abraham and Bernard van Linge (who flourished about 1620–40 and painted many windows in Oxford and elsewhere) were the chief of these immigrants, but they were ably imitated by native artists such as Henry Gyles of York (1645–1709), the Price and the Peckitt families of the same city, and by artists like Francis Eginton (1737–1805) and James Pearson (d. 1805), who carried the art right on to the threshold of the 19th century. At their best, the achievements of the glass painters in this genre have a miniature delicacy and a perfection of means beyond reproach. But it must be realized that we have entered on a new art, with its own aesthetic, and we must beware of any confusions of judgment in this respect. The unfortunate truth is that in England (and elsewhere in Europe) it was the glass painters who were guilty of confusion. Social custom never called upon them for a domestic art comparable with Swiss glass painting, but nevertheless they attempted to apply the methods of the domestic glass painters (but without their technical skill) to the quite different needs of ec-

	12th century and earlier	13th century	14th century	15th century	16th century	17th and 18th centuries
England	Canterbury, York, Lincoln, Beverley, Salisbury, Waterperry, Chartham, Westwell, West Horsley, Aldermaston, Upper Hardres	Canterbury, York, Lincoln, Beverley, Salisbury, Waterperry, Chartham, Westwell, West Horsley, Aldermaston, Upper Hardres	Oxford (Merton College and Christ Church), Wells, Gloucester, York, Hereford, Tewkesbury, Winchester, Ludlow, Deerhurst, Durham, Eaton Bishop, Arlingham, Notgrove, Fladbury, Edgeworth, York, Carlisle, Westonbirt, Chetwode, Beer Ferrers, Exeter, Norbury, Hereford, Selling, Stanford, Mancetter, Selby, Waterperry	Great Malvern, Gloucester, Little Malvern, York, Warwick, Coventry, Ludlow, Margaretting, Leicester (Museum), Long Melford, Hessett, Blakeney, Levrington, Martham, Canterbury, Norwich, Gresford, Windsor, St. Neots (Cornwall), Llandrynog, Norbury, Cirencester, Bristol, Battlefield, Durham, Thornhill, Almondbury, Orchardleigh, Westwood, Ettington, Bledington, Barkway Ash, Oxford (Bodleian Library), Ockwells Manor, Winchester, Berkhampstead, West Wickham, Stamford, East Harling, Bury St. Edmunds, Combe	Fairford, Cambridge (King's College), St. Margaret's, Westminster, The Vyne in Hampshire, Basingstoke, Winscombe, Llanrhaidr, Stoke D'Abernon, Gilling	Lydiard, Tregoze, York, Oxford, London (Lincoln's Inn Chapel, St. Andrews, Holborn), Guildford, Ashtead, Stoke Poges, Southwell, Hagley
France	Le Mans, Châlons-sur-Marne, Poitiers, Le Champ (Isère), Angers, Aubazine, Bonlieu, La Bénisson-Dieu, St. Denis, Lyon, Normée, Pontigny, Strasbourg, Varennes, Vendôme	Reims (Cathedral), Rouen Le Mans, Semur-en-Auxois, St. Sulpice-de-Favières, Châlons-sur-Marne, Montiers-Hubert, Bourges, Soissons, Strasbourg, Poitiers (Cathédrale and Ste. Radegonde), Chartres, Séez, Sens, Troyes, Lyons, Clermont-Ferrand, St. Martin-au-Bois, Orbais l'Abbaye, Tours, Châteauroux, Vendôme, Angers, Auxerre, Beauvais, Brie-Comte-Robert, Essomes, Dijon (Notre-Dame), Wissembourg, Paris (Notre-Dame, Ste. Chapelle), St. Julien-du-Sault, Gassicourt	Angers, Avioth, Beauvais, Chartres (Église-St.Pierre), Evreux, Colmar, Dol, Clermont-Ferrand, St. Dié, Embrun, Limoges, Lyon, Mulhouse, Carcassonne, Mussy-sur-Seine, Narbonne, Nieder-Haslasch, Albi, St. Pierre-sur-Dives, St. Quentin, Rouen (St. Ouen), Séez, Semur-en-Auxois, Strasbourg, Vieux-Thann, Tonnerre (Hôpital), Toul, Toulouse, Troyes, Villeneuve-sur-Lot, Westhoffen	Rochefort (Rhône), St. Romain-au-Mont-d'Or, Wissembourg, Villefranche-sur-Saône, Walbourg, Châlons-sur-Marne, Abbeville, Riom, Aigueperse, Alençon, Le Mans, Saint-Saens, Le Bourget, Ambierle, Villeneuve-sur-Lot, Vieux-Thann, Ambronay, Argentan, Saverne, Sélestat, Solignac, Verneuil, Obernai, Beaujeu, Bernay, Paris (Ste. Chapelle, St. Séverin), St. Quentin, Rouen, Beaune (Hospice), Bourges, Saint-Lô, Saint Nicolas-du-Port, Chartres, Thann, Carentan, Carcassonne, Metz, Strasburg, Chamelet, Ceyzériat, Séez, Coutances, Embrun, Evreux, Moulins, Eymontiers, Lantic, St. Léry, Lisieux, St. Martin d'Estréaux, St. Omer, Quimper	Alençon, Ambronay, Angers, Autun, Grand-Andely, Montmorency, Argentan, Troyes, Bayonne, Autun, Auch, Chantilly, Châlons-sur-Marne, Brou, Brissac, Champigny-sur-Vende (Chateau), La Ferté-Milon, Vic-le-Comte, Limours, Rigny-le-Ferron, Conches, Pontoise, Saint-Julien-du-Sault Triel, Saint-Firmin, Lisieux, Rouen, Bourges, Paris (St. Gervais) Gisors, Villy-le-Maréchal, Moulins, Limours, Pont-Audemer, Montreuil, St. André-d'Apchon, Andrésy, Aumale, Auxon, Baignon, Bannalec, Bar-sur-Aube, Bar-sur-Seine, Baux, Beaumont-le-Roger, La Baussaine, Beauvais, Bécherel, Bernay (La Couture), Bignan, Blanzy-le-Château, Blénod, Bordeaux (St. Michel), Bourg, Bourg-Achard, Bourgtheroulde, Brienne-le-Château, Brieune-la-Vecille, Chailly, Chambéry, Champeaux, Chartres (St. Aignan), Châtillon-sur-Seine, Daveoay, Dijon (Notre-Dame), Ecouen, Elbeuf, Epernay, Falaise, Fécamp, La Ferté-Bernard, Flavigny, St. Firmin, Les Ifs, St. Julien, Limoges, Lombez, Louviers, Metz, Nogent-le-Roi, Paris (St. Étienne-du-Mont, St. Germain-l'Auxerrois, St. Gervais, St. Médard, St. Merri), Pont-Audemer, Pont-de-l'Arche, Pont l'Évêque, St. Quentin, Quimper, Rennes, Rigny-le-Ferron, Riom, Roanne, St. Saens, Sablé, Sens, Tours, Vendôme, Verneuil, Villefranche-sur-Saône, Viocennes (Château), West-Capelle	Auch, Bourges, Caudebec, Charmes, Ecouen, Guingamp, Malestroit, Le-Mesnil-Aubry, Neuville-sur-Saône, St. Nicolas - de - la - Taille, Orleans (Ste. Croix), Paris (St. Étienne-du-Mont, St. Eustache, St. Germain l'Auxerrois, St. Germain-des-Près, St. Gervais, St. Médard, St. Merri, N. D. des Blancs, Manteaux, St. Nicolas du Chardonnet), Pont-de-l'Arche, Quillebeuf, Rouen, Savianges, Toulouse, Troyes, Versailles (Château)
Germany and Austria	Augsburg, Goslar, Helmstedt, St. Magdalena, Weitensfeld	Metz, Naumburg, Cologne, Soest, München-Gladbach, Marburg, Merseburg, Haina, Bücken-ander-Weser, Heimersheim, Heiligenkreuz, Bonn (Museum), Nuremberg, Coblenz, Friesach, Heiligenkreuz, Vienna, Stift Ardagger, St. Ruprecht ob Murau, St. Michael bei Leoben	Marburg, Esslingen, Freising, Cologne, Soest, Nuremberg, Lübeck, Erfurt, Regensburg, Herford, Altenburg, Markterlbach, Coblenz, Metz, Limburg - an - der-Lahn, Rothenburg-ob-der-Tauber, Prague, Klosterneuburg, Gurk, Lieding, Wels, Vienna, St. Florian, Wiener-Neustadt, St. Leonhard-im-Lavanttale, Graz, Gars, Strassengel, Pram	Ulm, Augsburg, Regensburg, Munich, Blutenberg, Soest, Freiburg, Dortmund, Xanten, Lüneburg, Tübingen, Straubing, Erfurt, Metz, Ingolstadt, Rothenburg o.d.T., Nuremberg, Stendal, Werben, Northeim, Halberstadt, Altötting, Friedberg, Cologne, Judenburg, St. Leonhard-im-Lavanttale, Tamsweg, Friedersbach, Göttweig, Wels, Nonnberg, Laxenburg, Wiener-Neustadt, Stifts Zwettl, Klosterneuburg, Vienna, St. Florian	Metz, Cologne, Nuremberg, Straubing, Xanten, Kyllburg, Ingolstadt, Wilsnack, Schwerte, Treves, Ansbach, Münster (Westphalia), Hanau, Immenhauseo, Herford, Munich, Freiburg i.B. Landsberg, a.L., Augsburg, Tölz, Salzburg, Wiener-Neustadt, Steyr	Schwerin, Brunswick
Italy		Assisi	Milan, Orvieto, Assisi, Florence, Siena	Florence, Lucca, Pisa, Milan, Pavia, Prato, Assisi, San Miniato, Arezzo, Saromno, Siena, Bologna, Val d'Ema, Grosseto, Venice (SS. Giovanni e Paolo)	Milan, Florence, Rome (Santa Maria del Popolo), Perugia, Val d'Ema, Calcinaio (near Cortona), Arezzo, Monte San Savino, Bologna, Lucignano, Siecina, Siena	Milan
Netherlands			Tournai	Lierre, Diest, Tournai, Mons	Brussels, Herenthals, Gonda, Antwerp, Bruges, Diest, The Hague, Liége, Hoogstraeten, Lierre	Antwerp, Brussels, Diest
Spain			Leon, Toledo, Gerona	Leon, Burgos, Avila, Toledo, Oviedo, Barcelona, Gerona	Leon, Granada, Segovia, Barcelona, Avila, Tarragona, Toledo, Cervera, Astorga, Gerona, Seville	Seville, Segovia
Switzerland		Zurich (Museum)	Königsfelden, Zurich (Museum)	Basle, Zurich, Wettingen, Hilterfingen	Basle, Berne, Zurich, Wettingen, Lucerne	Zurich, Wettingen

clesiastical buildings. The result was generally ludicrous, and always inappropriate. The crowning example of this misalliance of two inconsistent arts is to be found in the window designed by Sir Joshua Reynolds for New College chapel, Oxford, which was actually carried out by a china painter.

The art of glass painting in England during the 17th and 18th centuries does not offer sufficient material for any comparative criticism or positive judgment. We see nothing but scattered individuals, each expressing his own wayward fancy, displaying some ingenuity, supplying some passing need, but achieving nothing of significance.

Eighteenth and Nineteenth Centuries.—The Gothic revival that came as an offspring of the Romantic movement of the late 18th and early 19th centuries was not without its effect on the art of stained glass. The styles and methods of the early Gothic period were reconstructed, but devoid of all inner reason or inspiring sentiment. The art thus reduced to sterile formulas was easily commercialized, and factories for the manufacture of windows, any size, any subject and colours to taste, sprang up all over Europe, but especially in Germany. From this state of things the pre-Raphaelites and William Morris (*q.v.*), largely inspired by Ruskin, revolted, and whatever faults we may impute to them, we must nevertheless acknowledge that they made a feeling for design once again a vital force in the social life of England—and, indeed, eventually of Europe. We cannot enter here into the general character of Morris's achievements, but the basic fact upon which all his ideals rested was an intense aesthetic appreciation of mediaeval art. As much, perhaps, can be said of the initiators of the Gothic revival, though we may suspect that as an aspect of romanticism their enthusiasm sprang from a sense of the remoteness, the strangeness, and the mysterious gloom of this earlier art rather than from any real understanding of its rational basis or religious significance. But Morris had a more logical appreciation of the Gothic, and his attitude was as free from insincerity as it was devoid of any desire to imitate. He realized that without the spirit, the form could not exist, and his real greatness and overwhelming importance in the history of modern craftsmanship springs entirely from the fact that he attempted to evolve a style and invent a technique appropriate to the expressions of his age. Burne-Jones designed his first cartoons for stained glass as early as 1857, and the St. Frideswide window in Christ church, Oxford, dates from 1859. These early windows were carried out by Messrs. Powell of the Whitefriars Glasshouse, but from 1861 Burne-Jones worked exclusively for Morris. Stained glass was included in the first exhibition of the work of Morris, Marshall, Faulkner company in 1862. But the work of this period is weak in design and uncertain in technique. No attempt had been made to work out the proper function of the leads in the design, and the enamels are either poor in quality or badly fired. But even so, there was a certain freshness of attitude and a revolt against the dead conventions of the time. But as the demand for stained glass increased, so more thought was given to these matters and the improvement was rapid. In Morris's fully developed style, as in the cathedral at Birmingham, there is little to complain of in the technique of his glass. Time has yet to show whether the materials will weather so well as the 13th-century materials, but in this matter Morris was at the mercy of other people; he did not make his own glass, but obtained the best he could find in the market. In other technical matters Morris need not fear comparison with any age. His selection and disposition of colours is admirable, and he was not afraid of using new colours to achieve effects unknown to previous ages. In the use of leads to emphasize design he is masterly, and we must go back to the 13th century for an adequate comparison. The best examples of windows by Morris and Burne-Jones are to be found in Christ church, Oxford (the St. Cecilia and St. Catherine windows, 1874–75 and 1878), the "Judgment" window at Easthampstead (1875), the two windows in Salisbury cathedral (1879), the east window of St. Peter's, Vere street, London (1881), a window representing the Resurrection at Hopton (1882), the well-known windows in the cathedral at Birmingham, perhaps the finest of all (1887), and (perhaps the most important in London) the windows at Holy Trinity, Sloane street. What is reputed to be one of the finest examples of their work (1882) is a window at Biarritz in France, representing the marriage at Cana. In addition to all the work designed for churches, and confined in the main to Biblical subjects, Burne-Jones and Morris made a good deal of stained glass for domestic or secular purposes. In their choice of subjects for this profane glass, they perhaps unduly confined themselves to the romantic and literary predilections of their circle; but here again we cannot criticize their stained glass without entering into a discussion of their whole point of view. It is perhaps sufficient to say that it lacked the immediacy and the reality which we are entitled to expect from any contemporary expression of ideals.

The principles which William Morris established and followed give us an adequate criterion for the criticism of modern stained glass. It is sad to confess how little these principles have been followed in England itself. There is no longer any vital contact between the glass painters and the significant artists of the day. And in all directions there is a relapse into a servile and lifeless imitation of mediaeval mannerisms, due to a natural disregard for those self-sacrificing ordinances which Morris observed. It seems that only the complete filling of every window in every old church will ever terminate this atrocious vandalism. There might be some satisfaction in the prospect if it were certain that that would put a stop to what is known in the trade as "period work"; but these *pastiches* of the art of other epochs find their way into practically every new church that is built, and the only hope of any reform would seem to lie in the final triumph of those principles for which William Morris stood. For these principles are not altogether dead; they have merely migrated, and on the continent, especially in Germany, there is a school of glass painting which is not only modern in intention, but is inspired by all that is vital and significant in modern art.

Like the stained glass of William Morris, this modern continental stained glass is both religious and secular. The work of Jan Thorn-Prikker and of Karl Schmidt-Rottluff has given to the symbols and images of Christianity a new intensity and realism for which there is no parallel this side of the Renaissance; and the abstract designs in which these glass painters have experimented seem to open up infinite possibilities for this art of pure colour and light. These artists work under the direction of Gottfried Heinersdorff for the Berlin firm of Puhl and Wagner. In France, interesting experimental work is being done, for example, by MM. Maumejean of Paris and Hendaye. Stained glass, which dominates all other arts in this matter of colour, would seem to be peculiarly fitted for those experiments in abstract design and harmony which have been distinctive of a certain phase of modern art. The difficulty in all contemporary glass-painting of a nonecclesiastical kind, is not so much to extend the possibilities of the art in the direction of design and actuality, but rather to find a function for the art itself. For that we must look, with whatever faith we can summon, to those manifestations of civic and commercial enterprise for which our age is distinguished.

At Hagen, in Germany, the railway station has fine modern glass by Jan Thorn-Prikker; in Paris, the offices of a newspaper, *L'Intransigeant*, have windows with appropriate designs by Henri Navarre executed in leaded glass by Gaëtan Jeannin. But these are isolated instances of what seems to be the only chance of a revival of the art commensurate with its possibilities. There is nothing impossible in such a revival; the technical and material resources of the art are greater than they have ever been. All that is lacking is the desire; and to foster this we cannot do better than instil a wider consciousness of the unique and unrivalled beauties of the art as it has existed in the past.

(*See* Painting; Grisaille; Mosaic; Heraldry; Gothic Art; Genre; Religious Architecture; Glass Manufacture.)

Bibliography.—N. H. J. Westlake, *History of Design in Painted Glass* (1881–94); H. Oidtmann, *Die Glasmalerei, ihre Technik und Geschichte* (Cologne, 1893); H. Lehmann, *Zur Geschichte der Glasmalerei in der Schweiz.* (Mitteilungen der Antiquarischen Gesellschaft in Zürich, 1903–12); H. Oidtmann, *Geschichte der Schweizer Glasmalerei* (1905); L. F. Day, *Windows* (London, 1909); Leo Balet, *Schwäbische Glasmalerei* (Stuttgart, 1912); M. Drake, *History of English Glasspainting* (London, 1912); H. Schmitz, *Die Glasgemälde des Kgl. Kunstgewerbemuseums in Berlin* (Berlin, 1913); G. Heinersdorff and K. Scheffler, *Die Glasmalerei* (Berlin, 1913); P. Nelson, *Ancient Painted Glass in England* (London, 1913); J. L. Fischer, *Handbuch der Glasmalerei* (Leipzig, 1914); H. Schmitz, *Deutsche Glasmalereien der Gothik und Renaissance (Rund- und Kabinettscheiben)* (Munich, 1923); H. Lehmann, *Zur Geschichte der Glasmalerei in der Schweiz* (Leipzig, 1925); A. Hoff, *Thorn-Prikker und die neuere Glasmalerei* (1925); J. D. Le Couteur, *English Mediaeval Painted Glass* (London, 1926); H. Read, *English Stained Glass* (London, 1926); F. Kieslinger, *Gotische Glasmalerei in Österreich bis 1450* (Zurich, 1927); F. Gaudin, *Le Vitrail* (Paris, 1928). (H. Rd.)

STAINES, an urban district of Middlesex, Eng., which under the London Government Act 1963 was to become part of Surrey on April 1, 1965 (*see* London). It is on the Thames 17 mi. S.W. of Hyde Park corner. Pop. (1961) 49,838. Area 12.9 sq.mi. A Roman station was situated there, on the main road from London, and this Thames crossing was one of the earliest bridged. The present bridge, by George Rennie, was opened in 1832. The London Mark stone was placed there in 1285 to mark the western limit of the City of London's jurisdiction over the Thames. An annual fair was granted in 1228. The urban district includes Ashford, Laleham and Stanwell; it is mainly residential, but there are engineering works and some factories.

STAINLESS STEEL. Stainless steel basically is an alloy of chromium and iron containing more than 10% chromium. When chromium is added to iron in such quantities, it imparts throughout the entire body of the metal remarkable resistance to corrosion and heat. Other elements may be added to the iron-chromium combination to obtain steels with special characteristics. The most important of these is nickel, which is used to produce 18–8 (that is, 18% chromium, 8% nickel, balance iron). Other useful additions are columbium, manganese, molybdenum, phosphorus, selenium, silicon, sulphur, titanium and zirconium, all of which result in modifications for special service or fabrication. As is true of many scientific discoveries no single nation

can claim full credit for the development of stainless steel. Between 1903 and 1912 Harry Brealey of England, F. M. Becket of the United States and Benno Strauss and Edward Maurer of Germany shared in its initial development.

Selected scrap which contains only those elements desired in the finished product usually make up the bulk of the furnace charge used for producing stainless steel. If alloy scrap is not available, clean open-hearth scrap may be used, but more alloy is required to meet specifications. Pig iron containing 4% carbon, anthracite, or other carbonaceous material may be employed when a high-carbon melt is desired. Lime, silica, crushed coke, aluminum, silicon, fluorspar and iron ore are materials used for slag constituents and for controlling the composition of the metal bath. Because of the high temperatures and precise chemical control required, all stainless steel is produced in the electric furnace. Oxidizing and reducing slags may be used successively in the same heat, depending on the composition of the charge and the analysis of the product. Sulphur is removed from the metal with a reducing slag containing more than 2% calcium carbide, and silicon, phosphorus, manganese, chromium and carbon, with an oxidizing slag. However, carbon presents a problem in the presence of chromium, because of the chemical affinity chromium has for carbon. Carbon may be absorbed by the metal at the same time chromium is being oxidized and going into the slag. To oxidize carbon in a charge containing a high percentage of stainless steel scrap requires an excess of oxygen and extremely high temperatures (more than 3,200° F.). Oxygen supplied through a lance into the molten bath is used to a large extent to replace iron ore as a source of this element. Considerable chromium is oxidized along with the carbon, and the resultant slag will contain a high percentage of chromium oxide. Chromium oxide may be reduced and returned to the metal by the addition of lime and crushed ferrosilicon.

The stainless steels may be grouped according to chemical composition and response to heat treatment as follows: (1) ferritic steels, nonhardenable steels containing 15% to 30% chromium and with low carbon contents (from 0.08% to 0.20%), (2) martensitic steels, hardenable (by quenching) steels containing 10% to 18% chromium with varying carbon contents (0.08% to 1.10%), and (3) austenitic steel containing 16% to 26% chromium and 6% to 22% nickel, which is hardenable without quenching.

Ferritic Stainless Steels are better than the martensitic type and as good as the austenitic type in resistance to corrosion. However, because of their lower ductility they are not used to replace the austenitic type. Ferritic steels are magnetic and can be hot- and cold-worked but usually undergo excessive grain growth during prolonged exposure at elevated temperatures. They have lower high-temperature strength than the other two types and may develop brittleness after electric-arc, resistance or gas welding. Machinability is good and can be improved by adding sulphur or selenium. There is no tendency toward intergranular corrosion.

The production in the United States of ferritic stainless American Iron and Steel Institute (A.I.S.I.) type 430, widely used for automotive and architectural trim, was approximately a third of the total ingot stainless production in the 1950s.

Martensitic Stainless Steels are magnetic; they can be cold-worked without difficulty, especially when the carbon content is low; show satisfactory resistance upon exposure to weather, water and some chemicals; can be machined satisfactorily; have toughness; and are easily forged or otherwise hot-worked. The corrosion resistance is usually increased by hardening, which is accomplished by quenching. Welds made by the electric-arc, resistance or gas methods contain some martensite after cooling in air. A.I.S.I. type 410 is the most widely used of the martensitic grades. In sheet or strip form, this grade is used extensively in the oil industry for bubble trays, bubble caps and liners.

Austenitic Stainless Steels are not magnetic unless severely cold-worked. They can be hot-worked, and cold-worked if proper allowance is made for their rapid work hardening. Except for surface hardening, their alloys cannot be hardened by heat treatment. They are extremely shock-resistant and difficult to machine unless they contain sulphur or selenium. These steels can be welded with difficulty, and, if the welding operation is controlled properly, the resistance to corrosion is not impaired. These steels have the best high-temperature strength and resistance to scaling of the stainless steels. However, they are subject to intergranular corrosion at temperatures between 800° and 1,600° F. unless the structure is stabilized to prevent the presence of carbides at or near grain boundaries which are vulnerable to corrosion. Stabilization may be done by the addition of titanium (4 to 6 times the carbon content) or columbium (8 to 10 times the carbon content), which combine with carbon and prevent the formation of chromium carbides. The corrosion resistance of the austenitic stainless steels is usually better than that of the other two groups.

A.I.S.I. type 304, a companion grade to types 301 and 302 with somewhat lower carbon (0.08% maximum), is the most widely used of austenitic grades. It is well suited to withstand severe forms of corrosion encountered in the paper and chemical industries. The austenitic grades most resistant to oxidation are types 309 and 310, which are high in nickel and chromium. These steels resist scaling up to 2,100° F. and consequently are used for furnace parts and heat exchangers.

The heat treatment of stainless steels may be carried out with any conventional type of electric, gas-fired, oil-fired, salt-bath or induction furnace. Although all stainless steels are highly resistant to oxidation, they may be expected to scale slightly at high heat-treating temperatures. The scale formed on the stainless steels is tenacious and relatively impervious and, consequently, is more difficult to remove than scale on ordinary carbon steel. Two basic methods are used: mechanical, such as sandblasting and tumbling; and chemical, such as pickling in acid solutions.

An important development in the stainless steel industry was the increased use of austenitic stainless steels containing manganese instead of nickel as a major alloying element. These were substituted for A.I.S.I. type 301 (18% Cr, 8% Ni), because of the current shortage of nickel. Manganese, like nickel, has the property of producing an austenitic structure in steel, although it takes twice as much to do it. Steel containing 15% Cr minimum, 16.5% Mn, 0.10% carbon maximum, and 1% Ni as a substitute for type 301 stainless is being used.

BIBLIOGRAPHY.—Ernest E. Thum (ed.), *The Book of Stainless Steels* (Cleveland, 1935); American Iron and Steel Institute, *Annual Statistical Report*, pp. 45, 726 (New York, 1952); Allegheny Ludlum Steel Corporation, *Handbook on Stainless Steel* (Brackenridge, Pa., 1946); Allegheny Ludlum Steel Corporation, *Stainless Steel Handbook* (Brackenridge, Pa., 1951); Rustless Iron and Steel Corporation, *Heat Treatment of Stainless Steels* (Baltimore, 1944); American Society for Metals, *Metals Handbook*, pp. 325–332, 549–569 (Cleveland, 1948); J. H. G. Monypenny, *Stainless Iron and Steel*, 2 vol., 3rd ed. (London, 1951); D. K. Bullens, *Steel and Its Heat Treatment*, 3rd ed., pp. 280–292 (New York, 1935); J. M. Camp and C. B. Francis, *Making, Shaping and Treating of Steel*, United States Steel Corporation, ch. 35, pp. 1313–1337 (Pittsburgh, 1951); *Iron Age*, vol. 171, no. 1, pp. 135–138 (New York, March 12, 1953); U.S. Dept. of the Interior, Bureau of Mines films, *Stainless Steel*, no. 201.89, and *The Melting and Refining of Stainless Steel*, no. 240.33. (N. B. MR.)

STAIR, JAMES DALRYMPLE, 1st VISCOUNT (1619–1695), Scottish lawyer and statesman, was born in May 1619, at Drummurchie in Ayrshire. After seven years as regent of the University of Glasgow he resigned, going to Edinburgh, where he was admitted to the bar on Feb. 17, 1648. In 1649 he was appointed secretary to the unsuccessful commission sent to The Hague to treat with Charles II by the parliament of Scotland, and was sent in the following year to Breda, where the failure of Montrose's expedition forced Charles to change his attitude and to return to Scotland as the covenanted king. Stair met him on his landing in Aberdeenshire. He refused in 1654 to take the oath of allegiance to the Commonwealth. Three years later (1657), Stair was appointed a commissioner of justice in Scotland, on the recommendation of Monk. After the Restoration he was received with favour by Charles, knighted, and made a judge in the court of session. He refused to take the declaration that the National Covenant and the Solemn League and Covenant were unlawful oaths, and, forestalling deposition, he resigned. The king, however, summoned him to London and allowed him to take the declaration under an implied reservation. In 1669 a family calam-

ity, the exact facts of which will probably never be ascertained, overtook him. His daughter Janet, who had been betrothed to Lord Rutherford, was married to Dunbar of Baldoon, and some tragic incident occurred on the wedding night, from the effects of which she never recovered. (Sir Walter Scott took the plot of his *Bride of Lammermoor* from this incident, but he disclaimed any intention of making Sir William Ashton a portrait of Lord Stair.)

In 1670 Stair served as one of the Scottish commissioners who went to London to treat of the union. In Jan. 1671 he was appointed president of the court of session. In 1672 and again in 1673 he was returned to parliament for Wigtownshire. When the Test act was enforced Stair, dreading a fate similar to that of Argyll (*q.v.*), went to London to seek a personal interview with the king; but the duke of York intercepted his access to the royal ear, and when he returned to Scotland he found a new commission of judges issued, from which his name was omitted.

He retired to his wife's estate in Galloway, and occupied himself with preparing for the press his great work, *The Institutions of the Law of Scotland,* which he published in the autumn of 1681, with a dedication to the king. He repaired to Holland in Oct. 1684, and took up his residence, along with his wife and some of his younger children, at Leiden, where he published the *Decisions of the Court of Session Between 1666 and 1671* and a small treatise entitled *Physiologia nova experimentalis.*

In his absence a prosecution for treason was raised against him and others of the exiles by Sir G. Mackenzie, the lord advocate, but it was dropped because of the appointment of his son, the master of Stair, who had made his peace with James II, as lord advocate in place of Mackenzie. Stair returned from Holland in the suite of William of Orange. In 1689 he was again placed at the head of the court of session. He was attacked by an anonymous pamphleteer, perhaps Montgomery or Ferguson the Plotter, in a pamphlet entitled *The Late Proceedings of the Parliament of Scotland Stated and Vindicated.* He defended himself by publishing an *Apology.*

Shortly after its issue he was created Viscount Stair (1690). The massacre of the Macdonalds of Glencoe (Feb. 13, 1692), for which his son, the master of Stair, was largely responsible, naturally reflected on him.

On Nov. 29, 1695, Stair died in Edinburgh.

Stair's great legal work, *The Institutions of the Law of Scotland Deduced From Its Originals, and Collated With the Civil, Canon and Feudal Laws and With the Customs of Neighbouring Nations,* is important for the historical study of Scottish law.

BIBLIOGRAPHY.—J. Murray Graham, *Annals of the Viscount and First and Second Earls of Stair* (1875); A. J. G. Mackay, *Memoir of Sir James Dalrymple, First Viscount Stair* (1875); Sir R. Douglas, *Peerage of Scotland,* new ed. by Sir J. B. Paul.

STAIR, a series of steps arranged one above the other, occasionally varied by platforms known as landings. The total height of a stair is the rise, the total horizontal distance between the top and bottom steps is the run. A series of steps without a landing is called a flight. The horizontal surface of a step is called its tread, and the vertical front its riser. The projection of the tread beyond the face of the riser is termed a nosing. Staircase is a term for stairs accompanied by walls.

HISTORY

It is uncertain what is the oldest stair in existence. On the sacred road up the mountain of Tai Shan, in Shantung, China, there are many great flights of ancient granite steps. The earliest staircases seem to have been built with walls on both sides, as in Egyptian pylons which date from the 2nd millennium B.C. Contemporary models of houses, dating back to the middle kingdom, show a staircase leading up the side of a court to a roof terrace, with continuous wall below.

The origin of such stairs may have been earlier structures in wood, or even ladders cut out of logs. Egyptians may have used the self-supporting staircase in which each step is supported at its front by the step below and prevented from turning by having its end built solidly into a wall. The Assyrian ziggurat (*q.v.*) of the 9th or 8th century B.C. was often adorned with massive stairs; and the palace terrace at Persepolis has a double flight of steps, of great beauty, which shows the skill of the Persians in the 6th century B.C.

Farther west, stairs played an important part in the Cretan palaces, as in the palaces at Knossos and Phaistos (both *c.* 1500 B.C.). In Hellenic Greece the monumental stair was of less importance, due to the Greek love of informal and picturesque approaches. Thus in Athens the Propylaea of the Acropolis was probably entered by a winding path and the present great flight of stairs, dating from the Roman period, is not part of the original design. Interior staircases were common but unimportant; they were placed in out-of-the-way corners, and often built of wood. In the post-Alexandrian period there developed a new love of monumental grandeur, largely Asian in origin, which produced such magnificent stairs as those of the altar of Zeus at Pergamon (180 B.C.).

A similar monumental sense distinguished Roman stairs. Those of the propylaea at Baalbek (time of Hadrian), over 100 ft. wide, are typical, and again and again stairs are made essential parts of great architectural conceptions, as in the Temple of the Sun on the Quirinal hill at Rome, built by Aurelian (A.D. 273), some of the steps of which were taken, in 1348, to build the famous stairs of the Aracoeli on the Capitoline hill at Rome. Many of the apartment houses indicated on the marble plan of Rome (*c.* A.D. 205), now in the Palazzo dei Conservatori at Rome, seem to have had stair towers built in the centre of a courtyard. The Colosseum (*q.v.*) at Rome (completed A.D. 80) has elaborate and practical entrance and exit stairs, by which the crowds could be readily handled. These, in general, are supported and roofed by sloping vaults. Similar staircases were common in the theatres.

Merely inclined planes lead up within enormous buttresses at the ends of the narthex of Hagia Sophia (532) at Constantinople, so giving access to the galleries. In St. George at Salonika (late 5th century) a spiral staircase surrounds an open well. In the early Romanesque there is a growing use of the spiral staircase. To this development the great thickness of Romanesque walls and buttresses was particularly congenial, as it allowed the staircases to be built in solid masonry. Not only was the open circular well used, but the solid newel, in which each step had cut upon it a cylindrical form to act as part of the newel. This type of spiral staircase is still in use.

Although during the European middle ages stairs were used more for utilitarian than aesthetic purposes, where they were necessary they received adequate treatment, usually more direct and informal and more picturesque than those of the Roman tradition. Thus at Le Puy en Velay, in France, the 11th-century cathedral, which was built on a hilltop, is approached by a magnificent stair that provides a superb foreground to the polychrome façade. The steps that mount up to and past the apse of the 14th-century cathedral at Erfurt, in Germany, show the direct and beautiful relationship between the building itself and the stairs which is characteristic of the best Gothic usage.

The Gothic spiral staircase was cut with the utmost cleverness. Such circular staircases as those which lead up in the traceried turrets of the spire of Strasbourg cathedral in Alsace (1435) are miracles of daring design and delicate execution. The spiral staircase at the abbey church at St. Gilles, which is known as Vis St. Gilles, is of hewn stone built round a circular newel and covered with a semicircular rising vault.

The late Gothic period also produced a great number of beautiful interior staircases, like the "Stair of the Library" in Rouen cathedral (1477–79), which show that the possibilities of the staircase with straight runs were at last being realized. Of rich interior spiral staircases, the early 16th-century staircase leading to the organ in St. Maclou, in Rouen, and the contemporary staircase to the jube (*q.v.*), or rood loft, of St. Étienne du Mont in Paris (in which a certain amount of Renaissance feeling is present) are noteworthy. In the larger houses the lack of communicating passages led to many staircases. Thus in the house of Jacques Coeur at Bourges (*c.* 1450) there are eight separate spiral

staircases, whose exterior treatment, in rich, traceried towers, furnishes a great deal of the picturesqueness and varied beauty of such a building.

In France the spiral staircase added by Francis I to the château at Blois (1515–24) and the double circular staircase of the château of Chambord (begun 1519) are remarkable. The early Renaissance, however, produced its most remarkable results in the richly ornamented wooden staircases of Tudor and Jacobean houses of England. The usual arrangement was to have the staircase divided into comparatively short flights, at right angles to each other, around a central open well. The heavy newels at the corners were finished with urnlike finials at the top and carved drops at the bottom; the railing consisted of a large rail supported either by miniature arcades or square balusters, whose moldings were often sloped to follow the stairs; and the whole was covered with intricate surface ornament of strapwork. The staircase at Hatfield house (1611) is typical.

The early Renaissance staircases of Spanish palaces are in most cases placed in one corner of the courtyard and run up between walls to a landing with a return flight that opens out onto the upper-floor gallery. The walls, and sometimes even the treads and risers, are often cased in brilliant faïence tiles. Even in staircases of more Italian type, with balusters, the richness of wall surface remains, as in the staircase of the hospital of Santa Cruz at Toledo (1504–14), by E. de Egas. The climax of Spanish Renaissance monumental stair design is reached in the Escalera Dorada at Burgos cathedral (1519), by D. de Siloe, with its exquisite metal railing.

The Renaissance produced a recrudescence of the tradition of Roman formality in stair design. Despite such tours de force as G. Vignola's spiral stair in the palace at Caprarola (*c.* 1550), the general practice was to have important staircases run up in a straight flight, sometimes varied by landings, and sometimes with changes in direction between walls often crowned with a slanting vault. The most perfect developments of this type were made chiefly during the baroque period. The magnificence of the huge interior flights of the Genoese palaces (*e.g.*, the University, 1623, by B. Bianco) is famous. The so-called Spanish stairs at Rome in the Piazza di Spagna (1723–26), by F. de Sanctis, those of the Campidoglio, at Rome (begun 1547), by Michelangelo, and the infinitely varied terrace stairs of the Italian villas should also be mentioned. The inspiration of these Italian baroque stairs was felt throughout Europe and is largely the foundation of stair design in modern city and garden planning. Thus the enormous stairs of the gardens of Versailles (1667–88), by A. Le Nôtre, owe much to the Italian precedent, and were themselves imitated widely all over Europe.

Classicism combined with the baroque to give a new trend to staircase design throughout Europe. In France a gorgeous series of staircases was produced, of which the Escalier des Princes of Versailles is typical. The reaction to the lightness and gaiety of the Louis XV period (*see* Louis Styles) led to the development of many gracious and inviting house staircases of the type known as self-supporting, full of sweeping curves and with railings frequently of metal. In the Louis XVI period the new, popular classicism restrained the exuberance of the earlier staircases but retained their graciousness and lightness; of these, one of the loveliest is that of the Petit Trianon at Versailles (1766), by A. J. Gabriel, with a beautiful iron railing. Meanwhile, the classic trend had expressed itself in England by the substitution of open-string for closed-string stairs, that is, with the balusters coming down to the top of each tread rather than being supported on a slanting member, or string, which received the ends of the treads. This at once made for a more interesting treatment at the ends of the steps and led, eventually, to the development of the bracketed step end, in which scrolls and leafage of great richness were carved under the end of each riser. Moreover, the fact that English stairs were largely of wood inspired the design of turned balusters and newels of great variety. During the last half of the 18th century a growing trend toward lightness and delicacy of design led to the substitution of plain, tapered shafts in place of elaborate balusters, and the use of stairs rising in continuing curves instead of straight flights and landings. The same development took place during the 18th century in the colonial work of North America. Thus certain of the earliest houses, like the Capen house at Topsfield, Mass., 1683, have closed strings and heavy newels of Jacobean type, while the Lee mansion at Marblehead, Mass., 1768, has a magnificent open-string staircase, whose steps have paneled and scrolled ends and whose twisted balusters are as rich as any contemporary work in England. The curved staircase of the Valentine museum at Richmond, Va., 1812, is typical of the later delicacy.

During the middle and the third quarter of the 19th century staircase design reached its lowest ebb, not only in England and France, but throughout Europe, except in the case of those monumental and official buildings in Europe whose design never fell a victim to the current taste. Closed strings, balusters and newels of gargantuan proportions, and bulbous and meaningless elaboration again became the fashion. Yet one of the most gorgeous and effective staircases in the world dates from this period, that of the Paris Opéra (1861–74), by Charles Garnier.

Two elements have vastly affected modern staircase design, steel and reinforced concrete. The use of steel has led to simple plans of straight runs, and railings delicate and straightforward in design. Reinforced concrete, on the other hand, is a material so flexible that the most daring curves and fantastic sweeps become structurally sound. The masters of contemporary architecture have given new vitality to staircase design with such superb examples as the glass enclosed circular staircase of the Werkbund Exposition building at Cologne, Ger. (W. Gropius and A. Meyer), the interior stair of Frank Lloyd Wright's own house at Spring Green, Wis., and "Les Terrasses" at Garches, France, by Le Corbusier.

Contemporary architecture has largely de-emphasized the monumental stair and the monumental building. Most frequently entrances are effected at ground level. Emphasis has been placed on the small, light, open, free-standing stair (Gropius house, Lincoln, Mass., W. Gropius, 1939); on the intimate and picturesque (Wellesley college art centre, Mass., P. Rudolph, 1957); and on the short-run formalized axial solution (Architecture building, Illinois Institute of Technology, Chicago, Mies van der Rohe, 1955). There is, perhaps, some evidence that contemporary architects have given exterior stairs something less than adequate study.

In the orient the stairs leading down to the river at Benares, in India, of various dates, combine monumentality and picturesqueness; a similar skill is shown in the design of large numbers of temple and palace steps throughout the peninsula. Farther east the ruins of Angkor use wide flights of steps with magnificent effect in connecting the various levels of the terraced group. The Chinese have always been superb stair designers and builders, and there is hardly a temple or large house or palace which does not owe some of its effect to its stairs, either monumental and formal, like those of the Forbidden City or the Temple of Heaven, both in Peking, or winding and informal, as in Chinese garden design. Many of the temple stairs have, in the centre, an inclined plane running the entire height of the rise, richly carved with the symbolic dragon—the so-called spirit stairs by which the beneficent powers are supposed to enter. In Japan the tradition is much more informal, but the hilly nature of the country makes stairs an important feature of almost every temple layout.

STRUCTURAL DESIGN

Long use has demonstrated that certain proportions of riser and tread are safe and easy. For interior stairs twice the height of the riser, plus the length of the tread, should be 23, all dimensions being in inches. A variant is that the sum of riser and tread should be between 17 and 17½ in.; a third rule specifies that the product of riser and tread, in inches, should be not less than 70 nor more than 75. In general, English usage advocates a slightly lower rise for a given tread than U.S. usage, while in Europe the rise is frequently greater. For exterior stairs, the general usage is to keep the riser below 6 in. and the tread above 12.

Stone and Marble Staircases.—Modern stone and marble staircases are of two main types. In the first the actual exposed

treads and risers form, themselves, the structural elements. In the other type the treads and risers are supported upon a separate structural base which may consist of a masonry vault, a steel framework or a reinforced concrete slab. In the first class, in which the steps themselves are structural, the staircase is usually built into a wall at one side. Each step rests for a small distance upon the step below and is built into the wall for a distance sufficient to make the weight of the wall above prevent it from turning. The undersides of the steps, where exposed, are frequently cut to form an inclined surface, following the slope of the stairs, and may be decorated with paneling or carving or other ornament. This type of stair can be used both for straight and curved runs, and is particularly applicable to monumental entrance halls, where the stair, in a triple run with two landings, or in a continuous, curved sweep, fills a recess opposite the door. In such staircases the railing is usually of metal with the uprights securely doweled into the ends of the steps. Stairs somewhat similar can be built with the ends of the steps built into the walls on both sides, where a monumental run rises between two walls.

Wooden Staircases.—These can be divided into two classes, closed- and open-string stairs. In both, the essential support is given by rough timbers called carriages, which follow the slope of the stairs and are cut roughly to the shape of the underside of the steps.

In the case of closed-string staircases both sides of the staircase are finished by a sloping, straight-sided plank, approximately an inch and a half thick, on the inner faces of which grooves are cut, whose upper and outer faces exactly fit the top and side of the tread and riser; the under- and inner sides are, however, sloped, so that wedges may be driven in to force the tread and riser boards into position and hold them with perfect rigidity. Similar wedges are driven in between the rough carriages and the undersides of treads and risers. To give additional strength, the treads and risers are themselves rebated together, so that a small, projecting strip of the tread fits into a groove in the bottom of the riser, and a similar strip on the top of the riser fits into a groove on the bottom of the tread. The projecting portion of the tread is molded and an additional small cover mold placed below to cover the joint.

In the open-string staircase the strings at the end are cut to the shape of the steps, and the vertical cuts are usually mitred at 45°, so that the ends of the risers, also mitred, may form with them an invisible joint. In the ends of the treads dovetailed sockets are left to receive the dovetailed ends of the balusters; a molding, exactly similar to the nosing, is then applied across the end of the step and covers the bottom of the balusters. At times, bracketed and carved pieces are mitred at the ends of the risers and applied over the face of the string. It is this type of staircase that was the rule during the Georgian period in England and in the developed American colonial and early republic work. Where an open-string stair is built against a wall, it is usually finished at the wall end by a closed or housed string, similar to that described, forming a baseboard for the wall.

The structural strength for a wooden staircase is usually furnished by framing a heavy member, called a header, at the top and bottom of every run; to these headers the rough carriages are spiked, and the newels, if they exist, firmly fastened. The newels themselves are designed to receive the finished stair strings, and also the finished string member that is occasionally used at the edges of a floor in the stair-well opening. In modern staircases of this type additional strength is obtained by inserting iron or steel bands into the handrail; occasional balusters may be of steel, well fastened to the string.

Steel Staircases.—These have become almost universal in modern fireproof buildings of all types, except where unusual monumental effect is desired. The design of steel staircases is largely standardized. They consist of a steel string on each side, which takes the form of a plate or channel of sufficient rigidity to carry the weight.

Iron spiral staircases are usually formed of cast-iron members, each of which forms a tread and riser and has, at its inner end, a collar arranged to slip over a vertical pipe or cylindrical bar forming the central newel. Such spiral staircases are rigid independently of any wall, and are frequently used in open spaces, such as those existing in power houses, gymnasiums and the like. Reinforced concrete, sometimes finished as artificial stone, is used for stairs and permits of infinite variety of form.

Moving Staircases.—In many places where continuous transit between floors is necessary, as in large shops, underground railway stations and the like, moving stairs, sometimes known as escalators, are much used. (*See* ESCALATOR.)

Legal Regulation.—As stairs form one of the most necessary methods of exit from a building, their design has formed one of the most important portions of the building legislation of modern municipalities, especially in the case of public, commercial, industrial and educational buildings. Thus, in addition to the universal requirements for structural solidity, the attempt is generally made to furnish at least two stair exits to any large, enclosed, upper-floor area; and where fire risk and the danger of panic are unusually great, as in the case of theatres, to demand sufficient stairs to allow the building to be emptied in a very short time without disorder. Design is controlled by specifying the relationships of tread and riser, the distance apart and size of landings, whether or not curved stairs are allowed, etc., and often, because of the smoke danger, requiring exit stairs to be enclosed in a fireproof tower, with fireproof, self-closing doors at all entrances.

BIBLIOGRAPHY.—A. Swan, *The British Architect or the Builder's Treasury of Staircases* (1775); Viollet-le-Duc, *Dictionnaire Raisonné de l'Architecture* (1854–76); P. Le Tarouilly, *Edifices de Rome Moderne* (1840–55); Byne and Stepley, *Spanish Architecture of the 16th Century* (1917); P. G. Knoblauch, *Good Practice in Construction* (1923–26); T. F. Hamlin, *Forms and Functions of Twentieth-Century Architecture*, vol. i (1952). (T. F. H.; H. MN.)

STALACTITES AND STALAGMITES. Stalactites are elongated pendent forms of various minerals deposited from solution where very slow drip water may enter a void. Stalagmites are complementary forms of thicker proportions which grow up on the bottom of a cavity from the same drip-water source but whose mineral was deposited after free fall across the open space in the rock. An inclusive name for both forms is dripstone.

The dominant mineral in dripstone deposits is calcium carbonate (*see* CALCITE) and the largest displays are formed in caves in limestone and dolomite. Other minerals which may be deposited in these forms include other carbonates, opal, chalcedony, limonite and some sulfides. Icicles are stalactites of water. Ice in stalagmitic form is very rare.

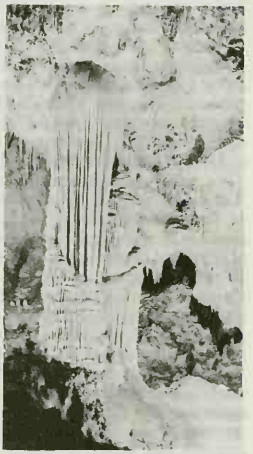
Conditions which favour dripstone deposition of calcium carbonates are (1) a source rock above the cavity; (2) downward percolation of water supplied from rain; (3) tight but continuous passageways for this water which determine a very slow drip; (4) adequate air space in the void to allow either (*a*) evaporation or (*b*) escape of carbon dioxide from the water which thus loses some of its solvent ability. Minor and imperfect stalactites may hang from the underside of masonry and concrete bridges and arches, and from ceilings of concrete arched tunnels and subways. One-way traffic through such structures may produce forms that hang out of plumb, inclining in the direction taken by recurrent air currents engendered by passing vehicles. Even in some caves, an air circulation is believed to have been the cause of inclined stalactites.

Stalactites, but not stalagmites, commonly exhibit a central tube or the trace of a former tube whose diameter is that of a drop of water hanging by surface tension. Examination will show that the drop on the tip of a growing stalactite leaves a deposit only around its rim. Downward growth of the rim makes the tube. The simplest stalactite form, therefore, is a thin-walled stone straw, and these fragile forms may reach lengths of two feet or more where air currents have not seriously disturbed the growth. The more common form of a downward-tapering cone is simply a thickening of the primitive straw type by deposition from a film descending the exterior of the pendant. This also makes the concentric structure seen in broken stalactites, slight variations in colour from impurities demarking successive additions. Neither these nor the comparable rings of stalagmite structure are annual rings.

Related to stalagmites are broader, flatter cones on a cave floor

which generally grade outward into a sheet of secondary lime known as cave travertine but more specifically as flowstone.

Stalactite forms may grade into very broad pendent leaves, sheets, blades and drapes, whose growth is entirely on the edges. Other variants of the conventional carrot-shaped stalactite may have a bulbous upper portion resembling the form of a radish, beet or turnip. More puzzling than this enlargement is a common ceiling attachment by a slender stem. Compound forms of much complexity may result from continued lateral enlargement of closely adjacent stalactites and stalagmites. Not every stalactite has a complementary stalagmite, and many of the latter may have no stalactite above them. Where the paired relation exists, however, continual elongation of one or both eventually results in a junction and the formation of a column. Large stalagmites and columns have a habit of developing shelves or brackets in vertical series, with depending drapes or relatively small stalactites growing down from the bottom of each.



BY COURTESY OF U.S. DEPARTMENT OF THE INTERIOR

JUNCTURE OF STALACTITES AND STALAGMITES FORM A COLUMN OF LIMESTONE IN THE GREEN LAKE ROOM, CARLSBAD CAVERNS NATIONAL PARK, CARLSBAD, N.M. Known as the Veiled Statue, this formation is the result of the disposition of carbonate of lime, carried in solution by descending ground water

The slow rate of growth of these calcium carbonate deposits is conditioned by a variety of variable factors. Deposition from drip water indicates saturation of the water under the existing conditions. But these conditions may change so much that deposition will cease although the drip continues. In some cases, solution of a previously growing stalagmite may occur and produce a cratered concavity where once was a projecting conical tip.

Lava caves containing lava stalactites have a wholly different genesis. The cave results from surface solidification of a lava flow during the last stages of its activity. Thus what is essentially a frozen crust may form over still mobile and actively flowing liquid rock. Dwindling of the supply may then leave long tunnels under this crust. Volcanic gases from bubbles in the lava collect under the tunnel roof. Mingled with air from vents in the roof, more intense heating from oxidation may raise the temperature of the gas sufficiently to re-fuse the ceiling rock, which then drips with the remelted lava, which may congeal in place to form rude stalactites.

See also CAVE and articles on various caves, such as LURAY CAVERN; POSTOJNA; etc. (J. H. Bz.)

STALACTITE WORK, in architecture, one of the most general characteristics of all of the Islamic styles, consisting of a series of little niches, bracketed out one above the other, or of projecting prismatic forms in rows and tiers, connected at their upper ends by miniature arches. Its infinite varieties may be classified into three groups, the first consisting of those basically niche shaped, in which the concave curve is the most important feature; the second group includes those in which the vertical edges between the niches are the most important feature, being set at all sorts of angles and having generally prismatic forms; the last group consists of elaborately intersecting, miniature arches. The first two groups occur commonly in Syrian, Moorish and Turkish work and in their simpler forms, in Persia; the last group is typically Persian and is found also in Mogul work in India.

Stalactites are of comparatively late development in Islamic art, the earliest buildings in Syria, Egypt and north Africa showing no traces of them. They seem to appear suddenly all over the Islamic world toward the beginning of the 12th century. Thus simple forms are found in the mosque at Ani in Armenia, built between 1072 and 1110, and they are common in Algiers and Sicily during the course of the century, as in the gate at Chella (1118–84), and in the building, which is known as La Ziza at Palermo (1154).

In Egypt the stalactite appears in panels flanking the main entrance of the El Akmar mosque at Cairo (1125–50). They reached their highest development in the 14th and 15th centuries, becoming the normal decoration for the heads of door niches and the bracketing under cornices and minaret galleries. The richest examples of the prismatic type are to be found in Moorish work in Spain, especially in the intricate wood and plaster ornament of such palaces as the 14th- and 15th-century Alhambra (*q.v.*) in Granada and the 14th-century Alcazar at Seville.

In Turkey, a peculiar type of faceted crystal shape is found and the form became the most common capital decoration.

See ISLAMIC ARCHITECTURE.

STALIN, JOSEPH VISSARIONOVICH (DZHUGASHVILI) (1879–1953), leader of the Communist party of the Soviet Union, prime minister, generalissimo and dictator of the U.S.S.R., was born on Dec. 21, 1879, in the town of Gori, Georgia, Transcaucasia.

His father Vissarion Dzhugashvili, a poor shoemaker, died when Stalin was 11 years old. His mother Ekaterina, the daughter of a serf, was an illiterate and deeply religious washerwoman. After the death in infancy of three children, Stalin was her only son. She brought him up with much tenderness and devotion and chose for him an ecclesiastical career in the Georgian Orthodox Church.

The Young Revolutionary.—Georgian folklore, native tribal traditions, fresh memories of serfdom and Georgian patriotic (anti-Russian) sentiment were the strongest influences in Stalin's childhood. He attended the parish school at Gori and, in the autumn of 1894, he matriculated at the theological seminary of Tiflis (Tbilisi). He was a very able pupil, a voracious reader and an ambitious debater. In 1898, while he was still at the seminary, he joined the ranks of Messame Dassy (the Third Group), a Georgian patriotic and socialist organization; and he came under the influence of Marxism, which was rapidly spreading among the Russian and Georgian intelligentsia. In 1899, when he was about to graduate, he was expelled from the seminary because of his "disloyal" views.

He plunged into revolutionary propaganda and agitation. At the turn of the century the revolutionary movement in Russia shed the ideas of agrarian socialism preached by the *Narodniks* and adopted the Marxist social-democratic program, advocated by G. V. Plekhanov and V. I. Lenin in *Iskra.* The young Stalin was a supporter of *Iskra.* In 1901 he was elected member of the clandestine Social Democratic committee of Tiflis. Police persecution forced him to "go underground." Arrested in April 1902, he remained in prison until the end of 1903. But while in prison he was elected member of the Social Democratic committee guiding the movement in the whole of the Caucasus. Deported to Siberia, he escaped from there and returned to Tiflis early in 1904, shortly after the Social Democratic party had split into bolsheviks and mensheviks. He joined the bolsheviks, who wielded little influence in Georgia. At this early stage he was already under the overwhelming influence of Lenin's writings; and in local clandestine sheets and periodicals he translated Lenin's ideas from the Russian into the Georgian.

In the revolution of 1905 Stalin played no conspicuous part, but was active in the background as an organizer of bolshevik fighting squads and partisan groups. He had a hand in organizing raids on treasury transports, the so-called "exes" (expropriations), the raiders "expropriating" government funds and using them to finance the party's revolutionary activity. Stalin first met Lenin at a party conference at Tampere (Tammersfors), Fin., in 1905. In 1906 and 1907 he participated in party congresses held in Stockholm and London; and in opposition to Lenin he advocated a policy of agrarian revolution identical with that later adopted by the bolsheviks in 1917. After the defeat of the 1905 revolution he

distinguished himself as the bolshevik leader of Baku. Many times arrested and deported, he succeeded each time in escaping from the places of exile.

In 1912 the bolsheviks finally broke with the mensheviks and constituted themselves into a separate party. Lenin then co-opted Stalin as member of the new central committee. Stalin (he began to use this pseudonym, the Man of Steel, early in 1913) was the first editor of *Pravda*, which began to appear on May 5, 1912. At the end of that year he attended important bolshevik conferences in Cracow, Pol., and then went to Vienna, Aus., to write *Marxism and the National Problem*. This important study was a theoretical argument in favour of the bolshevik program of "self-determination" of oppressed nations. The work established his reputation as a foremost expert in this field, but he was looked upon as a man of action and an organizer, not a theoretician. On his return to St. Petersburg he was again arrested, and deported to the Turukhan province in the polar circle. He remained in exile until the outbreak of the February (March, western calendar) revolution of 1917.

Stalin's Role in 1917.—Stalin's role in the events of 1917 was not prominent. This was the heyday of the brilliant revolutionary orators and agitators; and Stalin was not one of them. He was overshadowed not only by L. D. Trotsky, who was much more than an orator and agitator, but by G. E. Zinoviev, A. V. Lunacharsky, Alexandra Kollontai and other less important bolshevik leaders. After his return from Siberia in March, he became editor of *Pravda* and adopted a vaguely conciliatory attitude toward the provisional government of Prince Lvov. Soon afterward the bolsheviks, prompted by Lenin, embarked upon a revolutionary course of action; but Stalin remained one of their chief organizers and *Pravda's* coeditor. At the height of the antibolshevik campaign in July he organized Lenin's escape from Petrograd; and in the absence of Lenin, Trotsky, L. B. Kamenev and Zinoviev, he guided the 6th party congress immediately after the "July days." During the famous controversy which developed in September and October between Lenin, advocating an armed rising, and Zinoviev and Kamenev, opposed to insurrectionary tactics, Stalin was among Lenin's adherents, although in *Pravda* he attempted to bridge the gulf between the two factions. He was elected member of a bolshevik insurrectionist "centre" which was to join the Military Revolutionary committee headed by Trotsky; but he played no part at all in the insurrection itself.

The People's Commissar.—After the bolshevik victory he was appointed commissar of nationalities, and he held this post for nearly five years. Under Lenin's guidance he set out to put into effect the party's program of "self-determination" for oppressed nationalities. One of his first official acts was to proclaim on behalf of the new Russian government the independence of Finland. He participated closely in the working out of the first Soviet constitutions of 1918 and 1923. In 1919 he was, in addition, appointed commissar of the workers' and peasants' inspectorate, an important department whose function it was to supervise all other branches of the Soviet administration.

His voice carried much weight in the inner bolshevik councils during the civil war (1917–21), although he continued to be overshadowed by Trotsky, the founder and inspirer of the Red army. As member of the council of defense, political commissar and inspector of fronts, Stalin was in charge of the defense of Tsaritsyn (Stalingrad) in the summer and autumn of 1918; he organized the defense of Petrograd against Gen. N. N. Yudenich's offensive in May 1919, and the battle of Orel in October of the same year, when the Red army halted Gen. A. I. Denikin's advance on Moscow and passed to the counteroffensive. He served as political commissar with S. M. Budenny's cavalry corps during the advance on Lwow (Lemberg) in 1920; and he was in charge of the preparations for the final campaign against Gen. P. N. Wrangel's army and of the dramatic battle of the Perekop isthmus, the last battle of the civil war. Some of these episodes, especially his role in the defense of Tsaritsyn and in the march to Poland, were the subjects of much controversy. During the civil war an intense rivalry developed between him and Trotsky, the commissar of war. Not without ground Trotsky charged Stalin with surreptitiously supporting the left-wing Communists who were opposed to the employment of old army officers in the Red army and to the army's centralized structure and discipline established by Trotsky. Trotsky's influence remained paramount, but even in 1919 Stalin wielded enough power to bring about the dismissal of J. Vatsetis, Trotsky's protégé, from the post of commander in chief and his replacement by S. Kamenev. After the civil war, in Feb. 1921, Stalin initiated the bolshevik invasion of Georgia without the knowledge of Trotsky and the commander in chief.

In April 1922, after the 11th party congress, the last congress attended by Lenin, Stalin was appointed secretary general of the party and relinquished his posts in the council of people's commissars. This was a turning point in his career: as secretary general he soon came to control the whole machinery of the party and through it the government.

The Succession to Lenin.—Lenin's illness in 1923 and his death (Jan. 1924), coincided with a crisis of the revolutionary regime. The bolsheviks had from the beginning denied political freedom to the old ruling classes and their parties who opposed arms in hand the "proletarian dictatorship." But the bolsheviks had also pledged themselves to respect and secure "Soviet democracy," that is, freedom of political expression for the working classes and their parties. The proletarian dictatorship had been conceived as a system of representative government, but instead of a parliament, the councils of workers' and peasants' deputies were to form the representative organs. Toward the end of the Lenin era, however, the regime had already developed into a single-party system. All nonbolshevik parties had been banned, and only within bolshevik ranks was some freedom of expression preserved. Inevitably the question arose whether the bolsheviks themselves could enjoy the freedom they denied to others. The other critical development concerned the party's international orientation. The hope for an imminent spread of revolution abroad, which had inspired the bolsheviks, was decisively undermined in 1923 after the defeat of communism in Germany. Bolshevik Russia was at last recognized by most bourgeois governments, but as a Communist state it stood alone. Bolshevism was withdrawing into its national shell and was ill at ease with the revolutionary internationalism of the Lenin era. All these issues were brought to a head in the struggle over the succession to Lenin.

In 1923 Stalin, Zinoviev and Kamenev formed a triumvirate with the purpose of debarring from power Trotsky, who was generally regarded as Lenin's successor. In his last will Lenin advised his followers to "remove Stalin" from the general secretariat of the party on the ground that Stalin was rude and inclined to abuse power. But Stalin, supported by the other triumvirs, retained his office. In a constant struggle against his opponents he abolished whatever freedom of expression still existed within the party and transformed the party into a "monolithic" body. In the autumn of 1924 he expounded the theory of "socialism in one country," proclaiming the self-sufficiency of the Russian Revolution. This doctrinal innovation marked a departure from Leninist internationalism, although Stalin did his utmost to present his views as orthodox Leninism.

In April 1925, having defeated Trotsky, Stalin openly broke with the other triumvirs and, at the 14th party conference, joined hands with the party's right-wing led by N. I. Bukharin, A. I. Rykov and M. P. Tomsky. Zinoviev and Kamenev, on the other hand, made common cause with Trotsky. They denounced the theory of "socialism in one country," advocated a policy of industrialization and a struggle against the kulaks (well-to-do farmers), and demanded the restoration of "inner party democracy." At this stage Stalin, anxious to keep in line with Bukharin's moderate group, pursued a very cautious policy: he refused to undertake a drive against the kulaks and treated the demand for rapid industrialization as unreal. But in 1928–29, after he had defeated and expelled from the party Trotsky, Zinoviev, Kamenev and their followers, Stalin turned against Bukharin, Rykov and Tomsky. He then effected a most drastic change of policy and began to industrialize the U.S.S.R. and to collectivize agriculture with a speed and ruthlessness which horrified even the original advocates of these policies. In 1929 he expelled Trotsky from the U.S.S.R. and forced

all opposition leaders, right and left, to surrender and "recant." He had become undisputed master of the U.S.S.R.

Industrialization and Collectivization.—"We are 50 or 100 years behind the advanced countries. We must make good this lag in 10 years. Either we do it or they crush us." Thus Stalin explained the purpose of industrialization to a conference of industrial managers on Feb. 4, 1931. In the course of the decade the U.S.S.R. was indeed transformed from one of the most backward states into a great industrial power; this was one of the factors which was to assure Soviet victory in World War II. Stalin emphatically insisted on the need to develop primarily the heavy industries. Industrialization was consequently accompanied by an extreme scarcity of consumer goods and by widespread popular discontent. Stalin violently suppressed the discontent. With utmost determination he enforced a policy of incentive rewards to skilled workers, technicians and administrators, which provoked much indignation among party members brought up in an egalitarian spirit. They denounced Stalin as the leader of new privileged social groups.

Industrialization on the gigantic scale envisaged in the successive five-year plans encountered many other difficulties which would have baffled and defeated a government less ruthless than Stalin's. In the 1930s alone about 25,000,000 muzhiks were forcibly shifted from rural areas to the industrial centres and transformed into factory workers by means of intensive training and with the help of an extremely harsh industrial discipline. Throughout the Stalin era the U.S.S.R.'s urban population grew by about 45,000,000. In order to free labour for industry and to secure food for the swelling urban population, Stalin collectivized farming. In 1929 there were about 25,000,000 primitive rural smallholdings in the U.S.S.R. In 1952 there were 100,000 large and highly mechanized collective farms. The peasantry at first bitterly resisted collectivization. Stalin broke its resistance and ordered the deportation of many "kulaks" into the notorious labour camps. Despite initial failures and famines, the new system of farming achieved a high degree of consolidation in later years.

The Great Purges.—At the height of the industrial drive, shortly after he had introduced the quasi-liberal constitution of 1936, Stalin staged the great purge trials in which most of the old bolsheviks and some military leaders were charged with treason, terrorism and espionage, and brought to "confess" guilt. In this way Stalin exterminated the men who might have been able to overthrow him and form an alternative government during a national crisis. The purges, carried out on a mass scale, imparted to the Stalinist regime its peculiar terroristic character.

Foreign Policy Prior to 1941.—The national crisis which Stalin feared was approaching with World War II. Since Hitler's rise to power Stalin grappled with the dilemma between an anti-Hitler coalition with the west and agreement with Hitler. At first his diplomacy aimed at building up the anti-Hitler coalition, but finding the west reluctant to commit itself to an alliance with the U.S.S.R., Stalin hinted at his readiness to come to terms with Hitler in a speech at the 18th party congress in March 1939. In August he struck the bargain with Hitler under which the U.S.S.R. and Germany divided their spheres of influence in eastern Europe and the U.S.S.R. avoided immediate involvement in war.

The twists and turns of Stalin's foreign policy affected closely the attitude of the Communist International. After a long series of purges and expulsions Stalin succeeded in transforming the International into a "monolithic" organization, identifying the interests of world Communism with Stalin's dogmas and policies. Thus the Communist International supported Chiang Kai-shek during the Chinese revolution of 1925–27 when Russian diplomacy regarded the Chinese generalissimo as its ally. In 1935–39 the Communist parties organized the "popular fronts" to help the U.S.S.R. in its search for bourgeois democratic allies against Hitler. In 1939 they lent moral support to the Ribbentrop-Molotov pact.

Generalissimo.—The German attack on the U.S.S.R., started on June 22, 1941, caught Stalin unawares, as he himself later publicly admitted. In preparation for critical times he had assumed the premiership on May 6; but he still hoped to be able to postpone the inevitable conflict between the U.S.S.R. and Germany. After the German attack he formed the state defense committee consisting of himself, V. M. Molotov, G. M. Malenkov, L. P. Beria and K. E. Voroshilov. On July 3, 1941, he made the famous call on the Soviet people to "scorch the earth" occupied by the invader. "In occupied regions conditions must be made unbearable for the enemy and all his accomplices. They must be hounded and annihilated at every step, and all their measures frustrated."

Throughout World War II Stalin was personally responsible for the major military and diplomatic decisions of his government, although he left considerable freedom of initiative to his generals. After the initial period of confusion, his strategy, following the traditional Russian pattern of 1812, was to withdraw, to conserve strength, to gather new reserves and to force the attacker to spend his impetus and extend his lines—only after that could the U.S.S.R. pass to the offensive. He displayed extraordinary will power, tenacity and coolheadedness. He stayed on in the Kremlin when Hitler's armies stood at the gates of Moscow and the Soviet government was evacuated to Kuibyshev. From the Kremlin he organized a hasty yet effective evacuation from the threatened areas of industrial plants and of many millions of industrial workers. The shifting of industry from west to east enabled the U.S.S.R. to recover from the first blows and to build up a new and enormous armed power.

Stalin's political conduct of the war was characterized by strong emphasis on Russian patriotism, nationalism and traditionalism. Evoking the memories of 1812, he described the struggle as the Great Fatherland War, and he extolled the national heroes of Mother Russia and the military traditions of Peter the Great, Suvorov and Kutuzov. He rehabilitated the Orthodox Church and disbanded the Communist International. These moves, dictated primarily by considerations of domestic policy, were calculated to meet the needs of his foreign policy as well. He was anxious to preserve the alliance with the western powers, and his conduct of the war in a nationalist rather than in a revolutionary spirit helped to keep the grand alliance together, despite inner contradictions and tensions.

Some of the tensions were due to the fact that the western Allies repeatedly postponed the opening of the "second front" in western Europe. This was the main point in Stalin's controversies with Winston Churchill in 1942–43. But as the Soviet army broke Germany's military power in the "ten offensives" which followed the battle of Stalingrad in 1943, this issue lost importance. The course of the war revealed a profound shift in the international balance of power, a shift due primarily to Soviet industrialization. Stalin bargained adroitly during the Big Three conferences at Tehran and Yalta, and his diplomacy secured for the U.S.S.R. the predominance in eastern and central Europe which was in any case bound to result from the success of Soviet arms. The Tehran and Yalta agreements contained, however, the seeds of the discord which was to lead to the breakdown of the grand alliance after the war. Stalin was also the presiding spirit at the Potsdam conference of July–August 1945, at which he committed the U.S.S.R. to join the United States and Great Britain in the war against Japan, and at which he worked out together with the western Allies a rather impracticable armistice regime for occupied Germany.

Stalin's Last Years.—The year 1945 marked the climax in Stalin's crowded and chequered career. As generalissimo and uncontested leader he was surrounded by an adulation and a cult which assumed its most grotesque forms toward the end of his life. Soviet scientists, writers, musicians, linguists, philosophers and others were made to accept his judgment as final. Underlying the bizarre cult were Stalin's indubitable achievements. He was the originator of planned economy; he found Russia working with wooden plows and left it equipped with atomic piles; and he was "father of victory." But his achievement was marred by the despotism and cruelty of his dictatorship; and the patriarchal character of his rule, suited perhaps to the mentality of an illiterate and backward people, became an anachronism in the industrialized and modernized Russia of his own making.

Stalin's personal role in the conduct of Soviet foreign policy from 1945 to 1953 remained somewhat obscure. Communists and anti-Communists alike had seen in him the architect of the revolutions in eastern Europe and China. In the light of even the scanty documentary information available, this view appears to be illfounded. From the beginning Stalin was opposed to Tito's revolutionary policies in Yugoslavia; and he viewed without sympathy Mao Tse-tung's revolutionary strategy and tactics in China. On the other hand, he undoubtedly inspired some of the Communist revolutions in eastern Europe; and he expected all Communist revolutions, those which he had supported and those which he had obstructed, to adopt the Stalin cult. During his last years a struggle apparently developed in his entourage between the die-hards of Stalinism and those who stood for domestic reform, and also between adherents of a "tough" foreign policy and advocates of conciliation. Stalin tried to keep a balance between the conflicting attitudes and to shape Soviet policy as the resultant of the conflicting trends.

He died from cerebral haemorrhage in the Kremlin on March 5, 1953, at the age of 73—one of the most complex, powerful and controversial figures in world history. (I. D.)

BIBLIOGRAPHY.—J. V. Stalin, *Sochinenya* (Works), in progress (Moscow, 1946), *Marxism and the National and Colonial Question* (London, 1940), *Problems of Leninism* (Moscow, 1945), *History of the Communist Party of the Soviet Union (Bolsheviks)* (Moscow, 1948), *Concerning Marxism in Linguistics* (Moscow, 1950), *Economic Problems of Socialism in the U.S.S.R.* (Moscow, 1952); B. Souvarine, *Stalin: A Critical Survey of Bolshevism* (London, 1939); L. Trotsky, *Stalin: An Appraisal of the Man and His Influence* (London, 1947); I. Deutscher, *Stalin: A Political Biography* (London, 1949).

STALINABAD: *see* DUSHANBE.

STALINGRAD: *see* VOLGOGRAD.

STALINGRAD, BATTLE OF, the name given to the climax, between Aug. 1942 and Feb. 1943, of the fighting between the German and the Soviet armies in World War II, ending in the defeat of the German advance against Stalingrad (the former Tsaritsyn, later renamed Volgograd).

From June to Dec. 1941 the German armies had penetrated more than 500 mi. into Russia, reaching the outskirts of Leningrad and Moscow in the north and centre, while in the south they got well to the east of the river Dnieper. But Moscow, which covered the assembly of the Soviet reserves and the rail communications to the northwest and southeast, remained in Soviet hands and was still a factor of great military importance.

The German plan for 1942, however, ignored Moscow in favour of the economic and political advantages to be gained from a drive far to the southeast to capture the oil fields of the Caucasus and reach the Caspian sea. To cover this ambitious enterprise the Germans had to secure a line from the upper Don river through Stalingrad to the lower Volga in order to hold up possible Soviet counteraction from the north. Stalingrad was a large industrial city at the apex of this line, and its speedy capture was essential for the success of the whole German campaign.

Early in July 1942 the army group A, of about 30 divisions under the direct control of the Wehrmacht supreme headquarters (O.K.W.), duly set off from Rostov and the Crimea along the Black sea coast toward the oil fields. To the north of the army group A, the army group B, of about 60 divisions, attacked east from Kursk and Kharkov to secure the Don from Voronezh down to Rostov, so that the 6th army (Gen. Friedrich von Paulus) could be launched across the river for the capture of Stalingrad.

The 6th army, composed at first entirely of German troops, was of superlative quality. It consisted of some 20 divisions (250,000 men) of which three were armoured. Nominally under the operational control of army group B, actually it was directly under Hitler by wireless, and little initiative was left to General Paulus. The army crossed the Don on Aug. 22–23, 1942, about Vertyachi and Peskovatka. German panzers got to Erzovka, a suburb on the Volga north of Stalingrad, almost at once, but heavy fighting about Kalach with the Soviet 1st (Guards) army (Maj. Gen. K. S. Moskalenko) and 62nd army (Lieut. Gen. V. I. Chuikov) delayed the German advance in the centre. By the end of September, however, the Germans had a footing in Stalingrad, where N. S. Khrushchev, the political commissar, distinguished himself in its energetic resistance. Signs then became apparent of an impending Soviet counterattack southward by the Don army group (Lieut. Gen. K. K. Rokossovski) from its bridgehead at Kremenskaya, just west of the northern boundary of the 6th army. On Nov. 19 the blow fell and was made heavier by an unexpected thrust from the south by the Stalingrad army group (Col. Gen. A. I. Eremenko). On Nov. 23 the prongs of these two attacks met near Kalach, and the 6th army became surrounded.

The best course of action now would have been for the 6th army to abandon the Stalingrad enterprise and to break out southwest to the Don. Hitler at once forbade it. The Stalingrad pocket was to become a "fortress" and was to be supplied by air. Neither idea bore any relation to what was possible.

At this juncture Gen. Erich von Manstein, one of the ablest commanders in World War II, took over a new German army group "Don" with headquarters near Rostov. The relief of the 6th army was his chief, yet not his only object, for army group B on the upper Don was crumbling behind him and endangering the communications of army group A, still stretched out to the southeast. Nevertheless Manstein assembled a panzer corps at Kotelnikovo to attempt the relief of the beleaguered army. The corps drove north on Dec. 12, but by Dec. 19 was halted by units of the Soviet 2nd (Guards) army (Lieut. Gen. R. Ya. Malinovski) and 51st army (Maj. Gen. N. I. Trufanov), when it was still 20 mi. short of the Stalingrad pocket. By this time, moreover, the 6th army was in no state to attempt a break-out. Its troops were frozen and half starved and its vehicles had fuel only for 20 mi. Manstein had now done his utmost. He also wanted the panzer corps urgently for his own battle on the Don, so back it had to go. The Stalingrad "fortress" hung on for six more terrible weeks, until finally, on Feb. 2, 1943, Paulus, now a field marshal, surrendered with his whole army. During the battle the 6th army lost about 150,000; and of the 91,000 Germans who became captives only a small percentage saw Germany again.

See Ronald Seth, *Stalingrad: Point of Return* (1959). (B. T. W.)

STALINO: *see* DONETSK.

STALINOGORSK: *see* NOVOMOSKOVSK.

STALINOGROD: *see* KATOWICE.

STALINSK: *see* NOVOKUZNETSK.

STALL, in church architecture, is a fixed seat enclosed at the back and separated from adjacent stalls at the sides by high projecting arms. Stalls are placed in one or more rows on the north and south sides of the chancel in cathedrals, monastic churches and large parish churches. In an English cathedral the canons and prebendaries each have a stall assigned to them. In the chapels of the various knightly orders the stalls are assigned to the members of the order, thus, in St. George's chapel, Windsor, are the stalls of the knights of the Garter, in Henry VII's chapel in Westminster abbey are those of the knights of the Bath, adorned with the stall plates emblazoned with the arms of the knight occupying the stall, above which is suspended his banner. Architecturally considered, the stalls of a cathedral or church are a marked feature of the interior adornment. They are richly carved, and are frequently surmounted by canopies of tabernacle work. The seats generally can be folded back so as to allow the occupant to stand upright or kneel; beneath the seat, especially in monastic churches, is fixed a small bracket, a *miserere,* which affords a slight rest for the person while standing. Among specimens of carved stalls may be mentioned the Early Decorated stalls in Winchester cathedral (1296); the Early Perpendicular ones in Lincoln minster (*c.* 1370); the early 15th-century canopies in Norwich cathedral (15th century); and the towering corner stalls with their ornate carving filled with figures in Amiens cathedral (1508–20).

STALLBAUM, JOHANN GOTTFRIED (1793–1861), German classical scholar, was born at Zaasch, near Delitzsch in Saxony on Sept. 25, 1793, and died at Leipzig, where he was rector of the Thomasschule, on Jan. 24, 1861. His reputation rests upon his work on Plato, of which he published two complete editions.

STALYBRIDGE, a municipal borough in the Stalybridge and Hyde parliamentary division of Cheshire, Eng., 7½ mi. E. of Manchester. Pop. (1961) 21,940. Area 5 sq.mi. The whole district is densely populated. Stalybridge is one of the oldest cotton towns, the first cotton mill having been erected in 1776. It was incorporated in 1857 and its boundaries were extended in 1881 and 1936. Cotton and engineering are the principal industries. On the Derbyshire side there is a residential area.

STAMBOLISKY, ALEKSANDR (1879–1923), Bulgarian statesman, was born at Slavovitsa, Bulgaria, Sept. 1, 1879. He attended the agricultural college of Halle in Germany. On returning to Bulgaria, aged 18, he took up journalism, and in 1902 became editor of the organ of the newly-formed Agrarian league. Six years later he was elected to the Sobranye, where he led the Agrarian party. He now began to organize the peasant masses, who form 80% of the population of Bulgaria, into agricultural associations. In 1911, when the Grand Sobranye met at Trnovo to amend the constitution, his first conflict with King Ferdinand occurred. After the treaty of Bucharest (1913), Stambolisky and the Agrarians were unmerciful in their criticisms of Ferdinand's policy, though deterred from extreme measures by fear of external complications. His opposition to Ferdinand came to a head in 1915 during the negotiations that preceded Bulgaria's entry into World War I. Stambolisky strongly backed the Entente in spite of the king's anger, pressure from the government and the growing influence of the Central Powers. Summoned before the king, Stambolisky threatened him with personal violence if he should fight against the Allies, reminding him that he had a crown to lose. Ferdinand ordered his arrest. He was tried by court martial and condemned to penal servitude for life.

In Sept. 1918, when the resistance of the Bulgarian troops

began to slacken, Ferdinand released Stambolisky, who left for the Macedonian front. He returned at the head of the insurrectionary troops and their arrival at Sofia resulted in Ferdinand's abdication and flight. A partisan of the new ruler, King Boris, Stambolisky became a member of the cabinet in Jan. 1919 and prime minister in Oct.; as head of the government he went to Paris and signed the treaty of Neuilly in the same year. In Feb. 1920 he dissolved the Sobranye, and the following month was returned to power at the head of a homogeneous Agrarian majority. From 1920 to 1923, the Agrarians under Stambolisky ruled Bulgaria with a rod of iron. They had the Radoslavov cabinet, which had brought Bulgaria into World War I, condemned to death by popular vote. Stambolisky also carried measures based on the interest of the peasants (*see* BULGARIA), and declared: "Sofia, that Sodom, that Gomorrah, may disappear. . . . I shall not weep for her." Through his influence postwar Bulgaria loyally carried out the terms of the peace treaty, won the esteem of the Allies, and secured a reduction of reparations. Stambolisky also made persistent efforts to improve Bulgaria's relations with Yugoslavia, with which country he concluded an agreement at Niš (1922). His ultimate ideal was probably some kind of federal south Slav state, embracing Bulgaria.

The overthrow of the government by a military faction on June 9, 1923, was caused partly by dissatisfaction with Stambolisky's domestic policy and partly by the influence of the Macedonian party. The change of regime took place in one night, the Agrarian ministers being arrested in their homes. Stambolisky, pursued by an armed detachment to his native village of Slavovitsa, escaped and fled across the mountains, hoping to reach the frontier. After being followed by soldiers for three days, he was eventually surrounded and shot on June 12. A formidable personality, attractive in many ways, and made of finer clay than the average Balkan statesman, Stambolisky has been called a modern Rienzi. Tall and broad, his demeanour was fierce and his movements ungainly; his rough-hewn face was crowned by a mass of black hair; he impressed everyone with his strength and the sincerity of his advocacy of a "Green International," to unite the peasants of all countries. Stambolisky published *Authority, Anarchy and Democracy* and *What a Politician Ought To Be.*

STAMBULOV, STEFAN (1854–1895), Bulgarian statesman, was born on Jan. 31, 1854, at Trnovo, the ancient Bulgarian capital, where his father kept a small inn. He entered the seminary of Odessa in order to prepare for the priesthood but was expelled for his association with Nihilists, and, making his way to Rumania, he entered into close relations with the Bulgarian revolutionary committees at Bucharest and Giurgiu. In 1875, though only 20 years of age, he led an insurrectionary movement at Nova Zagora in Bulgaria, and in the following year organized another rising at Orekhovitsa. In the autumn of 1876 he took part as a volunteer in the Serbian campaign against Turkey, and subsequently joined the Bulgarian irregular contingent with the Russian army in the war of 1877–78. After the signature of the Berlin treaty in 1878 Stambulov settled at Trnovo as a lawyer, and was soon elected to the Sobranye, of which he became president in 1884. The revolution of Plovdiv, which brought about the union of Bulgaria with eastern Rumelia, took place on Sept. 18, 1885, and it was largely owing to Stambulov's urging that Prince Alexander decided to identify himself with the movement. War with Serbia followed, and Stambulov served as an ordinary soldier in the Bulgarian army.

After the abduction of Prince Alexander (Aug. 21, 1886), Stambulov, who was then at Trnovo, established a loyal government there, issued a manifesto to the nation, nominated his brother-in-law, Gen. Sava Mutkurov, commander in chief of the army, and invited the prince to return to Bulgaria. These measures resulted in the downfall of the provisional government set up by the Russophil party at Sofia. On the abdication of Prince Alexander (Sept. 8) Stambulov became head of a council of regency, with Mutkurov and Petko Karavelov as his colleagues; the latter, however, soon made way for Zhivkov, a friend and fellow townsman of the first regent. Stambulov frustrated the mission of Gen. Nikolay Kaulbars, whom the tsar dispatched as special commissioner to Bulgaria, held elections for the Sobranye, despite the interdict of Russia, and eventually secured the election of Prince Ferdinand of Coburg to the vacant throne (July 7, 1887). Under the newly-elected ruler he became prime minister and minister of the interior, and continued in office for nearly seven years. The aim of his foreign policy was to obtain the recognition of Prince Ferdinand, and to win the support of the Triple Alliance and Great Britain against Russian interference in Bulgaria. He gained the confidence of the Sublime Porte, and obtained concessions for the Bulgarian episcopate in Macedonia (*see* MACEDONIAN QUESTION). With the assistance of Austria-Hungary and Great Britain he negotiated large foreign loans which enabled him to develop the military strength of Bulgaria.

Stambulov's rule was despotic. He ruthlessly stamped out recurrent plots and armed risings engineered by equally ruthless political opponents with close support from influential Panslavist quarters in Russia. There was no clear distinction between party strife and sedition; Stambulov used his police, and sometimes the army, to deal with either. Maj. Konstantin Panitsa, after a conspiracy, was shot at Sofia in 1890; four political opponents were hanged in the following year, and Karavelov was sentenced to five years' imprisonment. Among the deadliest enemies of the regime were the Macedonian committees, whose effective part in the struggle was emphasized by the assassination of Stambulov's colleague Belchev, in 1891, and of the Bulgarian representative in Constantinople, Vlkovich, in 1892.

The arrogant way in which Stambulov used his practically unlimited power, and the growing violence of the opposition throughout the country, became intolerable to Prince Ferdinand who, moreover, wanted an accommodation with Russia. He contrived to get his unmanageable premier to resign, and accepted his resignation in May 1894. Stambulov was now exposed to the vengeance of his enemies. On July 15, 1895, he was attacked and barbarously mutilated by a band of Macedonian assassins in the streets of Sofia, and succumbed to his injuries three days later. His funeral was interrupted by disgraceful riots, and an effort was made to perpetrate an outrage on his remains. No attempt was made to arrest his murderers; two persons, however, arraigned for the crime in 1896, were subjected to almost nominal penalties.

(J. D. B.; N. I. M.)

STAMFORD, a municipal borough in the parts of Kesteven, Lincolnshire, Eng., on the River Welland, the boundary between Lincolnshire and Northamptonshire. Pop. (1961) 11,743. It lies 11 mi. N.W. of Peterborough in the Rutland and Stamford parliamentary division. Stamford is a very ancient market town, built of local gray stone and once famous for its cloth. Its position on the Great North road and on the borders of the Fens contributed to its importance. Its modern industries include general and agricultural engineering, brick and tilemaking and the manufacture of plastics and toys. Of the old churches six remain: St. Mary's with a Decorated spire, All Saints with fine brasses, and St. Martin's which contains the Burghley tombs, are the most noteworthy. There are remains of the 7th-century Benedictine priory of St. Leonard and a west gate of the Carmelite monastery (1291). Browne's hospital is one of the almshouses known as Calises, from the wool merchants of Calais. There was a Danish fort at Stamford, which later became one of the five Danish boroughs. A charter of Edgar dated A.D. 972 mentions a market and a mint. In the reign of Edward the Confessor Stamford was a royal borough governed by 12 lawmen. The Norman castle was built before 1086. Henry III gave the town a charter in 1256 and it was incorporated in 1461–62.

Stamford was known for its wool trade, its churches and its monastic schools, and in 1333 was chosen as the headquarters of the students who seceded from Oxford. An Early Decorated doorway (1300) remains of Brasenose hall. The attempt to establish a regular university was prohibited by royal authority. The history of the place centred around the family of Cecil when the great Lord Burghley built Burghley house (1577–85), a fine quadrangular mansion south of the town, now the seat of the marquess of Exeter. The house contains notable art collections.

STAMFORD, a city of Connecticut, U.S., is located 35 mi.

N.E. of New York city, on Long Island sound. Founded in 1641 by 29 secessionists from the Wethersfield church, the town was under the New Haven colony which later merged with the Hartford colony into what became the state of Connecticut. In 1644 one-third of the population of Stamford emigrated to Long Island but the town continued to grow as an agricultural, lumber, leather and shipbuilding community. By the time of the American Revolution it had approximately 5,000 residents. A city government was formed within the town in 1893 and the town and city merged in 1949. The population quadrupled in the first half of the 20th century, reaching 74,293 in the 1950 census. The population in 1960 was 92,713 for the city and 178,409 for the standard metropolitan statistical area, comprising Stamford city and Darien, Greenwich and New Canaan towns of Fairfield county. (For comparative population figures *see* table in CONNECTICUT: *Population.*)

The countryside still retains much of a rural New England atmosphere north of the Merritt parkway, which bisects the city, but rapid population growth marks its emergence as a suburban industrial, commercial and commuter centre. It is the site of major chemical research laboratories; manufactures include chemical products, postage meters, ball bearings and die castings. In addition to arterial highway and major rail facilities, the city is connected to New York by regular helicopter service and is near the Westchester county, N.Y., airport. It is served by public and private secondary schools, a branch campus of the University of Connecticut, a main library and branches and a museum and nature centre. Recreational facilities include municipal beaches, a municipal golf course, private and public marinas, about 600 ac. of park land and numerous private clubs. (W. L. T.)

STAMITZ, JOHANN WENZL ANTON (1717–1757), the eldest and most important of a family of Bohemian musicians famous in the 18th century. He was born at Deutschbrod on June 19, 1717. He was a solo violinist at the coronation of the emperor Charles VII (1742), and in 1745 the elector took him to Mannheim as director of the chamber of music. Stamitz and his successors are known as the Mannheim school. Although Phillip Emanuel Bach in Berlin and the young school in Vienna were arriving at similar results, it seems certain that Stamitz was the real pioneer who not only demonstrated the dramatic value of the "working-out" section of the first movement of a sonata or symphony but, by incorporating the minuet, definitely established the four-movement form. In addition he demanded of the orchestra a greater independence from the different instruments, singling out the much neglected viola. He died at Mannheim on March 30, 1757. His works include 27 concertos, sonatas and solos for violin, several sets of symphonies and harpsichord music.

Much confusion has arisen between Johann and his eldest son CARL STAMITZ (1746–1801), who sometimes used the name Johann. He produced an opera, *Der verliebte Vormund,* at Frankfurt and another, *Dardanus,* in St. Petersburg, and wrote symphonies.

STAMMERING: *see* SPEECH DISORDERS.

STAMMLER, RUDOLF (1856–1938), German jurist and the most influential legal philosopher of the early 20th century, especially in Germany, Spain and Latin America, was born on Feb. 19, 1856, in Alsfeld (Hessen). He was professor of law in Marburg (1882–84), Giessen (1884), Halle (1885–1916) and Berlin (1916–23). He broke the dominance of legal positivism and historicism and directed legal philosophy toward neo-Kantian ideas. He died April 25, 1938, at Wernigerode. (N. S. TF.)

STAMP, JOSIAH CHARLES STAMP, 1ST BARON (1880–1941), distinguished British economist and statistician, was born in London on June 21, 1880. He entered the inland revenue department in 1896 and became assistant secretary to the board in 1916. Meanwhile he had taken the external degree in economics of London university. In 1919 he left the civil service to become director and secretary of Nobel Industries Ltd., an explosives firm, and in 1926 he became president of the executive of the London, Midland and Scottish railway.

In 1928 he became a director of the Bank of England. Stamp sat on the royal commission on income tax (1919); on the committee on taxation and national debt (1924–27); on the Dawes committee on German currency and finance (1924); on the court of inquiry into the coal industry (1925); and on the Paris committee of inquiry into the reparations question of 1929. Among many honours he was knighted in 1920, awarded the Order of Merit in 1936 and made a baron in 1938. In 1939 he became adviser to the government on economic co-ordination. Stamp was the author of many books, his works on taxation, public finance and the national income being especially valuable. He and Lady Stamp were killed in an air raid at Shortlands, Kent, on April 16, 1941.

See J. H. Clapham, *Lord Stamp, 1880–1941* (1942). (T. W. H.)

STAMP: *see* POSTAL SERVICES; PHILATELY.

STAMP ACT, THE, one of the important causes of the American Revolution, was an act, introduced by George Grenville and passed by the British parliament in 1765, with the object of raising revenue in the American colonies. The act set stamp duties on every "skin or piece of vellum or parchment, or sheet or piece of paper" used for legal documents, academic degrees, liquor licences, various commercial instruments, etc. Pamphlets, newspapers, almanacs, etc., were also taxed, and duties were placed on all advertisements. Determined opposition, at once shown in all the colonies, was so great that the proceeds from the duty were less than the expenses of collection. The next year parliament repealed the act as a matter of expediency but passed another act declaring the right of parliament to bind the colonies by its enactments. *See* UNITED STATES: *History; see* also Index references under "Stamp Act, The" in the Index volume.

STAMP DUTY, a tax applied to written documents, particularly those arising from various financial transactions, collected by the use of stamps which must be imprinted on or attached to the documents. For many years Great Britain has applied flat rate stamp taxes to receipts and bank checks, and graduated ones to documents involved in the transfer of property, leases, mortgages, insurance policies and others. Failure to apply the stamps invalidates the document. Similar taxes are used in many other countries, particularly in southern Europe. Stamp duties in the United States are applied to new stock and bond issues of corporations, sales of securities and real estate and certain other transfers. Distinct from the stamp duty or tax, as such, is the stamp technique for collection of excise taxes upon commodities, whereby a stamp must be affixed to each article as evidence of payment of tax. This technique was used in the United States for collection of liquor and tobacco taxes prior to 1959. *See* also STAMP ACT, THE. (J. F. D.)

STANDARD. For the various meanings of this word, *see* DIMENSIONAL ANALYSIS; PHYSICAL UNITS; STANDARDIZATION; STANDARDS, NATIONAL BUREAU OF; TIME, STANDARD; WEIGHTS AND MEASURES.

STANDARD, BATTLE OF THE, a name given to the battle of Aug. 22, 1138, near Northallerton, in which the Scottish army, composed largely of wild and undisciplined Highlanders and Galloway Picts under King David, was defeated by the English levies of Yorkshire and the north Midlands. The latter formed a single deep line of dismounted men in front of a chariot carrying the consecrated banners of St. Peter of York, St. John of Beverley, St. Wilfred of Ripon and St. Cuthbert of Durham. King David planned to use his nucleus of knights and archers to lead the way and open a gap in the English line, but the fierce pride of the clansmen compelled him to give them precedence. Their headlong onset was repulsed. Left isolated except for his bodyguard of knights, King David retired from the field.

STANDARD DEVIATION. In statistics the term standard deviation is an index used to measure the dispersion of a sample of measurements about the arithmetic mean of these measurements. More precisely, suppose $x_1, x_2, \ldots, x_n$ are the n measurements in a sample and let

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

be their arithmetic mean. The standard deviation s of the sample is defined as

$$s = \frac{1}{n-1}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\right]$$

The quantity s^2 is called the variance of the measurements. The

divisor $n-1$ is used rather than n for the following reason: Suppose we have a very large population of measurements and we denote its variance by σ^2. Suppose a sample of n measurements is drawn from this population and consider all possible samples that could have been drawn. Each of these samples would have an s^2. The arithmetic average of all of these s^2's is exactly σ^2, a very simple result of basic importance in statistical inference that would not hold if we had divided by n in defining s.

Terms which have been used earlier to denote standard deviation include mean-square error and root-mean-square deviation. The term mean error was used by Gauss in the early 19th century. In these earlier definitions n rather than $n-1$ is used in defining s. The term standard deviation is now almost universally used rather than any of the earlier designations.

The term standard deviation is also similarly used as the name of a quantity to describe the dispersion or spread of a probability distribution. More precisely, suppose a discrete random variable x takes on values $x_1, x_2, \ldots, x_k$ with probabilities $p(x_1)$, $p(x_2)$, $\ldots, p(x_k)$. Let $\mu = x_1\,p(x_1) + x_2\,p(x_2) + \ldots + x_k\,p(x_k)$. The variance σ^2 of the random variable x is defined as

$$\sigma^2 = (x_1 - \mu)^2 p(x_1) + (x_2 - \mu)^2 p(x_2) + \ldots + (x_k - \mu)^2 p(x_k)$$

The standard deviation σ of the random variable x is simply the square root of the quantity on the right.

In the case of a continuous random variable x having probability density function $f(x)$ on the interval (a, b) we similarly have

$$\mu = \int_a^b xf(x)dx, \qquad \sigma^2 = \int_a^b (x - \mu)^2 f(x)dx$$

(S. S. W.)

STANDARDIZATION, the process of developing and bringing about the utilization of standards. While standardization of some kind or degree is part of almost every human activity, planned and systematic industrial standardization is meant in normal usage of the term. In the broadest sense a standard is anything used to measure; *e.g.*, a standard of conduct, a standard of value or a standard of weight or length. This discussion, however, deals mainly with the nature and importance of standardization in industry rather than the evolution of standards in social customs, language, games, music or law.

Classifications of Standards.—Since industrial standards may cover anything from product attributes to safety practices it is well to start by giving a definition for the word "standard" that is applicable in every context: A standard is that which has been selected as a model to which objects or actions may be compared. In every case a standard provides a criterion for judgment and its form thus depends on what is to be judged and how it is to be judged. Standards may be devices or instruments such as precision-gauge blocks, light meters or trip scales for making comparisons of size, weight, colour or other attributes. They may be physical models to which things are compared to determine whether, within agreed limits, they are identical. The term is also applied to written mathematical or symbolical descriptions, drawings or formulas setting forth the important features of objects to be produced or actions to be performed.

Company standards are those developed or adopted by a single enterprise for its own use. Industrial standards are those accepted and followed by a group of firms with similar activities and products. National standards are those promulgated by interindustrial or governmental standardizing bodies. International standards are issued by organizations such as the Universal Postal union, the International Wool Textile organization and the International Organization for Standardization. A business firm in its operations may find it useful to make reference to standards drawn from any or all of these sources.

From the standpoint of the individual company where industrial standards find their principal use, standards may be classified as: (1) technical if they deal with products, materials, methods and equipment; (2) managerial if they deal with administrative practices and procedures; (3) substantive if they deal with what is to be done; and (4) procedural if they describe how something is to be accomplished. Some standards are applicable throughout the company while others apply to only one department. From the point of view of their intended use, standards may be classified as follows:

a. Engineering standards applied in technical work and including properties of materials, fits and tolerances, terminology, drafting practices and the like.

b. Product standards intended to describe attributes and ingredients of manufactured items and embodied in drawings, formulas, material lists, descriptions or models.

c. Production standards for processes, methods, operations and equipment included in technical specifications, inspection and test methods, machine setup drawings and operation sheets.

d. Purchasing standards covering specifications for materials, parts, equipment and supplies bought by the firm.

e. Office and procedural standards referring to record-keeping, procedures for production and inventory control, report forms, methods of analysis, budgeting and cost accounting and other administrative practices.

f. Safety standards on machine operations, plant housekeeping, identification of hazards and personnel deportment.

Importance of Standardization.—Standards and the benefits derived from them are so much taken for granted that their importance is not appreciated until an annoyance or emergency stemming from incomplete standardization or utter lack of it is encountered. The traveler abroad finds an electrical appliance designed to operate on one cycle basis of electricity does not operate on another. The purchaser of an imported article discovers that metric sizes do not mesh with feet and inches. The homeowner learns that the replacement part manufactured for the current model of the dishwasher he owns does not fit his earlier model without costly adaptation. The housewife's food bargain does not turn out to be such a bargain when she sees that Grade A does not mean the same thing for butter that it does for canned fruit.

Modern mass production and mass marketing could not exist in anything like their present forms without comprehensive standardization. Without standards there could be no manufacture on the basis of interchangeable parts. Without standards, manufacturers could not accurately communicate about materials with scattered suppliers and would have to spend more time inspecting incoming items. Without product standards, buyers could not rely on brand names as indicators of consistent quality. The first cost of complicated production equipment would be much higher than it is now and would be much more expensive to maintain if standard equipment components were not available. Standards play a vital role at every point in the complex system of communication, transportation, production and exchange which must operate smoothly to produce and distribute large quantities of low-cost goods in an advanced economy.

Standards as Management Tools.—Standardization enables industry to retain and build on its experience to produce better products, for standards are one of the principal forms in which technical knowledge is recorded. By liberating effort that would otherwise have to be expended in rediscovering and reinventing, standards aid in increasing the productivity of industry on which material progress ultimately depends. The use of standards embodying the most economical available methods reduces costs and conserves resources. But while industrial standardization is responsive to consumer needs and desires, its utility for managerial purposes gives it its strongest impetus.

It is management's job to plan, organize and control company operations. Expression of plans via standards and application of systems of control in which deviations from standards cause corrective measures to be taken permit managers to delegate routine tasks and concentrate on major decisions. Extensive use of standards makes for clarity in communication and facilities coordination. It is generally accepted that standards help management find the best solutions to problems whether they concern product qualities, material properties, process characteristics or operating methods. As standards reach beyond individual companies to suppliers and users of products, further advantages appear.

Standards for products that prevent needless variation and avoid the waste of excessive quality lead in turn to production econo-

mies: cheaper materials may be made to serve, painstaking production steps eliminated, inspection requirements reduced and simpler equipment and operations employed. Reduction of product variety also brings dividends through design simplification, making the use of special equipment feasible by permitting longer production runs on fewer items, reduction of inventories and use of simpler planning and control procedures.

Company Standards Programs.—Every company has standards but not every company consciously pursues standardization as a means of cost reduction and improvement of operations. Many companies handle standards problems informally; they do not systematically review and codify their standards or use them as an integral part of the management process. Companies that do have formal programs for the development and administration of standards employ a variety of organizational arrangements and procedures but they usually have some things in common. Some have a separate group headed by a standards engineer devoting full time to standards work; in other companies the work may be assigned to permanent committees or to committees convened for particular standards projects. In many firms consultative or coordinative committees are established in addition to the staff groups or working committees which do the detailed work of standards preparation and distribution.

Regardless of the organizational scheme, those responsible for standards work are usually expected to advise and assist top management in the formulation of policies relating to standards and their use. They are called upon to draw up procedures for creating new standards or revising old standards, to report on standards practice and evaluate the company program, to act as coordinators and mediators as standards are being developed, to serve as company representatives to technical associations and other standardizing bodies, to provide technical assistance to standards users and to maintain standards manuals and other records.

Other Contributors to Standardizotion.—Individual firms secure significant benefits by conducting their own standardization programs but there is a point at which standardization on a wide scale becomes necessary to prevent wasteful conflict and duplication of effort among competing systems of standards. Certain fundamental standards should be developed in forms that make them universally usable, and mechanisms for coordination of intersecting standards being evolved in separate fields should be provided. The standards activities of governmental departments, trade associations, technical associations and other organizations serve in part to meet national standards needs, but the presence of a specialized standardizing organization which can coordinate the diverse standardization activities of many different types of organizations and promote general acceptance of basic standards can be a strong force for progress.

In the U.S. the American Standards association (A.S.A.), which has over 110 technical organizations and 2,300 companies as members, performs this function. The A.S.A. does not initiate or write standards but provides the means by which national standards can be developed and approved. All groups having substantial concern with a proposal under consideration in an A.S.A. project have the right to representation. The association makes no effort to enforce any standard but relies on the voluntary support of interested groups.

What the A.S.A. does nationally, the International Organization for Standardization (I.S.O.) does internationally. National and international standards are essentially the same and are developed in the same way. However, just as there are fewer industrial than company standards and fewer national than industrial standards, there are fewer international than national standards. Developing, writing and adopting an international standard is difficult because of the breadth of representation and the diversity of needs and viewpoints that must be reconciled.

To say that standardization is one of the most pervasive and crucial of industrial and scientific activities is no exaggeration. Everyone's life is touched continuously by standards directly or indirectly. Scientific research that widens basic knowledge depends on accurate standards for physical constants and properties of materials as does applied industrial research. Companies may buy and sell with clear understanding on both sides of transactions when they negotiate in standard terms. Producers can manufacture articles of uniform quality and operate more efficiently with standardized processes and products. Consumers may buy agricultural products, meat, processed foods, pharmaceuticals, textiles, clothing, house furnishings and appliances with the assurance that they are getting what they pay for and that the goods are safe to use. *See* Mass Production. *See* also Index references under "Standardization" in the Index volume.

Bibliography.—Dickson Reck (ed.), *National Standards in a Modern Economy* (1956); Benjamin Melnitsky, *Profiting from Industrial Standardization* (1953); Jack Rogers, "Industrial Standardization: Company Programs and Practices," National Industrial Conference Board, *Studies in Business Policy*, no. 85 (1957). (J. D. Rs.)

STANDARD OF LIVING. The precise meaning of this phrase must usually be judged from its context. In popular literature it is likely to imply the relative kind and quantity of consumer goods used. Social scientists use the phrase in a different sense and have supplementary concepts. The standard of living to them relates to conditions considered essential by a person or group. The standard of consumption relates solely to consumer goods, whereas the standard of living includes many other things—leisure time, working conditions, freedom to choose consumer goods and a job. In professional literature the range of things included by the phrase is not always spelled out. At times it seems to be restricted to consumer goods. The statement that people of the U.S. have a high standard of living may imply that the mass of the population has come to consider certain consumer goods as essential which are accepted in other countries as the right of a select few. The concepts, level of living and of consumption, relate to existing conditions—what is experienced in contrast with what is considered essential. Many so-called standard of living studies might more accurately be described as studies of consumption levels. They do, of course, give indirect information about standards since use tends to determine essentiality.

A few people undoubtedly have, at least for short periods of time, an excess of some of the things which they desire, in which case their level of living is higher than their standard. The standard of others may be higher than their level. Vague fleeting wants for things not available may occur. These do not constitute a part of the standard. It relates only to those wants which generate a strong drive for their fulfilment. Their achievement brings a marked sense of gratification and their continued lack leads to unrest and a sense of frustration.

Man has biological, social and aesthetic needs. The specific ways in which these are met are culturally determined. Inherited customs and traditions ensure continuity in a society. Nevertheless cultural change is universal, although its nature and extent differ among societies and from time to time. In the United States the experience of migration to a new environment and the mingling of people from many countries tended to increase the willingness to accept new goods. Scientific and other developments that reveal better ways of meeting recognized needs lead to qualitatively higher standards. If productivity increases at a faster rate than population, quantitatively higher standards are made possible. Important changes within societies may occur. The dominance of the democratic ideal and the decreasing cost per unit of output which characterizes mass production tends to reduce class differences. The mediaeval ideal was a stable class society, stable with respect to both occupation and consumption. In it demarcations in standards of living among groups were so important that sumptuary laws existed which prohibited spending beyond one's station in life. Freedom to compete for jobs and in consumption does not entirely wipe out class lines. In a study of the class structure of U.S. communities by W. Lloyd Warner and his associates, six social classes were identified in a typical U.S. city. Differences in pattern of consumption were found that were independent of income, especially for education, housing and the automobile. These differences led to the conclusion that the standard of living of the groups differed qualitatively as well as quantitatively. The study did not reveal, however, which groups were influencing most the direction of change.

Some attempts have been made to measure the extent and urgency of both met and unmet needs. Attitude studies, in which people have been asked about their needs and desires, were still in an experimental stage at mid-20th century. In addition, analyses have been made of change in purchases with change in income and price. These led to the classification of goods as necessities and luxuries—a classification which is, however, very crude. Studies over a period of time show that luxuries of one period are necessities in another; and furthermore the so-called luxuries in a given society may be looked upon as necessities by some of its members.

Standards of consumption affect population growth and long-run labour supply through their effect on the age of marriage and the birth rate. They also affect the total labour supply in the short run in that people weigh the relative merits of more income against more leisure or more direct production for the family; they affect the labour supply in a given occupation insofar as more income and the additional consumer goods thus made possible may compensate for less pleasant work. The existence of the desire for more consumer goods is also important in inducing families to migrate. Standards of living also have a bearing on the business cycle. Forecasters would like to know when income falls to what extent will families cut consumption? To what extent will they decrease savings and draw on assets in order to maintain customary consumption? And furthermore if production rises, will it find a ready market because of desires previously unmet? Or will the additional income go wholly or largely to savings? (*See* CONSUMPTION.) In the administration of a war economy or other emergency when customary levels of consumption cannot be maintained it is helpful to know where cuts will be least resented, and how best to carry on an informational program in order to make clear the need for the shortages and thus minimize unrest and frustration.

Occupation is an important part of the standard of living. Some types of work are so much preferred over others that a lower income will be accepted in lieu of a lower ranking job. Changes in income differentials during World War II in the United States brought considerable mobility of workers from more preferred to less preferred types of work. Ideas prevail based largely on customary income as to what the standard of living of a given occupational group should be. Jessica Peixotto, in *Getting and Spending at the Professional Standard of Living* (1927), emphasized the special consumption needs of professional people, and pointed out that the desire for high consumption would attract people away from academic professions unless earnings were made more nearly comparable to those of alternative jobs. Except in industries where wages are unusually low, the need to attract workers and ability of industry to pay rather than a minimum standard of living are the major factors of wage negotiations in the United States.

The standard of living has long been a matter of major interest to population analysts. Thomas Malthus, in his *Essay on Population* (1798), argued that population growth is determined by means of subsistence. He believed that it was held in check chiefly by starvation, pestilence and war, all of which were direct results of population pressure on means of subsistence. This view assumed a static standard of living. By the beginning of the 20th century, the decline in the rate of population growth in industrialized countries after decades of rising productivity shifted attention to the effect of a rising standard of living. Then two fears came to be expressed: (1) that families in general would come to place such a high value on material goods for themselves and their children that the birth rate would fall to a point where total population would decline; and (2) that the effect of a high standard of living on the birth rate was especially marked among families with high income; thus, the quality of the population was likely to be adversely affected. An upturn of the birth rate in some western countries after the middle 1930s, however, allayed the first fear. For example, the 1950 census of the United States reported a 55% increase since 1940 in children under 5 years of age, while the total population increased by only 14%.

An increase in real income or a change in relative values or both may have played a part. Joseph S. Davis considered that this change may have been due to "a change in *standards* of living . . . of such a nature that marriage and children are given a relatively higher priority in what is urgently desired, instead of continuing to lose ground to ever higher per capita consumption, comfort, and leisure" (J. S. Davis, *The Population Upsurge in the United States*, Food Research Institute War-Peace Pamphlets No. 12, Stanford University, Stanford, Calif., Dec. 1949, p. 57). Population analysts by 1950 also had evidence that the inverse relation of birth rate to income was tending to disappear, and that within social classes birth rate and level of income tended to be directly related. The highest birth rate was still found at the lowest income level. The factors responsible for this awaited investigation; there might be a relative lack of interest in consumer goods available in the market in contrast with desire for children, or an acceptance of a low level of education for children so the cost of rearing them is low; or children might be desired as a means of security in old age; or there might be lack of knowledge of birth control methods; or an assumption might exist that society would care "adequately" for the children.

Policies and programs of the United Nations brought to the forefront once more a consideration of the validity of the Malthusian doctrine. Among the nations of the world high birth rates occur where the level of consumption is low, and where death rates are high. The value system of these societies tends to stimulate high birth rates. Only thus can the population be maintained in the face of high death rates. Experience indicates that industrialization brings a reduction in mortality before a decline in fertility. To what extent is a decline in fertility dependent on a rising standard of living? What is the process of bringing this about? There are no clear-cut answers, in spite of the experience in western countries with falling birth rates which resulted, in part, from a rising standard of living.

Standards of living of individuals are a dominant driving force in a society. There are, in addition, social standards of living which are the basis of many types of policies. The levels of consumption provided those in prison or on relief are, for example, an expression of a society's minimum standard for its members. Social standards are implicit in minimum wages, unemployment benefits, public old age pension systems, family allowances and the provision of various goods from public funds. In addition to individual and social standards there are normative standards provided by scientists and other specialists which are used as yardsticks in evaluating consumption.

There are, for example, the recommended dietary allowances of the National Research council (1945), the major yardstick used in appraising the nutritional quality of U.S. diets. The American Public Health association, through its committee on the hygiene of housing, provided a yardstick for measuring adequacy of housing. Such standards as these are commonly utilized in formulating quantity budgets to measure the cost of a given level of consumption. The quantity of goods included in such budgets may exceed those consumed by families at the general level of the budget at the time it was prepared. Since normative standards incorporate the scientific development at a relatively early stage of its impact on consumer demand they tend to point the way in which consumption is developing. (*See* also CONSUMPTION; COST OF LIVING.)

BIBLIOGRAPHY.—Joseph S. Davis, "Standards and Content of Living," *American Economic Review* (March 1945); American Public Health Association, Committee on the Hygiene of Housing, *Basic Principles of Healthful Housing* (1939); John Dollard, *Caste and Class in a Southern Town* (1949); Graham Hutton, *Midwest at Noon* (1946); Clyde V. Kiser and P. K. Whelpton, "Social and Psychological Factors Affecting Fertility. Fertility Planning and Fertility Rates by Socio-Economic Status," *Milbank Memorial Fund Quarterly*, vol. 27 (1949); Hazel Kyrk, *A Theory of Consumption* (1923); Theresa S. McMahon, *Social and Economic Standards of Living* (1925); Margaret Mead, *And Keep Your Powder Dry* (1943); Frank W. Notestein, "Population—The Long View," in T. W. Schultz (ed.), *Food for the World* (1945); Jessica B. Peixotto, *Getting and Spending at the Professional Standard of Living* (1927); Thorstein Veblen, *The Theory of the Leisure Class* (1926); W. Lloyd Warner and Paul Lunt, *The Social Life of a Modern Community*, "Yankee City Series," vol. 1 (1941); Carle C. Zimmerman, *Consumption and Standards of Living* (1936); F. Zweig, *Labour, Life and Poverty* (1948).

(M. G. R.)

STANDARD OIL COMPANIES. These are corporations in the U.S. petroleum industry which were members of the Standard Oil Company (New Jersey) group of companies in the late 1800s and early 1900s. Many have retained the word "Standard" in their names, and the term "Standard Oil" sometimes is used in referring to them individually or collectively. Accurate reference, however, distinguishes between them by their individual designations, such as "Standard Oil Company (Indiana)," or "Standard Oil Company (Kentucky)," because they have been separate, independent and competing companies for many years.

Their history began with a petroleum refining firm in Cleveland, O., which in 1870 became The Standard Oil Company (Ohio). Many individuals and firms in the producing, refining, transporting or marketing branches of the oil business became affiliated with this company. The petroleum industry expanded rapidly, for kerosene at the time was replacing coal and whale oil. As the Cleveland enterprise grew more complex, a different form of organization became necessary, and in 1882 the 41 participants

pooled their interests and organized the Standard Oil trust. The nine leading stockholders became trustees, with John D. Rockefeller, a founder of the Ohio company, heading the group. The trust organized companies in several states and abroad.

In 1892 an Ohio state court decision made it impossible for the Ohio company to continue participation in the trust, and the latter, therefore, was dissolved. Following this action, Standard Oil Company of New Jersey, which had been formed by the trust in 1882, was empowered to exercise holding company functions; its name was changed to Standard Oil company. It is referred to as Standard Oil Company (New Jersey) to distinguish it from corporations bearing the same name but chartered in other states; and in 1899, it became the holding company for all the interests previously grouped in the trust.

The period was one of exceptional U.S. economic growth, marked by mergers, industrial combinations and often aggressive business methods. Consolidating facilities and financial resources advanced large-scale and efficient production and distribution of goods and services. But many people believed the resulting concentrations of economic power were excessive, and public pressure for governmental restraints on large business organizations brought about the enactment of new regulations. The United States government in 1906 brought suit against Standard Oil Company (New Jersey), under the Sherman act, and in 1911 following a supreme court ruling the parent company separated from 33 of its subsidiaries. At the time of dissolution, the organization as a whole had gross assets of $860,000,000.

The dissolution left the separated companies with serious dislocations in the functional operations of their business, throwing certain major operations out of balance: sources of crude oil were separated from refineries for processing the crude; refineries were left without markets able to absorb their potential output.

As a result of the introduction of mass production and consequent development of gasoline demand, followed by the immense petroleum requirements for World War I, the separated companies, in common with other large independent companies never associated with the Standard Oil group, experienced tremendous growth. After that war some former members of the pre-1911 organization began to seek crude oil wherever in the world there seemed a chance it might be found and to develop new markets abroad and at home. Several companies were purchased by, or merged with, other firms. Some became among the industry's most important companies. Today some engage in all major phases of the industry and in related businesses and have participated in the development of many of the world's principal oil fields. Affiliates of Standard Oil Company (New Jersey), the parent company in the pre-1911 organization, by the late 1950s produced in 17 countries, refined in 23 and marketed in 135 nations and dependencies.

Other members of the early organization which, although they became competitors, have kept the word "Standard" or an abbreviation of it in their names are Standard Oil Company of California, Standard Oil Company (Indiana), Standard Oil Company (Kentucky), The Standard Oil Company (Ohio) and Socony Mobil Oil Company, Inc., (merger of Standard Oil Company of New York and Vacuum Oil company). Former members also include The Atlantic Refining company, Borne Chemical company (formerly Borne-Scrymser company), The Buckeye Pipe Line company, Chesebrough-Pond's Inc. (Chesebrough Manufacturing Company Consolidated merged with Pond's Extract company), The Eureka Pipe Line company, National Transit company, The Ohio Oil company, Southern Pipe Line company, South Penn Oil company, Swan-Finch Oil corporation (formerly Swan & Finch company) and Union Tank Car company (formerly Union Tank Line company). (J. O. L.)

STANDARDS, NATIONAL BUREAU OF. Established March 3, 1901, by an act of the congress of the United States, the national bureau of standards is the principal agency of the federal government for basic and applied research in physics, mathematics, chemistry and engineering. Supplementing the enabling legislation, congress further prescribed the functions of the bureau in public law 619, July 22, 1950. Six major functions are listed: (1) maintenance and development of national standards of measurement and their application; (2) determination of physical constants and properties of materials; (3) development of methods of testing materials, structures, and mechanisms and the testing of materials, supplies and equipment; (4) scientific and technical cooperation with government and private agencies; (5) advisory services for federal agencies; and (6) invention and development of devices to serve special needs of the government. In addition to these general functions, other types of activities are specifically authorized, including research in nuclear physics, radio propagation and mathematics, as well as the classic divisions of the physical sciences.

The bureau is organized into sections, grouped according to subject matter into 15 major divisions: electricity, optics and metrology, heat and power, atomic and radiation physics, chemistry, mechanics, organic and fibrous materials, metallurgy, mineral products, building technology, mathematics, electronics, ordnance, missiles and radio propagation. The principal laboratories are in Washington, D.C. Field stations were maintained in a number of states and even outside the continental limits.

The basic research of the bureau is related to scientific standards and measurements. Continued research is conducted even in fields often assumed static (*e.g.*, length and time standards) in response to new or anticipated developments in other fields of science. Thus, the bureau developed atomic standards of length and time in order to provide more precise measurements needed in atomic and nuclear physics and high frequency communications.

Major projects involving pure and applied research and development are undertaken in accordance with needs of the federal government. Extensive work was under way, in the 1950s, in atomic and nuclear physics. The atomic bomb project was initiated by the bureau, upon presidential request, in 1939, and work in this area was expanded to include nonmilitary aspects, particularly in the field of radioactive isotopes. Contributions to the national defense also included projects in ordnance and missile devices. The development of special optical glasses, synthetic mica, radio beacons and the instrument landing system (ILS) for aircraft were other typical projects of a specific nature.

The manifold expansion of the radio-frequency spectrum after World War II created many problems concerning the behaviour and properties of high-frequency radio waves. Extensive studies of the ionosphere, of ionospheric propagation, ground effects and climatological factors have been conducted. These studies had direct applicability to commercial radio and television.

The work of the bureau of direct interest to the general public has to do with standards of measurement. The ordinary standards of weights and measures used by the public and industry, and which affect every commercial transaction, are dependent upon the basic work of the bureau. Every instrument—ranging from the electric meters in households to altitude indicators on supersonic aircraft—depends on the bureau's standards, measurements and calibrations for its precision and accuracy. Safety codes and safety measures concerned with housing, plumbing, electricity, fire, X-ray, radioactive isotopes, elevators, radio, etc., are developed by the bureau. Extensive studies of building materials and structures have been pursued. Studies of dental materials, hospital equipment, surgical tools, the structure of the human skeleton, cutting of tissue for microscopic examination and synthetic blood plasma are typical of the work conducted by the bureau in those areas of the biological sciences which have problems essentially physical in nature. (*See* STANDARDIZATION.)

The results of the bureau's work are available through three monthly periodicals (*Journal of Research, Technical News Bulletin* and *Basic Radio Propagation Predictions*) and a series of nonperiodical publications. An indexed list of publications (*Publications of the National Bureau of Standards*, circular 460, 375 pp.) is available from the superintendent of documents, U.S. government printing office, Washington 25, D.C. (HH. O.)

STANDISH, MYLES or MILES (*c.* 1584–1656), American colonist, was born, according to Nathanael Morton, in Lancashire, Eng., and became a soldier in the Netherlands, where he met the Pilgrim Fathers. He sailed with them in the "Mayflower" in 1620, and became the captain or military defender of New Plymouth. As soldier and linguist he was invaluable in dealing

with the natives. He rescued Weymouth colony from the Indians in 1623; returned to England on behalf of Plymouth colony in 1625–26, and later helped to buy out the Merchant Adventurers. He was treasurer of the colony 1644–49, and for a time assistant governor; but was never elected governor, perhaps because not a member of the Pilgrim Church. He removed to Duxbury, 1632, where he died in 1656. Longfellow's well-known poem, *Courtship of Myles Standish*, is based on a late tradition, published in Alden's work on Epitaphs in 1812–14.

See the histories of Plymouth colony by William Bradford and by Nathanael Morton. Also Porteus: *Captain Myles Standish* (1920) and *History of Standish* (1927). (T. C. P.)

STANDLEY, WILLIAM HARRISON (1872–1963), U.S. naval officer, was born in Ukiah, Calif., Dec. 18, 1872, and graduated from the United States Naval academy in 1895. He served as assistant chief of naval operations and then commanded destroyers of the Pacific fleet and later cruisers of the scouting force. From July 1, 1933, until his retirement, with rank of admiral, on Jan. 1, 1937, Standley served as chief of naval operations. He represented the U.S. at the conference on naval armament in London, 1935–36. He also was a member of the Roberts commission, which investigated the Pearl Harbor attack, but criticized the commission's report in later years. From April 1942 to Oct. 1943 he was ambassador to the U.S.S.R. In 1955 he published, in collaboration with Rear Adm. A. A. Ageton, *Admiral Ambassador to Russia*. He died Oct. 25, 1963. (J. B. Hn.)

STANFORD, SIR CHARLES VILLIERS (1852–1924), Irish composer, was born in Dublin on Sept. 30, 1852. After studying in London under Ernst Pauer he won (1870) a scholarship at Queen's college, Cambridge, whence he migrated to Trinity college in 1873, and succeeded J. L. Hopkins as college organist, a post he held till 1892. His appointment as conductor of the Cambridge University Musical society gave him great opportunities. In 1874 to 1877 he visited Germany each year, studying under Reinecke and Kiel. His first important composition was incidental music to Tennyson's *Queen Mary* (1876). In 1881 his opera, *The Veiled Prophet*, was given at Hanover; followed by *Savonarola* (1884), and *The Canterbury Pilgrims* (1884). A long interval separates these from his later operas, *Shamus O'Brien*, the least pretending but most successful of his dramatic works (1896), and *Much Ado About Nothing* (1901). Meanwhile he had been appointed professor of composition at the Royal College of Music, 1883; conductor of the Bach choir, 1885; professor of music in the university of Cambridge, succeeding Sir G. A. Macfarren, 1887; conductor of the Leeds Philharmonic society, 1897, and of the Leeds festival from 1901 on. He was knighted in 1902 and died in London March 29, 1924. His instrumental works include six symphonies, many chamber compositions (among them two string quartets), many songs, part songs, madrigals, etc., incidental music to the *Eumenides* and *Oedipus Rex* (as performed at Cambridge) and Tennyson's *Becket*. His last operatic work was *The Travelling Companion*, posthumously produced in 1926. His church music and editions of Irish and other songs are well known. *See* his *Studies and Memories* (1908) and *Pages from an Unwritten Diary* (1914).

STANFORD UNIVERSITY, a U.S. institution of higher learning, is near Palo Alto, Calif., in the beautiful Santa Clara valley. It was founded in 1885 by Leland Stanford (1824–93) and by his wife Jane Lathrop Stanford (1825–1905) as a memorial to their only child, Leland Stanford, Jr., who died in 1884 at the age of 15. The founder was governor of California in 1862–63 and United States senator from 1885 until his death. The doors were opened in 1891 to 559 students. The university campus consists of Stanford's former Palo Alto farm, comprising about 9,000 ac.

The buildings, designed originally by H. H. Richardson and completed by his successors, Shepley, Rutan and Coolidge, are of soft buff sandstone in a style adapted from the old California mission (Moorish-Romanesque) architecture, being long and low with wide colonnades, open arches and red-tiled roofs.

Inner and outer quadrangles of buildings, centring on Stanford Memorial church, form the original group of classrooms and offices. Buildings later erected in adjoining areas include the Thomas Welton Stanford art gallery, the main library, the Hoover Library on War, Revolution, and Peace, Memorial hall, a museum of art and archaeology, gymnasiums, the student union, alumni headquarters, Stanford University press, and laboratories of engineering, chemistry and physics. Outdoor Laurence Frost amphitheatre seats about 8,000 persons. There is an arboretum of about 300 ac.

The Hopkins marine station, which is maintained at Pacific Grove on the Bay of Monterey, has two main buildings, one of them, the Jacques Loeb laboratory, a gift of the Rockefeller foundation. After 1908 the university maintained the Stanford medical school, the Lane and Stanford hospitals, the Stanford school for nursing and the Lane medical library in San Francisco, although a long-term program was begun in 1953 to move these facilities to the Palo Alto campus.

The departments of the university are organized into schools of law, medicine, education, engineering, business, mineral sciences, and humanities and sciences.

In 1921 the Food Research Institute of Stanford university was organized, with the help of the Carnegie Corporation of New York. In 1926 the Harris J. Ryan high voltage laboratory was established. Through the generosity of the Daniel Guggenheim Fund for the Promotion of Aeronautics, Inc., an experimental laboratory in aeronautics (the Daniel Guggenheim Aeronautic laboratory) was established in 1926. Following World War II, the Salvatori Geophysics laboratory, the Electronics Research laboratory, an organic chemistry laboratory and the W. W. Hansen laboratories of physics were built.

The influence of sectarianism and politics was barred from the university by its charter, and by its private origin and private support. The president of the university has the initiative in all appointments and in all matters of general policy. Within the university the faculty power lies in an academic council. An advisory board of seven professors, elected by the academic council without nomination, accords its approval or disapproval to all appointments and all major policies before they are presented to the board of trustees by the president.

The university received from its founders securities and land worth about $30,000,000. Most of the large ranch holdings of the university were sold in 1919. The Hoover Library on War, Revolution, and Peace, founded by Herbert Hoover during World War I, contains more than 1,000,000 printed and manuscript items dealing with World Wars I and II, with revolutionary movements of the 20th century and with international relations. The Hoover library building, with a 280-ft. tower, was completed in 1941.

Four thousand of the university's acres were set aside for the educational campus. In the late 1940s a program was begun for developing the remaining 5,000 ac. for residential, light industrial, commercial and professional uses on a leasehold basis.

The university campus was zoned and space reserved for an extension of the residence hall system for men and for women, together with the athletic fields. The large extent of the campus made it possible to develop athletic fields of considerable size and diversity, together with open-air swimming pools and other athletic and recreational facilities. Part of the campus was reserved also for experimental and ornamental purposes and as residence sites for faculty members. David Starr Jordan (*q.v.*) was the first president. (R. L. W.; P. C. A.)

STANGERIA: *see* Gymnosperms.

STANHOPE, CHARLES STANHOPE, 3rd Earl (1753–1816), English statesman and scientist, an eccentric with a touch of genius, who made a name for himself not only as a thoroughgoing radical but also as an experimental scientist, was born in London, Aug. 3, 1753, and educated at Eton and Geneva. As Lord Mahon, he was M.P. for Chipping Wycombe from 1780 until his succession to the peerage in 1786. He opposed the American war, but also the Fox-North coalition. In 1783 he declined office under William Pitt, whose sister he had married in 1774, but remained his supporter until the French Revolution. He sympathized with the aims of the French republicans, loved to call himself "Citizen Stanhope," and was chairman of the Revolution society, one of the popular societies advocating the democratization of parliament. He opposed the war with France, and deprecated interference with its internal affairs. Being "in a minority of one" in the house of lords he seceded from parliament for five

years (1795–1800).

Stanhope's inventions included the printing press and the microscopic lens which bear his name, two calculating machines, a steam carriage and a steamship for conveying coal from Newcastle to London, which, however, said Lord Holland, "would have consumed its cargo before it could have reached its destination." He made a new kind of cement, much more durable than ordinary mortar, and artificial slates or tiles composed of tar, chalk and fine sand. He projected a canal to link up his Devonshire estate at Holsworthy with the Bristol channel, and experimented with new methods of raising and lowering the barges. His writings include *Considerations on the Means of preventing Fraudulent Practices on the Gold Coin* (1775), *Principles of Electricity* (1779), *A Letter to Burke, containing a Short Answer to his Late Speech on the French Revolution* (1790), and pamphlets supporting Fox's libel bill (1792) and opposing the union with Ireland (1800). Lady Hester Stanhope (*q.v.*) was his eldest daughter by his first wife. He died at Chevening, on Dec. 15, 1816.

See G. P. Gooch, *Life of Charles, 3rd Earl Stanhope* (London, 1914). (A. Al.)

STANHOPE, LADY HESTER LUCY (1776–1839), the eldest of the 3rd earl Stanhope's three daughters by his first wife, Lady Hester Pitt, was born at Chevening, Kent, on March 12, 1776. In 1803 she went to keep house for her uncle, William Pitt, the prime minister, and after his death a pension of £1,200 a year was settled on her. Disliking the conventionalities of English society, this eccentric lady went abroad in 1810 and never returned. After many wanderings in the near east she settled down among the half-savage tribes near Mount Lebanon, and acquired a reputation as a kind of prophetess. Her excessive liberality brought her into debt, and in 1838 Palmerston decided to appropriate the bulk of her pension to the settlement of her creditors' claims. She died on June 23, 1839, without an English attendant, her servants robbing her of practically all her possessions as soon as she was dead.

Bibliography.—Several years after her death there appeared three volumes of *Memoirs of the Lady Hester Stanhope as related by herself in Conversations with her Physician* (*Charles Lewis Meryon*) (London, 1845), and these were followed by three volumes of *Travels of Lady Hester Stanhope, forming the Completion of her Memoirs narrated by her Physician* (London, 1846). They presented a lively picture of this strange woman's life and character, and contained many anecdotes of Pitt and his colleagues in political life for a quarter of a century before his death. *See* also Mrs. Charles Roundell, *Lady Hester Stanhope* (London, 1909); the Duchess of Cleveland, *Life and Letters of Lady Hester Stanhope* (London, 1914); and F. Hamel, *Lady Hester L. Stanhope* (London, 1913). (A. Al.)

STANHOPE, JAMES STANHOPE, 1st Earl (1673–1721), English soldier and statesman, was a grandson of the first earl of Chesterfield, and eldest son of Alexander Stanhope, a diplomatist, from whom he acquired both strong Whig principles and a knowledge of languages and diplomacy. In 1691 he went to Italy to study the art of war under the duke of Schomberg. Subsequently he served in Flanders, Spain and Portugal under such distinguished generals as Marlborough and the duke of Ormonde. He accompanied the earl of Peterborough to Spain in 1705, supporting his plan to divert the campaign to Italy, and distinguished himself in the attack on Barcelona (*see* Spanish Succession, War of the). In 1706 as minister to the titular Charles III of Spain, he advocated an aggressive strategy, but allied dissensions led to defeat at Almanza (1707). Commander in chief in Spain in 1708, Stanhope captured Port Mahon to provide a winter anchorage for the Mediterranean fleet, and in 1710 his bold methods won victories at Almenara and Saragossa. Stanhope, however, occupied Madrid instead of excluding from Spain the French reinforcements which, under the duc de Vendôme, defeated and captured him at Brihuega (Dec. 1710). The Tory government in England did little for the release of so pronounced a Whig, and Stanhope returned only in Aug. 1712.

Stanhope had become M.P. for Newport (Isle of Wight) in 1701, and occupied the duke of Somerset's seat at Cockermouth, 1702–13. In 1710 he was among the chief managers of the Sacheverell trial (*see* Sacheverell, Henry), and the Whigs strove furiously to get him elected in his absence for Westminster at the ensuing general election. He accepted the Utrecht treaty once it was signed, but contrived to defeat the projected Tory commercial treaty with France. Stanhope rebuffed Tory accusations of speculation in Spain, and was now so prominent in opposition that when defeated at Cockermouth in 1713, he was overwhelmed with offers of seats and chose to sit for Wendover; the opposition moreover gave him supreme charge of the secret military measures designed to secure the Hanoverian succession.

To the general surprise, Stanhope was made southern secretary on the accession of George I. sharing with Robert Walpole the leadership of the house of commons. It was in foreign policy, however, that Stanhope revealed his genius. He at once renounced the Tory policy of alliance with France which had not ended with the Austro-Spanish conflict, in favour of the Whig tradition of agreement with the emperor, and, if necessary, war against France. By prolonged negotiation Stanhope persuaded the emperor to concede a Barrier treaty to the Dutch in 1715, which foreshadowed the ending of British diplomatic isolation by the treaty of Westminster with the emperor and Dutch in 1716. He also secured a very favourable commercial treaty with Spain in May 1716. The regent, the duke of Orleans, now hoped that Britain would support his succession to the French crown should Louis XV die; Stanhope was at first very hostile to his overtures, but was won to them in 1716 by Russian threats to the German interests of Hanover, and by a projected Russo-Swedish *rapprochement* which might intensify Swedish intrigues with the English Jacobites. The Triple alliance (1717) with France and the Dutch minimized foreign support for the Jacobites, and gained for England the valuable diplomatic influence and organization of France in northern Europe. Stanhope now sought to reconcile Spain and the emperor by offering the succession to the Italian duchies to Don Carlos, the queen of Spain's son, and compensating the emperor by exchanging his island of Sardinia for Sicily, then held by Savoy. In 1718 Stanhope brought the emperor into the Quadruple alliance to impose peace on these terms, terms which Spain accepted under duress in 1720. Stanhope's peace plan was not completely fulfilled for a decade after his death, but his French alliance made Britain for 15 years the diplomatic arbiter of the west, and earned him a viscountcy in 1717 and an earldom in 1718.

In the north Stanhope was less successful, for England was at loggerheads not only with Sweden (for its attacks on English shipping and Jacobite conspiracies) and Russia (for its incursions into Germany), but also with its ally France which desired to save the German territories of Sweden, coveted by Hanover and its allies, by sacrificing to Russia Sweden's Baltic lands from which the British navy derived essential stores. The king and the Hanoverian ministers were tempted to divert the fleets (sent to the Baltic every year after 1715 to protect British shipping) to illegal Hanoverian purposes, thus provoking the resignation of the Townshend-Walpole connection from the government in 1717; Stanhope was now chief minister. On the death of Charles XII in 1718, he sought favourable terms from Sweden and offered to defend it against Russia by a coalition of the other Baltic powers. To form this coalition he himself broke the resistance of the Hanoverian ministers to an alliance with Prussia in 1719, and in the following year finally ended Hanoverian interference in English affairs. Under his direction Sweden was reconciled with all its Baltic neighbours except Russia, but in 1721, after Stanhope's death, Peter the Great was able to dictate his own terms to Sweden.

Stanhope had no fixed home in England until after he was 40, and never rivaled Walpole's grasp of domestic politics, but as secretary of state and first lord of the treasury (1717–18) he held a leading role in home affairs. He co-operated cordially with Sir Robert Walpole and Lord Townshend in suppressing the Jacobite revolt of 1715, softened Townshend's fall in the winter of 1716–17, and carried through the comprehensive financial scheme introduced by Walpole in 1717. Always the champion of religious minorities, he repealed the Occasional Conformity and Schism acts in 1718, but further concessions to Protestant dissenters and Roman Catholics were prevented by the split in the Whig party. In 1719 he unreservedly backed the earl of Sunderland's plan to reunite the Whigs by promoting the Peerage bill to

fix the number of peers, state control of the universities and the repeal of the Septennial act which he himself had passed in 1716. By defeating the first of these measures, however, Walpole ensured both the abandonment of the rest, and his own return to office. The last year of Stanhope's life was darkened by the South Sea company scandal (*see* SOUTH SEA BUBBLE), and, though in no way implicated, he was about to retire from politics into Marlborough's place of captain general when in Feb. 1721 he died of a blood vessel burst in anger during a South Sea debate in the house of lords.

See the standard life of Stanhope by B. Williams (New York, Oxford, 1932). (W. R. WD.)

STANHOPE, PHILIP HENRY STANHOPE, 5TH EARL (1805–1875), English historian, better known as Lord Mahon, son of the 4th earl, was born on Jan. 30, 1805, was educated at Christ Church college, Oxford, and entered parliament in 1830. His chief interest was in literature. He was a trustee of the British Museum, and in 1856 proposed the foundation of the National Portrait Gallery, whose subsequent creation was due to his executors. He also assisted in the establishment of the Historical Manuscripts Commission, and was president of the Society of Antiquaries from 1846 onwards. In 1855 he founded the Stanhope Essay prize at Oxford. His works include: *Life of Belisarius* (1829); *History of the War of Succession in Spain* (1832); *History of England from the Peace of Utrecht to the Peace of Versailles* (1836–53); *Life of William Pitt* (1861–62); and *History of England, comprising the reign of Queen Anne until the Peace of Utrecht* (1870, new ed. 1908), the last two of which remain standard works. Stanhope had unique access to ms. authorities, a clear style, and a general impartiality in criticism. He died on Dec. 24, 1875, and was succeeded by his son, Arthur Philip (1838–1905).

STANISLAUS I. [LESZCZYNSKI] (1677–1766), king of Poland, born at Lemberg in 1677, was the son of Rafael Leszczynski, palatine of Posen, and Anne Catherine Jablonowska. He married Catherine Opalinska by whom he had one daughter. In 1697, as cupbearer of Poland, he signed the confirmation of the articles of election of Augustus II. In 1703 he joined the Lithuanian Confederacy against Augustus, and in the following year Charles XII. selected him to supersede Augustus, and by bribery and intimidation secured his election (July 2, 1704). A few months later Stanislaus was forced by a sudden inroad of Augustus to seek refuge in the Swedish camp, but finally on Sept. 24, 1705, was crowned king with great splendour. Stanislaus at once concluded an alliance with Charles, and rendered him some small assistance against the tsar; but he depended entirely on the success of his ally's arms, and after Poltava (1709) the vast majority of the Poles hastened to repudiate him and make their peace with Augustus, and Leszczynski, henceforth a mere pensioner of Charles XII., retreated to Swedish Pomerania. On the restoration of Augustus, Stanislaus resigned the Polish Crown (though he retained the royal title) in exchange for the little principality of Zweibrücken, residing at Weissembourg in Lorraine.

In 1725 Louis XV. of France married Stanislaus' daughter Maria, and supported his claims to the Polish throne after the death of Augustus II. in 1733, which led to the war of the Polish Succession. On Sept. 9, 1733, Stanislaus himself arrived at Warsaw, having travelled through central Europe disguised as a coachman, and on the 10th was once more elected king of Poland. Russia, however, protested, proclaimed the Saxon candidate king as Augustus III. (June 30, 1734), and in October besieged Stanislaus with his partisans in Danzig. The expected French help did not arrive until May 20, 1735, when a fleet disembarked a small force, which, however, could do little. On June 30 Danzig capitulated unconditionally after a stubborn resistance of 135 days. Stanislaus, disguised as a peasant, had contrived to escape two days before. He attempted to rally his partisans from Königsberg, and to secure fresh help from France and from Count Potocki in the Ukraine, but without avail.

In 1736 Stanislaus again abdicated the throne, but received by way of compensation the dukedom of Lorraine and Bar, which was to revert to France on his death. He settled at Lunéville, founded there the *Academia Stanislai*, and devoted himself for the rest of his life to science and philanthropy. He died in 1766 at the age of 89. Among his works may be mentioned: *Oeuvres du philosophe bienfaisant* (Paris, 1763; 1866).

See Robert Nisbet Bain, *Charles XII.* (1895); Louis Lacroix, *Les Opuscules inédites de S. L.* (Nancy, 1866); *Lettres inédites de S. L.*, ed. P. Boyé (1901); Marchioness Des Reaulx, *Le Roi Stanislas et Marie Leszczynski* (1895).

STANISLAUS II. AUGUSTUS [PONIATOWSKI] (1732–1798), king of Poland, the son of Stanislaw Poniatowski, palatine of Cracow, the friend of Charles XII. of Sweden. Through the influence of his uncles the powerful Czartoryski, he was sent to St. Petersburg in the suite of the English ambassador Hanbury Williams. Subsequently, through the influence of the Russian chancellor, Bestuzhev-Ryumin, he was accredited to the Russian court as the ambassador of Saxony. Through Williams he was introduced to the grand duchess Catherine, who was irresistibly attracted to the brilliant young nobleman, for whom she abandoned her other lovers. Poniatowski was concerned in the conspiracy to set aside the succession of the grand duke Peter and his son Paul in favour of Catherine, a conspiracy frustrated by the unexpected recovery of the empress Elizabeth and the consequent arrest of the conspirators.

Stanislaus returned to Warsaw much discredited, but nevertheless was (Sept. 7, 1764) elected king of Poland through the overwhelming influence of Catherine, and crowned on November 25 to the impotent disgust of his uncles. He was hated by the nobility, yet he tried to do his best. He inaugurated some economical reforms. After the first partition he sought to restore the power of their country, while his eloquent oration before the Diet on taking the oath on May 3, 1791, moved the deputies to tears. But when the confederation of Targowica was formed against the constitution, he was one of the first to accede to it, in this way completely paralysing the action of the army which, under his nephew Prince Joseph and Thaddeus Kosciuszko, was performing prodigies. On the outbreak of the insurrection of 1794 he was obliged to sue for his life to Kosciuszko, and his effigy was expunged from the coinage a year before his abdication. His last years were spent in his sumptuous prison at St. Petersburg (where he died in 1798) in writing his memoirs. He contracted a secret marriage with the countess Grabowska.

See R. N. Bain, *The Last King of Poland and his Contemporaries* (1909).

STANISLAVSKI, KONSTANTIN SERGEEVICH ALEKSEEV (1863–1938), great Russian actor, producer, and theoretician of the art of acting, was born in Moscow on Jan. 17 (new style; 5, old style), 1863. His father, S. V. Alekseev, was a wealthy industrialist, and his grandmother, a French actress. He began his acting career at 15 in the Alekseev Circle, an amateur group consisting mostly of the members of his family. In December 1888 he founded the Society of Art and Literature, one of Moscow's most popular amateur theatrical companies. In June 1897 V. I. Nemirovich-Danchenko (*q.v.*), playwright and director of the drama school of the Moscow Philharmonic Society, invited him to form a new theatre. They drew up the program of the Moscow Art Theatre (*q.v.*), Nemirovich-Danchenko undertaking responsibility for literary and administrative matters and Stanislavski for staging and production. Stanislavski scored his first great success as producer with A. P. Chekhov's *The Seagull*, in December 1898, and through his genius the Moscow Art Theatre became world famous. Chekhov wrote *The Three Sisters* (1901) and *The Cherry Orchard* (1903) specially for it, and in 1923 Stanislavski toured the United States with it. Stanislavski's fame, however, rests chiefly on his theory of acting, his celebrated system or method: "What is important to me [he wrote in *My Life in Art*] is not the truth outside myself, but the truth within myself . . . the truth of my attitude towards anything on the stage, towards my partners who represent other characters in the play and towards their thoughts and feelings." (Theatre Arts Books, New York, 1924; renewed 1952 by Elizabeth Reynolds Hapgood.) Other writings include *An Actor Prepares* (Eng. trans. 1936) and *Building a Character* (Eng. trans. 1949). He died in Moscow on Aug.

7, 1938. *See* also ACTING, DIRECTION AND PRODUCTION: *Acting: Stanislavski.*

See D. Magarshack, *Stanislavsky: a Life* (1951). (D. MK.)

STANLEY, the name of a famous English family which derived its name from Stanley near Leek, in the Staffordshire moorlands. Its first known ancestor is Adam de Stanley, who lived in the time of King Stephen. His descendant, William de Stanley, acquired the forestership of Wirral by marrying an heiress in 1284, and was ancestor of two brothers, Sir William and Sir John Stanley. The former married the heiress of Hooton in Wirral and was ancestor of the Stanleys of Hooton, whose baronetcy, created in 1661, became extinct in 1893. The younger brother, lieutenant of Ireland under Richard II and Henry IV, obtained from the latter the Isle of Man in fee. His grandson Thomas was father of the 1st earl of Derby and of Sir William Stanley of Holt, whose great wealth led to his execution for treason in 1495, and also of Sir John Stanley, ancestor of the Stanleys of Alderley in Cheshire. The earls of Derby are noticed under DERBY, EARLS OF.

Thomas Stanley of Alderley obtained a baronetcy in 1660; and the barony of STANLEY OF ALDERLEY was created in 1839 for Sir John Thomas Stanley, 7th Bart. (1766–1850).

EDWARD JOHN STANLEY, 2ND BARON (1802–69), entered the house of commons in 1831 and became undersecretary to the home department in 1834, patronage secretary to the treasury from 1835 to 1841, paymaster general in 1841 and undersecretary for foreign affairs from 1846 to 1852. In 1848, two years before he succeeded to the barony of Stanley, he was created Baron Eddisbury of Winnington. He was president of the board of trade from 1855 to 1858, and postmaster general from 1860 to 1866. Though never in the front rank of politicians, he was active behind the scenes in promoting combinations likely to assist his Whiggish friends—an important role in an era of loose party ties.

His wife, Henrietta Maria, a daughter of Henry Augustus Dillon-Lee, 13th Viscount Dillon, had great influence in social and political circles. When he was patronage secretary her husband was described by Lord Palmerston as "joint whip with Mrs. Stanley." Later in life Lady Stanley of Alderley helped to found the Women's Liberal Unionist association, and she was a strenuous worker for the higher education of women, helping to establish Girton college, Cambridge, the Girls' Public Day School trust and the Medical College for Women. She died in 1895. (M. R. D. F.)

STANLEY, ARTHUR PENRHYN (1815–1881), English divine, dean of Westminster, was born on Dec. 13, 1815, at Alderley, Cheshire, the son of Edward Stanley (1779–1849), afterwards bishop of Norwich. He was educated at Rugby under Arnold, and at Balliol college, Oxford. In 1839 he was elected fellow of University college, also taking orders. In 1840 he travelled in Greece and Italy, and for ten years he was tutor of his college. In 1845 he was select preacher, and published in 1847 *Sermons and Essays on the Apostolic Age.* He was a strong advocate of toleration and used his influence to protect from formal condemnation the "Tractarian" party. In 1847 he resisted the movement set on foot at Oxford against R. D. Hampden's appointment to the bishopric of Hereford. Finally, in 1850, in an article in the *Edinburgh Review,* in defense of the "Gorham judgment" he asserted two principles which he maintained to the end of his life—first, "that the so-called supremacy of the Crown in religious matters was in reality nothing else than the supremacy of law," and, secondly, "that the Church of England by the very condition of its being, was not High or Low, but Broad, and had always included and been meant to include, opposite and contradictory opinions."

He was greatly interested in university reform and acted as secretary to the royal commission reporting in 1852. Stanley was also appointed to a canonry at Canterbury. During his residence there he published his *Memoir* of his father (1851), and completed his *Commentary on the Epistles to the Corinthians* (1855). In the winter of 1852–53 he made a tour in Egypt and the Holy Land, the result of which was his well-known volume on *Sinai and Palestine* (1856). In 1857 he travelled in Russia, and collected much material for his *Lectures on the Eastern Church* (1861).

At the close of 1856 Stanley was appointed regius professor of ecclesiastical history at Oxford, a post which, with the attached canonry at Christ Church, he held till 1863. He published the first two volumes of his *History of the Jewish Church* in 1863 and 1865. In the storm which followed the publication of *Essays and Reviews* Stanley opposed the High Church Party. In 1836 he published a *Letter to the Bishop of London,* advocating a relaxation of the terms of clerical subscription to the Thirty-nine Articles and the Prayer Book. An act amending the Act of Uniformity, and carrying out in some degree Stanley's proposals, was passed in the year 1865. In 1862, Stanley accompanied the prince of Wales in Egypt and Palestine.

In 1863 he was appointed dean of Westminster. In December he married Lady Augusta Bruce, sister of Lord Elgin, then governor-general of India. He wrote a third volume of his *History of the Jewish Church,* a volume on the *Church of Scotland,* another of *Addresses and Sermons* preached in America, and another on *Christian Institutions* (1881). He was constantly fighting for the interests of the poor. He gave offence by defending Bishop Colenso, although he disapproved of Colenso's views, and, still more, by his invitation to the Holy Communion of all the revisers of the translation of the Bible, including a Unitarian among other Nonconformists. He desired that the Athanasian Creed be optional instead of imperative in the Church of England. In 1874 Stanley, who was much esteemed by Queen Victoria, conducted the Anglican ceremony at the marriage of the duke of Edinburgh and the grand duchess Marie. He died on July 18, 1881, and was buried in Henry VII.'s chapel.

Stanley's other works include: *Life of Dr. Arnold* (1844); *Essays on Church and State* (1870); *Memorials of Westminster Abbey* (1865).

See G. G. Bradley *Recollections of A. P. Stanley* (1883); R. E. Prothero and G. G. Bradley, *Life and Correspondence of Dean Stanley* (2 vol., 1893).

STANLEY, SIR HENRY MORTON (1841–1904), Welsh explorer of Africa, a naturalized American who later resumed his British nationality, is popularly famous for his relief of David Livingstone in 1871, but his true importance lies in his discovery and development of the Congo. He was born on Jan. 28, 1841, at Denbigh, North Wales, the illegitimate son of John Rowlands and Elizabeth Parry, and was registered as John Rowlands. His mother left him to the care of her father, Moses Parry, and he grew up partly in the charge of reluctant relatives and partly in the St. Asaph workhouse, where he suffered from the severity of the workhouse master, James Francis. He acquired a reasonable education at St. Asaph, developed a strong religious faith and learned to value order and cleanliness; but the humiliations of institutional life left deep marks on his character, as did his mother's continued neglect. At the age of 15 he ran away from the workhouse and his escape was followed by an unhappy interlude of dependence on his mother's relations which only ended when he sailed from Liverpool as a cabin boy to New Orleans, La., where he landed in 1858.

There Rowlands met Henry Morton Stanley, a benevolent merchant who took charge of him, gave him his own name and declared his intention of providing for him. Although Stanley died abroad soon afterward, leaving his protégé once more alone in the world, his kindness gave a new direction to the boy's life; henceforth young Stanley felt that at least one person had loved him and believed in him, and in his own prosperous middle age he too adopted a son.

For some years Stanley led a roving life: a soldier in the American Civil War, a seaman in merchant ships and in the federal navy, a journalist in the early days of frontier expansion—he even included a trip to Turkey. Twice he visited his mother in Wales only to encounter the same indifference she had shown since his birth.

In 1867 Stanley offered himself to James Gordon Bennett Jr. of the *New York Herald* as special correspondent on the British Abyssinian campaign, and was first with the news of the fall of Magdala in 1868. He was in Madrid covering the Spanish Civil War for the *New York Herald* when, in 1869, Bennett summoned him to Paris and commissioned him to "find Livingstone."

Relief of Livingstone, 1869–73.—Little had been heard of Livingstone after he left for the interior of Africa in 1866 to explore the central African lakes and to ascertain the source of the Nile,

but he was thought to be in the neighbourhood of Ujiji on Lake Tanganyika. Bennett instructed Stanley to travel to Africa by way of Egypt (for the opening of the Suez canal), Syria and Palestine, and it was not until Jan. 6, 1871, that he reached Zanzibar. Intent on a "scoop," Stanley was secretive with the Zanzibar authorities, including the British consul, Sir John Kirk, who was responsible for Livingstone's supplies. Stanley set out for the interior on March 21, and with a well-equipped caravan forced his way through a countryside disturbed by fighting and stricken with sickness. He reached Ujiji in November, where he found Livingstone, greeting him with the famous words "Dr. Livingstone, I presume." Together they explored Lake Tanganyika, proving that the Rusizi river flowed into and not out of the northern end of the lake and so could not be the Nile as had been suggested by those geographers who doubted J. H. Speke's choice of Lake Victoria as the main source. Stanley arrived back in Zanzibar in May 1872, leaving Livingstone to pursue his search for the Nile further west in the region of what, as Stanley himself was later to prove, was the Congo. Stanley was not well received in England where it was felt that he had anticipated Livingstone's sponsors and the feeling against him was increased by the knowledge that Livingstone had been in real need of relief and that his own friends had failed where a stranger had succeeded. In the same year (1872) Stanley published his story in *How I Found Livingstone*.

Stanley did little to conciliate either Sir John Kirk or Sir Henry Rawlinson, president of the Royal Geographical society, nevertheless, the society awarded him their Patron's medal for 1873. He addressed the British association at Brighton in October, and he was received by Queen Victoria who gave him a gold snuffbox and her thanks for his services. In 1873 Stanley went to Ashanti as war correspondent for the *New York Herald* and in 1874 he published *Coomassie and Magdala: the Story of Two British Campaigns in Africa* describing his experiences there and in Abyssinia.

Through the Dark Continent (1874–77).—Livingstone died in 1873, and Stanley resolved to take up the exploration of Africa where he had left off. The problem of the Nile sources and of the position and size of the central African lakes had been only partly solved by Livingstone, Sir Richard Burton, Speke and Sir Samuel Baker, and Stanley secured financial backing from the *Daily Telegraph* and the *New York Herald* for an expedition to Africa. Well manned and provisioned and with three white companions, Stanley left Zanzibar on Nov. 11, 1874, and headed for Lake Victoria. His visit to King Mtesa of Buganda during 1875 paved the way for the admission of Christian missionaries in 1876 and for the eventual establishment of a British protectorate. Circumnavigating Lake Victoria in his boat, the "Lady Alice," Stanley confirmed Speke's estimate of its size and importance. Skirmishes with suspicious natives on the lake shore, and especially the fight at Bumbireh which resulted in a number of casualties, gave rise to much criticism in England of this new kind of traveler with his journalistic outlook and forceful methods. On this voyage were sown the seeds of Stanley's later unpopularity and of the rough political campaigns when he was heckled as a "man of blood." His claim to be acting in self-defense was put forward in aggressive terms which did not reassure public opinion in England.

Stanley was prevented by tribal unrest from investigating Lake Albert, and he turned south to Lake Tanganyika, which he also circumnavigated finding no outlet to connect it with Lake Albert and the Nile system.

By the end of 1876 Stanley was ready for what he called "the grandest task of all"—the identification of the Lualaba river which Livingstone had hoped might be the Nile but thought more likely to be the Congo. In 1876 the expedition reached Nyangwe, Livingstone's furthest point north on the Lualaba, and set off down the river reinforced by Tippoo Tib, the leading Arab trader of the region.

When the Arabs turned back at the limit of their trading area, Stanley launched the "Lady Alice" and sailed down the unknown river to the lake he called Stanley pool; from there he marched overland to the sea which he reached on Aug. 12, 1877. His three white companions and half his Africans were dead, but he had traced the Congo. Stanley described this tremendous journey in *Through the Dark Continent* (1878).

The Founding of the Congo Free State (1879–84).—Failing to interest the British government in developing the Congo, Stanley accepted the invitation of King Leopold of the Belgians to lead an expedition there under the auspices of the Comité d'Études du Haut Congo, later merged in Leopold's International association for exploring and developing Africa. From Aug. 1879 to June 1884 Stanley was in the Congo basin where he made a road from the lower Congo up to Stanley pool, launched steamers on the upper river and built a chain of stations. His perseverance in the face of difficulties earned him from his men the nickname of Buta-Matari—the "rock breaker." The Berlin conference of 1884–85, which Stanley attended, recognized the Congo Free state under the sovereignty of King Leopold, representing the International association.

Relief of Emin Pasha (1887–1889).—Stanley's last expedition in Africa was undertaken for the relief of Emin Pasha, governor of the Equatorial province of Egypt. Emin had been cut off by the Mahdi revolt of 1882 in the neighbourhood of Lake Albert. Sir William MacKinnon, chairman of the British East Africa company, launched a rescue fund and invited Stanley to lead an expedition. Stanley chose to travel by way of the Congo; he knew the route and relied on Tippoo Tib to supply porters. He left England in Jan. 1887 and, stopping to recruit men in Zanzibar, arrived at the mouth of the Congo in March. The expedition reached the navigable head of the river in June and there, at Yambuya, Stanley left a rear column under Maj. E. M. Barttelot and J. S. Jameson, with orders to await Tippoo Tib's porters and to follow on if these failed to arrive.

Stanley, with Lt. W. G. Stairs, Capt. R. H. Nelson, Surgeon T. H. Parke and A. J. Mounteney Jephson, led the advance guard out of camp on June 28, reaching Lake Albert five months later after a grueling march through dense forest. They did not make contact with Emin until April 1888 when Stanley found, to his astonishment, that the pasha was unwilling and unready to evacuate his province. Leaving him to assemble his troops, Stanley then returned to look for the missing rear column. He reached Balanya, just short of Yambuya, on Aug. 17, where he found the remains of the column, depleted and demoralized. Tippoo Tib had not sent the porters, Barttelot had been murdered and Jameson died soon after from fever.

Stanley blamed Barttelot and Jameson for the disaster, and on his return to England a controversy arose about the fate of the rear guard. This was the most bitter of the many quarrels that did so much damage to Stanley's reputation.

Back at the lake in Jan. 1889, Stanley heard that Emin and Jephson had been detained by disaffected troops, but eventually the party was reunited and on April 10, 1889, some 1,500 persons set out, arriving at Bagamoyo on the east coast on Dec. 4. Stanley cleared up some geographical points on this expedition. The mist-covered Ruwenzori range which had been concealed when Baker discovered Lake Albert in 1864, and which Stanley himself had passed unnoticed in 1876 was identified as the legendary Mountains of the Moon; the size of Lake Albert was correctly estimated and its tributary the Semliki river traced south to the lake which Stanley named Albert Edward on the homeward journey. *In Darkest Africa* (1890) is Stanley's account of this arduous and unlucky expedition.

The Last Years.—Stanley married Dorothy Tennant on July 12, 1890, and they adopted a son, Denzil. Stanley was renaturalized a British subject in 1892, and sat in parliament as Liberal Unionist member for North Lambeth from 1895 to 1900. In 1897 he visited South Africa and wrote *Through South Africa* (1898). He was awarded a grand cross of the Bath in 1899. His last years were spent mainly at Furze Hill near Pirbright, Surrey, a small estate bought in 1898.

Stanley died in London at 2, Richmond Terrace, Whitehall, on May 10, 1904, and was buried at Pirbright after a service in Westminster abbey.

Summary.—From a modern viewpoint, much contemporary criticism of Stanley appears unjust. To the select band of travelers whose centre was the Royal Geographical society, dedicated to the pursuit of knowledge, Stanley appeared as an outsider, an American journalist, with huge funds at his disposal. His motives in exploring were suspect and his methods were considered to be rough. His humble origin probably mattered less than he supposed to men who came, in fact, from a variety of social backgrounds, but his uncouth and unco-operative manners and his tendencies to self-righteousness and to boastfulness, ensured him an unpopularity which injures his reputation to this day. Harsh and humourless, touchy and intolerant, Stanley antagonized many who should have been his friends, and the controversies which arose around him have tended to obscure his achievements.

Nevertheless, to Stanley belongs the credit of having assembled the geographical discoveries and guesses of an earlier generation into the coherent whole which is the map of central Africa as it is now known. He is the bridge between the old and the new; at first the intrepid explorer of the adventure stories, the friend of Livingstone and the discoverer of the Congo; then the pioneer who made the Congo Free State; finally, the elder statesman who lived into the age of Cecil Rhodes and of the opening up of modern Africa. *See* also Index references under "Stanley, Sir Henry Morton" in the Index volume.

BIBLIOGRAPHY.—Cadwallader Rowlands, *H. M. Stanley* (1872); A.-J. Wauters, *Stanley au secours d'Émin-Pacha* (1890); Mrs. J. S. James, *Story of the Rear Column* (1890); W. G. Barttelot, *The Life of E. M. Bartlelot* (1890); Sir Harry Hamilton Johnston, *The Nile Quest* (1903); H. Brode, *Tippoo Tib, the Story of His Career in Central Africa* (1907); Dorothy Stanley (ed.), *The Autobiography of Henry Morton Stanley* (1909); Olivia Manning, *The Remarkable Expedition* (1947); Ian Anstruther, *I Presume* (1956); Brian Farwell, *The Man Who Presumed* (1957). (D. MN.)

STANLEY, THOMAS (1625–1678), English poet, translator and the first English historian of philosophy, was born at Cumberlow, Hertfordshire, in 1625, and educated by William Fairfax, son of the translator of Tasso. He became a good classical scholar and an enthusiastic student of French, Italian and Spanish poetry.

Stanley entered Pembroke hall (later college), Cambridge, in 1639 and studied there and at Oxford, graduating M.A. in 1641. He traveled on the continent during the Civil War and on his return lived in the Middle Temple, London, where he devoted himself to literary work. He died in London on April 12, 1678.

Stanley was the friend of many poets and himself a prolific writer of original and translated verse. His first volume of poems appeared in 1647. Subsequent volumes included translations from Guarini, Marino, Petrarch, Ronsard, Ausonius, Oronta, Montalvan and others. His classic renderings of the Anacreontic poems appeared in 1651 and the same collection contains his version of Pico della Mirandola's *Platonick Discourse Upon Love*. His *History of Philosophy* was published in 1655–62 and his edition of Aeschylus with Latin translation and commentary in 1663. He had a graceful if tenuous gift as a lyrical poet and was a versatile and accomplished translator. His *History of Philosophy*, based largely on Diogenes Laertius, long remained a standard work.

BIBLIOGRAPHY.—*Original Lyrics*, ed. by L. I. Guiney (1907); *Minor Poets of the Caroline Period*, ed. by G. Saintsbury, vol. iii (1921); *A Platonick Discourse Upon Love*, ed. by E. G. Gardner (1914); M. Praz, "Stanley, Sherburne and Ayres as Translators and Imitators of Italian, Spanish and French Poets," *Modern Language Review*, xx (1925); H. Thomas, "Three Translators of Góngora," *Revue Hispanique*, xlviii (1920). (V. DE S. P.)

STANLEY, WENDELL MEREDITH (1904–), U.S. biochemist and Nobel prize winner, was born on Aug. 16, 1904, in Ridgeville, Ind. After receiving his Ph.D. degree from the University of Illinois, Urbana, in 1929, he became a research associate and instructor in chemistry there. In 1931 he entered the Rockefeller Institute for Medical Research, and in 1940 became a member of the institute. In 1948 he became director of the laboratory for virus research at the University of California. In research on the virus which caused the mosaic disease in tobacco plants, Stanley found that while it appeared to act like an inanimate chemical it presented evidence of being a living and growing organism. For "their preparation of enzymes and virus proteins in a pure form," Stanley and John H. Northrop divided equally one-half of the Nobel prize in chemistry in 1946, the other half being awarded to James B. Sumner for his discovery "that enzymes can be crystallized."

STANNARIES, tin mines (late Lat. *stannum*, Cornish, *stéan*, tin). Stannary courts exercised a jurisdiction peculiar to Cornwall and Devon. By ancient charters, the tinners of Cornwall were exempt from all other jurisdiction than that of the stannary courts, except in cases affecting land, life and limb. Tin mining in Cornwall, from the very earliest times was always prosecuted in accordance with a particular code of customs; the earliest charter which embodies them is that of Edmund, earl of Cornwall, but it is impossible to say how far these customs go back.

Twenty-four stannators were returned for the whole of Cornwall. Their meeting was termed a parliament, and when they assembled they chose a speaker. In earlier times the combined tinners of Devon and Cornwall assembled on Hingston down, a tract of highland on the Cornish side of the Tamar. After the charter of Earl Edmund, the Cornish stannators met (apparently) at Truro; those of Devon at Crockern Tor on Dartmoor. An officer was appointed by the duke of Cornwall or the crown, who was lord warden of the stannaries. The last Cornish stannary parliament was held at Truro in 1752. A committee was appointed to report on the subject, and an act of parliament was passed (1836), suppressing the law courts of the stewards of the different stannaries and giving to the vice-warden their jurisdiction. By the Stannaries act, 1855, the respective parliaments or stannaries courts of Cornwall and Devon were consolidated. By the Stannaries Courts Abolition act, 1896, the jurisdiction was transferred to the county courts. The most important customs were: (*a*) "free tinners" had the right to work upon rendering the "toll-tin," usually one-fifteenth of the produce, to the owner or lord of the soil; (*b*) the right of "tin-bounding," that is, the right of bounding waste lands, subject to tin-toll.

See G. R. Lewis, *The Stannaries* (1908); W. S. Lewis, *West of England Tin Mining*, with bibliography (Exeter, 1923); A. K. Hamilton Jenkin, *The Cornish Miner* (1927).

STANTHORPE, a tourist and health resort in southeast Queensland, Austr., 198 mi. S.W. of Brisbane by rail (103 mi. direct). Altitude 2,650 ft. Pop. (1951) 2,907. The town, situated in the "granite belt," originated as a tin mining centre in 1872 and tin sluicing is practised on a large scale. The granite soils produce vast quantities of stone fruits, apples, pears and grapes, and fruit forms the principal wealth of the state.

STANTON, EDWIN McMASTERS (1814–1869), American statesman, was born at Steubenville, O., on Dec. 19, 1814. He attended Kenyon college at Gambier, O., and was admitted to the bar in 1836. In 1847 he removed to Pittsburgh, Pa. As counsel for the State, he invoked successfully the aid of the Federal Government in preventing the construction of a bridge over the Ohio river at Wheeling, Va., (now West Virginia)—on the ground that the structure would interfere with the navigation of that stream by citizens of Pennsylvania. He removed to Washington in 1856. In 1858 he was sent to California by the United States attorney general as special Federal agent for the settlement of land claims. Before the Civil War, Stanton was a Democrat, opposed to slavery, but a firm defender of the constitutional rights of the slaveholders, and was a bitter opponent of Lincoln, whose party he then hated and distrusted. In the reorganization of President Buchanan's cabinet in 1860 Stanton became attorney general, and he did what he could to strengthen the weak policy of the President. Although he had often violently denounced President Lincoln, the latter thought he saw in Stanton a good war minister, and in January, 1862, invited him into his cabinet. Stanton was often harsh, and his peremptory manner was the cause of friction with the generals, an instance being his controversy with Gen. Sherman over the terms of surrender granted to J. E. Johnston's army. But he removed a horde of fraudulent contractors, kept the armies in the field well equipped and infused energy into procrastinating generals. Not the least of his achievements was the peaceable disbandment of 800,000

soldiers at the end of the war.

Remaining in Pres. Andrew Johnson's cabinet, Stanton exerted all his energies toward thwarting Johnson's policies, especially those related to the reconstruction of the southern States. He expressed disapproval of the Tenure of Office act, making the consent of the senate necessary for the removal of civil officers, and drafted the supplementary act on reconstruction, passed over the President's veto on July 19, 1867. Stanton was finally asked to resign, and on his refusal to do so the President suspended him from office and appointed General Grant (who had disapproved of the secretary's removal) secretary *ad interim*. When the senate, however, under the terms of the Tenure of Office act, refused (Jan. 13, 1868) to concur in the suspension, Grant left the office and Stanton returned.

On Feb. 21, 1868, Johnson appointed Gen. Lorenzo Thomas secretary of war *ad interim*, and ordered Stanton to vacate, but on the same day the senate upheld Stanton, who swore to a complaint against Thomas for violating the Tenure of Office act, and invoked military protection from General Grant. Congress impeached the President, the principal article being that based on the removal of Stanton.

When impeachment proceedings failed (May 26), Stanton resigned and returned to the practice of law. Stanton had a violent temper and a sharp tongue, but he was courageous, energetic, thoroughly honest and a genuine patriot. He died in Washington, D.C., on Dec. 24, 1869.

See F. Pratt, *Stanton, Lincoln's Secretary of War* (1953).

STANTON, ELIZABETH CADY (1815–1902), U.S. leader in the women's rights movement, was born in Johnstown, N.Y., on Nov. 12, 1815. The daughter of Judge Daniel Cady, she learned in his law office of the discriminatory laws under which women lived, and observed many tragic cases which determined her to win equal rights for her sex. In 1840 she married the abolitionist Henry Brewster Stanton. In 1848 she circulated petitions to secure the passage in New York of a law giving married women property rights. Also in 1848 she and Mrs. Lucretia Mott (*q.v.*) were the leaders of the first women's rights convention in the U.S. at Seneca Falls, N.Y. For it Mrs. Stanton drew up her famous woman's bill of rights and resolutions demanding redress of wrongs, which were adopted. All these demands have since been extensively granted. A resolution, introduced without Mrs. Mott's approval, favouring women's suffrage (*q.v.*), was the first organized demand for it in the U.S. In 1850 Mrs. Stanton became associated with Susan B. Anthony (*q.v.*), and for 40 years they worked together, Mrs. Stanton writing and Miss Anthony managing business affairs. They co-operated in editing *The Revolution* (1868–70), a women's rights newspaper, and with Matilda Gage were joint editors of the first three volumes of *The History of Woman Suffrage* (6 vol., 1881–1922). Mrs. Stanton died in New York city on Oct. 26, 1902.

See T. Stanton and H. S. Blatch (eds.), *Elizabeth Cady Stanton as Revealed in Her Letters, Diary and Reminiscences*, 2 vol. (1922); Alma Lutz, *Created Equal* (1940).

STANYHURST, RICHARD (1547–1618), English translator of Virgil, was born in Dublin. At University college, Oxford, his commentaries on Porphyry won him a reputation for precocious learning. He read for the bar and after a stay in Ireland became a Roman Catholic. Most of his subsequent life was spent in the Netherlands where he became a priest (probably *c.* 1607) and chaplain to Archduke Albert of Austria. He died in Brussels.

Stanyhurst wrote the *Description* and *History* of Ireland found in Holinshed's *Chronicles* (1577) and Latin works including *De rebus in Hibernia gestis* (1584). His chief work, however, was his translation of *The First Four Books of Virgil His Aeneis* (Leiden, 1586), which attempted to prove Gabriel Harvey's theory that classical prosody could be applied to English poetry.

Thomas Nashe said that his translation "trod a foule, lumbering, boystrous, wallowing measure." But though forced to the point of absurdity and unintelligibility, from a prosodic viewpoint his translation, with its attempt to indicate quantity by spelling, is interesting.

See editions of the *Aeneis* by E. Arber (1880) and D. van der Haar (1933).

STAPLE, originally a prop or support; in current usage a U-shaped piece of metal, generally galvanized or wrought iron, in wire form with sharpened points, and used as a fastener by driving into wood; also the chief or ordinary bulk articles of trade. In England the term was applied to towns which were appointed by the king to be the merchants of the staple. These merchants had a monopoly in the staple commodities, viz., wool, wool fells, leather, tin and lead, and were the origin of all English trading companies. *See* also NAIL.

STAPPEN, CHARLES VAN DER (1843–1910), Belgian sculptor, was born in Brussels, Dec. 19, 1843. His first contribution to the Brussels Salon was "The Faun's Toilet" of 1869, and after this he soon became recognized as the leader, along with Paul de Vigne, of the new Belgian school of sculpture, which, while remaining true to life, drew inspiration from Greek art and the Italian Renaissance.

Examples of Stappen's decorative sculpture are seen in the decoration on the Palais des Postes (1872), the statues for the Alhambra theatre, the pediment "Orchestration" for the Conservatoire Royal de Musique, and the noble bronze group, "The Teaching of Art," on the façade of the Palace of Fine Arts, all in Brussels. His best-known monuments are those to "Alexandre Gendebien" (1874) and "Baron Coppens," at Sheel (1875).

STAR. The stars are self-luminous objects, shining by radiation derived from energy sources within themselves. By contrast, planets shine by reflected light only, while gaseous and diffuse nebulae may shine either by reflected light or (essentially) by fluorescence. As far as its properties can be compared with those of other stars, the sun is a typical star. It has a mass more than 300,000 times that of the earth, 1.99×10^{30} kg., a radius of 696,000 km. (about 432,200 mi.) and a power output of 3.8×10^{23} kw. The masses, radii and luminosities of other stars are often expressed in terms of that of the sun.

Table I gives data pertaining to the 20 brightest stars or, more precisely, stellar systems, since some of them are double or even triple stars. Note that all of these stars are as bright or brighter than the sun; some of the companions are fainter. Stars such as Altair, α Centauri and Procyon are called dwarf stars; their dimensions are roughly comparable with that of the sun. Sirius and Vega are also dwarf stars; their higher temperatures give a large rate of emission per unit area. Aldebaran, Arcturus and Capella are examples of giant stars. Thus, measurements with the interferometer (which gives the angular diameter) combined with parallax measurements give diameters of 24 and 45 solar diameters for Arcturus and Aldebaran, respectively. Betelgeuse and Antares are examples of supergiants. The latter has a diameter 200 times that of the sun whereas the diameter of Betelgeuse (if we choose a distance 1,700 light-years) is 250 times. Other examples of supergiants include α Herculis and the variable Mira whose diameters are 580 and 460 times as great as that of the sun. Several white dwarf stars of low luminosity and high densities are listed.

Table II gives the 20 nearest known stars. Only α Centauri, Procyon and Sirius appear in both lists. All other stars are fainter than the sun and most of them can be seen only with the telescope. The highly luminous stars can be seen at great distances; intrinsically faint stars can be seen only if they are very near the sun.

DISTANCES OF THE STARS

Before any adequate discussions of stars as celestial bodies can be made, it is necessary to know their distances. Several methods have been devised, some of them of a statistical nature, which permit the distances of various types of objects to be ascertained.

Trigonometric Parallaxes.—If the position of a nearby star is measured from two points on opposite sides of the earth's orbit, a small angular displacement will be observed. Using the earth's orbit as the base line, the distance of the star can be found from the angular size of the parallax, p (*see* fig. 1). If $p = 1''$, the distance of the star would be 206,-

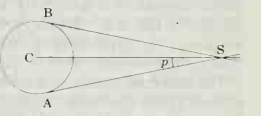


FIG. 1.—PARALLAX OF A STAR
Circle AB represents orbit of earth; angle CSA is heliocentric parallax, p

TABLE I.—*The Twenty Brightest Stars*

Name	Visual magnitude and spectrum			Distance in light-years	Visual luminosity ⊙		
	A	B	C		A	B	C
Sirius	−1.6 AO	7.1 A5	—	8.7	23	0.008	—
Canopus	−0.9 FO	—	—	180:	5,200:	—	—
α Centauri	+0.3 GO	1.7 K5	11 M	4.29	1.0	.28	.000052
Vega	0.1 AO	—	—	26.5	48	—	—
Capella	0.2 GO	10.0 M1	13.7 M5	45	120	.014	.0005
Arcturus	0.2 KO	—	—	36	76	—	—
Rigel	0.3 B8p	—	—	550–1,100	23,000:	—	—
Procyon	0.5 F5	10.8	—	11.3	5.8	.00044	—
Achernar	0.6 B5	—	—	140:	800:	—	—
β Centauri	0.9 B1	—	—	200:	1,300:	—	—
Altair	0.9 A5	—	—	16.5	8.3	—	—
Betelgeuse	(0.9) M2	—	—	650:	(13,000:)	—	—
α Crucis	1.4 B1	1.9 B1	—	220:	900:	600:	—
Aldebaran	1.1 K5	13 M2	—	68	120	—	—
Pollux	1.2 KO	—	—	35	30	.0021	—
Spica	1.2 B2	—	—	160:	600:	—	—
Antares	1.2 M1	5.2 B4	—	170:	700:	17:	—
Fomalhaut	1.3 A3	—	—	23	11	—	—
Deneb	1.3 A2p	—	—	540:	6,000:	—	—
Regulus	1.3 B8	7.6 K2	13	84	140	.44	.0030

Source: After Peter van de Kamp in *Publications of Astronomical Society of the Pacific*, vol. 65, p. 30 (1953).

265 times the distance of the earth from the sun, or 3.26 light-years. This unit of distance is called the parsec; hence

$$r = 1/p \quad (r \text{ in parsecs if } p \text{ is in seconds of arc}) \tag{1}$$

The nearest star (the triple-star system of α Centauri) has a parallax of 0".76. Its distance is 1.31 parsecs, or 4.29 light-years. Trigonometric parallaxes can be measured only for relatively close stars. The probable error of such a parallax is 0".005. This means that there is a 50–50 chance that a star whose parallax is 0".01 lies between 67 and 200 parsecs (corresponding to $p = 0.015$ and 0.005, respectively and an equal chance that it lies outside that range.

For measuring the distance of more remote stars other methods must be used.

Statistical Methods Based on Stellar Motions.—In spite of their designation as "fixed stars," the stars are not fixed in positions on the celestial sphere, nor in space. Edmund Halley, from a comparison of the positions of Arcturus, Aldebaran and Sirius, in the 18th century showed that there was an actual motion of the stars across the line of sight (proper motion). The amount of the proper motion (in seconds of arc per year) depends on two factors, the distance of the star and its velocity in km. per second in the plane of the celestial sphere (tangential velocity)

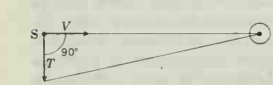


FIG. 2.—COMPONENTS OF STELLAR MOTION

V represents radial velocity; *T* is tangential velocity of the star S with respect to the sun ⊙

$$T = 4.74 \frac{\mu \text{ sec.}}{\text{parallax}} \tag{2}$$

$$\text{or tangential velocity} = \frac{\text{annual proper motion}}{\text{parallax}} \times 4.74 \text{ km. per second.}$$

The motion in the line of sight (radial velocity) is obtained directly from spectroscopic observations (fig. 2). If λ is the wave length of a characteristic spectrum line emitted by the star (*e.g.*, a line of iron) and λ_L is the wave length of the same line measured in a laboratory on the earth, the difference $\Delta\lambda = \lambda - \lambda_L$ is related to the radial velocity V, by Doppler's principle (*see* LIGHT: *Waves and Interference*) by

$$\Delta\lambda/\lambda = V/c \tag{3}$$

where c is the velocity of light. Shifts of the spectral line (positive V) toward red indicate recession. If the parallax is known, measurements of μ sec. and V enable one to determine the space motion of the star. Radial velocities must be corrected for the motion of the earth around the sun so that they will refer to the relative line-of-sight motion of the star with respect to the sun.

The observed motion of a given star is made up of two effects, the motion of the star with respect to the average of the nearby stars (the local standard of rest) and the motion of the sun with respect to this same standard of rest. The first problem that has to be solved is the determination of the sun's motion with respect to the nearby stars, both in velocity and direction. This vector quantity is established most reliably from radial velocity data. Stars in the region of the sky toward which the sun is moving (apex of the sun's way) show a predominantly negative velocity (approach) whereas those in the opposite region of the sky show a positive velocity (recession). Stars at position 90° away from the apex or antapex show a random mixture of positive and negative radial velocities. A statistical analysis of the observed radial velocities permits the determination of the apex and the velocity of the sun $V_\odot$ with respect to the average of the nearby stars. From the radial velocities of 2,149 bright stars, W. W. Campbell and J. H. Moore found the position of the apex to be $\alpha = +271°$, $\delta = 29°$ (constellation of Hercules not far from star μ Herculis). The space velocity of the sun was found to be 19.5 km. per second.

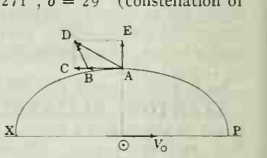


FIG. 3.—PERSPECTIVE DIAGRAM SEEN FROM ABOVE OF EFFECTS OF SOLAR MOTION

P is apex, X is antapex, A is position of a star where proper motion is observed. AE (perpendicular to lines A⊙ and AC) is tau component of star's motion; it lies perpendicular to plane of great circle, PAX, passing through apex, star and antapex. AC is upsilon component; AB, the reflection of the sun's motion $V_\odot$. Vectors BD and CB represent motion of star with respect to local standard of rest

Once the direction of the solar motion is known, the proper motion of any star can be divided into two components, one along the great circle passing through the apex, the star, and the antapex and the other perpendicular thereto. The first mentioned component is called the upsilon component, the other is called the tau component (fig. 3). Now the upsilon component contains the effects of both the star's own peculiar motion and the reflection of the sun's motion. The backward drift caused by the sun's motion will be a maximum for stars 90° from the apex, and zero at both the apex and antapex. The tau component is unaffected by the solar motion. Let us suppose that $|\bar{V}_s|$ is the average radial velocity (without regard to sign and corrected for the solar radial velocity) of all the stars of a particular kind; *e.g.*, giant stars like Arcturus. Now $|\bar{V}_s| = |\bar{V}_\tau|$, the average tau component of velocity if the stellar motions are at random. Hence, using equation (2)

$$\bar{p} = 4.74 \frac{|\bar{\tau}|}{|\bar{V}_s|} \tag{4}$$

where $(\bar{\tau})$ is the average of tau components without regard to sign. Notice that what is obtained is the average parallax; the parallax cannot be obtained for any individual star, since in any particular instance V_s will not be the same as V_τ. Nevertheless, if attention is confined to groups of stars of the same intrinsic luminosities (and not over too great a spread in distance) the method will give good results. It is assumed that the stars are moving at random and, strictly speaking, motion of the sun should be determined with respect to this particular group of stars. Statistical parallaxes fail for very remote objects and require reliable proper motions. Since the accuracy of a proper motion depends on the time interval between two epochs of observation, an increasing base line in time will serve to improve the data.

Moving Cluster Methods.—If a group of stars constitute what is called a moving cluster their space motions are parallel to one another. Provided the cluster fills a substantial area of the sky, the place in the celestial sphere toward which the cluster appears to be moving, *i.e.*, its convergent point, can be found from a careful observation of the proper motions (*see* fig. 4). The radial velocities are also measured. The angular distance θ between the

TABLE II.—*The Twenty Nearest Stars*

Name	Visual magnitude and spectrum			Distance in light-years	Visual luminosity ⊙		
	A	B	C		A	B	C
Sun	−26.9 GO	—	—	—	1.0	—	—
α Centauri	0.3 GO	1.7 K5	11 M5e	4.3	1.0	0.28	0.000052
Barnard's star	9.5 M5	*	—	6.0	.00040	*	—
Wolf 359	13.5 M6e	—	—	7.7	.000017	—	—
Luyten 726-8	12.5 M6e	13.0 M6e	—	7.9	.00004	.00003	—
Lalande 21185	7.5 M2	*	—	8.2	.0048	*	—
Sirius	−1.6 AO	7.1 wd	—	8.7	23.	.008	—
Ross 154	10.6 M5e	—	—	9.3	.00036	—	—
Ross 248	12.2 M6e	—	—	10.3	.00010	—	—
ε Eridani	3.8 K2	—	—	10.8	.25	—	—
Ross 128	11.1 M5	—	—	10.9	.00030	—	—
61 Cygni	5.6 K6	6.3 MO	*	11.1	.052	.028	*
Luyten 789-6	12.2 M6	—	—	11.2	.00012	—	—
Procyon	0.5 F5	10.8 wd	—	11.3	5.8	.00044	—
ε Indi	4.7 K5	—	—	11.4	.12	—	—
Σ 2398	8.9 M4	9.7 M4	—	11.6	.0028	.0013	—
Groombridge 34	8.1 M2e	10.9 M4e	—	11.7	.0058	.00044	—
τ Ceti	3.6 G1	—	—	11.8	.36	—	—
Lacaille 9352	7.2 M2	—	—	11.9	.013	—	—
B.D. +50°1668	10.1 M4	—	—	12.4	.0010	—	—

*Unseen components.
Source: After Peter van de Kamp in *Publications of Astronomical Society of the Pacific*, vol. 65, p. 73 (1953).

centre of the cluster and its convergent point being known, the tangential velocity T can be computed in km. per second from the radial velocity V. Since the tangential velocity expressed in seconds of arc is the total proper motion μ sec., the parallax can be found from equation (2). This method has been applied to the Ursa Major cluster (which includes five of the seven stars of the Big Dipper) and to other groups such as the Hyades.

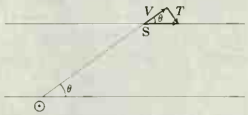


FIG. 4.—METHOD FOR DETERMINING THE DISTANCE OF A MOVING CLUSTER

S is a star in a moving cluster, θ is angle between cluster and convergent point. Tangential velocity $T = V \tan\theta$

Dynamical Parallaxes.—The distance of a double star whose orbit is known can be found from its apparent brightness provided the star follows the mass-luminosity correlation (*see* below). This method is useful for double stars in clusters (such as the Pleiades).

Membership in Galactic Clusters.—A star cluster such as the Pleiades contains stars similar in intrinsic brightnesses and colours to the sun and its neighbours. Hence if the apparent brightnesses of these stars are measured and compared with their true brightnesses, the distance of the cluster can be found from an application of the inverse-square law. If absorption of light in interstellar space between the cluster and the sun occurs, its amount has to be determined and the apparent brightnesses of the stars corrected to what they would have been for no interstellar absorption. This method is standard for establishing the distances of remote clusters.

Galactic Rotation.—The Milky Way is in slow rotation about a centre, some 28,000 light-years distant from the sun. This rotation is like that of the planetary system in that the stars more distant from the centre travel more slowly. The general nature of the galactic rotation can be established from the gaseous component of the interstellar medium with the aid of both optical and radio-frequency observations. From the radial velocities of very remote stars their distances may be obtained with the aid of equations derived by J. H. Oort.

Remote Systems.—All of these methods fail for remote systems such as the Magellanic clouds or the Triangulum or Andromeda galaxies. For these systems other methods must be used. Certain types of objects, blue stars, variable stars are recognized in some of them. The intrinsic brightnesses of these objects being known, their apparent brightnesses in the stellar systems are measured under investigation and thus their distances are obtained. Again, corrections for space absorption must be applied if necessary.

Checks on the methods of distance determinations can sometimes be made when two or more different techniques can be applied to the same stars or system of stars. One example is provided by the Ursa Major moving cluster where R. M. Petrie and B. Moyls obtained distances in good agreement with those found from the direct trigonometric parallaxes of the same stars. A second illustration is supplied by the binary Procyon. The orbit of the bright star about the centre of gravity of the system and its inclination to the plane of the sky are known. K. Strand measured the radial velocity of the star from plates secured at Mount Wilson and converted them to actual orbital velocities. With the aid of the known period, he then found the average radius (semi-major axis) of the orbit, a. Since a is also known in seconds of arc, the parallax is found directly. Again, results in good agreement with the trigonometric measurements are found. Occasionally discrepancies are found as in the Pleiades, but such discordances sometimes turn out to yield other dividends such as clues to stellar evolution.

To summarize, stellar distances are accurately known for nearby objects. Stars in remoter parts of the galactic systems do not yield such satisfactory values for their distances.

Finally, it should be mentioned that the distance, or more accurately speaking the intrinsic brightness of a star whose apparent brightness can be readily measured, can sometimes be found from its spectrum. Such spectroscopic parallaxes are becoming increasingly important in astronomy, but they have to be calibrated by measuring the distances of the prototype of each spectroscopic species of stars by some method such as those described in the preceding sections.

OBSERVABLE DATA FOR STARS

Nomenclature.—The earliest designation of the stars still in common use are the Arabic names. These follow no particular system, are often fanciful in character and are rarely used except by navigators and amateurs. The brighter stars in each constellation are now denoted by a system introduced in 1603 by Johann Bayer. The brightest star in the constellation is denoted as α, the next brightest as β, etc. When stars are of comparable brightness, as in the Big Dipper, they are lettered according to their position in the constellation. Later John Flamsteed introduced a system of numbers assigned in order of right ascension.

Stars too faint to be given a Bayer letter or Flamsteed number are listed in star catalogues, the most extensive of which is the *Bonner Durchmusterung* (B.D.). The first B.D. catalogue prepared by Argelander gives the positions and rough brightnesses of more than 324,000 stars north of −2° declination. Schönfeld extended the catalogue to 23° south and added 133,000 stars while the rest of the southern hemisphere was covered by the *Cordoba Durchmusterung* (C.D.) and the *Cape Photographic Durchmusterung* (C.P.D.). Star charts are provided for each catalogue so that the identification of a given star is not difficult.

A somewhat less extensive catalogue but one that gives accurate positions for about 30,000 stars is the Boss *General Catalogue*. The stars are listed according to their positions for 1950. Perhaps the most frequently used catalogue is the *Henry Draper Catalogue* (H.D.), which gives the brightnesses, spectral classes and rough positions for 1900 for 225,000 stars (*Harvard Observatory Annals*, vol. 91–99). Extensions have been published for certain regions in the sky. The roughly 6,000 naked-eye stars plus many telescopic stars are included in the H.D. catalogue.

For yet fainter stars, *e.g.*, objects 100 times as faint as the dimmest star that can be seen with the unaided eye, the *Astrographic Catalogue* or *Carte du Ciel* can be used, which is, however, not yet complete for all parts of the sky.

Star charts or photographs of large regions of the sky have also been published. The most extensive series of these charts

is the National Geographic society–Palomar survey carried out with the 48-in. Schmidt camera. This survey covers only the part of the sky observable from Palomar mountain but includes photographs taken in both the blue spectral region and the red spectral region with an appropriate filter. Hence, strongly coloured stars and gaseous nebulae can be readily identified. More within resources of amateurs are the B.D. charts or the Skelnate Pleso charts.

Stellar Positions.—Accurate measurements of stellar position play a basic role in many problems of modern astronomy. The positions of the brighter stars in the equatorial co-ordinate system (right ascension α and declination δ and given for some epoch) are measured with a meridian circle. Fainter stars are measured upon photographic plates with respect to these brighter stars and finally the entire group is referred to the system of extragalactic nebulae. The reason for this last step is that these external galaxies define an essentially fixed or immovable system, whereas both the bright and faint stars are affected by large-scale stellar motion connected with galactic rotation.

For problems concerned with the structure of the galaxy, the positions are often given in terms of galactic co-ordinates, galactic longitude, l, and galactic latitude, b, in which the plane of the Milky Way is the analogue of the plane of the equator.

Stellar Motions.—From accurate measurements of stellar positions one can obtain the proper motions whose fundamental role in the determination of stellar distances has already been indicated. Special efforts have been made to detect stars of large proper motion since these tend to be nearby, often intrinsically very faint stars. Parallax programs are often based on lists of stars of large proper motion. Although proper motions can be measured for very faint stars, radial velocities can be determined only for stars whose spectra can be photographed with a slit spectrograph. This limits our measurements to fairly bright stars. (*See* SPECTROHELIOGRAPH.)

Brightness of the Stars.—The brightnesses of the stars are usually expressed by means of their magnitudes, a usage which has come down from classical antiquity. A star of the 1st magnitude is 2.5 times as bright as one of the 2nd magnitude, which in turn is 2.5 times as bright as one of the 3rd magnitude, etc. A star of the 1st magnitude is 100 times as bright as one of the 6th magnitude. Sirius, apparently the brightest star in the sky is -1.6. Canopus the second brightest is -0.9, while the faintest star that can normally be seen without telescopic aid is of the 6th magnitude. Stars as faint as the 23rd magnitude have been measured with the Hale 200-in. telescope.

The scale of magnitudes constitutes a geometrical progression. To convert from magnitudes to light ratios the formula

$$\log \frac{l_m}{l_n} = 0.4\,(n - m) \qquad (5)$$

is used where l_n and l_m are the brightnesses of stars of magnitudes n and m, respectively. To the extent that brightness ratios can be regarded as power ratios we can think of one magnitude as corresponding to four decibels. Note that magnitudes are actually defined in terms of observed brightnesses and this quantity will depend on the light detecting device employed. Thus we may speak of visual magnitudes as measured with the eye, photographic magnitudes as measured with a photographic plate, and also of various types of photoelectric magnitudes.

So far the magnitude scale has been defined; the zero point of the scale also needs to be given. This point is fixed arbitrarily with regard to the star catalogues of Ptolemy and Hipparchus.

The brightnesses which are actually measured give the apparent magnitudes. These cannot be converted to intrinsic brightnesses until the distances of the objects concerned are known. The absolute magnitude of a star is defined as the magnitude it would have if it were placed at the standard distance of 10 parsecs, or 32.6 light-years (2,062,650 astronomical units). The apparent visual magnitude of the sun is -26.9, that of the full moon -12.7 and the brightest star is -1.6. The absolute magnitude of the sun corresponds to a diminution in brightness by a factor $(2{,}062{,}650)^2$ and is therefore $-26.9 + 2.5 \times \log\ (2{,}062{,}650)^2 = -26.9 + 31.6 = 4.7$. The corresponding absolute magnitude of the full moon would be about $+32$ and would require a telescope with an aperture of about a quarter-mile. The brightest stars such as Rigel (β Orionis) have absolute magnitudes of -8 or even -9. The faintest known star, the companion to B.D. $+4°4048$, which was discovered by G. van Biesbroeck, has an absolute visual magnitude 18.1 which makes it 200,000 times fainter than the sun. A number of such stars may exist in space, but they are extremely difficult to detect. Included among the faint stars are the white dwarflike objects that are extremely dense stars near the end of their evolution.

Stellar Colours.—Even the casual observer knows that all stars are not of the same colour. Consider the constellation of Orion, for example. Most of the naked-eye stars are blue-white in colour, but Betelgeuse (α Orionis) is a deep red. As seen in the telescope, Albireo (β Cygni) consists of two stars: one blue, the other orange. Some quantitative means of measuring stellar colours is needed. One method is to compare the visual magnitude of the star with a magnitude measured with a photographic plate. The photographic plate is insensitive to yellow and red light but sensitive to blue and violet, so that hot blue stars appear bright, and yellow and red stars are faint. The scale of photographic and all other magnitudes is the same as for visual magnitudes, *i.e.*, one magnitude step corresponds to a brightness ratio of 2.5 but the zero point is adjusted so that white stars with surface temperatures of about 10,000° K. have the same visual and photographic magnitudes. A blue star with a temperature of 20,000° K. is then brighter photographically than visually by about half of a magnitude, whereas a cooler star like the sun ($T = 5{,}700$° K.) is about half a magnitude fainter photographically than visually. The colour index is defined as the photographic magnitude minus the visual magnitude. The colour index of the sun is $M_p(\odot) - M_v(\odot) = +5.26 - 4.73 = 0.53$.

Sometimes visual magnitudes are determined photographically with the aid of orthochromatic or yellow-sensitive plates and yellow filters. Such magnitudes are referred to as photovisual magnitudes and differ from visual magnitudes less than visual magnitudes (as found by different observers) differ among themselves.

As a fundamental light or colour measuring device, the photographic plate has certain disadvantages. In particular, the amount of blackening of the emulsion is not proportional to the intensity of the light (or to the exposure time). The connection between blackening and intensity can be calibrated but it varies with the colour of the light, development and exposure time. Fundamental measurements of stellar magnitudes and colours are now made by photoelectric photometry although the photographic plate continues to play a vital role as an interpolation device. For example, it may be necessary to measure the magnitudes and colours of hundreds of stars in a star cluster. Such a task would be prohibitive if one had to observe each individual star with a photoelectric photometer. Instead, the observer selects a sequence of 20 or 30 stars covering the entire range of magnitude and colour in the cluster, measures them carefully with the photoelectric photometer and compares with them all the other cluster stars recorded on the photographs. (*See* PHOTOMETRY, CELESTIAL: *Photoelectric Photometry;* PHOTOGRAPHY, CELESTIAL.)

The photoelectric photometer has several great advantages. First, it permits measurements to be made with an accuracy limited only by the atmosphere. Second, the relation between the number of electrons ejected from the photocathode and the light intensity is a strictly linear one. Stellar colours can be measured very quickly and conveniently by placing appropriate filters in front of the cell.

The older system of stellar magnitudes employed only photographic and photovisual (or visual) magnitudes. It encountered a number of severe difficulties. Only two of them will be mentioned here. The first grows out of the fact that most stars do not radiate like black bodies. In particular, stars with temperatures in the range 8,000° K. to 20,000° K. have a very strong drop in their energy outputs near λ3,700 Å and continuing to the ultra-

violet. When the magnitudes of these stars were determined with photographic lenses, this region was cut out by the flint glass components of the achromatic objectives. But when the same stars were observed with reflecting telescopes with aluminized mirrors, radiation from the ultraviolet region was included, with the result that photographic magnitudes highly discordant with respect to previous determinations were obtained. A second difficulty is that when only one colour is observed ($M_p - M_v$), it is impossible to distinguish between the effects of reddening due to space absorption and reddening due to a low temperature. That is, if a star is measured and its colour is found to be red, it is not possible to say whether the star is red because it is cool or whether it is really a hot star whose colour has been reddened by the absorption and reddening of light in space. By measuring the magnitudes of the same stars in three colours it is possible to overcome these difficulties. W. Becker used the methods of photographic photometry to establish the amount of interstellar reddening.

J. Stebbins and A. E. Whitford used a photoelectric photometer with colour filters to measure six colours extending from the infrared to the limit of transmission of the earth's atmosphere. The most satisfactory system of stellar magnitudes and colours, however, is the UBV photoelectric system of Harold Johnson and W. W. Morgan. The ultraviolet or "U" filter has a broad band pass centred near 0.35 μ (3,500 Å). It transmits the region of the spectrum beyond the big intensity drop at λ3,700 in the moderately hot stars. The blue or "B" filter has a broad maximum centred near λ4,200 with cutoffs in both the ultraviolet and green. It renders magnitudes roughly similar to (but not identical with) the old system of photographic magnitude. The visual or "V" filter transmits over a broad band pass from about 0.49 μ to 0.65 μ with a maximum near 0.55 μ. It yields essentially visual magnitudes. Now the Johnson-Morgan system of colours and magnitudes is precisely defined in terms of stars of accurately known spectral classes and intrinsic luminosities; *i.e.*, surface temperatures and surface gravities. Furthermore, the stars cover a sufficiently wide range in colour, absolute magnitude and distribution over the sky to permit the observer to use them most effectively.

Stellar Spectra.—Probably the largest amount of information that is attainable for a single star is obtained from its spectrum. The spectrum of a star gives at once its temperature and information about its chemical composition. Stellar spectrograms secured with a slit spectrograph on a sufficient scale reveal also whether the star is a member of a close binary system, whether it is a highly luminous supergiant, a moderately bright giant or a dwarf star like the sun.

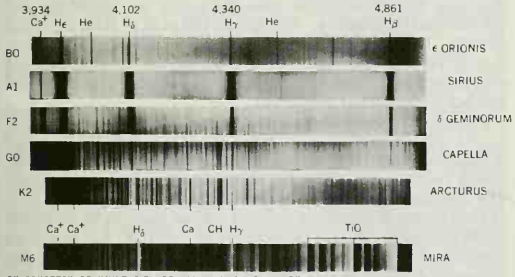


FIG. 5.—REPRESENTATIVE SPECTRAL TYPES

Principal Draper classes are represented. Note the great strength of the hydrogen lines in spectral class A1, (Sirius), the helium lines in B0 (ε Orionis), the metallic lines in G0 (Capella) and K2 (Arcturus) and the molecular bands in the long period variable, M6 (Mira)

Such spectrograms may permit a determination of the quantitative chemical composition of the star. They will also tell if the star is in rapid rotation, if it has an extended atmosphere and if it has a strong magnetic field.

(*See* SPECTROSCOPY, ASTRONOMICAL.)

The vast multitude of stars fall in a small number of spectral classes. The system of spectral classification embodied in the *Henry Draper Catalogue* embraces the following classes (ranging from the hottest to the coolest stars): O B A F G K M. (*See* fig. 5.) This group is supplemented by the R and N stars (sometimes referred to as the carbon C stars) and the S stars. The R, N and S stars differ from the others in chemical composition; they are invariably giant or supergiant stars and include no dwarfs.

The spectral sequence, O-M, represents a group of stars of essentially the same chemical composition but of differing temperatures and atmospheric pressures. This simple interpretation was put forward in the early 1920s by Meg Nad Saha and has provided the basis for future progress. The spectral sequence is also a colour sequence; the O and B stars are intrinsically the bluest and hottest; the M, R, N and S stars are the reddest and coolest. When the effects of space reddening are eliminated and stars of a given kind; *e.g.*, dwarf stars are considered, there is a one-to-one correspondence between colour and spectral class.

Consider first the cool stars of class M. They exhibit spectral lines of familiar metals such as iron, calcium, magnesium, etc., and also (particularly in the red and green part of the spectrum) strong bands with sharp edges on the red side caused by titanium oxide. In the cooler stars of class M these bands become very strong and smother the continuous spectrum of the star. In spectral class K, represented in fig. 5 by Arcturus, the titanium oxide (TiO) bands disappear, and the spectrum exhibits a wealth of lines of the metals. A few especially refractory fragments of molecules such as cyanogen (CN), the hydroxyl radical (OH) and CH persist in these stars and even in G stars such as the sun. The spectra of the G stars are dominated by the metallic lines, particularly those of iron, calcium, sodium, magnesium and titanium. The λ4,227 line of calcium (produced in the flame or bunsen burner) is much weaker than in the M stars, but the λ3,968 and λ3,933 (H and K) lines of ionized calcium (Ca^+) are much stronger than any other feature. This behaviour of calcium illustrates the phenomenon of thermal ionization. At low temperatures the calcium atom retains all its electrons and radiates the spectrum characteristic of the neutral or normal atom. If the temperature is raised, collisions between atoms and electrons and the absorption of radiation tend to detach electrons and produce ionized Ca^+ atoms. At the same time ions and electrons may recombine to produce neutral calcium atoms. The situation may be expressed by the reversible reaction:

$$Ca \leftrightarrows Ca^+ + \epsilon \qquad (6)$$

If the temperature is high or the electron pressure low, or both, the atoms tend to be mostly ionized. At low temperatures and high densities the equilibrium favours the neutral state. The concentrations of ions and neutral atoms can be computed from the temperature, density and the ionization potential (the energy required to detach an electron from the atom; it is usually expressed in electron volts). In the stars of spectral class F, the lines of the neutral atoms become relatively weak in comparison with those of the ionized atoms. The hydrogen lines become stronger, attaining their maximum intensities in spectral class A, where the temperature is around 9,000° K. Thereafter they gradually fade as hydrogen becomes ionized.

Spectral lines are produced by transitions between energy levels. The yellow D lines of sodium or the H and K lines of ionized calcium are produced by jumps from the lowest energy levels of these atoms. The observable hydrogen lines, however, are produced by jumps from atoms in the second energy level, which lies far above the ground level. Only at high temperatures are sufficient numbers of atoms maintained in this level by collisions, radiations, etc., to permit an appreciable number of absorptions to occur. At the low temperatures of the M or even of the K stars there are few atoms in the second level. On the other hand, at the very high temperatures of the B and O stars, the hydrogen atoms are mostly ionized and therefore cannot absorb any radiation.

Another feature to be noticed is that the lines of the ionized metals, such as iron, in the F stars never attain the strength that the neutral lines do in the K stars. This effect is partly due to the

circumstance that these lines involve higher energy levels which are more sparsely populated than the lower levels. Another factor, however, is that the general fogginess or opacity of the atmospheres of these hotter stars becomes greatly increased. For example, above each square inch of the visible surface of the sun there is about 20 times as much material as there is above each square inch of the surface of Sirius, an A1 star.

The continuous (as distinct from the discrete or line) spectrum of the sun is produced primarily by the photodissociation of the negative hydrogen ion. In the solar atmosphere a small number of hydrogen atoms pick up extra electrons to form negative ions. When these are subsequently destroyed by photodissociation they produce a continuous absorption of radiation. In the hotter stars the principal source of absorption is the photoelectric effect on hydrogen atoms.

Photoelectric absorption by normal hydrogen atoms is confined to the remote ultraviolet, but if the hydrogen atoms can become excited to their second or third energy levels, the photoionization energies then required are much lower and the continuous absorptions fall in the normally observable range. Other sources of opacity exist but are relatively minor except in the coolest stars with dense band absorption.

The hotter stars of spectral class B, *e.g.*, ϵ Orionis, are characterized by lines of helium, singly ionized oxygen, nitrogen and neon. In the O stars, the lines of ionized helium appear. Other prominent lines include those of doubly ionized nitrogen, oxygen and carbon and trebly ionized silicon. In the hottest stars of spectral class O, the lines of hydrogen and helium are still visible.

The physical processes behind the formation of stellar spectra are well enough understood to permit determinations of temperatures, densities and quantitative chemical compositions of stellar atmospheres. The star studied in greatest detail is, of course, the sun, but other stars also have been investigated in detail.

Although the general features of the spectral sequence can be understood in terms of a variation of temperature, the general features are also spectral features that depend on the densities of their atmospheric gases. These densities depend in turn on the surface gravities. Stars of high surface gravities (dwarfs) tend to have high densities; those of low surface gravities, giants and supergiants, have relatively low surface densities. The hydrogen lines provide an illustration in point. Normally, an undisturbed atom radiates a very narrow line. If its energy levels are disturbed by charged particles passing nearby it will no longer radiate at its characteristic frequency ν_0 but at some nearby frequency $\nu_1 = \nu_0 + \Delta\nu_0$. In a hot gas, charged particles will pass radiating atoms in a random fashion and at different distances, for each of which there is a corresponding value of the frequency shift $\Delta\nu$. The spectrum line radiated by the whole mass of gas accordingly will be blurred out, the amount of the blurring depending on the density of the gas in a known fashion. The dwarf stars of spectral class A such as Sirius show broad hydrogen lines with extensive "wings" whereas supergiant A stars such as Deneb (α Cygni) show relatively narrow hydrogen lines. Another example is the λ4.227 line of neutral calcium which shows broad wings in dwarf M stars and is very narrow in supergiants. The ionization of the calcium in the supergiant is greater than in the dwarf, but the principal effect arises from density broadening. Luminosity effects in stellar spectra have to be established empirically. The procedure is (1) to observe a group of stars of different but known absolute luminosities; (2) to select among these stars spectral lines that are sensitive to luminosity but are not used for assigning the Draper class. Molecular bands (such as those of CN) as well as atomic lines are used for absolute magnitude criteria. Much effort has been expended on this problem by E. Hertzsprung, W. S. Adams and A. Kohlschütter, A. H. Joy, B. Lindblad and his associates in Sweden, and many others. In the modern system of spectral classification introduced by W. W. Morgan and P. C. Keenan, the luminosity class of the star is assigned along with the Draper class. Thus α Persei is classed as F5Ia, which means that it falls about halfway between the beginning of class F (FO) and of class G (GO). The Ia means that it is a particularly luminous supergiant with an absolute visual magnitude of the order of -7. The star π Cephei G2III is a giant falling between GO and KO but much closer to GO. The sun is classified as G2V. A star of luminosity class II falls between supergiants and giants; one of class IV is a subgiant.

Positions, magnitudes and colours can be measured for vast numbers of stars. Spectrograms can be secured for certain of the brighter stars. If the star is close enough, its trigonometric parallax can be measured. If it is a member of a visual binary system whose orbit and parallax are known, the mass can be found. The orbit of each component must be known with respect to the centre of gravity in order to determine the individual masses. The masses of components of eclipsing binaries can also be found if both spectroscopic and photometric data are available.

Stellar Diameters.—Eclipsing binary systems provide the most extensive data on stellar dimensions. Fortunately, eclipsing systems include not only dwarf stars such as the sun and Sirius but also supergiant stars such as 31 Cygni and VV Cephei. The angular diameters of bright red giant and supergiant stars were measured in the 1920s with the Michelson stellar interferometer. The principle upon which this instrument is based goes back to the interference pattern produced by two slits when they are illuminated with light from a single narrow slit source. In the stellar interferometer the two slits are replaced by two mirrors. The disk of the star can be regarded as being equivalent to two narrow sources placed side by side. Each source produces its own interference pattern; the observer sees the patterns superposed on one another. By varying the distance between the two movable mirrors the two interference patterns from the two sides of the star can be made to cancel one another. From the known separation of the mirrors and the effective wave length of the light, the angular diameter of the stars can be computed. Unfortunately, the instrument is sensitive to atmospheric seeing and only the largest stars can be measured by this method. R. Q. Twiss and Hanbury Brown developed another type of interferometer that depends on photon correlations. With it they measured the angular diameter of Sirius.

Finally, if the absolute magnitude of a star and its temperature are known, its size can be computed. The temperature determines the rate at which energy is emitted by each unit of area, and the total luminosity gives the total power output. Thus the surface area of the star and its diameter may be estimated. This method is actually the only one available for estimating the dimensions of white dwarf stars. The chief uncertainty is often in choosing the temperature that represents the rate of energy emission.

True Stellar Luminosities; Bolometric Magnitudes.—In theoretical studies of stellar structure and evolution the quantities that must be known are the mass, the radius, the luminosity and the chemical composition. The masses and radii of stars are found directly. The chemical compositions are found only for the surface layers and may differ from those of the interior, at least with respect to the hydrogen-helium ratio. Hydrogen is converted into helium by nuclear processes.

Consider the determination of the true luminosity of the star; *i.e.*, its actual energy output. When the energy radiated by a star is observed, only that portion to which the energy detection device is sensitive and which the atmosphere transmits is obtained. The luminous efficiency for stars like the sun is very high; most of their energy is emitted in spectral regions that can be observed. An M star with a surface temperature of 3,000° K. has an energy maximum (on a wave length scale) at 10,000 Å; *i.e.*, far in the infrared. A cooler M star radiates most of its energy in the 1.5–3 μ region. Only a small portion of the actual stellar spectrum is normally observed. Bright cool stars can be observed, however, with the thermocouple. If two dissimilar metals are fastened together to form a continuous circuit and one junction is heated while the other junction remains cool, a small current will flow. The amount of the current will depend on the temperature rise and on the nature of the metals employed in a known fashion. As used at the telescope, one junction of the thermocouple is actually heated by the radiation of the star; hence the sizes of the thermoelements must be very small. Nevertheless, E. Pettit and S. B. Nicholson succeeded in measuring the energy output of

certain cool, bright stars such as Betelgeuse, Antares and the long period variables χ Cygni and Mira (o Ceti). Corrections for the heavy absorption by molecular bands of water vapour and other constituents of the earth's atmosphere have to be determined and applied. Magnitudes measured with a thermocouple are referred to as radiometric magnitudes.

The problem posed by the hottest stars is more difficult. Whereas there are numerous "windows" in the infrared, the transparency of the earth's atmosphere in the ultraviolet diminishes to zero at 2,900 Å (due to the extinction by ozone). A star whose surface temperature is 20,000° K. or above will radiate most of its energy in the inaccessible ultraviolet. Furthermore, the shape of the energy distribution may deviate widely from that of a black body. Continuous absorption caused by the photoionization of the hydrogen atom from its normal (lowest) level reduces the emergent energy flux to a very low value in the region beyond about 920 Å. On the redward (longward) side of this limit the atmospheric gases have a high transparency so that energy can flow outward relatively easily. The only defensible procedure for calculating this energy distribution is to construct a theoretical model of the star's atmosphere that will reproduce the features of the spectrum that can be observed, particularly the energy distribution in the observable range. If a satisfactory fit can be obtained for the observable part of the spectrum, some estimate of the energy outside the observable range can be made.

Magnitudes that refer to the total energy radiated by the star, rather than simply the energy emitted over a narrow wave length range, are called bolometric magnitudes. The bolometric corrections required to reduce visual magnitudes to bolometric magnitudes are thus large for very cool and for very hot stars but are relatively small for stars such as the sun.

DOUBLE STARS

Binary stars are considered in three categories, depending on the mode of observation employed. Visual binaries are objects that can be seen as double stars in the telescope. True doubles, as distinguished from optical doubles, display a common space motion (as revealed by proper motions and radial velocities) and sometimes a common orbital motion as well. Spectroscopic binaries are found from radial velocity observations. A component of such a binary shows a continuously changing velocity with a fixed period; the velocity curve repeats itself exactly from one cycle to the next, and the motion can be interpreted as orbital motion in a close binary. If two stars of approximately equal brightness are observed, as in Mizar, two superposed spectra are found. The lines caused by the individual stars indicate velocity variations that can be interpreted as orbital motion. Eclipsing binaries are detected from the light variations they show when one star passes in front of the other. The best-known example is Algol (β Persei), and good data are known for many such systems.

Visual Binaries.—The earliest observers of double stars tended to regard them as optical pairs; *i.e.* as two independent stars seen along the same line of sight. W. Herschel measured a number of these pairs hoping to find evidence for parallactic displacements caused by the earth's orbital motion. Instead, he found displacements in several of these systems indicating a true orbital motion of the stars themselves. In the 19th century a great deal of effort was expended in discovering and cataloguing binary stars. R. G. Aitken (1927) listed 17,180 pairs north of −30°. E. Rossiter added 7,350 more, while the catalogue of Van den Bos listed 16,000 pairs south of −19°. The surveys appear to be complete to the 9th magnitude and include a total of about 40,000 objects. Most of these objects show little or no orbital motion, but a few are interesting systems. Visual binaries must be relatively near the sun in order to be detected, whereas spectroscopic binaries and particularly eclipsing binaries can be found at great distances.

The importance of visual binary systems lies in the fact that they permit the determination of stellar masses when the parallax is known. If the period, P, of the binary system is expressed in years and the radius of the orbit, a (more accurately its semimajor axis), is expressed in astronomical units, the sum of the masses are given by Kepler's law:

$$m_1 + m_2 = \frac{a^3}{P^2} = \frac{a''^3}{p''^3 P^2} \qquad (7)$$

where p'' is the parallax in seconds of arc and a'' is the semimajor axis in the same units. For example, for the well-known binary 70 Ophiuchi, $P = 87.85$ years, $p'' = 0''.199$, $a'' = 4''.55$, so that $a = 22.8$ astronomical units, and $m_1 + m_2 = 1.55$ solar masses. Furthermore, from a measurement of the motions of the two components with respect to the background stars, K. Strand and R. Spong were able to determine the orbit of each star with respect to the common centre of gravity. They thus found a mass ratio of 0.422 and assigned individual masses of 0.90 and 0.65 solar masses. (*See* ORBIT: *Kepler's Laws.*)

The faintest known binary for which accurate masses are available is Krüger 60. The two stars, spectral classes M4 and M6, are separated by 9.5 astronomical units, have a period of 44.6 years and have masses 0.27 and 0.10 that of the sun. The binary with the shortest known period (4.56 years) is H.D. 9,770, for which Dawson finds a mean separation of 4.7 astronomical units and a combined mass of about 5 solar masses.

A number of multiple systems are known. These consist of a close spectroscopic binary and a more distant visual companion. The Mizar-Alcor system is the classical example. Mizar and Alcor (the naked-eye pair) are so far apart that they show no orbital motion, only a common motion through space. Mizar is a visual double, but each component is itself a spectroscopic binary.

In some respects 61 Cygni is the most interesting visual binary system. It was the first star whose distance was measured and the first stellar system found to contain a planet. In the telescope it is a double star consisting of two K-type components separated by 837 astronomical units and moving about one another with a period of 720 years. One of these stars has an invisible companion with a mass $\frac{1}{60}$ that of the sun which moves about it with a period of about five years. According to all accepted criteria this companion must be regarded as a bona fide planet. Careful examination of the motions of other stars in the neighbourhood of the sun has suggested that other planetary companions may exist. For example, the single star called Ross 614, distant some 13 light-years from the sun, has an invisible companion of $\frac{1}{12}$ the sun's mass.

Spectroscopic Binaries.—J. H. Moore and F. J. Neubauer (1948) listed 480 systems for which orbital elements had been determined. Unfortunately, it is not possible to find the semimajor axes or the masses for most spectroscopic binaries since the angle between the plane of the stellar orbit and the plane of the sky cannot be determined. If two spectra can be observed, mass ratios can be determined. If only one spectrum can be observed the best that can be done is to find a quantity called the mass function which enables one to calculate a lower limit to the stellar masses. If the spectroscopic binary is also observed to be an eclipsing system, the inclination and often the masses of the individual stars can be found.

Eclipsing Binaries.—The most important type of double star is the eclipsing binary, which consists of two close stars moving in an orbit so oriented in space that the light of one may be occulted or eclipsed by the other. Depending on the orientation of the orbit and the sizes of the stars, the eclipses may be total and annular or both eclipses may be partial. The best-known example of eclipsing binaries is the famous β Persei or Algol, which has a period of 2.87 days (interval between eclipses). The brighter star (of spectral class B8) contributes about 92% of the light of the system and the eclipsed star much less than 8%. The system contains a third star that is not eclipsed. There are about 17 eclipsing binary systems visible to the naked eye.

The basic observation datum for an eclipsing binary is the light curve. It displays magnitude measurements of the system over the length of a complete light cycle. Usually the measurements are made by comparing the light of the variable star with that of a nearby comparison star of presumably fixed brightness. Consider fig. 6(A). The deep or primary minimum is produced when the star of higher surface brightness is eclipsed. It represents the total eclipse and is characterized by a flat bottom. The

shallower secondary eclipse occurs when the star of higher surface brightness passes in front of the other; it corresponds to an annular eclipse (or transit). In a partial eclipse (fig. 6[B]) neither star is ever completely hidden and the light changes continuously during an eclipse; there are no flat bottomed minima.

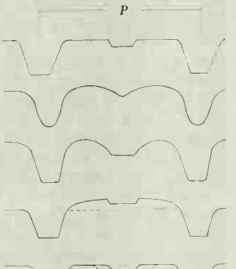


FIG. 6.—SCHEMATIC LIGHT CURVES OF ECLIPSING BINARIES

An analysis of the shape of the light curve during an eclipse, following the methods of H. N. Russell, gives the ratio of the radii of the two stars, r_1 r_2 and also r_1 in terms of the size of the orbit, the ratio of the luminosities l_1 l_2, and the inclination of the plane of the orbit to the plane of the sky.

If radial velocity curves are available, *i.e.*, if the system can be observed as a spectroscopic binary as well as an eclipsing binary, additional information is found. In the fortunate circumstances where both velocity curves are observable, the absolute dimensions of the orbit and the sizes, masses and densities of the stars can be found. Furthermore, if the parallax of the system can be found the temperatures of the individual stars may be estimated from their luminosities and diameters. These procedures can be carried out for the faint binary Castor C and the bright B-star μ Scorpii.

Light curves of the types as shown in figs. 6(A) and 6(B) usually occur in systems where the companions are moderately distant from one another. If the stars are close to one another, additional effects occur that cause the light to vary between eclipses. One of these effects is the tidal distortion produced by the gravitational attraction of one star upon the other. The stars become elongated along the line joining them. Hence, midway between eclipses they appear broadside to the observer and the light of the system is a maximum. (*See* fig. 6[C].)

The reflection effect is exhibited in fig. 6(D). Suppose that a small, high temperature star is paired with a larger object of low surface brightness. If the distance between the stars is small, the hemisphere of the cool star presented toward the hotter one will be substantially illuminated. Just before (and just after) secondary eclipse this illuminated hemisphere is pointed toward the observer, and the total light of the system is a maximum. The reflection effect may be large for two stars of greatly differing surface temperature. It vanishes if both stars are of the same surface temperature.

If a light curve is affected by reflection effects and tidal distortion as exhibited in figs. 6(C) and 6(D), it is necessary to correct these variations before the light curve can be analyzed. Such a process is called rectification. If the stars are oblate in shape due to rapid rotation or if the surface brightness is altered as a consequence of the variation of gravity over the surface, it is necessary to allow for these effects in the interpretation of the light curve. Eclipses of oblate stars are necessarily more troublesome to handle than those of spherical stars.

If the disk of the sun is photographed with a high-speed camera, the centre is found to be brighter than the limb. The difference in brightness depends on the wave length of the light used. It is large in the violet, small in the red and almost vanishes in the infrared. This limb darkening arises from the fact that the sun gets hotter toward its core. At the centre of the disk radiation is received from deeper and hotter layers (on the average) than from the limb. A similar situation holds for the stars; for any given colour of the spectrum the degree of limb darkening will depend on the temperature of the star. Limb darkening must be taken into account in the analysis of the light curve. If it is neglected, the calculated size of the eclipsed star will be too small.

Modern light curves obtained with the photoelectric cell are sometimes so accurate that the tiny differences produced by different degrees of limb darkening can be found and the correct law of darkening as well as the stellar sizes accurately established.

One other effect on the light curves must be mentioned. Some stars move about one another in elliptical orbits. If the major axis of the orbit is at right angles to the line drawn to the observer the stars will take a different length of time to go from primary minimum to secondary minimum than from secondary minimum to primary minimum because they move faster when close together than when far apart (more precisely, the radius vector of the orbit sweeps out equal areas in equal times). Hence the secondary minimum will not fall precisely at the midpoint between primary eclipses (fig. 6[E]–A). If the system is observed at a later date the secondary minimum may be symmetrically placed, but now the two eclipses are of unequal duration (fig. 6[E]–B). At this time the major axis is pointed toward the observer. One eclipse occurs when the occulting star is moving most rapidly, the other when it is moving most slowly. At a still later date the secondary minimum is displaced to the opposite side from the midpoint (fig. 6[E]–C). Observation of these effects in any eclipsing system demonstrates that the major axis of the orbit is rotating and establishes the rate of rotation. What causes this rotation of the orbit? A third body might do it and possibly sometimes does, but such a body would produce other effects that could be observed; in most instances it must be concluded that only two stars are present.

Two isolated spherical stars attract one another as point masses and move about one another in elliptical orbits whose orientation remains fixed in space. If, however, the stars are distorted by their mutual tidal attractions, the elliptical orbit will slowly rotate, the speed of rotation will depend in a precisely known fashion on their separation, masses, radii and the degree of mass concentration to the centre. The orbit would turn at a maximum possible speed for stars that were homogeneous; it would not rotate at all for stars whose masses were essentially all concentrated to the centre. All the theories of stellar structure predict large mass concentrations toward the centre. (For the sun the central density is about 80 times the mean density.) These predictions can be tested by the eclipsing binaries, and it turns out that in all instances the central condensations are large (as required by theory) but that in some systems the central density may be great. This latter result can be understood in terms of the more modern work on stellar structure.

Thus, as Russell called it, the "royal road of eclipses" provides information on the densities, radii and masses of the stars as well as on their internal structures. It is becoming evident that eclipsing stars are going to provide many useful clues to the problem of stellar evolution.

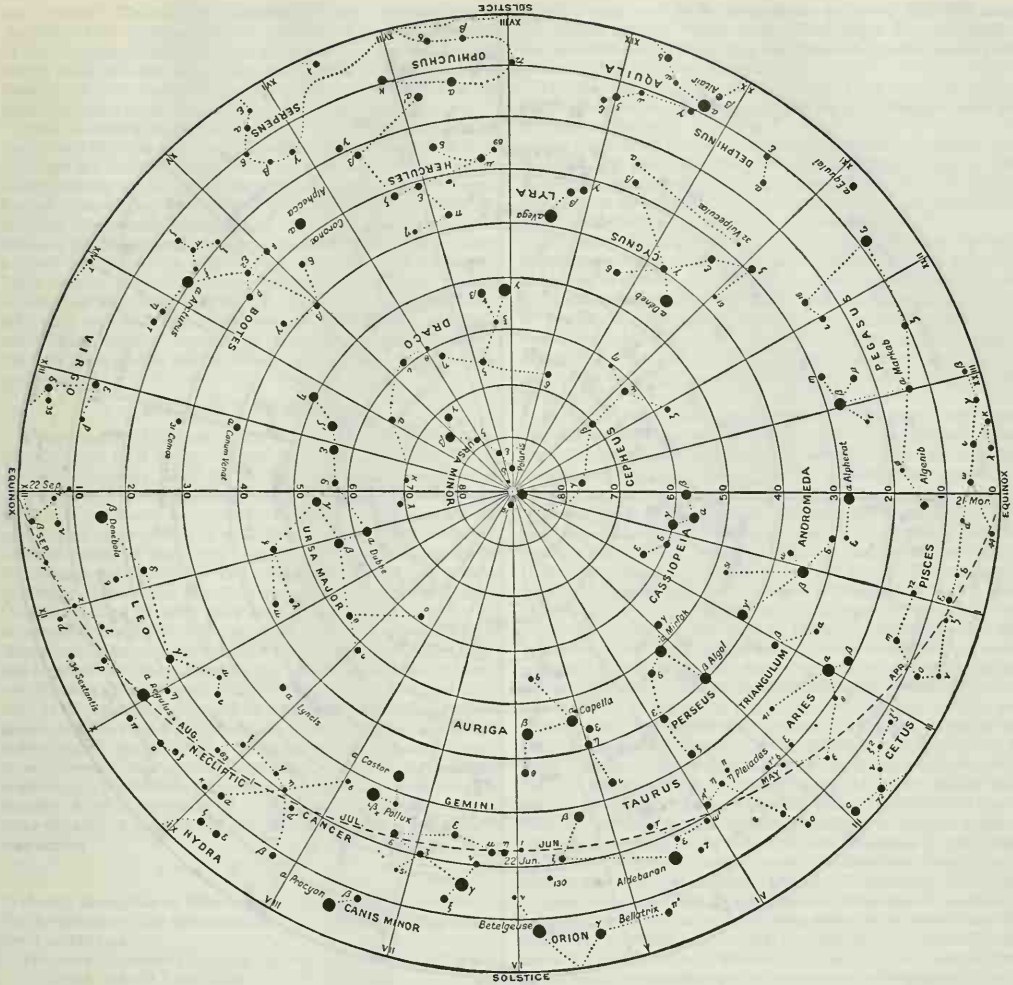
Since the individual stars are often close to one another, extreme examples of tidal distortion, etc., are found. The commonest type of eclipsing binaries, those of the W Ursae Majoris type, are pairs of stars apparently similar to the sun in size but revolving about one another while almost touching.

Although eclipsing binary systems give much detailed data on stars, they do not necessarily give good statistics applicable to isolated single stars. Systems in which a smaller hotter star is accompanied by a larger cooler object are easier to detect than are systems that contain, for example, two main sequence stars. Furthermore, the evolutionary development of two stars near one another does not exactly parallel that of two well-separated or isolated stars.

About 17 or 18 naked-eye eclipsing binaries are known. They include combinations of a variety of stars ranging from objects smaller than the sun (Castor C) to huge supergiants (such as VV Cephei) which would engulf Jupiter and all the inner planets if placed at the position of the sun.

Statistics of Stellar Systems.—If the known data for all binaries are combined, some interesting statistics emerge. In the neighbourhood of the sun most of the stars are members of binary systems, and some of the nearest single stars are suspected of having companions.

Although some binary systems consist of stars separated by hundreds of astronomical units while others are contact binaries, the most probable binary system is built on the same scale as the

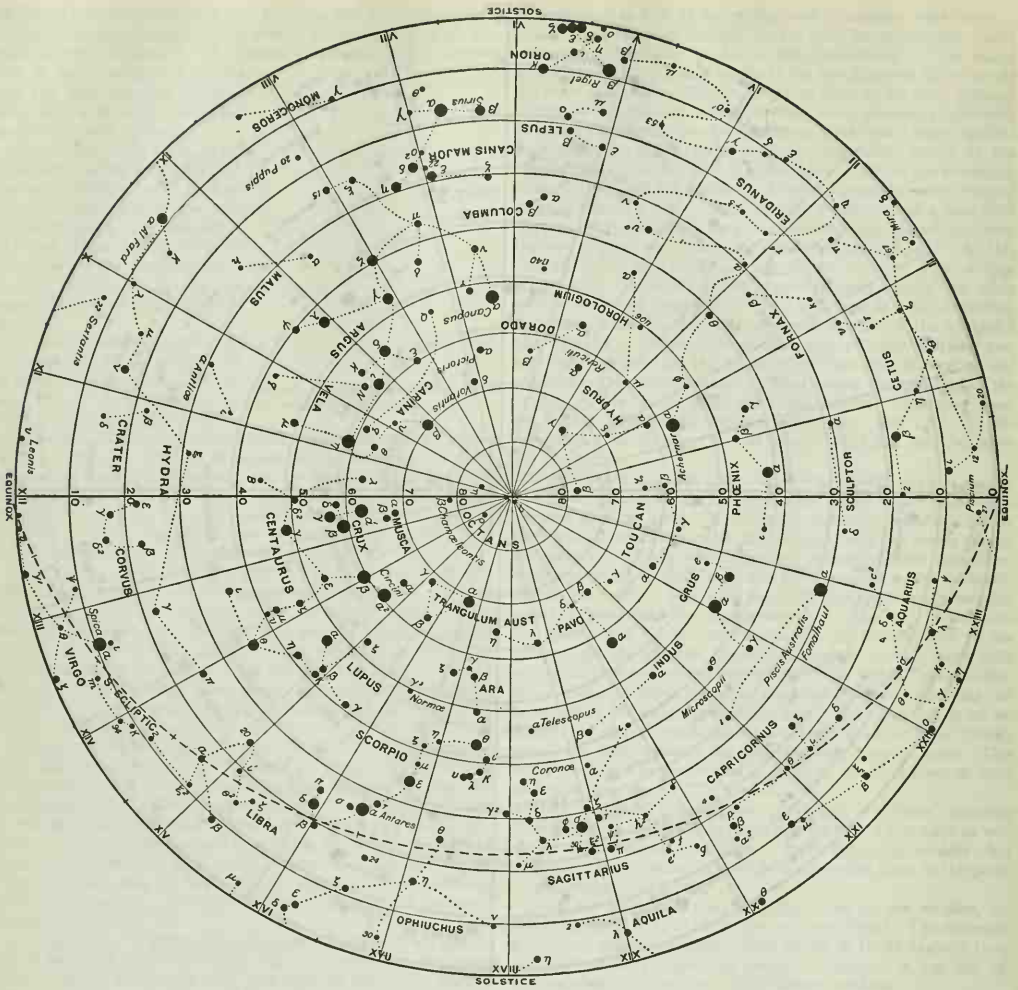


MAGNITUDE: 1● 2● 3● 4●

BY COURTESY OF J. C. CLANCEY, I.S.O., F.R.A.S., FROM "SIMPLIFIED STELLAR MAPS" PUBLISHED BY THE SUPERINTENDENT, GOVERNMENT PRINTING AND STATIONERY, RANGOON

STAR CHART OF NORTH POLAR HEMISPHERE

Most stars visible to the naked eye and north of the celestial equator are shown in this chart. The North Star, Polaris, is shown slightly to the right of the north celestial pole. This star is not exactly true north. At the bottom of the chart is the first magnitude star Betelgeuse in the constellation of Orion. This was the first of the large stars to have its diameter measured. At the left of the north celestial pole is the "Great Bear" with its pointer stars in line with the North Star. In the upper part of the chart is the constellation Hercules toward which the entire solar system is moving at the rate of about 12 miles per second. Owing to the rotation of the earth on its axis, all stars shown in this chart appear to rotate around the north celestial pole in a counter clockwise direction



MAGNITUDE: 1● 2● 3● 4●

BY COURTESY OF J. C. CLANCEY, I.S.O., F.R.A.S., FROM "SIMPLIFIED STELLAR MAPS" PUBLISHED BY THE SUPERINTENDENT, GOVERNMENT PRINTING AND STATIONERY, BANGOOK

STAR CHART OF THE SOUTH POLAR HEMISPHERE

Showing all of the principal stars visible to the naked eye found south of the celestial equator. To the left of the south celestial pole in the chart is the Southern Cross (Crux). Notice the 2 pointers, α and β Centauri, just below Crux; α Centauri is the nearest star to the earth. Below the south celestial pole and slightly to the left is Scorpio which contains the supergiant red star Antares, which is one of the largest stars that has been measured up to the present time. It is visible from practically every point in the inhabited world. To the right of the south celestial pole is Fomalhaut, a first magnitude star in the constellation "The Southern Fish." Sirius, of the constellation "The Big Dog" and the brightest star in the sky, is near the top of the chart. All stars shown in chart appear to revolve around the south celestial pole in a clockwise direction due to the rotation of the earth on its axis

solar system. The average separation between components is of the order of ten astronomical units. Studies of the relative masses of the two components show that there is a tendency for the division in mass between two components of a binary to be nearly at random. A mass ratio of about $\frac{1}{2}\frac{1}{0}$ could occur about 5% of the time, and under these circumstances a solar system might form.

Thus, according to the work of G. P. Kuiper, the formation of double and multiple stars on the one hand and of solar systems on the other are different facets of the same process. A solar system could certainly not be formed by an encounter between two stars. Rather, planets may be produced by a process akin to that depicted in the nebular hypothesis of Laplace except that most of the mass and angular momentum is swept out of the system. Only a small fraction of the original material is retained as planets. Solar systems may be frequent in the earth's galaxy; perhaps 100,000,000 stars have bona fide planets. Among these 100,000,000 solar systems there are perhaps 2,000,000 or 3,000,000 which have planets capable of supporting higher forms of life. Of these planets man is destined probably to remain forever ignorant.

STELLAR VARIABILITY

A fair number of the stars in the sky are intrinsically variable; that is, their total energy output fluctuates with time. Some of these stars were found by accident, but many were detected as a result of carefully planned searches. Variable stars are important in astronomy for a number of reasons. They usually appear to be stars at critical or short-lived phases of their evolution; detailed studies of their light and spectral characteristics, spatial distribution and association with other types of stars may provide important clues to the life histories of various classes of stars. Certain kinds of variable stars such as Cepheids and novae are extremely important in that they make it possible to establish the distances of remote stellar systems such as the Magellanic Clouds and some of the nearer extragalactic nebulae or external galaxies (*see* NEBULA: *The Extragalactic Distance Scale*). The principle is this: If the intrinsic luminosity of a particular kind of variable is known and this kind of variable star can be recognized in a distant stellar system, the distance of the latter can be estimated from a measurement of the apparent luminosity. If l is the apparent brightness, L the intrinsic brightness corresponding to the standard distance, R, of 10 parsecs or 32.6 light-years, $l/L = R^2/r^2$ where r is the distance of the star. If the stellar brightness is expressed in magnitudes

$$y = m - M = 5 \log r - 5 \qquad (8)$$

If there is absorption of light in space, the apparent magnitude, m, for the amount of the extinction must be corrected. (*See* INTERSTELLAR MATTER.)

The most important classes of variable stars are the following:

Cepheids and RR Lyrae stars	SS Cygni stars
Long-period variables	Supernovae
Semiregular variables	R Coronae Borealis stars
Irregular red variables	β Canis Majoris stars
Novae	T Tauri stars
Spectrum and magnetic variables	

The long-period variables, the novae and the Cepheids were the types earliest recognized. The first four classes probably owe their variability to a common cause-pulsation. The next three are sometimes referred to as catastrophic variables, while the remaining four types owe their variability to less dramatic but perhaps just as bizarre causes.

The spectrum and magnetic variables, spectral class A, show small amplitudes of light variation but pronounced spectroscopic changes. Their spectra are characterized by strong lines of heavy metals, particularly rare earths, which vary in intensity in a periodic fashion. H. W. Babcock discovered that these stars have strong magnetic fields, up to 6,000 gauss. Not all magnetic stars are known to be variable in light, but the variable ones seem also to possess variable magnetic fields.

The T Tauri variables are dwarf stars of spectral classes G and K. They invariably fall in regions of high obscuration; *i.e.*, dense clouds of absorbing material. Their changes in brightness are erratic; some such as R Monocerotis are associated with small fan-shaped nebulae. Their spectra show bright lines and washed out absorption lines suggesting that they are rapidly spinning objects in clouds of diffuse material.

Various lines of evidence suggest that these stars are objects either recently formed or in the process of formation from the interstellar medium.

The β Canis Majoris stars are high temperature (spectral class B) stars that often show complicated variations in spectral line shapes and intensities, velocity curves and light. They often have two periods of variation so similar in duration that complicated interference or "beat" phenomena are observed.

R Coronae Borealis variables are giant stars of about the sun's temperature whose atmospheres are characterized by excessive quantities of carbon. Their brightness remains constant until suddenly the star dims by several magnitudes and then slowly recovers its original brightness; the star's colour remains the same during the changes in brightness. The dimmings occur in a random fashion. L. Berman and J. A. O'Keefe have suggested that the clue to the variability was to be sought in the huge concentrations of carbon.

At times the carbon vapour literally condenses into soot, and the star is hidden until the smog blanket is evaporated. Similar veiling may sometimes occur in other stars, particularly the long-period variables, thus providing an engaging problem in chemistry.

The novae or temporary stars are discussed in detail elsewhere (*see* NOVA AND SUPERNOVA) so only a few summarizing remarks need be made here. Normally they are small blue stars much fainter than the sun although very much hotter. When an outburst occurs, the star may brighten very suddenly, ten magnitudes or even more. Thereafter it fades, sometimes slowly and sometimes rapidly. The rate of fading is connected with the brightness of the nova. The brightest novae, which reach absolute magnitudes about -8, fade the most rapidly, whereas a typical slow nova, which reaches an absolute magnitude -5, may take ten times as long to decline in brightness. The changes in light are accompanied by pronounced spectroscopic changes that can be interpreted as arising from an ejected shell that dissipates slowly in space. The mass of the shell is presumably rather small, comparable with the mass of the earth and presumably much smaller than the mass of Jupiter. Presumably only the outer shell of the star is affected; the main mass settles down until a new outburst appears. The existence of repeating novae such as T Coronae Borealis suggests that perhaps all novae repeat at intervals ranging from thousands to perhaps millions of years. B. Kukarkin and R. R. Parenago suggested that the amplitude of the outbursts is probably connected with the intervals between them in the sense that the larger the explosion the longer the interval. Associated with novae are stars such as Z Andromedae which in their normal condition show remarkable spectra combining the characteristics of cool M-type stars, gaseous nebulae and hot stars with bright emissions of hydrogen and helium. They occasionally show novalike outbursts.

Stars of the SS Cygni type show novalike outbursts but with a much smaller amplitude and with intervals between outbursts of a few months to about a year. These novalike stars often are binaries, and it may be that the development of this particular type of star occurs in a close binary system.

A distinct type of phenomenon is provided by the supernovae. These stars become ten or even several hundred million times as bright as the sun and then gradually fade away. Such outbursts occur once every 400 or 500 years in a given galaxy, but three are known to have occurred in our own galaxy—Tycho's star of 1572, Kepler's nova of 1604 and the Crab Nebula. The latter object has been identified with a temporary star observed in China and Japan in 1054 and now observed as a strong radio source and an unusual optical object. Supernovae are believed to be stars that actually explode as a consequence of thermonuclear reactions. (*See* ASTRONOMY: *Radio Astronomy.*)

Stellar pulsations are believed to account for the variability of the Cepheids, the long-period variables, the semiregular vari-

ables and even the irregular red variables. Of this group the Cepheids have been studied in greatest detail, both theoretically and by observation. These stars are regular in their behaviour; some repeat their light curves with great faithfulness from one cycle to the next.

Much confusion existed in the study of Cepheids until it was recognized that there existed two families of stars called population types I and II. The population type I is that associated with the spiral arms of this and other galaxies. It contains stars of all ages, from those in the process of formation to defunct white dwarfs. The best examples of type I populations are galactic or open clusters such as the Pleiades or the beautiful double cluster in Perseus (h and χ Persei). Population type II comprises the stars of the central bulge of our galaxy, elliptical nebulae, globular clusters and most of the stars moving between the spiral arms. These stars are all very ancient; their ages are probably of the order of 5,000,000,000 years.

Corresponding to these population types are separate species of variable stars. The Cepheids belonging to the spiral arm population (type I) are characterized by regularity in their behaviour. They show continuous velocity curves indicative of a regular pulsation. They exhibit a relation between period and luminosity in the sense that the longer the period of the star the greater its intrinsic brightness. This period-luminosity relationship has been used to establish the distance of remote stellar systems. Cepheids are also found in the type II population, *e.g.*, away from the Milky Way in globular clusters, but these type II Cepheids exhibit very different properties from the classical type I Cepheids. They are bluer than classical Cepheids of the same period and their light curves have different shapes. Studies of the light and velocity curves indicate that shells of gas are ejected from the stars as discontinuous layers that later fall back toward the surface. Most disappointing of all, however, is the fact that these stars exhibit no well-defined relation between period and luminosity. Consequently, they are worthless as distance indicators. The distance of a Cepheid in a distant stellar system can be determined only if it is known to be a classical Cepheid.

Closely associated with the type II Cepheids are the cluster-type or RR Lyrae variables. All these stars have periods less than a day, and there is no correlation between period and luminosity. Their absolute magnitudes are about 0; *i.e.*, they are about 100 times as bright as the sun. The most recent work indicates a magnitude of about +0.6. They are useful for determining the distances of star clusters and some of the nearer external galaxies, since their short periods permit them to be detected readily.

The possibility of stellar pulsations was suggested long ago by A. Ritter, direct applications to the Cepheids were made first by H. Shapley, while Sir Arthur Eddington developed the theory necessary for the problem. If all stars have roughly the same kind of internal structure, the period and density are related by the condition

$$P\sqrt{\rho} = \text{constant} \qquad (9)$$

where the value of the constant depends on the internal structure; *i.e.*, the "model" of the star. Since the masses of Cepheids are not known, there is no means of testing this hypothesis.

Furthermore, while the star pulsates, all parts of the main body move in synchronism, but the outer parts get out of step or lag behind the pulsation of the inner core.

What causes the pulsations is not known, and all that can be asserted is that the mechanism that generates the pulsations has a period that exactly matches the free period of oscillation of the stars.

The long-period variables also probably owe their variations to pulsations. Here the situation is complicated by the vast extent of their atmospheres. These stars appear to fall into two groups; (1) those with shorter periods (of the order of 200 days) tend to be associated with the type II population; (2) those of periods of the order of a year belong to the type I population. These stars are invariably red cool giant and supergiant stars of spectral classes M (normal composition) R and N (carbon rich) or S (heavy-metal rich). The range in visual brightness may be a hundredfold, but the range in total energy output is much less because at very low temperatures (1,500° K.–3,000° K.) most of the energy is radiated as heat rather than as light. A slight change in temperature can produce a huge change in the output of visible radiation. At these temperatures, compounds and probably solid particles are formed so that the light change may be profoundly affected by chemical as well as physical factors.

Red semiregular variables such as the RV Tauri stars show complicated light and spectrum changes. They do not repeat themselves from one cycle to the next and suggest the simultaneous operation of two or more modes of oscillation. Betelgeuse is an example of an irregular red variable. In these stars the free period of oscillation does not coincide with the periodicity of the driving mechanism.

Variable stars may provide very important clues to stellar evolution. Probably all of them represent more or less ephemeral phases in the evolution of a star. Aside from such catastrophic events as a supernova in which an entire star is torn asunder, some phases of stellar variability may be of such brief duration as to permit detectable changes during an interval of 50 or 100 years. Other stages may require many thousands of years. For example, the period of δ Cephei has not changed by a detectable amount since it was discovered.

CORRELATIONS BETWEEN LUMINOSITY, SPECTRUM AND MASS

Some of the most important generalizations concerning the nature and evolution of stars can be made from correlations between different observable properties and from certain statistical results. One of the most important of these correlations concerns the spectrum and luminosity.

If one plots the absolute magnitudes of the stars (or their intrinsic luminosities on a logarithmic scale) against the spectral types, as Russell was the first to do in 1913, the stars do not fall at random on the diagram but tend to congregate in certain restricted domains (*see* fig. 7). Most of the congregated stars are dwarfs belonging to what is called the main sequence. These stars range from hot blue objects 10,000 times (or more) brighter than the sun, down through white A stars such as Sirius, yellow-white F stars such as Procyon, through the sun, down to orange K stars such as ϵ Eridani, and finally to faint red dwarfs thousands of times fainter than the sun. The sequence is continuous, the luminosities fall off smoothly with decreasing surface temperature, the masses and radii decrease but at a much slower rate and the densities gradually increase.

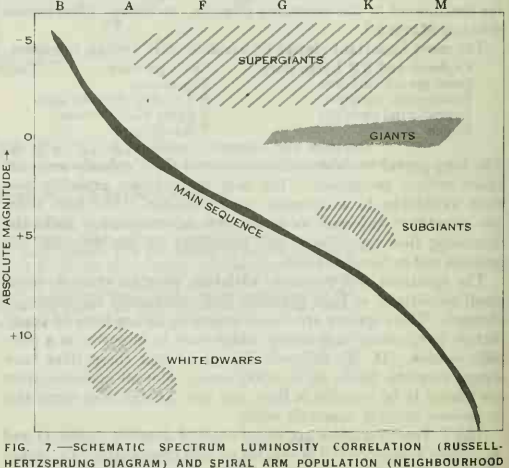


FIG. 7.—SCHEMATIC SPECTRUM LUMINOSITY CORRELATION (RUSSELL-HERTZSPRUNG DIAGRAM) AND SPIRAL ARM POPULATION (NEIGHBOURHOOD OF THE SUN)

The second group of stars to be recognized was the giants, objects such as Capella, Arcturus and Aldebaran, yellow, orange or red stars about 100 times as bright as the sun and with diameters of 10,000,000 or 20,000,000 mi. The supergiants include stars of all spectral classes; they show a big spread in intrinsic brightness; some even approach absolute magnitudes -7 or -8. Some red supergiants such as the variable VV Cephei exceed in size the orbit of Jupiter or even Saturn, although most of them are smaller. Supergiants are rare in space; their great luminosity enables them to be seen at great distances.

The subgiants are stars that are redder and larger than main sequence stars of the same luminosity. Most of the known examples are found in close binary systems, particularly in eclipsing systems where conditions favour the detection of doubles in which a subgiant is paired with a main sequence star.

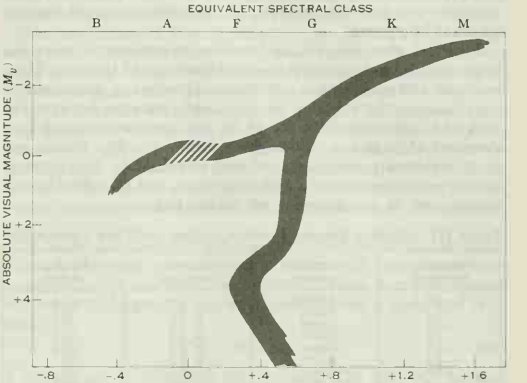
The white dwarfs comprise a group of stars falling about ten magnitudes below the main sequence. Numerically, they exceed the representatives of any group except the main sequence. These objects are of interest for several reasons, not the least of which is that they represent the last stage in stellar evolution, the last feeble glow of a dying star. Their masses are comparable with that of the sun; their densities exceed anything that can be produced in the laboratory. The first of these remarkable objects to be recognized was the companion to Sirius, originally detected by its gravitational attraction upon the larger, brighter star and only later discovered visually as a faint object, about 10,000 times fainter than Sirius. Its mass was about equal to that of the sun. Later work showed that the colour and spectrum of this star corresponded to about spectral class A; *i.e.*, it had a surface temperature of 8,000° K. or 10,000° K. Hence the rate of emission of energy from the surface must be much greater than that of the sun. Since it was about 120 times fainter than the sun, it must have a much smaller surface area and therefore a much smaller volume. Thus the density would have to be a great deal higher than that of the sun. The average density is about 10,000 times that of water. Under these conditions the electrons are nearly all stripped off the atoms and form what is called a degenerate gas. Such material has remarkable properties. It does not obey the ordinary gas laws and is an almost perfect conductor of heat.

Notice that the spectrum luminosity diagram has numerous gaps. For example, above the white dwarfs and to the left of the main sequence there are very few stars. The main sequence and the giants are separated by a gap named after E. Hertzsprung, who was the first to clearly recognize (in 1911) the difference between main sequence and giant stars. The actual concentration of stars in different parts of the diagram changes considerably. Highly luminous stars are rare; those of low luminosity are very numerous.

One feature of fig. 7 must be emphasized. It applies to the spiral arm in the neighbourhood of the sun and represents what would be obtained if Russell-Hertzsprung diagrams were constructed for a large number of open clusters such as h and χ Persei the Pleiades, Coma Berenices and the Hyades and added all together. It includes very young stars (only a few million years old) as well as very ancient (3,000,000,000 or 4,000,000,000 years old) stars. This spiral arm population is called by Walter Baade a type I population to distinguish it from the type II population associated with the globular clusters and the central bulge of the galaxy.

Fig. 8 exhibits the type of spectrum-luminosity relation characteristic of the globular clusters, the central bulge of the galaxy and the elliptical external galaxies. Since these systems are very remote from the observer, the stars are very faint and their spectra cannot be observed. Instead it is necessary to resort to measurements of their magnitudes and colours. Since the colours are closely related to the spectral classes, equivalent spectral classes may be used, but it is to be noted that stellar colours, not spectral classes, are observed in this instance.

The difference between fig. 7 and 8 is striking. In fig. 8 there are no supergiants, and instead of falling at $M_v = 0.0$ the giant branch starts around -3.5 for very red stars and forms a continuous sequence until an absolute magnitude of about -1 is reached. Then there is a band of stars, all about the same colour, proceeding downward (*i.e.*, to fainter stars) to about $M = +3$. The giant branch is then connected to the main sequence at about $+3.5$ by a narrow band. The main sequence extends to fainter and redder stars in much the same way as in the spiral arm population. The main sequence ends just above the sun, however, and does not extend up through the A, B and O spectral classes. Occasionally a few stars are found in the region normally occupied by the main sequence.



FROM "ASTROPHYSICAL JOURNAL"; REPRODUCED BY PERMISSION OF THE UNIVERSITY OF CHICAGO PRESS

FIG. 8.—COLOUR-MAGNITUDE ARRAY FOR A TYPICAL TYPE II POPULATION AS REPRESENTED BY A GLOBULAR CLUSTER

Absolute visual magnitude is plotted against the colour index of the star in the sense C.I. $= M_v - M_{pv}$. It is assumed that the RR Lyrae stars have an absolute magnitude of 0. Equivalent spectral types are plotted at top of diagram

The horizontal branch falls near magnitude 0 and fills the Hertzsprung gap. The striped area in fig. 8 indicates the domain of the RR Lyrae stars. M. Schwarzschild found that every short-period cluster variable fell in this small region of the diagram. Conversely, every star in this range of magnitude and colour has been found to be a variable. The horizontal branch extends beyond the RR Lyrae stars; among these blue stars are found novae and the nuclei of planetary nebulae. Not all globular clusters show identical colour-magnitude arrays. Some clusters such as M3 have many RR Lyrae stars; others such as M13 have very few. Some have numerous giants; others have numerous blue stars on the horizontal branch. In M92 the whole giant branch is bluer than in M13.

Spectroscopic observations indicate that stars in globular clusters do not all have the same hydrogen-metal ratios. Among the extreme examples of Baade's population type II stars found in the neighbourhood of the sun are the subdwarfs which fall about one magnitude below the main sequence. The abundance of metals in these stars is about $\frac{1}{10}$ that in the sun. Stars of a somewhat similar composition are found in certain globular clusters, particularly M92. Many globular clusters appear, however, to have compositions intermediate between that of the sun and these extreme subdwarfs.

Not all stars of Baade's population type II are confined to globular clusters and the central bulge of the galaxy. Many are found scattered throughout the galactic system; some are encountered in the neighbourhood of the sun. High velocity stars are of Baade's type II, but not all type II stars display a high velocity. Numerically the ancient type II population probably greatly exceeds the type I population in our galaxy.

Another item of great statistical interest is the relationship between the luminosities of the stars and their frequency of occurrence. The naked-eye stars are nearly all brighter than the sun. On the other hand, if a census is taken of the known stars within 20 light-years of the sun it is found that the opposite situation is

true. Most of these stars are fainter than the sun (*see* Table II). The bright stars are easily seen at great distances; the faint ones can be detected only if they are close. For example, a star of absolute magnitude −4.0 would be visible to the unaided eye at a distance of 1,000 parsecs, whereas a star of magnitude +11 would be visible only if it fell closer than 1 parsec. That is, only if stars of magnitude +11 were 1,000,000,000 times more abundant in space than stars of magnitude −4 would one observe them in equal numbers, provided one observed all stars to some fixed limit of apparent brightness.

Table III gives the relation between the numbers of stars and their luminosities for stars in the neighbourhood of the sun. Thus, in a cube 100 parsecs on a side one could expect to find about 18 stars between the 4th and 5th magnitude, 80 stars between the 10th and 11th, but there would be less than one chance in a thousand of finding a star brighter than −5. The fainter end of the luminosity function is not well established. It appears to have a maximum near the 14th magnitude and to fall off to yet fainter stars in a fashion not yet established.

TABLE III.—*Relation Between Numbers of Stars and Their Luminosities*

Absolute magnitude	Numbers of stars	Absolute magnitude	Numbers of stars
−5.5	0.00040	+ 3.5	10.5
−4.5	0 0024	+ 4.5	18.5
−3.5	0.0085	+ 5.5	27.0
−2.5	0.030	+ 6.5	36.0
−1.5	0.093	+ 7.5	46.0
−0.5	0.270	+ 8.5	56
+0.5	0.80	+ 9.5	68
+1.5	2.20	+10.5	80
+2.5	5.2	+11.5	91

Source: Compiled from data of W. J. Luyten and P. J. van Rhijn.

The luminosity function depends on the part of the galaxy under consideration. The neighbourhood of the sun includes examples of both populations type I and II. For a pure population type II the luminosity function will differ substantially from that of a pure population type I. Also, the luminosity function changes with distance from the galactic plane.

One of the most important correlations is that between mass and luminosity for main sequence (dwarf) stars. If one takes the visual binaries for which good parallaxes and masses are known and plot the mass of each star against its bolometric luminosity one finds that the points tend to fall on a curve which shows that the more massive stars are simultaneously the more luminous. The discussion by K. Strand, P. van de Kamp and H. N. Russell and Sitterly showed that for stars in the neighbourhood of the sun the luminosity is very nearly proportional to the fourth power of the mass, the actual relation being

$$L \cong \text{constant } M^{3.8} \tag{10}$$

where L and M are expressed in terms of the luminosity and mass of the sun respectively. Later work indicates that the constant is very nearly unity; *i.e.*, the sun lies on the mass-luminosity curve defined by the nearby stars. The relationship is not a precise one, because for each mass there does not correspond a unique luminosity. In fact, a small range in luminosity exists for main sequence stars as a consequence of stellar evolution.

The mass-luminosity correlation is derived from visual double star systems and applies to main sequence stars. It fails for giants and supergiants and for the faint components of eclipsing binaries that usually are subgiants. Also it does not apply to any of the stars in a globular cluster brighter than absolute magnitude $M = +4.0$, or in other words stars other than those that fall on the main sequence.

The white dwarfs have long been known to fall off the mass-luminosity correlation; these stars are abnormally faint for their masses. They represent defunct stars near the end of their evolution, objects which have exhausted their nuclear fuel and presumably now are simply cooling down.

The mass-luminosity correlation was predicted theoretically by Eddington. It is a general relationship which holds for all stars built on essentially the same internal density and temperature distributions; *i.e.*, the same models. The types of stars for which it fails are those built on different models than the main sequence stars.

STELLAR STRUCTURE

The internal pressures, temperatures and densities of main sequence stars and particularly white dwarfs can be calculated with greater confidence than can the internal temperature and density of the earth. This is true because of the following basic physical principles:

First, we suppose that the star is in mechanical equilibrium. That is, at each point the pressure exactly balances the weight of the overlying layers. If g is the acceleration of gravity, ρ is the density and ΔP the increase in pressure over a small increment in distance Δr, then

$$\Delta P = g\rho\Delta r \tag{11}$$

A second equation expresses the continuity of mass. That is, if M_r is the mass of material lying in a sphere within a radius r, the amount of mass added when one goes a distance Δr through a shell of volume $4\pi r^2\Delta r$ is:

$$\Delta M = 4\pi r^2\rho\Delta r \tag{12}$$

The third relation involved is what is called an equation of state; it expresses the relation between the pressure, density and temperature of the material. Throughout the body of the star the material is entirely gaseous. In most stars it obeys closely the perfect gas law:

$$P = \frac{\Re\rho T}{\mu} \tag{13}$$

Here P is the pressure (usually expressed in dynes per square centimetre although it may also be given in atmospheres), ρ is the density (g. per cubic centimetre) and T is absolute temperature (° K.). $\Re$ is the gas constant (defined for a gram molecule of material). Finally, μ is the molecular weight, 2 for molecular hydrogen, 4 for helium and 56 for iron vapour. In the interior of a typical star, however, almost all of the material is completely ionized. Under these circumstances the hydrogen molecules are not only broken down into individual atoms but the atoms themselves are broken down (ionized) into protons and electrons. Hence, the molecular weight of ionized hydrogen is $\frac{1}{2}$, since the average mass of a proton and an electron is $\frac{1}{2}$ on the atom weight scale. On the other hand, a completely ionized helium atom contributes a mass of 4 with 2 electrons of negligible mass; hence the molecular weight is $\frac{4}{3}$. A completely ionized nickel atom contributes one nucleus of mass 58.7 plus 28 electrons, hence its molecular weight is $\frac{58.7}{29} = 2.02$. Since stars contain a preponderance of hydrogen and helium which are completely ionized throughout the interior,

$$\mu = \frac{1}{2X + (3/4)Y + \alpha(1 - X - Y)} \tag{14}$$

where X and Y are the concentrations by weight of hydrogen and helium, respectively, and α is nearly 2. Hence the molecular weight will depend critically on the chemical composition, particularly on the ratio of hydrogen to helium and the total content of heavier material.

If the temperature is sufficiently high, the radiation pressure, which is numerically equal to one-third of the energy density, must also be taken into account. Hence the total equation of state is

$$P = P_g + P_r = \frac{\Re\rho T}{\mu} + \frac{1}{3}aT^4 \tag{15}$$

where $a = 7.56 \times 10^{-15}$ is called the radiation density constant. Set $\mu = 2$ (upper limit), $\rho = 1.4$ g. per cubic centimetre (mean density of sun) and calculate the temperature at which the radiation pressure would equal the gas pressure. With $R = 8.32 \times 10^7$ the required temperature is 28,000,000°, which is much hotter than the centre of the sun, where the density is about 100 times that of water. Hence radiation pressure in the sun may be

neglected, but it cannot be neglected in hotter, more massive stars. Radiation pressure may indeed set an upper limit to stellar luminosity (as Eddington suggested long ago).

Certain stars, notably the white dwarfs, do not obey the perfect gas law. Instead, the pressure is almost entirely contributed by the electrons, which obey a law called the degenerate gas law. The pressure and density are related by a law of the form

$$P = 9.9 \times 10^{12}\left(\frac{\rho}{\mu'}\right)^{\frac{5}{3}} \tag{16}$$

where μ' is the average mass per free electron of the totally ionized gas. Notice that the temperature does not enter at all. At yet higher densities the equation of state becomes more intricate, but S. Chandrasekhar showed in elegant fashion how even this complicated equation of state could be employed in the calculation of the internal structure of these white dwarf stars. There are probably better models for the white dwarfs than for most other objects.

In order to make further progress it must be known how energy is transported in the interior. Thermal conduction cannot play any significant role inasmuch as the heat conductivity of gases is very low. One mode of transport is the actual flow of radiation through the star. Starting out as X-rays the radiation is gradually "softened" as it works its way to the surface to emerge as ordinary light and heat. The rate of flow of radiation is proportional to the temperature gradient, $\Delta T/\Delta r$, the rate of change of temperature with distance. Specifically,

$$\text{Rate of energy flow} = \frac{4acT^3}{3\kappa\rho}\left(\frac{\Delta T}{\Delta r}\right) \tag{17}$$

Here a is the radiation density constant, c is the velocity of light, ρ is the density and κ is a measure of the opacity of the material. That is, the larger the value of κ, the smaller the transparency of the material and the steeper the temperature fall required to push the energy outward at the required rate. The opacity κ can be calculated for any temperature, density and chemical composition. The actual computations are difficult; κ is found to depend in a complicated way on the density and temperature.

In many of the outer layers of the sun and particularly in certain giant stars, transport of energy takes place by quite another mechanism: large-scale mass motions of gases or convection. Large volumes of gas deep within the star become heated, rise to higher layers and mix with their surroundings, thus releasing great quantities of energy. The extremely complicated flow patterns cannot be followed in detail, but convection results in a relatively simple relation between pressure and density, namely

$$P = K\rho^{\gamma} \tag{18}$$

where γ is $\frac{5}{3}$ for a completely ionized gas and K is a constant of proportionality.

Finally, the most fundamental property of stars must be considered, namely, that they radiate energy and hence must derive their energy from sources within themselves. For normal main sequence stars such as the sun, the source of the energy lies in the conversion of hydrogen into helium. The reaction believed to occur in the sun is the proton-proton reaction proposed first by H. Bethe and R. Critchfield and later revised by C. C. Lauritsen and W. A. Fowler and independently by E. Schatzman. The reactions are:

$$\begin{aligned} {}_1H + {}_1H^1 &\rightarrow {}_2H^2 + (\epsilon^+) \\ {}_1H^1 + {}_1H^2 &\rightarrow {}_2He^3 + \gamma \\ {}_2He^3 + {}_2He^3 &\rightarrow {}_2He^4 + 2H^1 \end{aligned} \tag{19}$$

Two protons collide to form a deuteron (heavy hydrogen nucleus) with the liberation of a beta ray, or positive electron (positron). This positive electron meets a negative electron and the two annihilate one another with the liberation of energy in accordance with the Einstein relation

$$E = mc^2 \tag{20}$$

where m is the mass annihilated, c is the velocity of light and E is the energy released. Here m is the mass of the two electrons. The annihilation energy amounts to 1.02 Mev. This process is one of low probability, and many proton-proton collisions occur before a positive electron is created.

Next, a proton collides with the deuteron to form the nucleus of the light helium atom of atomic weight 3. Finally, two light helium nuclei collide to form a nucleus of ordinary helium and two protons. In the course of this reaction four protons have been consumed to form one helium nucleus, and two negative electrons have perished.

The total mass of four hydrogen atoms is $4 \times 1.008 = 4.032$ while that of a helium atom is 4.004. Hence 0.028 atomic mass units or 0.7% of the original mass has disappeared. Some of this has disappeared in the form of the elusive neutrinos, but most of it has been converted into radiant energy. To keep shining at its present rate, the sun has to convert 674,000,000 tons of hydrogen into 670,000,000 tons of helium every second. More than 4,000,000 tons of matter literally disappears into radiation every second. The rate of energy generation according to the proton-proton reaction varies as about the fourth power of the temperature.

In hotter stars the principal source of energy is the carbon cycle, in which hydrogen is converted into helium with carbon serving as a catalyst. The reactions are as follows: First, a carbon nucleus captures a hydrogen nucleus or proton to form a nucleus of N^{13}, thus

$$C^{12} + H^1 \rightarrow N^{13} + \gamma \tag{21}$$

A gamma ray is emitted in the process. The light N^{13} nucleus is unstable, however. It emits a positive electron which encounters a negative electron and the two annihilate one another. The resulting C^{13} nucleus is stable. Eventually the C^{13} nucleus captures another proton, forms N^{14} and emits another gamma ray.

$$\begin{aligned} N^{13} &\rightarrow C^{13} + \epsilon^+ \\ C^{13} + H^1 &\rightarrow N^{14} + \gamma \end{aligned} \tag{22}$$

The N^{14} is stable, ordinary nitrogen. When it captures a proton to form O^{15}, the resulting nucleus is unstable against beta decay. That is, it emits a positive electron.

$$\begin{aligned} N^{14} + H^1 &\rightarrow O^{15} + \gamma \\ O^{15} &\rightarrow N^{15} + \epsilon^+ \end{aligned} \tag{23}$$

That is, the O^{15} nucleus emits a positive electron that meets a negative electron and is annihilated. Eventually the N^{15} nucleus encounters a fast-moving proton and captures it. It might be supposed that at this point a nucleus of ordinary O^{16} would be formed. Actually this process occurs rarely, the favoured reaction being

$$N^{15} + H^1 \rightarrow C^{12} + He^4 + \gamma \tag{24}$$

The capture of the proton causes the resultant nucleus to break down into the original C^{12} nucleus and an alpha particle, or helium nucleus of mass 4. Thus, the original carbon nucleus reappears, and the four protons that have been added permit the formation of an alpha particle. The same amount of mass has disappeared although a different fraction of it may have been carried off by the neutrinos.

The physicist with his particle accelerator produces these reactions by firing protons with hundreds or even thousands of kilovolts of energy at targets of graphite or nitrogen compounds. Since both the heavy nucleus and the proton have positive charges, they repel one another by Coulomb's law. Occasionally the particles come close enough together for short-range attractive forces in the nucleus to allow the proton to stick. The higher the energy involved, the greater the probability of proton capture.

In the interior of a star the particles are moving rapidly about as a consequence of their high energies. Once in a while a proton is directed sufficiently close to a nucleus for it to be captured so that a nuclear reaction can occur. Only the protons of very highest energy (many times the average energy in a star like the sun) are capable of producing reactions. Since the energies of the protons depend very strongly on the temperature, the rate of energy pro-

duction rises steeply with the temperature. The temperature dependence is much steeper for the carbon-nitrogen cycle than it is for the proton-proton reaction.

The hottest stars of the main sequence (and probably also the supergiants) shine with energy produced by the carbon cycle. The faint red dwarf stars use exclusively the proton-proton reaction, whereas the sun shines mostly by the proton-proton reaction but has some contribution from the carbon cycle.

With the aid of the aforementioned mathematical relationships including the law of energy generation it is possible to attack the problem of stellar structure. In the first approximation it is usually assumed that the stars have a uniform chemical composition throughout their interiors. This assumption simplifies the calculations greatly, but it has to be abandoned when stellar evolution is concerned. The mathematical and physical relationships involved are such that solutions expressible in algebraic form or as functions of types known to mathematicians are not obtainable. The solutions have to be carried out numerically by a step-by-step process known as mechanical integration. At the surface the pressure and density vanish; at the centre of the star the pressure, density and temperature must remain finite. A frequent procedure is to carry the calculations from the surface inward and from the centre outward, the condition being that the two solutions must fit together, giving the same pressure, density, temperature, etc., when they meet. One of the most important generalizations that can be drawn from such calculations is the famous mass-luminosity relation, discovered by Eddington. It may be written in the form

$$L = C\frac{M^a\mu^b}{R^c} \tag{25}$$

where M is the mass of the star, μ the molecular weight, and R the radius. C is a constant that depends on the particular numerical choice of physical relationships involved, for example, on the exact value of the opacity. Here a, b and c are constants that also depend for their precise values on the exact form of the law of opacity, for example. Numerically, a is of the order of 5, b of 7 or 8 and c about 1.5. This means that the luminosity depends steeply on the mass but even more steeply on the molecular weight and therefore on the chemical composition.

Still another relationship between luminosity, mass, radius and molecular weight is derived from temperature, density distribution and the known nuclear energy sources. For a star shining on the carbon cycle a relation of the form is obtained:

$$L = B(1 - X - Y)\mu^d\frac{M^e}{R^f} \tag{26}$$

where X and Y are the proportions by weight of hydrogen and helium, respectively. Here d is of the order of 20, e about 22 and f about 23 since the carbon cycle depends on a high power of the temperature. This equation is called the energy output relation and was used by Schwarzschild to demonstrate how, for a given model, one could obtain both the hydrogen and the helium contents.

The calculation of stellar models has occupied the attention of astronomers for a number of years. For many main sequence stars it is possible to obtain satisfactory solutions with a chemical composition that is uniform throughout the star. In the sun and fainter stars of the main sequence energy is transported throughout the outer layers by convection currents whereas in the deep interior energy is transported by radiation. Among the hotter stars of the main sequence the reverse appears to be true. The deep interiors of stars which derive their energy primarily from the carbon cycle are in convective equilibrium whereas in the outer parts the energy is carried by radiation. It is possible to reproduce the observed masses, luminosities and radii of main sequence stars with reasonable assumptions concerning their chemical composition.

Chemically homogeneous models of giant and supergiant stars cannot be constructed. That is, if it is assumed that a yellow giant star such as Capella is built on the same model as the sun or other main sequence stars, the central temperature will turn out to be so low that no known nuclear process can possibly supply the observed energy output. Progress has been made only by assuming that these stars were once main sequence objects which, in the course of their lives, exhausted the hydrogen in their deep interiors. As a consequence these stars developed inert cores composed mostly of helium, which is the ash of the hydrogen burning processes. Since helium is impervious to nuclear reactions at the few tens of millions degrees and the densities likely to prevail in these interiors, no thermonuclear energy could be obtained from these depleted cores. Energy is presumed to be generated in a thin shell surrounding the inert core and is presumably produced by the carbon-nitrogen cycle. Such models are called shell-source models. As the star uses up more and more of its hydrogen supply, the core increases in mass but the outer envelope of the star continues to expand in size. These shell-source models appear to be capable of explaining the observed luminosities, masses and radii of giant and supergiant stars.

The depletion of hydrogen fuel is appreciable even for a dwarf star such as the sun. The sun appears to be about 5,000,000,000 years old and to have been shining at its present rate for about 1,000,000,000 years. On the assumption that the sun was initially composed of about $\frac{4}{5}$ hydrogen by weight, Schwarzschild, Howard and Härm, found that at the present time the concentration of hydrogen in the centre of the sun must be only about 30% by weight. Furthermore, most of the sun was in radiative equilibrium, leaving only about 0.33% in the outer convection zone. In order to maintain its observed luminosity, the sun's central temperature has increased noticeably since the time the solar system was formed, largely as a consequence of the depletion of the hydrogen in the interior with the accompanying increase in molecular weight and temperature. As the sun has evolved it has also become brighter. Schwarzschild, Howard and Härm concluded that over the past 5,000,000,000 years the sun brightened by about half a magnitude, and they suggest that in the early Pre-Cambrian era (about 2,000,000,000 years ago) the solar luminosity was about 20% less than now. The average temperature on the earth's surface must have been about the freezing point of water. It is possible, of course, that the earth's atmosphere might then have acted as a more efficient heat trap and permitted the surface temperature of the earth to be actually higher than this rather cool value.

Table IV, giving the internal structure of the sun, takes into account the carbon cycle as well as the proton-proton reaction. At the centre of the sun the temperature is 14,600,000° and the density is about 134 times that of water. Nevertheless, in spite of the high density the material is very nearly a perfect gas. Weymann's model takes into account the depletion of hydrogen in the interior of the sun.

TABLE IV.—*Internal Structure of the Sun*

r/R	Temperature (°A.)	Density (water = 1.0)	r/R	Temperature (°A.)	Density (water = 1.0)
0.00	14,600,000	134	0.60	2,500,000	0.405
0.10	12,650,000	85.5	0.70	2,270,000	0.124
0.20	9,350,000	36.4	0.80	1,270,000	0.0354
0.30	6,650,000	12.9	0.90	605,000	0.00945
0.40	4,740,000	4.13	0.96	226,000	0.00217
0.50	3,430,000	1.30	0.98	111,000	0.00073

Source: After R. Weymann, Princeton University Observatory.

STELLAR EVOLUTION

From what has been said it is clear that one must anticipate some evolution among the stars. The aforementioned calculations by Schwarzschild and his co-workers of Princeton university show that these evolutionary effects have not been negligible even for the sun. For a more massive star the effects must be even more spectacular since the rate of conversion of mass into energy is increased. The sun produces energy at the rate of about 2 ergs per gram per second, but a luminous main sequence star may produce energy at the rate of 2,000 ergs per gram per second. Consequently, while evolutionary effects may require billions of years to be manifest in a star such as the sun, they may occur within a few millions of years in highly luminous stars. A supergiant star

such as Antares, a bright main sequence star such as one of the belt stars of Orion or even a more modest star such as Sirius cannot have endured as long as the sun. Therefore, if it has existed for as long as 5,000,000,000 years it must have spent most of its life as a nonstellar object. Alternatively, these stars may have been formed relatively recently in the history of the galaxy. Thus, one might expect to find within the galaxy (and even within the neighbourhood of the sun itself) stars that are well evolved and/or even approaching extinction, as well as occasional stars that must be very young or even in the process of formation. First, let us examine the life history of a star as it can best be envisaged on the basis of modern theories of physics and stellar evolution. Stars are presumably formed from the grains and gas in the interstellar medium. The material between the stars is not uniformly distributed in space but is spread in a patchy fashion. Occasionally a cloud will form sufficiently large for gravitational attractions to draw it together.

As it begins to pull itself together, the internal temperature and density start to rise until finally the star breaks into incandescence with a faint red glow. At this stage the star is still shining by its gravitational contraction; *i.e.*, by the conversion of potential energy into heat. As the internal temperatures rise to a few million degrees, deuterium (heavy hydrogen) is first destroyed. Then lithium, beryllium and boron are broken down into helium as their nuclei are bombarded by protons moving at high velocities.

Detailed calculations by E. Salpeter and particularly by L. Henyey and his associates show that the star originally falls above the main sequence because it is too bright for its colour. As it contracts it moves to the left toward the main sequence.

Finally, as the central temperature and density continue to rise, a state is reached where the proton-proton or carbon cycle reaction becomes active and the development of the star is stabilized. The star then reaches the main sequence, where it remains for most of its active life. The time required for the contraction phase depends on the mass of the star. A star of the sun's mass may require millions of years to reach the main sequence, while one of much greater mass may reach the main sequence in a few hundred thousand years.

By the time the star reaches the main sequence it is still chemically homogeneous. As time goes on and the hydrogen in the core is converted into helium, the temperature slowly rises to compensate for the depletion of the fuel. If the star is sufficiently massive to have a convective core, the material in this region has a chance to be thoroughly mixed, but the outer region does not mix with the core. In the sun, on the other hand, there is no convective core and the helium-hydrogen ratio is maximum at the centre and decreases outward. During the lifetime of the sun there has been a steady depletion of the hydrogen so that at the centre the concentration of hydrogen is probably only about one third of the original amount. The rest has all been converted to helium.

The rate of evolution depends strongly on the mass of the star. In the sun the effects are only slightly manifest (they could not have been found if observers had not been seeking them). A star of twice the sun's mass burns its energy at such a rate that it cannot last more than about 3,000,000,000 years, whereas a star of ten times the sun's mass has a lifetime measured in tens of millions of years.

As the composition of the star's interior becomes seriously modified it begins to move away from the main sequence, slowly at first, and then more rapidly. When about 10% of the star's mass has been converted into helium, the structure of the star changes drastically. The core becomes inert. That is, all the hydrogen has been burned out and this central region is composed almost entirely of helium with small admixtures of heavier elements. The energy production now takes place in a thin shell where hydrogen is consumed and added to the inert core. The outer portion of the star expands outward, and as the star swells up it may gradually rise in luminosity. The details of the evolution depend on the hydrogen-metal ratio, and the course of evolution will differ for stars of the two population types.

Consider first a globular cluster (fig. 8). The main sequence does not extend to stars brighter than absolute magnitude +3.5. The giant sequence begins at this point and extends on the colour magnitude diagram to the right and vertically upward for several magnitudes. Then it gradually bends over, the most luminous stars in the cluster being the red giants. This colour-magnitude diagram exhibits the effects of stellar evolution. All main sequence stars more massive than about 1.5 solar masses have evolved off the main sequence and most of them have died. The colour-magnitude array then represents an evolutionary picture of the cluster. The giant branch is not, strictly speaking, the path of evolution but rather the locus or envelope of evolutionary paths. It represents a snapshot of stars of different masses but of the same age. Presumably the stars leave the main sequence and evolve into stars of steadily increasing luminosity until they reach the peak of the diagram. Then what? Presumably later stages of evolution are connected with the horizontal branch at absolute magnitude zero where the RR Lyrae stars and other blue objects are found. The last stages of evolution are represented by the white dwarfs which are too faint to be detected in globular clusters, but they must be present there in great numbers.

F. Hoyle and Schwarzschild have calculated the evolutionary track of a star from the point where it leaves the main sequence until it reaches the top of the giant branch. When the core of the star becomes depleted of hydrogen, the energy generation is confined to a narrow shell and the outer part of the star becomes distended. That is, the star swells up and moves to the right in the colour-magnitude array. The energy is carried outward by large-scale convection currents. By the time about 20% of the total mass of hydrogen has been converted into helium, these convection currents have become so rapid that they approach the velocity of sound. Then as the star evolves upward in the diagram, the temperature in the energy-producing shell begins to rise, slowly at first, then more rapidly until near the top of the giant sequence the central temperature reaches a value of about 40,000,000° K. While this is happening the total energy output of the star rises, and the core shrinks by gravitational contraction. The compression of the core raises its temperature and density until the core reaches a temperature of about 120,000,000° K. At that temperature, the previously inert helium is consumed in the production of heavier elements.

If two helium nuclei are jammed together one might expect them to form a nucleus of beryllium of mass 8, thus

$$ {}_2He^4 + {}_2He^4 \rightarrow {}_4Be^8 \qquad (27) $$

Actually ${}_4Be^8$ is unstable and breaks down again into two helium nuclei. If the density and temperature are high enough, however, the ephemeral Be^8 nucleus may capture another helium nucleus to form a nucleus of C^{12}, thus

$$ {}_4Be^8 + {}_2He^4 \rightarrow {}_6C^{12} \qquad (28) $$

and reactions such as

$$ \begin{aligned} {}_6C^{12} + {}_2He^4 &\rightarrow {}_8O^{18} \\ {}_8O^{16} + {}_2He^4 &\rightarrow {}_{10}Ne^{20} \\ {}_{10}Ne^{20} + {}_2He^4 &\rightarrow {}_{12}Mg^{24} \end{aligned} \qquad (29) $$

may also occur. When the helium burning stage is reached, about half the original hydrogen mass of the star has been converted into helium. The star has two energy sources at the time but soon thereafter this helium burning would appear to dominate.

At this point the star again has a central energy source. The shell-source type of model no longer holds and the star presumably moves quickly to the horizontal branch of the colour-magnitude diagram. The theory of evolution for this stage is very uncertain but some general conclusions may be easily drawn. The energy output of helium burning (per ton of material) is much less than for hydrogen burning. The lifetime of the star on the horizontal branch will not be long. Presumably it evolves to bluer colours and lower luminosities but certainly ends as a white dwarf. The energy output of the white dwarfs is so low that they may go on shining by radiating away their store of energy until none is left to emit. The final stage of a star is a ball not much larger than the earth, but with a density 50,000 times that of water, covered with a thin layer of ice and surrounded by an

atmosphere of hydrogen and helium a few feet thick.

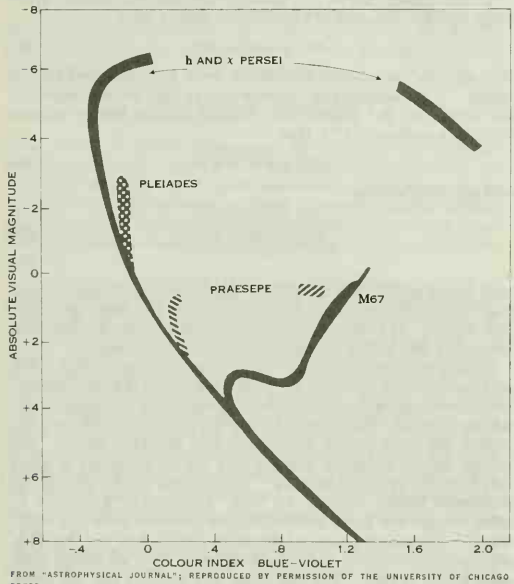
One of the interesting results of the calculations by Hoyle, A. R. Sandage and Schwarzschild is a determination of the ages of the globular cluster. The turnoff point at the upper end of the main sequence lies just above the position occupied by the sun. Therefore, stars more massive than about 1.3 solar masses have evolved away from the main sequence. The time required for such a star to exhaust the hydrogen in its core is of the order of 5,000,000,000 or 6,000,000,000 years. Hence the globular clusters are very ancient objects, probably among the oldest in our galactic system.

Turning to the population type I a very different situation is found. If the colour-magnitude arrays for different galactic clusters are compared, the situation is that exhibited in fig. 9. In the double cluster h and χ Persei the main sequence extends to about magnitude -6 where it turns off abruptly to the right. There are also a few red supergiant stars. The Pleiades main sequence turns upward from the usual main sequence at about magnitude O but there are no giant stars while the Praesepe (Beehive cluster in Cancer) shows a giant sequence and a main sequence that is curved up at about $+2$. The cluster M67 has a colour-magnitude array that resembles superficially that of the giant branch of a globular cluster, except that there is no horizontal branch and the most luminous giants are much fainter than the globular cluster giants.

The feature distinguishing these different clusters is their different ages. The youngest cluster, the pair h and χ Persei, contains stars of a great range in luminosity. Some stars have already evolved into the supergiant stage so that the top end of the main sequence is bent over. The stars of luminosity greater than 10,000 times that of the sun have already largely depleted the hydrogen in their cores and are leaving the main sequence.

The brightest stars of the Pleiades (Alcyone, Maia, Taygeta, etc.), have started to leave the main sequence and are approaching the critical phase when they will have exhausted all the hydrogen in their cores. There are no giants in the Pleiades. Presumably the cluster contained no stars as massive as some of those found in h and χ Persei. The cluster is also older than the double cluster.

The Beehive or Praesepe cluster is still older than the Pleiades.



FROM "ASTROPHYSICAL JOURNAL"; REPRODUCED BY PERMISSION OF THE UNIVERSITY OF CHICAGO PRESS

FIG. 9.—COLOUR-MAGNITUDE ARRAYS FOR FOUR GALACTIC CLUSTERS OF DIFFERENT AGES

All stars much more luminous than Procyon ($+3$) have started to leave the main sequence, and there are some giants. The Hyades display a similar colour-magnitude array. These clusters contain a number of white dwarfs, indicating that the initially most luminous stars have already run the full gamut of evolution. Finally, M67 must be the oldest cluster of all; the best data of stellar structure indicate that it is several thousand million years old. All the bright main sequence stars have long since disappeared.

Some estimated ages of typical galactic clusters, judged from the turnoff points of the main sequence, are given in Table V.

Hoyle and Schwarzschild were also able to establish that the difference between the giant branches in M67 and in the globular clusters was to be attributed to differences in chemical composition, in particular to the hydrogen-helium ratio. In the type II stars, they took the ratio of hydrogen to the metals by numbers as 135,000 (in accordance with the spectroscopic data for the subdwarfs) whereas for the type I stars they took a ratio of 10,000. The population type I giant stars, as a consequence of their larger metal contents, produced a more extended envelope and hence later spectral classes.

The giant stars of the Praesepe are comparable with the brightest stars in M67. Now the M67 giants have evolved from the main sequence near absolute magnitude $+3.5$, whereas the Praesepe giants must have evolved from main sequence stars of absolute magnitude about $+0.5$. Consequently, the Praesepe giants must have masses about twice as great as those of the M67 giants.

TABLE V.—*Estimated Ages of Typical Galactic Clusters*

Cluster	Age (millions of years)	Cluster	Age (millions of years)
Orion association (1)	1	Coma Berenices	3,000
Perseus association (1)	10	Hyades	4,000
Pleiades	200	Praesepe	4,000

Source: After A. R. Sandage.

Thus the failure of the mass-luminosity law for giant stars is easily understood. Giant stars of the same luminosity may therefore have appreciably different masses.

An example in point is provided by the components of the giant star Capella, for which K. O. Wright found masses three times that of the sun, whereas each component of the giant binary TW Cancri has a mass 1.3 that of the sun, according to D. M. Popper. Thus Capella is presumably a relatively young star, TW Cancri an ancient one.

Several other differences between the various galactic clusters are to be noted. The M67 giant sequence is continuous with the main sequence, whereas there is an appreciable gap between the giant and main sequence stars in the Praesepe and a much wider one in h and χ Persei. One suspects that these vacant regions are connected with domains of instability in the structure of the star.

One interesting fact noted by R. Kraft and John Irwin was that several Cepheids are found in galactic clusters, and in each instance where the colour-magnitude arrays can be constructed the Cepheid appears to fall in the gap.

If some star clusters such as h and χ Persei are only a few million years old, whereas others such as the Hyades are a thousand million years old, the process of star formation in clusters must be going on at the present time, and we ought to be able to catch occasional stars or groups of stars in the process of formation. G. Herbig and G. Haro have found evidence that the T Tauri stars, variables found only in regions where there are great quantities of interstellar dust and gas, are stars in the process of formation. They probably are still contracting and have not yet started to shine by thermonuclear hydrogen burning reactions.

Some of the most striking evidence for star formation was found by Parenago in the region of the Orion nebula and by M. F. Walker in the cluster associated with S Monocerotis NGC 2264.

FORMATION OF THE ELEMENTS IN STARS

Many years ago in an anonymous communication to a scientific journal William Prout proposed the hypothesis that all elements

originated from hydrogen. Although this suggestion gathered dust for many years it has been revived in connection with the possible origin of elements within stars. Astronomers and nuclear physicists have devoted much effort to this problem.

The first investigation of the abundances of elements was undertaken by F. W. Clarke and Washington, who analyzed samples of the earth's crust. Later, V. M. Goldschmidt pointed out that a much more representative sample of the probable composition of our part of the universe could be obtained from meteorites, at least as far as nonvolatile constituents were concerned. Finally, in his pioneering analysis of the sun, Russell gave the first clear perspective of the relative abundances of the elements in our part of the universe. Subsequent work has revised and extended Russell's early results; in particular, it has demonstrated the overwhelming preponderance of hydrogen and helium, compared with which all other elements can be regarded as mere impurities. Table VI, based largely on work at the University of Michigan on the composition of the sun and early-type stars, gives the most abundant elements (by numbers) in a representative sample of our part of the universe.

Among the elements lighter than iron which have even atomic numbers, those with atomic numbers divisible by four tend to be most abundant (rule of Oddo and Harkins). Other abundance effects have been noted; the abundances of light elements tend to be greater than those of heavier elements. That is, a plot of abundance against atomic weight, A, shows a decline with increasing A to about $A = 100$. Thereafter the abundance is more nearly constant. A plot of numbers of atoms against atomic weight or number is far from smooth. Aside from the irregularities produced by the rule of Oddo and Harkins there are huge fluctuations for certain elements. Thus lithium, beryllium and boron are very rare compared with carbon, nitrogen or oxygen. There is a pronounced abundance peak at iron and less conspicuous peaks at other elements, often connected with the so-called magic number nuclei. There is also a maximum at lead, the most stable of heavy elements.

Astronomers and physicists believe that the abundance curve of the elements provides definite clues as to their origins. The preponderance of hydrogen suggests that perhaps all nuclei were built up from protons. G. Gamow suggested that the elements were created as the consequence of the expansion of a huge mass consisting of neutrons. As the density decreased the neutrons decayed into protons and electrons. The protons then captured neutrons, one after another, underwent beta decay (ejection of electrons) and built up heavy elements. The big difficulty with this hypothesis is that atomic masses 5 and 8 are unstable and there is no way to build nuclei beyond helium. Other difficulties have been pointed out, and the Gamow hypothesis must be greatly modified or discarded.

TABLE VI.—*The Most Abundant Elements (by Numbers)*
(Silicon = 10,000)

Element	Abundance	Element	Abundance
H	2.5×10^8	P	79
He	4.5×10^7	S	14,000
C	63,000	Cl	6,300
N	102,000	Ar	3,160
O	330,000	K	23
Ne	500,000	Ca	600
Na	500	Ti	23
Mg	4,800	Cr	250
Al	416	Fe	1,440
Si	10,000	Ni	150

The alternative hypothesis to the effect that heavy elements are built in stars has much to recommend it. Consider, first of all, the following observation facts. Not all stars have the same chemical composition. A long time ago R. H. Curtiss pointed out that the splitting in the spectral sequence among the cooler stars could be understood in terms of composition differences. The R and N stars contain more carbon than oxygen, whereas the S stars appear to have an enhanced content of zirconium as compared with titanium. From a more detailed spectroscopic study of the cool stars, P. W. Merrill has found evidence for differences in both the carbon-oxygen ratios and in the ratio of the zirconium group to titanium group of metals. The M stars appear to have a "normal" (*i.e.* Solar) composition with oxygen more abundant than carbon and the zirconium group of elements much less abundant than the titanium group.

Other abundance anomalies are found in the Wolf-Rayet stars where objects containing predominantly helium, carbon and oxygen are distinguished from objects containing helium and nitrogen, some carbon but no observable oxygen.

Significantly, all these abundance anomalies are found in stars believed to be well advanced in their evolutionary development. No main sequence dwarf stars show such effects.

The most critical observation, however, was Merrill's detection of the unstable element technitium in the S stars. The element has been produced on the earth synthetically. The longest lived isotope Tc^{99}_{43}, which has a half life of 200,000 years, is produced by the following reaction:

$$\begin{aligned} Mo^{98} + n &\rightarrow Mo^{99} \\ Mo^{99} &\rightarrow Tc^{99} + \beta^- \end{aligned} \qquad (30)$$

That is, molybdenum of mass 98 captures a neutron to form molybdenum of mass 99, which decays (in about 67 hours) to form technetium of mass 99 with the ejection of a negative electron. Furthermore, the Tc in the S stars appears to be about as abundant as the neighbouring elements, ruthenium and molybdenum. The implication is that all these elements were built in the S stars and that these objects have lifetimes of about 200,000 years. How the star gets the heavy elements from the core to the surface without exploding is unknown at present.

Work by the Burbidges, F. Hoyle, W. Fowler and A. G. W. Cameron has demonstrated how elements may be built in stars by nuclear processes occurring at very high temperatures and densities. No one mechanism can account for all the elements, and several distinct processes have been proposed.

The great abundance of helium is easily understood. It is the normal "ash" of hydrogen consumption. In the dense cores of massive stars the helium itself may be consumed to form C^{12}, O^{16}, Ne^{20} and Mg^{24}. At yet higher temperatures a host of nuclear reactions may occur to produce the elements up to iron. Heavy elements such as zirconium cannot be produced in this manner. Apparently, the only way they can be formed is by the capture of neutrons.

Cameron suggested that the light isotope of carbon, C^{13}, could get mixed into the helium core and combine with alpha particles (helium nuclei) to produce O^{16} and neutrons in accordance with the following reaction:

$$C^{13} + He^4 \rightarrow O^{16} + n + (2.6\ \text{Mev})\ \text{energy} \qquad (31)$$

These neutrons are available for capture by heavier nuclei, particularly those with large target areas, to form yet heavier elements. Thus the elements of the zirconium group, for example, could be built from lighter elements by successive neutron capture. Fowler and the Burbidges suggested other nuclear processes that could result in the production of neutrons and the subsequent formation of heavy elements. They concluded that about 40% of the iron in a mixture will be transformed into heavier elements if about 20 neutrons could be produced for every iron atom in the mixture.

After a given nucleus captures a neutron, it may find itself too heavy for its charge and decay into the nucleus of the element next to it in atomic number. Beta decay is often very slow, and if the flux of neutrons is high the nucleus may capture another neutron before it can undergo decay. Hoyle suggested that in supernovae explosions vast quantities of neutrons are produced and these result in the rapid build-up of massive elements. One interesting feature of the building up of heavy elements by neutron capture at a high rate in a supernova explosion is that nuclei much heavier than lead or uranium can be built. In particular, nuclei of the synthetic element Californium 254 may be produced; these nuclei decay by spontaneous fission with a half life of 55 days.

Hoyle, the Burbidges and Fowler suggest that the vast supplies of energy liberated in supernovae outbursts may be connected with

this reaction since the light curve of a supernova closely resembles the decay curve of Cf^{254}. In any event it is not possible to explain all the different kinds of isotopes of all the elements with their proper abundance ratios without invoking several different mechanisms.

Thus helium clearly comes from hydrogen burning; carbon, oxygen, neon and magnesium come from the subsequent consumption of helium in three particle reactions.

The abundance peak at iron is presumed to be produced by an equilibrium between various types of nuclei set up at enormously high temperatures and densities—possibly in supernovae.

Some nuclei must be produced by neutron captures that proceed at a leisurely rate. Thus, whatever produces the neutrons in the cores of the S stars does so in such a way that the neutrons are captured by nuclei at such a rate that these new nuclei may undergo beta decay (ejection of positive or negative electrons) before a new neutron is captured. Other nuclei must be produced by the rapid capture of neutrons by nuclei in such a way that beta decay cannot occur before another neutron is captured.

No known nuclear process is capable of producing lithium, beryllium and boron in stellar interiors. They are probably produced by the breaking down of heavier elements, such as iron and magnesium, by high energy particles in stellar atmospheres. Presumably these high energy particles originate as a consequence of electromagnetic disturbances in the neighbourhood of sunspots and are probably connected with such things as flares.

Finally, in certain of the magnetic stars, some elements such as the rare earths have enormously enhanced abundances. The nuclei of these elements also are presumably produced by electromagnetic effects of the magnetic fields.

Nuclear astrophysics is a new field, but it promises to be one of the most exciting aspects of stellar astronomy. The connections between stellar evolution, stellar composition differences and element building appear to be dimly perceived and should provide an enduring impetus to future research. *See* ASTROPHYSICS; COSMOGONY; COSMOLOGY; CELESTIAL MECHANICS; *see* also Index references under "Star" in the Index volume.

BIBLIOGRAPHY.—R. H. Baker, *Introduction to Astronomy* (1957); W. T. Skilling and R. S. Richardson, *Brief Text in Astronomy* (1954); P. van de Kamp, *Basic Astronomy* (1952); L. Goldberg and L. H. Aller, *Atoms, Stars and Nebulae* (1943); O. Struve *et al., Elementary Astronomy* (1959); C. H. Payne-Gaposchkin, *Stars in the Making* (1953); F. Hoyle, *Frontiers of Astronomy* (1955); L. H. Aller, *Astrophysics: Nuclear Transformations, Stellar Interiors and Nebulae* (1954); M. Schwarzschild, *Structure and Evolution of the Stars* (1958).

(L. H. A.)

STAR APPLE (*Chrysophyllum cainito*), an evergreen tree of the family Sapotaceae (*q.v.*), called also cainito, native to tropical America and cultivated for its edible fruit. The pleasantly flavoured fruit, 2 to 4 in. across, is apple shaped.

In a cross section of the unripe fruit the seed-bearing cells form a star.

STARAYA RUSSA, a town of the Russian Soviet Federated Socialist Republic, U.S.S.R., in the Leningrad *oblast*, in 58° N., 31° 22′ E., on the Polista, by means of which and Lake Ilmen it is brought into steamer communication with Leningrad. Pop. (1926) 21,511. Brine springs on the east of the town were used as a source for the supply of salt as late as 1865; afterward they were used only as mineral waters (temperature 51°–54° F.), having a great resemblance to those of Kreuznach in Germany. Thousands of visitors were accustomed to resort to them every summer, and because of this circumstance Staraya Russa was better built and better kept than any other town in the government of Novgorod. The inhabitants are supported chiefly by the summer visitors. There is a trade in rye, oats and flax.

Staraya Russa belonged to the republic of Novgorod, and suffered in the wars between Russia, Lithuania and Livonia. It was afterward annexed to Moscow.

STARA ZAGORA, the capital of a district of southern Bulgaria lying on the southern slope of the Karaja Dagh, 70 mi. N.W. of Adrianople, with which it is connected by railway. Population (1956) 55,322. The city is surrounded by vineyards, and has flour mills, breweries, brandy distilleries, copper foundries and tanneries. The production of silk and attar of roses is carried on in the district, which contains numerous mineral springs, and there is a trade in corn and wine. It is the junction of the trans-Balkan and Burgas-Philippopolis railway lines. It was destroyed during the Russo-Turkish War (1877–78) and rebuilt on modern lines. During the rebuilding, important Thracian, Roman, Byzantine and Turkish antiquities were discovered.

Stara Zagora, founded probably by the Thracians, was known to the Romans as Augusta Traiana, but afterward, to distinguish it from a Macedonian town of this name, it was named Beroe or Berrhoea. By the Turks the name was changed in the 17th century to Eski-Zagra or Eski-Zaara, whence its Bulgarian name.

STARCH. This is perhaps the most widely distributed substance in the vegetable kingdom and occurs, often in great abundance, in almost every plant. Pure starch has the appearance of a white, glistening, friable powder, and possesses a harsh feel when rubbed between the fingers; it is tasteless, has no smell, is not soluble in cold water and undergoes no change when exposed to the air. When examined under the microscope it is found to consist of granules of definite shape, size and appearance. These show very wide variations, some being minute while others attain a comparatively large size; they are, nevertheless, quite characteristic of the plant from whence they are obtained, and the identification of any particular starch by comparison with those of known origin is a more or less simple matter. Granules or cells which are usually oval in shape are found to be composed of a series of concentric layers arranged around a nucleus or hilum which appears as a dark spot, the outer layers being the oldest in point of growth.

Starch belongs to the group of carbohydrates in which are included the sugars, gums and cellulose, and is composed of the elements carbon, hydrogen and oxygen. Its constitution is very complex, but it is built up in the plant from two simple compounds, water and the carbon dioxide gas present in the atmosphere, through the agency of the green colouring matter, chlorophyll, under the activating influence of sunlight. The rôle of starch in plant economy is that of a reserve material, and as such it must be transferred from the leaves and other green parts of the plant, where it is formed, to more permanent quarters. To facilitate this migration it is broken down into sugars which dissolve in the cell juice, in this way passing through the cell walls of the plant finally to become reconverted into starch, either in the fruit or seeds, or in the bulbs and tubers as the case may be. It is thus rendered available as a source of sustenance for the young plant until it is sufficiently advanced to carry on the process of starch formation itself. Starch as it occurs in the granular packages visible under the microscope, is not present as simple molecules of the empirical formula $(C_6H_{10}O_5)_n$, but probably *as sheaves or groups of molecules of varying density of packing*, the individual molecules being associated together by secondary valence forces. Starch is a highly organized or "associated" substance. When heated with water it forms a paste, the physical characteristics of the resulting paste being different for each type of starch. Aside from the differences in these pastes due to the variety of starch grains, the properties are further modified by mechanical and chemical action. When starch is heated in plain water, it takes on water and hydrates. The granular packages first swell and then rupture to form a paste. The body and consistency of the pastes will vary greatly, depending on the severity and intensity of the cooking and agitating treatments. These modifications are due mainly to mechanical means—that is, merely tearing the granules without materially altering the chemical structure. In the presence of chemicals, however, particularly acids and enzymes, the molecule itself is degenerated, the products formed representing a breakdown of the original molecule. The type of starch degeneration product formed in the presence of these chemicals and the extent to which the original molecules are degenerated will depend on the nature of the chemical substance, concentration, temperature, etc. In this manner are made the thin boiling (low viscosity) starches, dextrins, syrups and sugars. This process is employed in the manufacture of corn or glucose syrups, used extensively in confectionery, and of glucose sugar used in

brewing. Hydrolytic degeneration is also effected by certain natural ferments or "enzymes" such as those in saliva, pancreatic juice, malted grain (diastase)—all germinated seeds, and it is the latter agent which is employed in the fermentation industries—brewing, distilling, etc.—converting starch in the materials used into material capable of fermentation by yeast.

Development of starch in the growing plant already referred to, and the rebuilding of the soluble products into the cellulose which constitutes the tissues, is carried out by ferments or enzymes elaborated by the plant itself, while a similar process goes on in the human system when starchy foods are eaten and converted into soluble sugars capable of assimilation.

Dry starch heated to upwards of 320° F. is transformed into dextrin, a pale, yellowish powder soluble in water, known as British gum. Dextrins of lighter colour are also made by spraying with aqueous acids and heating to lower temperatures.

A very sensitive reaction of starch is the production of an intense blue coloration with a solution of aqueous iodine, a characteristic much used in its identification. The average starch content of the principal starch yielding plants, together with the temperatures at which the purified starch gelatinizes when heated with water, are, according to Lippmann, as in the following table:

	Average Starch content %	*Gelatinizing temp.° F*		*Average Starch content %*	*Gelatinizing temp.° F*
Potato	18–20	145	Millet	53–55	..
Wheat	54–58	152	Peas	39–40	..
Barley	40–46	145	Lentils	39–40	..
Oats	35–38	..	Arrowroot	20–39	158
Rye	44–46	131	Tapioca	25–40	156
Maize	55–65	145	Yams	25–35	..
Rice	70–79	142	Sago	..	158

The storehouse, or starch-containing part of the plant, consists of a network of cells, within the walls of which the starch granules are closely packed together. The process of manufacture involves the rupturing of the enveloping cell walls and separation of the starch from the associated material by levigation.

Manufacture.—The manufacture of starch from roots and tubers is usually carried on in the locality in which they are grown; that from potatoes on the continent of Europe, principally in Germany, France and the Netherlands; arrowroot, cassava and tapioca are prepared from the roots and tubers of the maranta and manioc plants in the West Indies and South America; sago is manufactured from the pith of the sago palm and also from the yams or sweet potatoes in the East Indies and Borneo. Cereals can be more readily transported and stored, so locality is a matter of convenience; thus rice is used as a raw material in England, France, Germany and Italy and other parts of Europe, while maize is used very extensively in North America, and in England, France and Germany. Wheat is not largely employed.

The purposes for which starch is manufactured can be classified under three heads: (*a*) *Industrial.* These include sizing yarn and cloth in the textile industries; dressing cloth; thickening mordants and colours in calico printing; the preparation of British gums and dextrin, confectioners' glucose or corn syrups, and sugars for use in the fermentation industries. (*b*) *Food.* Used alone in the form of cornflour, arrowroot, tapioca, sago, etc., or in conjunction with other substances in the preparation of custard and blanc mange powders, sauces, cocoa, confectionery, etc. (*c*) *Laundry.* In pipe or crystal form and powders, both "thick" and "thin boiling" preparations.

The suitability of starches from various sources for these purposes depends on their cost, purity, the stiffness or viscosity of the paste they make with hot water and the size of the granules. Potato, maize, rice and wheat starches are those chiefly used for industrial purposes; rice, maize, arrowroot, cassava and sago for foodstuffs; and the small granuled starches, rice and maize, are used extensively in the laundry owing to the fact that the granules are required to enter the texture of the linen or cloth before becoming gelatinized by the hot iron.

The general principles for dealing with all types of roots and tubers are substantially the same, but of course vary in their elaboration and efficiency in different parts of the world, and they also only differ to some extent in detail when cereals are being treated. They are briefly as follows:—The tubers first go through washing machines where earth, stones, sand and grit are removed; they then pass to a rasping machine where all the cells are completely ruptured. The resulting pulp is washed in a fine stream of water on fine-mesh brass sieves, which separate the starch from vegetable fibre, etc.; the resulting milky starch liquor is run into vats to separate fine sand and grit by settlement, through fine silk sieves for further purification, and is finally allowed to run very slowly over long, shallow, wood troughs, runs or tables, where the pure starch settles out and the impurities pass off at the ends with the water. The pure white starch is washed from the runs by a jet of water and is pumped to either filter presses or hydro extractors, which remove all except about 40–45% of the water, leaving the starch in the form of a damp cake.

The cereals contain varying proportions of protein or gluten material associated with the starch, and special methods have to be employed in their preliminary treatment. The protein contained in rice is found to be rendered soluble by very weak caustic soda solution, and advantage is taken of this in the production of starch from the raw material. Maize in addition to protein contains an oil-containing germ which it is necessary to remove. In this case the grain is first steeped for 2 to 4 days in warm water containing 1% of sulphur dioxide, ground in burr stone mills, and the resulting diffluent mass diluted with water is passed through a long V-shaped tank where the germ floats and is skimmed off.

Wheat contains gluten which forms an elastic paste with water and cannot be removed by levigation. The procedure carried out in this case is to reduce the grain to flour which is then made into a dough with water, divided into small pieces and placed in a semicircular silk or fine brass sieve where a travelling roller presses out the starch which is removed by a fine spray of water. The starch liquor goes through a very similar purification process to that employed for maize, often being treated with dilute caustic soda solution to swell the gluten and assist its removal between the sieving processes. It is necessary to carry out the final drying process at very low temperatures to prevent the gelatinization of the starch granules. Several methods, similar in principle for starches from all sources, are in use in the large factories, and consist of either bringing the starch lumps or blocks in contact with a continuous current of warm air, or evaporating off the excess of moisture in a vacuum chamber. For producing the starch in the form of powder or flour the moist lumps are either placed on an endless band running backwards and forwards in an enclosed chamber through which warm air is drawn, or they are placed on trays mounted on a trolley and passed through a tunnel in contact with warm air. In the tropics drying is usually carried out in the open air.

Prism or crystal starch is prepared from maize or rice starch for laundry purposes, and is produced by forming the damp starch into blocks by draining off the water from the starch liquor in boxes with perforated bottoms covered with filtering cloths; the blocks are cut into two or more sections which are partly dried in a warm chamber; the yellowish brown crust which forms on the outside is scraped off, and the blocks wrapped in paper, returned to the drying chamber and dried very slowly. Under this treatment the blocks break up into irregularly shaped prisms technically termed "crystals." Pearl tapioca is made by heating the slightly moist starch from manioc in a steam-heated copper pan with constant stirring, some of the cells becoming gelatinized and adhering together in small, irregular masses. Further drying follows and the crude lumps are passed through rollers and graded.

Granulated or pearl sago is prepared by drying the starch in such a way that it forms a plastic dough in which a portion of the granules have become gelatinized, forcing this through sieves, and drying the resulting granules in the air.

Bibliography.—T. C. Taylor and R. P. Walton, "Characterization of Certain Starches and Their Amyloses," *Jour. Amer. Chem. Soc.*, vol. 51: 3431–3440, Nov. 1929; T. C. Taylor and G. M. Salzmann, "The action of aqueous alkali on starches, amyloses and modified starches," *Ibid*, vol. 55: 264–275, Jan. 1933; G. V. Caesar and E. E. Moore, "Consistency changes in starch pastes; tapioca, corn, wheat, potato, and sweet potato," *Ind. and Eng. Chem.*, vol. 27: 1447–1451, Dec. 1935; G. V. Caesar and M. L. Cushing, "Classification of starch derivatives; significance of the barium hydroxide method," *Ibid*, vol.

31: 921–924, July 1939; C. G. Caldwell and R. M. Hixon "A study of the periodic acid oxidation of starches and dextrins as a means of determining molecular size," *Jour. Biol. Chem.*, vol. 123: 595–606, April 1938. (H. Her.; G. V. C.)

STAR CHAMBER was the name given, because of the stars depicted on the ceiling of the main room, to a building in the king of England's palace of Westminster. It was constructed in the reign of Edward III and apparently from the first was one of the regular meeting places of the king's council. In the 16th century the name came to be applied also to the meetings there of councilors and judges sitting as a court.

This court inherited its jurisdiction from the medieval king's council (*see* Privy Council). The king of England, despite the evolution of the various courts of common law from the 12th century onward, had never parted with supreme jurisdiction or the right to supplement and reinforce the action of the ordinary courts. This right he exercised largely through his council, which dealt also with judicial business that fell outside the normal scope of those courts. Thus, to the council were referred, from at least the early 14th century, the petitions of suitors who sought the special intervention of the crown either because the ordinary courts had no remedy to offer in their cases or because the remedies were being frustrated by force or fraud or corruption. The council as a body did not itself deal with all these matters. Cases directly involving life or property were regarded as outside its scope and were generally left to the common-law courts. Many petitions, too, were about private matters between subject and subject, civil causes in which the crown's interests were not directly involved.

The council, while retaining undisputed control over all complaints addressed to it, fell into the habit of referring more and more of these cases of little public concern, and often great legal complication (*e.g.*, trusts and uses, or breach of oral contract), to the chancellor and a few other councilors and then, from the later 15th century, to the chancellor sitting alone. A similar course was adopted, rather more erratically, with the growing number of petitions from poor men and royal servants. Richard II gave the lord privy seal some special responsibility for these and by Richard III's time certain councilors were acting as "masters of requests"—perhaps as a "court" of requests—to deal with them.

The council as a body, however, generally retained for its own peculiar province the handling of cases involving public order, such as riot, breach of the peace, the maintenance of lawbreakers, the keeping of retainers and the corruption of officials in which the crown had a more direct interest. Here its procedures gave it considerable advantages over the ordinary courts. It was less bound by rigid forms. It did not depend upon juries either for indictment or for verdict. It could act upon the petition of an individual complainant or upon information received. It could put an accused person on oath to answer the petitioner's bill and to reply to detailed interrogatories. It had the help of the judges in reaching its decisions and its sentences were backed by the full majesty of the state and the law. On the other hand its methods lacked the safeguards which common-law procedures provided for the liberty of the subject, and parliaments in the 14th and 15th centuries, while recognizing the occasional need for and usefulness of those methods, had attempted to limit their use to causes beyond the scope or power of the ordinary courts. Thus the judicial activity of the king's council in these matters continued to be in the nature of a special intervention from outside the regular course of justice.

Under the Tudors, however, this activity became so frequent and continuous that it developed into a regular part of the normal machinery for enforcing the law. Henry VII (*q.v.*), having a council fully obedient to the royal will, made great use of its jurisdiction to repress lawlessness, particularly riot, livery, maintenance, embracery and corruption of sheriffs and jurors. His famous act of 1487, later misleadingly titled *pro camera stellata*, did not create a court of star chamber. But it did reassert statutory sanction for the regular employment of the council's procedures in such cases as those just mentioned. Also, using a common Tudor device, it placed upon certain councilors a special responsibility for dealing with them. Moreover, the councilors named included the lord chancellor and the two chief justices (or two other judges) of the courts of king's bench and common pleas. As these were all tied to Westminster during term time, the act confirmed the council's habit of dealing with such cases in term time at its headquarters, the star chamber. For it did not, of course, prevent other councilors, if available, from dealing with those causes, nor did it in any way limit the meetings in the star chamber to these matters or indeed to judicial business. Nevertheless, the presence of the judges helped to distinguish these meetings from others; and in general the act helped to give greater fixity of place and greater regularity of time to this aspect of the council's judicial activity.

It was during Thomas Wolsey's chancellorship (1515–29) that this activity, as well as the activity of chancery in civil equity (*see* English Law: *The Growth of Equity*), grew with greatest rapidity. He not only used it, with as much vigour as Henry VII, against riot, maintenance and so forth; he also used it with increased vigour against perjury, forgery, fraud, slander and any actions that could be considered as tending to breach of the peace. And he encouraged suitors to appeal to it in the first instance and not merely after failing to find an efficient remedy in the ordinary courts. Thus, with Wolsey this activity of the king's council began to take its place as a regular part of the administration of justice, as a regular "court" alongside the ordinary courts of law. At the same time, the continued growth and increasing definition of chancery's equity jurisdiction in civil suits between subject and subject; the final emergence of a distinct court of requests for poor men's causes; and such other developments as that of the high court of admiralty—all helped to define, as it were by subtraction, the function of the meetings of councilors and judges in public session in the star chamber. They became meetings, above all, for the exercise of the semicriminal jurisdiction of the king-in-council. Not exclusively so, for "all cases may be here examined and punished if the King will"; but matters of riot or tending to riot, robbery, libel, slander, perjury and offenses against royal proclamations were their primary concern.

The final stage in the evolution of this semicriminal jurisdiction of the king's council into the court of star chamber came with the emergence of a more clearly defined privy council. This was, again, the result of pressure of business, this time of executive and administrative business following upon the breach with Rome. The privy council, while retaining and frequently exercising (under the name of the council table), a "supereminent" jurisdiction, was evolved primarily to perform the advisory and executive functions of the older king's council. It attended upon the king; it met in private, all the year around, and often daily. In 1540 it began to have its records kept in a separate book by a clerk (later four clerks) of its own, distinct from the senior clerk of the council (who came to be called clerk of the star chamber). From that date its activities can thus be fairly clearly differentiated from the public sessions of councilors and judges held in the star chamber at Westminster, as a rule during term time only, and usually (at least by the time of Elizabeth I) on Wednesdays and Fridays. These meetings then gradually ceased to be thought of as meetings of the king's council and came to be regarded as sessions of a court of star chamber.

Even so, the distinction between this court and the privy council remained one of function and procedure rather than of persons: Certainly under Elizabeth I and the Stuarts membership of the court was, except for the presence of the judges, almost entirely confined to privy councilors. As a result, it retained much of the majesty, even if it lost the name, of the king's council. It also, of course, continued to use the conciliar procedures. Cases began upon petition or information. The accused, summoned by writ of subpoena, had to answer on oath and could be examined upon interrogatories framed by the prosecution. Depositions were taken from witnesses, but no jury was used. Proceedings were in English and much was done in writing. The punishments, which were arbitrary and sometimes made to fit the crime, included

imprisonment, fine, the pillory, whipping, branding, mutilation, but never death.

The court did useful work in enforcing the law where other courts were subject to corruption or perversion, in providing remedies where others were inadequate, and in supplementing local justice. Its popularity in the 16th century was great—it has been estimated that about 43,000 cases may have come before it during the reign of Elizabeth I. That popularity lasted well into the Stuart period, despite the court's increasing fees and the growing number of its officials. Sir Edward Coke, for example, described it as "the most honorable court (our parliament excepted) that is in the Christian world." However, its employment to enforce unpopular policies, particularly in ecclesiastical matters, made it increasingly hated by the parliamentary and Puritan opponents of Charles I and of Archbishop Laud. As a result it was abolished by an act of the Long parliament in 1641.

(R. B. Wm.)

STAR CLUSTER. In the astronomical sense, a star cluster is a stellar organization held together by the mutual gravitational attraction of its members, which are physically related through common origin. They move through space together like a swarm of bees. Binary and multiple stars, star clusters and star clouds are all examples of gravitational organizations of generically related stars. Star clusters are distinguished from star clouds by their subordinate size. In some instances, clusters appear as condensations or subsidiary organizations within star clouds.

On the basis of definition, there is no strict separation of multiple stars and very poor star clusters. The astronomer employs the descriptive term that is most convenient. As a working rule, a star cluster should contain enough stars to be describable statistically.

Related to star clusters are stellar associations, loose space groupings of stars of common origin without sufficient total mass in the group to remain together as an organization. In some cases, star clusters or multiple stars appear as nuclei within associations. The expansion rates of some associations have been measured. Associations are, in terms of the usual astronomical time scale, disintegrating rapidly. The stars composing them must have been formed comparatively recently. The measured rate of expansion of the ζ Persei association, for example, indicates that the stars composing the association must have been formed approximately 1,300,000 years ago.

Classes of Star Clusters.—In the galaxy, the huge stellar system of which the solar system is a member of insignificant size, two general classes of star clusters are found: globular clusters and galactic clusters. These two classes of objects may be briefly contrasted as follows: The globular clusters contain very large numbers—tens or hundreds of thousands—of stars. The galactic clusters typically contain rather few stars—only a few tens or hundreds. The brightest stars in a typical globular cluster are red. They are distended stars of low surface temperature. Generally, the brightest stars in a galactic star cluster are blue.

These two kinds of clusters, which differ so markedly in their physical properties, are found in characteristically different parts of the galaxy. Fig. 1 shows a schematic side view of the galaxy, with its in excess of 100,000,000,000 stars. It is primarily a flat system, roughly circular in plane view, with a diameter of the order of 100,000 light-years. Spiral arms, similar to those seen in many external galaxies, marked out by highly luminous blue stars, gas and dark, nonluminous material, exist in this plane. The galaxy is rotating around an axis perpendicular to the flat plane and passing through the centre, or nucleus, of the galaxy. At the sun's position, about 30,000 light-years from the galactic nucleus, approximately 200,000,000 years are required to complete one rotation. The galaxy does not rotate as a solid disk. Shorter or longer rotation periods are observed for objects nearer to, or more distant from, the centre.

The galactic star clusters are found primarily in the galactic plane and there predominantly within the spiral arms. As judged by the kinds of stars they contain, as well as by their motions and space distribution, the galactic clusters are related to the Milky Way star clouds. They, like the star clouds and other objects associated with the spiral arms, must have originated in the highly flattened outer parts of the stellar system. They move in orbits that are approximately circles centred on the galactic nucleus and lying in planes close to and nearly parallel to the galactic plane.

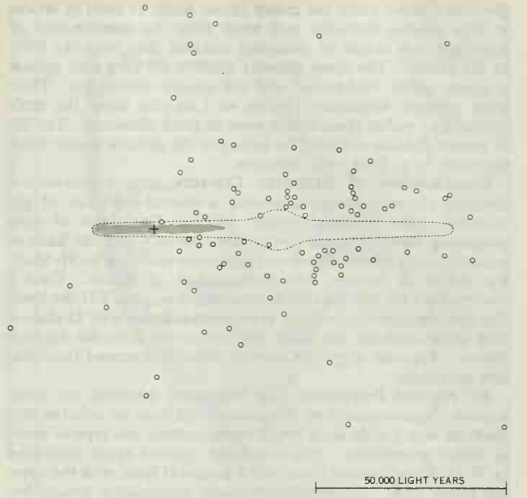


FIG. 1.—SPACE DISTRIBUTION OF GALACTIC AND GLOBULAR STAR CLUSTERS
Our stellar system as it would appear to an outside observer. Dotted line indicates probable outline of flattened lens-shaped system formed by the stars, as seen edgewise. Eccentric position of our sun is shown by cross. Known galactic star clusters are scattered among the stars in shaded region. Small circles represent globular clusters surrounding the star system like satellites

The globular star clusters, of which more than a hundred are now known to be members of the galactic system, differ markedly from the galactic star clusters in their distribution in space. The globular clusters are distributed in the form of an extensive halo about the galaxy (*see* fig. 1). They have their greatest space concentration in the region of the galactic nucleus and are probably generically related to the stars in the nucleus and the galactic halo. As a system, they rotate about the axis through the galactic centre, but much more slowly than the stars forming the main body of the galaxy. Individual globular clusters move around the galactic nucleus in greatly elongated orbits that lie in planes of various orientations.

Lists and Catalogues of Clusters.—Many of the most outstanding examples of star clusters of both classes were discovered by the early telescopic observers of the 17th and 18th centuries. In 1874 the French astronomer Charles Messier published a catalogue of nebulae and clusters in which he tabulated 57 star clusters. A systematic search of the whole heavens undertaken by Sir William Herschel and continued by his son Sir John Herschel led to the discovery of the majority of clusters known at present. The cluster catalogue published by Alter, Ruprecht and Vanysek (1958) listed 117 globular clusters and 576 galactic clusters.

Whenever possible, a cluster is designated by its number in the *New General Catalogue of Nebulae* (N.G.C.) or *Index Catalogues* (I.C.) compiled by John L. E. Dreyer. Less frequently, it is designated by its number in the catalogue published by Messier, or its name, such as the Pleiades, if it is a well-known cluster.

Clusters exist in external galaxies, as well as in our own. This discussion will be limited to clusters that are members of our own galaxy.

GLOBULAR STAR CLUSTERS

Total Number in the Galaxy.—The halolike distribution of

globular clusters about the galaxy causes many of them to appear at high galactic latitudes, well away from the great masses of stars and dark clouds of obscuring material that form the body of the galaxy. The dense globular clusters are thus seen against a sparse stellar background and are readily discovered. Their great intrinsic brightness (10,000 to 1,000,000 times the sun's luminosity) makes them visible even at great distances. The list of known globular clusters that belong to the galactic system must therefore be at least 50% complete.

Classification of Globular Clusters.—*By Appearance.*—Globular clusters, though displaying a general similarity of appearance, differ in stellar content and the distribution of stars composing them. Cluster compactness has provided the basis of a classification scheme based on appearance and used by H. Shapley and H. B. Sawyer, who distinguished 12 classes. Class I clusters have the highest central concentration, class XII the least. The globular clusters are fairly evenly divided among all 12 classes, with some tendency for fewer clusters to occur in the extreme classes. Apparent ellipticity may be added as a second classification parameter.

By Physical Properties.—The integrated spectrum or, alternatively, measurements of integrated brightness at selected frequencies over a wide wave-length range provide one type of index of stellar population. The integrated spectral types estimated by N. U. Mayall extend from mid-A to mid-G type, with the average being F7.6. Published multicolour photometric results are insufficient to provide analogous colour statistics. Other physical properties of the clusters that might be utilized as classification parameters are numbers and properties of variable stars (*see* below) within the cluster, character of the magnitude-colour diagram of the cluster, etc. Much remains to be learned from such classification parameters and combinations of them.

Stellar Content of the Globular Clusters.—*Colour-Magnitude or Magnitude Spectral-Type Diagrams.*—The stellar content of a cluster can be determined if for large numbers of individual stars in the cluster there are established at least two properties: magnitude (or brightness) and colour, or magnitude and spectral type. Fig. 2 displays a schematic magnitude-colour diagram for a typical globular cluster. The brightest stars in the globular cluster are red, cool stars. There are no very bright blue, hot stars so characteristic of the great majority of galactic clusters. The hot, blue stars (some are exceedingly blue) in the globular clusters are intrinsically very much fainter than stars of the same colour located in the vicinity of the sun. In fact, in the globular clusters (*see* fig. 2) the bluer a star, the fainter it is. In the solar neighbourhood, the bluer a star, the brighter it is. The magnitude-colour diagram of a globular cluster bears little resemblance to the magnitude-colour diagram of stars found in the vicinity of the sun. It is, presumably, characteristic of the nuclear region of the galaxy.

Variable Stars.—The presence of great numbers of variable stars in globular clusters was demonstrated by S. I. Bailey of Harvard College observatory. The overwhelming majority of the more-than-a-thousand variables known in globular star clusters have periods of less than a day. These are the cluster type or RR Lyrae variables; they occur only in the small region of the magnitude-colour plane (fig. 2) labeled variable star gap. All stars that fall in this narrowly defined region are variable stars. The light variation of a cluster-type variable is caused by pulsation; *i.e.*, by rhythmic contraction and expansion of the star.

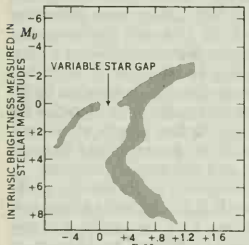


FIG. 2.—SCHEMATIC MAGNITUDE-COLOUR DIAGRAM OF GLOBULAR STAR CLUSTER
Shaded areas indicate regions of the plane occupied by globular cluster stars. Variable star gap indicates region in which cluster variables occur

Globular clusters differ markedly in their variable star content. Thus, the well-known cluster ω Centauri contains in excess of 160 cluster variables, while 47 Tucanae contains almost none.

Over 100 periodic or semiregular variables with periods longer than one day are known in globular clusters; they are brighter than the cluster-type variables. Approximately half of these stars have periods between 1 and 40 days; they may be regarded as Cepheidlike variables characteristic of the globular-cluster population.

It is remarkable that in all the known globular clusters only three eclipsing variable stars have been discovered. Conditions within globular clusters, unlike conditions within galactic clusters and the neighbourhood of the sun, are apparently unfavourable for the formation of double stars.

The cluster-type variables in a cluster are all of very nearly the same brightness. Under the assumption that the brightnesses of these stars are the same from one cluster to another and, further, that they are the same as the brightnesses of the RR Lyrae stars in the solar vicinity, the astronomer possesses a powerful means of determining globular-cluster distances. He can, from the measured motions and apparent brightnesses of the RR Lyrae stars in the vicinity of the sun, calibrate or determine the brightnesses of these nearby variables. He can then measure the apparent brightnesses of the cluster-type variables in a globular cluster and from the difference in the apparent brightnesses of the variables in the cluster and the derived brightnesses of the nearby variables infer the distance of the globular cluster. The cluster distances measured in this way range from 12,000 light-years to more than 100,000 light-years.

The determination of these enormous distances with a reasonable degree of accuracy constitutes one of the triumphs of modern astronomical investigation. A new era in the astronomer's understanding of the galaxy began when Shapley used such distances to investigate the space distribution of the globular clusters and, from that distribution, derived the direction of and distance to the galactic centre.

Spectral Characteristics of Stars in Globular Clusters.—The spectra of stars in globular clusters differ from those of stars in the solar neighbourhood or in galactic clusters. In globular cluster stars the spectral lines arising from the metals are very weak. This fact has caused difficulties in assigning spectral types to individual stars in clusters, especially when small dispersion spectra have been employed. The degree of line weakness appears to be consistent within any one globular cluster but to differ from one cluster to another.

The phenomenon of the weak spectral lines of the metals in stars in globular clusters may be explained in terms of an underabundance of metals (taking the local stars as normal), ranging from a factor of 20 to 200 or more in different globular clusters. Such a difference in the abundances of the elements in stars in globular clusters and in galactic clusters (and in different parts of the galaxy) must arise from differences in the composition of the primeval material from which the stars were formed. The differences constitute an important fact to be included in any theory of the formation and evolution of stars in the galaxy.

Size and Structure of Globular Clusters.—Specification of a precise diameter for a globular cluster is difficult. The stars simply thin out at greater and greater distances from the cluster centre. The linear diameters of different globular clusters range from 50 to 300 light-years, not counting the outermost stars. Clusters rich in stars are generally larger than those poor in stars. The space density of stars in all globular clusters is very much higher than it is in the general region of space around the sun or even in galactic star clusters.

A typical globular cluster is slightly flattened, as though the cluster were rotating. The flattening reveals itself through the elliptical form of the projected image of the cluster, a directly observable quantity. It is not possible to determine independently the true flattening of a cluster and the inclination of its equatorial plane to the line of sight. However, it is possible to say, from the results at hand, that some clusters have equatorial diameters at least 25% greater than their polar diameters.

The kinetic theory of gases has stimulated much work on the

statistical laws governing the distribution of stars in globular clusters. No completely satisfactory representation of an observed distribution law has yet been found. The studies that have been made, of course, have involved only a few thousands of the brightest (and more massive) stars. The fainter objects in the globular star clusters have not been adequately considered in any of these treatments.

GALACTIC STAR CLUSTERS

Number of Galactic Clusters in the Galaxy.—The galactic clusters located in the flat-disk portion of the galaxy and concentrated in the spiral arms of the galaxy are far more numerous than the globular clusters. The catalogue published by Alter, Ruprecht and Vanysek listed 576 galactic star clusters, but in contrast to the situation of the globular clusters only a very small portion of the existing galactic star clusters have been found. The reason is, of course, that the galactic star clusters are located in the part of the Milky Way in which much dark obscuring material exists; we can therefore survey only a small portion of the total volume of the Milky Way in which these clusters are located. It is estimated from statistical studies that at least 40,000 galactic clusters exist throughout the entire galaxy.

Classification of Galactic Clusters.—*By Appearance.*—The general purpose of any classification of clusters made on the basis of appearance is to make possible selection of homogeneous groups of objects. Among the classification schemes proposed for the galactic clusters, the one devised by Trumpler has been the most successful. He assigned three-symbol designations to each cluster on the basis of: (1) the degree of central concentration (consideration was given to the fact that the cluster was "detached" or involved in a dense star field); (2) the range in luminosity of the cluster stars; (3) the number of stars contained in the cluster. Further notations were employed to indicate peculiarities (elongated or unsymmetrical form, nebulosity, etc.) of a few clusters. Clusters that have the same three-symbol designation in this system appear to be quite similar in linear dimensions.

By Physical Characteristics.—Trumpler also devised a classification scheme based upon the nature of the magnitude-spectral-type diagrams of the clusters. Typical diagrams are exhibited in fig. 3. A two-symbol designation characterizes the cluster.

Class 1 includes those clusters that contain main sequence stars. Class 2 clusters have a small or moderate number of red giant stars in addition to main sequence stars. Class 3, of which there are few examples, is composed of clusters in which the number of giant stars is very considerable.

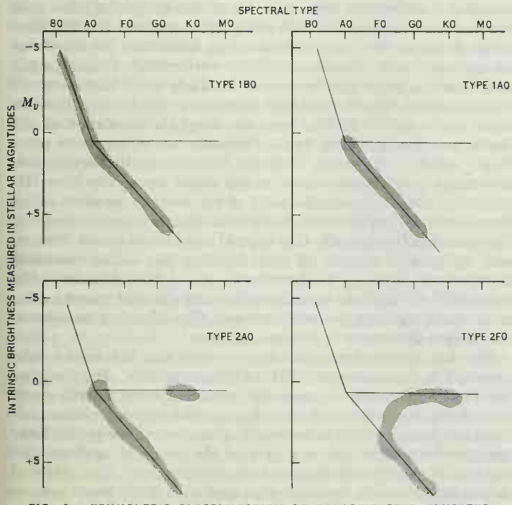


FIG. 3.—TRUMPLER'S CLASSIFICATION OF GALACTIC STAR CLUSTERS

Diagrams showing two principal classes 1 (above) and 2 (below), with two spectral types, or subclasses, for each class. Shaded areas indicate areas of the plane occupied by stars in galactic clusters

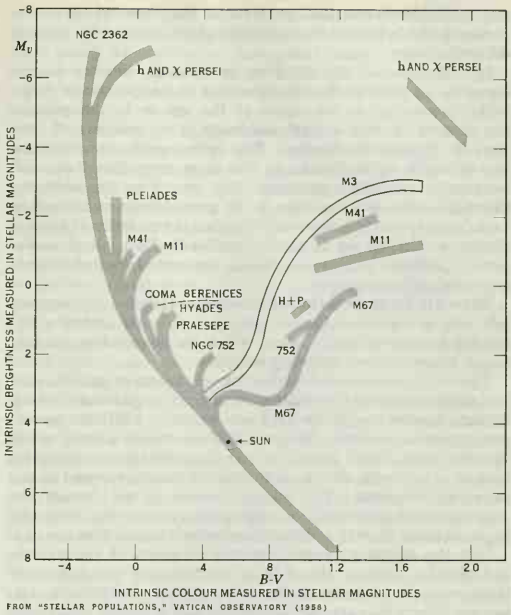


FIG. 4.—COMPOSITE MAGNITUDE-COLOUR DIAGRAM OF GALACTIC AND GLOBULAR STAR CLUSTERS. M3 IS A GLOBULAR CLUSTER; OTHERS ARE GALACTIC CLUSTERS. THE POSITION OCCUPIED BY THE SUN IS INDICATED

In each of these three classes the clusters may again be distinguished by the earliest spectral type found in the cluster. For example, 1B3 designates a cluster composed of main sequence stars, the earliest of which is B3.

Trumpler made the significant discovery that in this classification scheme not all classes are occupied. He found no class 2 objects with stars earlier than B4, for example. This classification scheme relates to the larger, more general questions of the stellar content of the clusters, which, we know now, relate directly to cluster age. Trumpler's discoveries have given impetus to studies of stellar evolution.

Stellar Content of the Galactic Clusters.—An outstanding characteristic of the galactic star clusters is that they are much richer in highly luminous stars than the general solar neighbourhood, even though they are found in the same region of the galaxy as the sun. Stars in clusters are often thousands of times more luminous than the sun; in fact, the clusters contain some of the most luminous stars known. The clusters are also found to be much more frequently associated with gas and dust, which appear as bright nebulosity when illuminated by the cluster stars, than would be expected on the basis of chance alone.

Recent observations of the colour-magnitude diagrams of galactic clusters (*see* fig. 4) indicate the kinds of stars found within the clusters and illustrate in detail Trumpler's classification scheme as well. Of particular interest is the position of the turnup of the upper portions of the main sequences of the clusters, a feature early noted by Trumpler and now known to be related to cluster age. The great majority of galactic clusters observed are far younger than the globular clusters or, presumably, than the general group of stars forming the main body of the galaxy. As a star grows old, its brightness and colour change. The forms of the main sequences of the various clusters have given us our

most valuable observational evidence of the shapes of the evolutionary paths in the colour-magnitude plane followed by stars of different masses.

Investigations of the numbers of stars of different brightnesses in the galactic clusters show that, in comparison with the stellar population in the region of the sun or in the globular star clusters, there is a great deficiency in the numbers of faint stars in the galactic cluster. This phenomenon, reflecting the way in which matter breaks up into stars, must arise from differences in conditions obtaining: (1) very long ago when the globular clusters and the stars in the general body of the galaxy were formed; and (2) more recently when the majority of galactic clusters condensed out of the interstellar medium. Such evolutionary problems as these are among the central ones of modern astronomical investigation.

Variable Stars in Galactic Clusters.—Galactic clusters contain very few intrinsically variable stars; they do contain a significant number of double stars, which are observed as spectroscopic binary stars or eclipsing variables.

The small population of intrinsic variable stars in galactic clusters differs radically from that observed in the globular clusters. Galactic clusters contain no RR Lyrae variables, which are numerous in globular clusters. However, they do contain a few Cepheid variable stars. These galactic-cluster Cepheids appear to be identical in nature with the classical Cepheid variables found in the solar neighbourhood. The galactic clusters do not contain any counterparts of the periodic or semiregular variables that are found in the globular clusters and that have periods longer than one day.

Like the differing colour-magnitude diagrams of the galactic and globular clusters, the differing variable-star populations of these two types of clusters indicate that globular and galactic clusters are objects that are very different in stellar content and physical characteristics. Their ages and their past histories must also be quite different.

Strange Stellar Types in Young Clusters.—The stars in galactic clusters are generally similar to the stars in the vicinity of the sun with respect to spectral and other physical characteristics. Notable exceptions to this rule occur in some very young clusters, which characteristically contain very bright stars and are frequently associated with extensive clouds of dust and gas. They also contain strange stellar types—stars that, as judged on the basis of the common variety of nearby stars, are too red (too cool) for their brightness (roughly solar brightness or fainter). In some instances these young groups contain examples of the very interesting T Tauri variety of erratic intrinsic variable stars. The T Tauri variables have intrinsic brightnesses similar to, or fainter than, the sun; their spectral types are G or later. It is presumed that these strange stellar types represent stars less fully developed than stars of similar masses in older clusters or in the vicinity of the sun. Much remains to be learned from the study of such stars.

Sizes and Structure of Galactic Clusters.—The limiting diameters of galactic clusters vary between 15 and 75 light-years; the richer clusters are larger than the poor ones. Stars of different brightnesses or masses do not show the same space distribution throughout the cluster. The brighter (more massive) stars within any cluster are generally found concentrated near the centre of the cluster. The fainter (and less massive) stars are spread over a larger volume of space and indicate a larger diameter. The apparent diameter of a cluster estimated on a photograph is generally no more than one-third of the limiting cluster diameter for the faint stars. The galactic clusters are far less regular in form than the globular clusters. Diameters are therefore especially difficult to estimate.

In the central, and therefore denser, parts of galactic clusters, each star is, on the average, separated from its nearest neighbour by a distance of approximately 1½ light-years. However, in the central portion of rich clusters the space density of stars is at least a hundred times as great as the space density of stars in the solar neighbourhood. If we were to construct a model of an average galactic cluster, using pinheads for stars, we should require a few hundred pins. We should then have to scatter these throughout a sphere approximately 150 mi. in diameter. On the average, the pinheads would be some 4 or 5 mi. apart.

Disintegration of Galactic Star Clusters.—The stars in a cluster continuously interact gravitationally; they are in continuous motion. Chance close encounters between stars produce strong perturbations in their motions. As a result of such perturbations a star may be given sufficient velocity to escape from the gravitational field of the cluster. Through such "evaporation" of stars a cluster shrinks both in linear size and in stellar content. Very general considerations show that the stars of small mass evaporate from the cluster more rapidly than those of large mass. Even if they do not gain enough energy to evaporate from the cluster, the small-mass stars will tend to move to greater-than-average distances from the cluster centre because of perturbations; they form an extended halo about the cluster. This theoretical prediction may provide at least a partial explanation of the observation that the fainter members of a cluster occupy a larger volume of space than the brighter and more massive members.

The cluster tends to be disrupted by the differential rotation of the galaxy, as well as by the chance encounters that occur between a cluster and the complexes of dark material and stars as the cluster moves in its orbit around the galactic nucleus. The encounters between the cluster and cloud-star complexes produce more important disintegration forces than the differential galactic rotation. The smaller the diameter of the cluster and the greater the total mass of stars in the cluster the longer the time required for disruption. For a cluster like the Pleiades, the disruption time is of the order of 1,000,000,000 years.

See STAR; NEBULA; ASTROPHYSICS; ASTRONOMY.

BIBLIOGRAPHY.—P. Ten Bruggencate, *Sternhaufen* (1927), a monograph dealing critically with only a part of the field; Harlow Shapley, "Star Clusters," *Harv. Obs. Monogr.* (1930), a general treatise on globular and galactic star clusters, containing full bibliography prior to 1930, "Stellar Clusters," *Handbuch der Astrophysik*, vol. ii, ch. 5 (1933); R. J. Trumpler, "Preliminary Results on the Distances, Dimensions and Space Distribution of Open Star Clusters," *Lick Obs. Bull.*, vol. xiv, p. 154 (1930); Per Collinder, "On Structural Properties of Open Galactic Clusters and Their Spatial Distribution," *Ann. Lund Obs.*, no. 2 (1931); H. Mineur, "Equilibre des nuages galactiques et des amas ouverts dans la voie lactee, evolution des amas," *Ann. Astrophys.*, vol. ii, no. 1 (1939); A. J. Deutsch, "Principes fondamentaux de classification stellaire," *Colloq. int. Cent. nat. Rech. sci.* (1955); L. Spitzer, "Disruption of Galactic Clusters," *Astrophys. J.*, vol. 127, p. 17 (1958); G. Alter, J. Ruprecht and V. Vanysek, *Catalog of Star Clusters and Associations* (1958). (H. F. WR.)

STARETS (plural *startsy*), a holy man, usually a monk, commanding wide popular respect for his gift of spiritual guidance, in the Russian church of the 18th and 19th centuries. The fact that most of the *startsy* were monks meant that their influence was unconnected with priesthood. They were mainly simple people who relied on direct guidance from the Holy Spirit and were not connected with the official church hierarchy. The Church Slavonic word *starets* in the Middle Ages was simply a translation of the Greek word *geron*, which, in the Orthodox world, meant no more than "monk." However, with the advance of the hesychastic movement (*see* HESYCHASM), which aimed essentially at a reformation after the monastic ideal of the ancient eremitic settlements in the Egyptian and Palestinian-Syrian deserts, the word began to signify especially that type of aged monk whose "inward activity" (*noera ergasia*) of contemplation was so far developed that he was chosen by younger monks as a spiritual leader. The way of life of the skete (ascetic settlements grouped round a common church), arising in connection with this reforming movement, encouraged the *starets* in this special sense.

The first great *starets* on Russian soil was Nil Sorski, who preached the skete manner of life to Russian monks. His disciples, the Transvolgian *startsy*, concerned with the purely ascetic, moral way of life, were defeated in a dispute with the Josephites (followers of Joseph, abbot of Volotsk), who clung to the public functions of the monasteries; *e.g.* care of the poor and aged and the training of young men for the church hierarchy. The ideals of ascetic sanctification, however, were kept alive in the desert monasteries through the writings of Nil Sorski, and in the 18th century they produced the true Russian *starets* in the narrow sense of the word. This development was assisted by the neohesychast ideas

from Mt. Athos, which the *starets* Paissy Velichkovsky (d. 1794) introduced to the Slav world through his Church Slavonic translation of the hesychast *Philokalia* (*q.v.*) of Nikodimos Hagioritis.

A further step was taken when *startsy* became not merely spiritual leaders inside the monasteries but also pastors and counselors of the laity (*see* SERAPHIM, SAINT). People now began to make pilgrimages to the *startsy*, the most famous of whom were those of the monastery of Optina Pustynye, in central Russia: Leonid (d. 1841), Makarios (d. 1860), and Ambrosius (d. 1881). Also important for Russian spiritual life in the 19th century were the two former bishop *startsy*, Ignatius Bryanchaninov (d. 1867) and Feofan Zatvornik (Theophanes the Recluse; d. 1894), the first for his ascetic writings and letters and the second for his modern Russian edition of the *Philokalia*. A fine example of the type of spirituality produced by the *startsy* teaching is the anonymous 19th-century *Way of a Pilgrim*. Dostoevski made the *startsy* famous in world literature by his *starets* Zossima in *The Brothers Karamazov*. *See* also MONASTICISM: *Christian Monasticism: Eastern Monasticism*.

See I. Smolitsch, *Leben und Lehre der Starzen*, 2nd ed. (1952).

(F. v. L.)

STARFISH. This name covers three classes of the echinoderms, (1) sea stars or Asteroidea, (2) brittle stars or Ophiuroidea and (3) feather stars or unstalked Crinoidea. The Asteroidea and Ophiuroidea are free-moving and live mouth downward; the feather stars, though they can swim or crawl, live most of the time attached to some object by their dorsal cirri, mouth upward or outward.

1. Sea stars, to which the name starfish is usually restricted, include many different forms, of which the most familiar are *Asterias* of the North Atlantic coasts and its relatives in the North Pacific. These animals consist of a central body radiating out into five tapering arms. The flattened upper surface is covered with a leathery skin, in which are small lobed plates of carbonate of lime, many bearing prickles, together forming a jointed reticulate skeleton. Multitudes of small pincerlike pedicellariae are present. A sea star of this type can crawl over any surface and can squeeze its supple body through incredibly narrow crevices. The rate of progress is about six inches a minute. The common sea star of the Atlantic coast of North America south of Cape Cod (*Asterias forbesi*) is very destructive to mussels and oysters, which form its favourite food. In attacking oysters and similar bivalves the sea star fixes the suckers of one or two arms to one valve, and those of the opposite arms to the other. It then begins to straighten out its rays. The oyster can withstand a strong sudden pull but not a continued pull, and eventually it opens its shell. The sea star then extrudes its stomach through its mouth, digests the oyster and after the meal is finished withdraws its stomach. Sea stars begin to eat voraciously when very young. One less than $\frac{3}{8}$ in. across ate over 50 young clams of half that length in six days. A sea star may become sexually mature in less than a year, then producing many thousands of young.

These familiar sea stars, however, are scarcely representative of the group as a whole. Sea stars vary in shape from sharply stellate with long tapering rays, or pentagonal with the angles of the pentagon produced into slender spikelike arms, to pentagonal or almost circular; but most of them are stellate. The great majority have normally 5 rays, but some have 6, and a number have more, in one type (*Heliaster*) from 20 to 44. In size they vary from $\frac{1}{2}$ in. to more than 3 ft. across. Most of them are predacious, feeding on mollusks, barnacles, worms, crustaceans and other smaller creatures which they take into their stomachs, but many are scavengers, feeding on detritus or swallowing mud and digesting out of it the organic matter. Most commonly seen of the exotic sea stars is a large one more than 1 ft. across, common in tropical American waters, in shape a low pyramid with a stellate base (*Oreaster*), which is widely sold as a curio. The cushion stars, from the Indo-Pacific region, are pentagonal to circular, 6 to 8 in. across, greatly swollen, with broadly rounded edges, and look like small cushions. Along the sandy or muddy coasts of all warmer seas live two types of sea stars which, though common, are usually unfamiliar. One of these (*Astropecten*) is regularly star-shaped, with stiff arms bordered with large plates. It usually lives just beneath the surface of the sand. The other (*Luidia*) has 5 or more broad, flat and brittle sharp-edged arms. It is said sometimes to swim by means of the podia. Sea stars are eaten by gulls, crows, ravens, foxes and various fishes. Their chief parasites are crustaceans.

2. Brittle stars or Ophiuroidea (snake tails) differ from sea stars in having long and slender arms, which are quite distinct from the usually pentagonal or circular body. The digestive and genital organs are wholly within the body, except in a few cases where the latter extend into the arm bases. There is no vent, and pedicellariae are never present. They are the most active of all echinoderms. They live among seaweeds or eelgrass, in chinks and crevices of rocks or corals, buried in mud, or lying on the sea bottom in deep water. Some are symbiotic or semiparasitic in sponges or on various coelenterates, feather stars, or sea urchins. In the tropics many are brilliantly coloured. Several brittle stars are hermaphroditic, some live in pairs, male and female, and some brood their young. Many are important as food for fishes, in the North Atlantic particularly cod and haddock.

3. Feather stars (unstalked Crinoidea) are distinguished as Comatulida (hair stars) from various other crinoids that at earlier periods of the world's history lost the stem. The stem is not really absent but is reduced to a knob, from which arise numerous jointed appendages (cirri) by which the animal clings to some support. From the small body extend 5 arms (in 2 genera 10) which may remain undivided, but most commonly fork near the base or branch several or many times. The arms are fringed with slender branches (pinnules), giving a feathery appearance. Though usually sedentary, feather stars can crawl about and some can swim. There are about 600 kinds arranged in 140 genera. They occur from the shore line down to about 18,045 ft. and are most diversified in the Indo-Pacific region, most abundant in the north. For further details of structure *see* ECHINODERMATA.

See also references under "Starfish" in the INDEX.

(A. H. CK.)

STARK, HAROLD RAYNSFORD (1880–), U.S. naval officer, was born in Wilkes-Barre, Pa., Nov. 12, 1880, and graduated from the United States Naval academy in 1903. He served as chief of the bureau of ordnance, in the navy department, 1934–37; he commanded cruiser division 3, U.S. fleet, until May 1938, and then cruisers, battle force. On Aug. 1, 1939, Stark became chief of naval operations with the rank of admiral, serving in that post until the spring of 1942, when he was appointed commander of U.S. naval forces in Europe. He left that command in 1945 and retired April 1, 1946. Stark was awarded the army distinguished service medal and was twice given the navy distinguished service medal. (J. B. HN.)

STARK, JOHANNES (1874–1957), German physicist who in 1919 received the Nobel prize for physics for his discovery that spectral lines are distorted in an electrical field—the so-called "Stark" effect (*q.v.*). He was born April 15, 1874, at Schickenhof, Bavaria. He studied at Munich and in 1900 became an assistant at the University of Göttingen. He went as professor in 1909 to the technical college at Aachen, in 1917 to Greifswald and in 1920 to Würzburg. He retired in 1922. Stark devoted himself principally to the study of the modern theory of radiation and the atomic theory. He discovered the Doppler effect in parallel rays, for which the Vienna academy awarded him the Baumgartner prize in 1909. His essays include "Prinzipien der Atomdynamik" (1910–15); "Die elektrischen Quanten, Die elektrische Strahlung, Die Elektrizität in Gasen" (1902); and "Atomstruktur und Atombindung" (1928). Stark was the founder of the *Jahrbuch für Radioaktivität und Elektronik* (1904), which he edited until 1919. A supporter of Hitler and an anti-Semitic racial theorist, he was president of the Reich Physical-Technical institute, 1933-39. In 1947 a denazification court sentenced him to four years in a labour camp. He died at Traunstein, Bavaria, June 21, 1957.

STARK, JOHN (1728–1822), U.S. soldier, was born at Nutfield, now Londonderry, N.H., on Aug. 28, 1728. During the Seven Years' War he served under Robert Rogers, first as a lieutenant and later as a captain. At the beginning of the Revolu-

tionary War he raised a regiment and as colonel did good service in the battle of Bunker Hill, the Canadian expedition and Washington's New Jersey campaign in the winter of 1776–77. In March 1777 he resigned his commission. Later in the year he was placed in command, with the rank of brigadier general of militia, of a force of militiamen, with whom on Aug. 16 near Bennington, Vt., he defeated two detachments of Burgoyne's army. For this victory, which did much to bring about the capitulation of General Burgoyne, Stark received a commission as brigadier general in the continental army (Oct. 4, 1777). He took part in the operations about Saratoga, and for a short time was commander of the northern department. In Sept. 1783 he was breveted major general. He died at Manchester, N.H., on May 8, 1822.

STARK EFFECT is the splitting of spectral lines observed when the radiating atoms, ions or molecules are subjected to a strong electric field. Soon after P. Zeeman discovered the magnetic splitting of spectral lines the electric analogue of the Zeeman effect (*q.v.*) was sought by many physicists, but success came first to Johannes Stark (*q.v.*) in 1913. The earlier experiments failed because a strong electric field could not be maintained in conventional spectroscopic light sources (*see* SPECTROSCOPY) because of the high electrical conductivity of luminous gases or vapours. Stark observed the hydrogen spectrum emitted just behind the perforated cathode in a canal ray tube. By placing a second charged electrode parallel and close to this cathode he was able to produce a strong electric field in a space of a few millimetres. At low gas pressure the mean free path of the atoms was much greater than the distance between the plates so that few conducting ions could be produced by collisions. In an electric field of the order of 100,000 v. per centimetre Stark observed, with a spectroscope pointing across the field, that the Balmer lines of hydrogen were split into a number of symmetrically spaced components, some of which were linearly polarized parallel to the lines of force, the remainder being polarized perpendicular to the direction of the field. This transverse Stark effect resembles in some respects the transverse Zeeman effect, except that the separations (in wave numbers per volt in the first case, per oersted in the second) are larger, the number of Stark levels increases with electron-shell number (*i.e.*, principal quantum number) and the intensity distribution among the components is different. A satisfactory explanation of the Stark effect in hydrogen, given in 1916 independently by K. Schwarzschild and by P. Epstein, was a great triumph for quantum mechanics (*q.v.*).

The Stark effect was also observed in the spectra of metals by employing special (LoSurdo) tubes containing a cathode of the metal under investigation, and viewing the light emitted near the cathode where a large potential gradient exists. In general, the electric splitting in spectra of metals is small compared with that in hydrogen, and unlike the Zeeman effect, which is symmetrical in spacing and intensity distribution of components, the Stark effect is usually only a displacement of unresolved components toward greater or smaller wave lengths. This effect does not depend in a simple way on the nature of the combining atomic-energy levels and for this reason the Stark effect in comparison with the Zeeman effect has little practical value in the analysis of complex spectra or of atomic structure. (W. F. M.)

STARLING, ERNEST HENRY (1866–1927), English physiologist, one of the foremost of his age, was born in London, April 17, 1866. Educated at King's College school, he entered Guy's hospital in 1882 and graduated M.D. in 1890. In that year he was appointed lecturer in physiology at Guy's. He was elected a fellow of the Royal society in 1899, and in 1900 became Jodrell professor of physiology at University college, London, where he worked until, in 1922, he retired from the Jodrell chair and was appointed Foulerton research professor of the Royal society.

The subjects for investigation that particularly attracted Starling were those physiological processes that seemed capable of interpretation in terms of chemistry and physics. The conditions determining transudation from the vessels and lymph flow occupied his attention for several years, and he showed that the hydrostatic and osmotic pressures within the vessels supplied the balance of force necessary to explain the previously perplexing experimental facts. His researches on the movements of the intestines, in conjunction with Sir William Bayliss (*q.v.*), demonstrated the neuromuscular mechanisms involved. Their discovery of secretin not only laid bare the way in which the secretion of the pancreas was called forth and adjusted, but also stimulated further research on the chemical integration of the body functions. By ingenious experimental methods Starling was successful in maintaining the mammalian kidney, isolated from all connection with the body, in a state of functional activity, thereby bringing to light new facts concerning renal secretion. His most important researches, however, were those dealing with the heart and circulation. He demonstrated the mechanism by which the heart is able to increase automatically the energy of each contraction in proportion to the mechanical demand made upon it and, apart from the nervous system, to adapt its work to the needs of the body.

During World War I Starling was director of research at the Royal Army Medical Corps college and engaged in devising defensive methods against poison gas. Subsequently, 1917–19, he was chairman of the Royal society's food committee, scientific adviser to the ministry of food and British scientific delegate on the Inter-Allied Food commission. He died on May 2, 1927.

STARLING, the name for birds of the passerine family Sturnidae, which contains about 103 species. The normal range of the family is Africa, Europe and Asia, to Australia, with most species in the tropics. Certain species have been introduced into various other parts of the world, such as Madagascar and Hawaii, and into North America, where the common starling of Europe, *Sturnus vulgaris*, is widespread.

JOHN H. GERARD

STARLING (STURNUS VULGARIS), COMMON TO EUROPE AND NORTH AMERICA (WINTER PLUMAGE)

Dark, iridescent plumage characterizes many starlings, but duller birds exist, and one species is mostly white. Bare, coloured areas of skin, or wattles, decorate the heads of some. Certain species of the tropical forest are mostly arboreal and eat fruit; those of open country walk about on the ground and eat insects. Several species are familiar birds, coming about houses and gardens, where they may cause annoyance by usurping the nesting sites of other birds.

Gregariousness is a common trait. The birds may feed or roost in large flocks, and some nest in colonies. The nest may be in a hole in a tree, bank or building or may be an oval, pensile structure. The eggs are plain or spotted.

A number of starlings are kept as cage birds, notably the talking mynas and several other southeast Asian species. The African oxpeckers, genus *Buphagus*, that eat ticks off the skin of hoofed animals are aberrant starlings. *See* also BIRDS. (A. L. RD.; X.)

STAR-NOSED MOLE (*Condylura cristata*), a North American species. In habits it resembles the other moles (*q.v.*) but is distinguished by the presence around the nostrils of a ring of rose-coloured feelers or tentacles, probably sensitive to touch and sound. *See* also INSECTIVORE.

STAR-OF-BETHLEHEM (*Ornithogalum umbellatum*), a hardy, spring-blooming, bulbous herb of the lily family, with narrowly oblong tapering perianth segments or floral leaves that are wide-spreading, white on the inside, green with white margins on the outside and longer than the stamens. A native of the Mediterranean region, it is distributed over Europe, Asia and Africa and has become naturalized in eastern North America. (J. M. BL.)

STARVATION is a state of extreme malnutrition caused by the long-continued deprivation of essential nutrients. It usually results from insufficient food intake, either because food is not available or because the person is unable to eat, due to illness or other factors. Starvation may also result from disturbances in digestion, absorption or utilization of food. *See* MALNUTRITION.

STASSFURT, a town in the former Prussian province of Saxony, Ger., and after 1952 in Magdeburg district; one of the chief seats of German salt production, it is on the Bode river, 20 mi.

S.W. of the city of Magdeburg by rail to Aschersleben. Pop. (1961 est.) 26,017. The systematic exploitation of the salt beds to which the town owes its prosperity dates from 1856 only.

Besides the rock salt, the saline deposits yield deliquescent salts and other saline products, which encouraged the foundation of numerous chemical factories. For many years the bulk of the world's potash supply originated in the Stassfurt mines.

STATE. The first studies of organized political society were made in Greece, mainly by Plato and Aristotle, in the 4th century B.C. The unit of study was the polis or city-state. The Romans, adapting Greek political ideas, applied them to a larger territorial unit, the *respublica*, which may be freely translated as public concern; or the *status rei publicae*, state of the public concern. Roman studies, however, emphasized the legal aspects of the political order. It is altogether probable that we derive the word state from the Roman *status*. In the 15th century Machiavelli used the term "state" extensively in his works.

Authorities agree on certain essential properties of a state—population, territory, a government clothed with a monopoly of force for the preservation of peace and order and having a plenitude of authority within the state independent of external control except that of international law. Since a state cannot function under that law unless generally recognized by the members of the community of nations, such recognition is often considered another property of the state. In addition most political theorists emphasize purpose or end in their definition of the state; this purpose is stated as the promotion of the common good or general welfare. This is a good which everyone shares and is distinct from the concept of the greatest good for the greatest number. Two terms are to be distinguished from the term "state," namely, society and nation. The free association of people in families and in social and cultural groupings constitutes a society; these groups lie largely outside of political control. The nation is a unit of society with a common language, tradition and culture which may not coincide with state boundaries. The states of the United States do not constitute states in the accepted meaning of the term. Nor may the term be applied to protectorates or semi-independent commonwealths, such as Puerto Rico. The term "states-general" (*q.v.*) derives from the French *états* ("estates"), referring to the various orders or conditions of society, as the clergy, nobility and third estate or common people, which were represented in European states-general of the medieval period.

See also SOVEREIGNTY.

BIBLIOGRAPHY.—Henry Sidgwick, *The Development of European Polity* (1893); Edward Jenks, *Law and Politics in the Middle Ages* (1905); W. W. Willoughby, *Fundamental Concepts of Public Law* (1924); R. M. MacIver, *The Modern State* (1926); H. J. Ford, *Natural History of the State* (1915); J. Maritain, *Man and the State* (1951).
(J. G. KN.)

STATE, ACT OF, the plea of act of state in English law means primarily a defense raised by the crown on behalf of a servant of the crown, who is sued in the English courts for some wrongful act ratified by the crown. In this sense the plea is not recognized by the English courts who, in rejecting it, uniformly held that, as the king cannot do a wrong so also "it follows that the king cannot authorize a wrong." The only exception to this rule is in the case of a wrongful act done to an alien who, in such a case, cannot sue (*Buron* v. *Denman* [1848] 2 Ex. 167) if the act was done outside British territory and was ratified by the crown, but later it was held in *Johnstone* v. *Pedlar* 90 L.J.P.C. 181 that if the act complained of was done to him in British territory he had the same right to sue as a British subject and the defense of act of state was, in that case, equally inapplicable. But the term is also used in another and less questionable sense. It will be entertained by English courts where the acts in respect of which an action is brought are "transactions of independent states between each other," such as those arising out of the execution, or non-execution, of a treaty, the succession of the crown to a foreign territory by cession or conquest, and the like. For example, a concession granted by the chief of native territory is not enforceable whether by Petition of Right or otherwise in British courts by the grantees against the crown, on succeeding to the territory by annexation (*Cook* v. *Sprigg* [1897] A.C. 572). So too the contractual obligations incurred, whether toward a British subject or an alien, by a foreign state are not binding on the crown, after its conquest and annexation by the crown, unless the crown expressly elects to adopt them (*West Rand Gold Mining Co.* v. *the King* [1905] 74 L.J.K.B. 753); for annexation is an act of state of an international character. In such cases the title of the subject has originated with a foreign government, not with the crown. If the contract or concession in question had originally been made by the crown itself, it would be another matter. Even in the case of a British protectorate the crown may plead act of state as a defense to claims in respect of expropriation of native territory because the crown in such a case is acting in territory which is not British (*see* the case of *Sobhuza II* v. *Miller* [1926] A.C. 518). So too, claims arising under a treaty are not enforceable against the crown unless and until the treaty, which is an international act, becomes part of English municipal law by being incorporated in a statute. The negotiation of the treaty is an act of state into which, and its meaning, the courts cannot inquire. In this sense the declaration of war and the making of peace by the sovereign is also an act of state. As Lord Somer put it in *Johnstone* v. *Pedlar* (Op. cit. at p. 794) "Municipal Courts do not take it upon themselves to review the dealings of State with State and Sovereign with Sovereign." On the other hand, the mere fact that some act, wrongful per se in English law, was done in execution, or alleged execution, of a treaty made by the crown against the person or property of a British subject is a defense that would never be entertained by the courts in action against a servant of the crown, for, while a treaty cannot confer new rights, enforceable at law, upon a British subject, neither can it deprive him of any of his existing rights under English law. (J. H. Mo.)

STATE, DEPARTMENT OF: *see* GOVERNMENT DEPARTMENTS.

STATE GOVERNMENT (U.S.). In 1776 the 13 British colonies in America proclaimed themselves to be independent states, and in 1781, under the Articles of Confederation, they formed a "perpetual union." The perpetuality of the union was not assured, however, for the articles which established that "firm league of friendship" provided also that each state should retain its "sovereignty, freedom, and independence," and in the following years the states showed a strong disposition to exercise the sovereignty thus retained. The federal constitution, drafted in 1787 and put into effect two years later, established a stronger national government, the "more perfect" union that was needed. The constitution made no reference to the troublesome question of sovereignty, leaving that to be disposed of by time—to be denied to the states, as it turned out, by the Civil War.

The federal constitution divides the powers of government between the national (commonly called federal) government and the states. The federal government has those powers which are expressly delegated to it by the constitution and the authority to make all laws that may be "necessary and proper" to implement the powers so delegated. The delegated powers are not numerous, but they are fundamental, including those to make war, conduct foreign relations, regulate interstate and foreign commerce and, of course, to levy taxes. The last-named power may be exercised not only to pay the debts and provide for the common defense, but also for the "general welfare." Federal power has been extended little by constitutional amendment, but it has been expanded by the liberal use that congress has made of the "necessary and proper" clause and by the judicial sanction that has commonly followed such use. It has been said that wars, emergencies and scientific developments and inventions do not increase federal power, but only furnish the occasion for the exercise of existing powers. As a legal proposition the statement is accurate, yet it would appear to be unrealistic to assume that the almost constant application of power to new conditions and subject matter does not in fact result in increasing power.

In some matters, not all of which are within the legislative power of congress, federal power and influence have been extended through a system of grants-in-aid to the states. The basic principle is that congress, acting under its authority to appropriate money for the general welfare or under some more specific authorization,

makes funds available to the states for a particular purpose, on condition that the states make appropriations for the same purpose and meet a standard set by congress in the use of the federal-state funds. The largest federal grants have been made for highway construction, social security and education.

There are some exceptions, but generally speaking the powers conferred by the constitution upon the federal government may not be exercised by the states. The constitution further limits the powers of the states by imposing upon them certain specific prohibitions. For example, the states are forbidden to levy import or export duties or pass any law impairing the obligation of contracts. By far the most significant limitations are those of the 14th amendment to the constitution, which stipulates that no state shall "deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws." The force of this restriction becomes apparent when it is remembered that in the United States substantive law as well as procedures must meet the requirements of due process of law.

The states have all powers not conferred upon the federal government by the constitution and not prohibited by it to the states. The powers of the states are thus residual. Despite the fact that federal power has been steadily increased, the powers left to the states are numerous and comprehensive. The national congress enacts statutes which define and provide for the punishment of crimes relating to interstate commerce, the currency, national security and a few other subjects; but the U.S. criminal codes are essentially those of the states. It is true that congress through its war, currency, banking, taxing and interstate and foreign commerce powers exercises a wide control over business, but the chartering and regulation of corporations, the laws of contracts, torts, negotiable instruments, sales and others closely related to the conduct of business are enacted by the state legislatures. Property laws are within the special province of the states, and so are those respecting wills and inheritance. Marriage, divorce and domestic relations fall within the jurisdiction of the states. The states determine who may vote, provided only that they make no discrimination because of race or sex. Most of the laws regulating political parties and elections are enactments of state legislatures. Local government is established and controlled by the states. Public education is a leading state function. The state is close to the individual. His birth certificate and his burial permit are granted by the state; and as he progresses from the cradle to the grave, the state protects his personal rights and secures his property; carries the major part of the cost of his education; licenses him to enter a calling or a profession; ministers to his convenience in numerous ways; and penalizes his delinquencies.

Each state has a written constitution of its own design, subject only to the requirement that it shall not violate the federal constitution. State constitutions follow the federal pattern in that they contain bills of rights and adhere to the cherished American principle of the distribution of powers among the legislative, executive and judicial branches. Most of the state constitutions provide in considerable detail for both the organization and functions of government and in consequence they are rather long documents, in contrast to the national constitution. Such elaboration of detail makes frequent amendment necessary, a process that has been pursued with vigour in a number of states. By contrast, the federal constitution of 1787 survived with only 24 amendments, whereas over half the states amended their constitutions more than 50 times.

The bicameral (two-house) legislature is used in all the states except Nebraska, which instituted the unicameral system in 1937. During and for some years following the Revolutionary War, the legislatures enjoyed public confidence and practically unlimited powers in state government. Chiefly because a number of legislatures failed in their trust, the states developed the practice of placing limits on their powers to deal with revenue, appropriations, borrowing, local government and certain other subjects. Furthermore, in the course of time, the state courts assumed a restrictive view of legislative powers. And finally, between 1898 and 1918 general dissatisfaction with legislatures led more than a third of the states to adopt the initiative and referendum. These limitations do not deprive legislative bodies of essential powers, and they may have the effect of improving their work.

In colonial America the governor exercised wide powers; in revolutionary times he was stripped of much of his authority; but from the mid-20th century, at least, the tendency was to make him an effective force in legislation, and, since 1915, to strengthen his control over administration. Although his authority over administration has often been impaired by the fact that much of the work of the state has been parceled out to boards, commissions and other relatively independent agencies, in perhaps half the states this situation has been partially corrected by the establishment of a limited number of administrative departments and the placement of a group of related services in each of them. In such states the department heads are commonly appointed by the governor and serve under his immediate direction, thus giving him an administrative position somewhat analogous to that of the president. In the great majority of the states the governor's influence over legislation is particularly significant, and it is explained by a variety of factors, including his political position, his permanent and sometimes professional staff, his veto power and the time limit under which a majority of the legislatures must work. The hand of the governor has been incidentally strengthened by the trend to increase the length of his term.

The system of law and justice in the states was inherited from Great Britain, but it has been modified in the states to fit their requirements. One such modification is in the matter of the selection of judges, most of the states following the elective system and limiting the term of office to four, six or some other number of years. The courts in all the states follow the practice of judicial review; that is, they pass upon the constitutionality of statutes where such determination is necessary to decision in cases appropriately before them. The administration of justice has fallen short of a desirable standard in some states because of the absence of any central directing authority. Several states rectified this condition by the establishment of a unified court system or by giving a judge of the supreme court the authority to assign extra judges in districts with congested dockets and to take other action to expedite the administration of justice.

With some notable exceptions, the states have not been conspicuously successful in meeting the responsibilities imposed by modern developments. For example, many of them neglected to provide for effective local government, to improve standards of administration or to modernize the judicial system.

For shortcomings in those services which call for large expenditures, such as highway construction, education and social security, the states with low per-capita income are only partially responsible themselves. And since federal taxes absorb a large part of taxable income, most states have difficulty in finding adequate revenue.

The states are hardly open to criticism for their failure in matters removed from their control by the revolutionary expansion of commerce and business. Until about 1890 they wrestled alone in the attempt to regulate railroads and other large interstate business. It became necessary for the national government to enter the field, using its power over interstate commerce. It may not be said, however, that the states are losing their place in the federal system. If the national government has assumed functions that were formerly left to the states, the states also have greatly expanded their functions, particularly in the direction of education, health, welfare and public convenience. The states remain vital units of government.

See also STATES' RIGHTS; UNITED STATES (OF AMERICA): *Administration and Social Conditions;* and sections on government in separate articles on each of the states.

BIBLIOGRAPHY.—The Council of State Governments, *The Book of the States, 1956–57* (1956); Austin F. Macdonald, *American State Government and Administration* (1955); F. G. Bates and O. P. Field, *State Government*, ed. by Pressly S. Sikes and John E. Stoner (1954); W. Brooke Graves, *American State Government* (1953); The Commission on Intergovernmental Relations, *A Report to the President for Transmittal to the Congress* (1955). (C. O. J.)

STATEN ISLAND, an island of New York, U.S., lying in New York harbour south of Manhattan between New Jersey and

Brooklyn, it forms Richmond county and Richmond borough, the third largest of the five New York city boroughs. Triangular in shape, the island has about 35 mi. of waterfront and an area of about 60 sq.mi. Although principally a residential suburb of the Borough of Manhattan and neighbouring New Jersey cities, Staten Island has considerable manufacturing, shipbuilding, oil refining and metalworking. The borough is connected with the surrounding areas by ferries and bridges. Population (1960) 221,991.

When discovered by Europeans, Staten Island was inhabited by a few Aquehonga Indians, a branch of the Raritans. Beginning in 1630 and during the next three decades several settlements were attempted by such Dutch leaders as Patroon D. P. de Vries, Gov. Willem Kieft (or Kiefft) and Cornelis Melyn. These attempts at settlement proved unsuccessful because of Indian attacks which, particularly in 1655, laid waste the white settlements on the island. The first permanent settlement was made in 1661 when the Dutch West India company, having purchased the Staten Island estate of Melyn, granted land to French Waldenses and Huguenots. This settlement was established at Oude Dorp (Old Town), a few miles south of The Narrows.

The island was named during the period of Dutch control for the states-general of the Dutch Republic. Following the acquisition of New Netherland in 1664 by the duke of York, Welsh and English farmers established homes and farms on the island. During the American Revolution Staten Island was the scene of talks between the three representatives of the Continental congress—Benjamin Franklin, John Adams and Edward Rutledge—and Lord Howe in an unsuccessful attempt at reconciliation. This meeting was held at the Billopp or "Conference" house in Tottenville on the southern shore of the island on Sept. 11, 1776. The island was the residence of Giuseppe Garibaldi, the Italian patriot, in the early 1850s during his exile from Italy. After the American Civil War industrialization increased while farming and fishing waned in importance. In 1898 Staten Island became one of New York city's boroughs. (*See* also NEW YORK [CITY]; RICHMOND.)

The island has numerous points of interest which include Barrett park, containing the Staten Island zoo; the museum of the Staten Island Institute of Arts and Sciences; the Staten Island Historical museum; the Sailors' Snug Harbor (a home for retired U.S. seamen); and the Billopp or "Conference" house. There are two liberal arts colleges on the island, the larger being Wagner college, a Lutheran coeducational institution founded in 1883, and Notre Dame College of Staten Island, a four-year Roman Catholic college for women, established in 1931.

There are many historic churches and houses still standing, among which is the Voorlezer's house (1696), one of the earliest known elementary school buildings in the country. (D. L. D.)

STATES-GENERAL, the English translation of (1) the *états-généraux* of France and (2) the *staten-generaal* of the Dutch Netherlands. The name in both cases signifies the assembly of the estates of the realm, as distinct from provincial assemblies of estates; it is thus the equivalent of the English parliament, of the Cortes in Spain or the diet of the Holy Roman empire, which the states-general resembled in their original constitution.

In France the states-general were first summoned in 1302 by Philip IV for the purpose of giving him moral support in his quarrel with Pope Boniface VIII. They consisted of representatives of the three orders of clergy, nobles and commons (*tiers état,* third estate), an arrangement which survived to the end, though their composition, as well as their effective powers, varied greatly at different times. For various reasons (notably the disorganization occasioned by the Hundred Years' War and the religious wars of the 16th century) they never secured the power over taxation and legislation early obtained by the English parliament, and after 1615 they were never summoned until 1789, when they transformed themselves into the National Constituent assembly. This assembly consisted of the representatives of the three orders sitting together, but the states-general had ceased to exist (*see* FRANCE: *History*).

In the Netherlands the convocation of the states-general, consisting of delegates from the provincial estates, dates from about the middle of the 15th century, under the rule of the dukes of Burgundy. The name was transferred, after the separation of the northern Netherlands from the Spanish dominions, to the representatives elected by the seven sovereign provincial estates for the general government of the United Provinces. The states-general, in which the voting was by provinces (each province having one vote), was established from 1593 at The Hague. The states-general came to an end after the revolution in 1795, with the convocation of the National assembly (March 1, 1796); but the modern Netherlands continues to use their name for its national legislature, which consists of the first chamber and the second chamber. *See* NETHERLANDS, THE: *History.*

STATES OF THE CHURCH: *see* PAPAL STATES.

STATES' RIGHTS, a phrase used to describe those governmental powers retained by the individual states of a federal union under the provisions of the federal constitution. In the United States, Switzerland and Australia, these powers tend to be those remaining after the powers of the central government are enumerated in the constitution. In Canada and the German Federal Republic, on the other hand, the powers of both the levels of government were defined by specific constitutional provisions.

In the United States the term also has been widely used for a variety of political programs. Before the Civil War it was the rallying cry of southern opponents of northern-inspired tariffs and northern proposals to abolish or restrict slavery (*see* NULLIFICATION). By the second half of the 20th century, it had become the catchword of opposition to integration of Negroes and whites in public schools, and frequently of critics of the growing power of the federal government.

The governmental authority actually remaining in the member governments varies greatly from federal system to federal system. It is perhaps greatest in the United States and weakest in countries such as Mexico and Brazil, where most power is concentrated in the central government and the states are little more than administrative agencies. States' rights (or authorities) dwindled rapidly after 1900 in the United States, and somewhat similarly in Canada and Australia. The major centripetal forces in the U.S. were in the increased activity of the central government through the power of congress to control interstate commerce, financial subsidies to the state governments or their subdivisions, concerted national measures for combating depression and participation in World Wars I and II.

In general, states' rights suffer more from financial subsidies of the federal government than from the commerce clause. These subsidies sometimes involve substantial central control in fields that are not included in the federal government's enumerated powers. If the state government rejects the subsidy, it presumably damages its area financially. The result of the subsidy process in the second half of the 20th century in the United States was that the federal government exercised varying degrees of influence, ranging from considerable to determinative, in public policy involving vocational education, public housing, airport construction, sewage plant construction, state and local road construction, fish and game conservation and other fields which a reading of the constitution would clearly indicate were states' rights. Canadian and Australian experience in the 20th century was similar, the central governments concerned moving through subsidies into many fields of state or provincial action.

Some political scientists believe that a federal system of government (*see* FEDERAL GOVERNMENT) and states' rights are merely transitional forms of government which will be ultimately succeeded by unified government. A quick glance at 20th-century federal systems would seem to confirm this view. However, a more thorough analysis indicates that states' rights may persist a long time. In three of the federal countries, the United States, Australia and Canada, representative study commissions in the decades of the 20th century concluded that it was desirable to maintain the autonomous authority of the state or provincial governments. The reasons given included the training of the citizenry in democratic processes through state and local self-government, the greater opportunity under a federal system to adapt governmental policies to the needs of particular areas and the deconcentration of power resulting from a federal system. There was

considerable evidence that for these reasons a large number of thoughtful people in the three countries wished to maintain states' rights.

BIBLIOGRAPHY.—K. C. Wheare, *Federal Government*, 3rd ed. (1953); report and studies of the U.S. Commission on Intergovernmental Relations (1955); report and studies of the Royal Commission on Dominion-Provincial Relationships (1938); S. Seabury, *The New Federalism* (1950); C. J. Bloch, *States' Rights* (1958); Arthur M. Schlesinger, "The States' Rights Fetish," *New Viewpoints in American History* (1922).

(G. C. S. B.)

STATE SUCCESSION: *see* SUCCESSOR STATE.

STATIC, a term in radio denoting conduction or charging current in an antenna (*q.v.*) resulting from physical contact between the antenna and charged bodies or masses of gas. In the United States this term is synonymous with atmospherics.

STATICS is a branch of mechanics concerned with the conditions under which a body remains at rest. Although statics relates strictly to both solid and fluid systems, it is common to give special names to the statics of some particular types of deformable bodies. In particular, elasticity (*q.v.*) and plasticity theory concern specific types of solids, while the statics of liquids is known as hydrostatics.

This article is devoted to a brief discussion of the statics of rigid bodies, a subject whose considerable technical importance will first be explained, and is to be regarded as a sequel to the article MECHANICS.

A Simple Problem of Rigid Statics.—Consider the simplest type of railway bridge, consisting essentially of a metal beam supported by two masonry abutments. Suppose that it is wished to estimate what supporting force is transmitted to the stationary beam at one of the abutments. This information might be needed for a variety of reasons, although the designer of the structure would normally postulate some suitable loading of the beam when making strength calculations. This required force can be found from consideration of the statics of the beam, isolated for the purposes of analysis as in fig. 1.

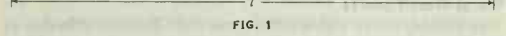


FIG. 1

This beam has forces applied to it: they are its weight W (a distributed body force) and also the vertical forces F_1 and F_2 transmitted by the supports. It is shown in MECHANICS that the conditions under which this beam can be in equilibrium are: (1) the resultant force applied to it, in any direction, shall be nil; and (2) the resultant moment of all the applied forces about any fixed axis shall be nil. The beam in fig. 1 is at rest, and mathematical expression may now be given to this fact. Requirement (1) demands that

$$W - F_1 - F_2 = 0$$

To apply the moment relation, the moments about the left-hand support can be reckoned; if the centre of mass of the bridge (*see* MECHANICS) is midway across the span, then

$$\frac{Wl}{2} - F_2 l = 0$$

From these two equations the expected result can be arrived at, that

$$F_1 = F_2 = \frac{W}{2}$$

Thus the required force is equal to half the weight of the beam in this case.

In this simple calculation, no allowance is made for the fact that the bridge is to some extent flexible. On the contrary, the bridge is assumed to be rigid inasmuch as use is made of a characteristic length l which could be measured before the beam is lifted off the ground and placed on its supports. Obviously, if the bridge were made of rubber, the distance between the supports would be much less than the length of the unsupported span, so that mathematically the required length l would depend upon W.

Since it is reasonable to assume that the variation of l is too small to be of consequence, this structure is said to be statically determinate. The effect of statical determinacy is to separate the mathematical problem of equilibrium from that of compatibility (*see* MECHANICS). Assumptions of this sort account largely for the great importance of the statics of rigid bodies: it is not that the bodies are really rigid but that they may be assumed so for the purposes of calculation.

The Statics of a Rigid Body.—The conditions under which a body, supposed rigid, can remain at rest must next be considered in more detail. Expressed algebraically, the problem is the examination of equations (21) and (22) of the article MECHANICS, when all the time derivatives are 0. That is, for any rigid body at rest it is necessary to satisfy force equations

$$\left.\begin{aligned} F_x &= \Sigma f_x = 0 \\ F_y &= \Sigma f_y = 0 \\ F_z &= \Sigma f_z = 0 \end{aligned}\right\} \qquad (1)$$

and also moment equations

$$\left.\begin{aligned} M_x &= \Sigma(xf_y - yf_x) = 0 \\ M_z &= \Sigma(yf_z - zf_y) = 0 \\ M_y &= \Sigma(zf_x - xf_z) = 0 \end{aligned}\right\} \qquad (2)$$

In words, equation (1) states that the sums of the resolved parts of the external forces in the directions of fixed orthogonal axes OX, OY and OZ shall be 0. The quantities M_x, M_y, M_z are the moments about OX, OY and OZ of the external forces, and these, too, must vanish.

In this article, the meanings of these equations will be discussed further and, in particular, the use of vectors (introduced in the article MECHANICS) will be extended.

While equations (1) and (2) govern the forces acting on a body at rest, it is not necessarily possible to solve these equations explicitly. Statically determinate systems only are of concern here, so that this difficulty will not arise.

Static Equivalence; Transmissibility.—Suppose that a rigid body is under the action of forces, one of which is of magnitude f and is applied to it at some point P (fig. 2).

Let the component of f in the plane XOY be r; it is obtained by projecting the force vector parallel to ZO. Now the state of rest of the body will not be altered if two equal and opposite forces f are applied at Q (as in fig. 2), Q being a point lying on the line of action of f. They too will have components r.

The effect on the equilibrium of the outer pair of these components r (shown dotted in fig. 2[B]) is nil, since they are equal and opposite and have no net moment about the axis OZ. If removed, these components leave only the one component r shown. Similar conclusions are reached by projecting the original forces at P and Q on to the planes YOZ and ZOX. If the three remaining components (of which r in fig. 2[B] is one) are now

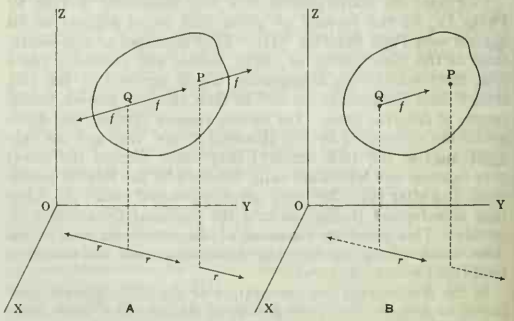


FIG. 2

compounded, they give the force f at Q, as shown in fig. 2[B]. The remaining force f at Q is statically equivalent to the original force f at P in that substitution of one for the other has no effect on the state of equilibrium of the body. To this extent, then, the point of application of f may be transmitted along the line of action; this is known as the property of transmissibility of a force.

Parallel Forces and Couples.—Consider two parallel forces which act on a body; their lines of action define a plane. The addition of these forces is clearly a case which cannot be dealt with by the parallelogram law of vectors, since the lines of action do not intersect. The addition of concurrent forces is explained in the article MECHANICS.

A simple extension of the above reasoning shows that the statical effects of a given system of forces acting on a body are unaltered by the imposition of two equal and opposite forces having the same line of action. If, therefore, f_1 and f_2 (shown in fig. 3) are two parallel forces (acting in the plane of the paper) whose resultant is required, the equal and opposite forces P can be superposed as shown. The resultant of P and f_1 is R_1 and that of P and f_2 is R_2. Let the lines of action of R_1 and R_2 intersect at A; this point may be regarded as the point of application of R_1 and R_2 and hence of f_1 and f_2 without the forces P.

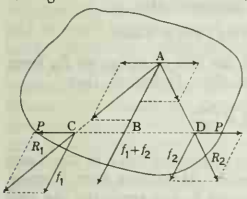


FIG. 3

The forces R_1 and R_2, supposed acting at A, may now be replaced by their components P-plus-f_1 and P-plus-f_2, respectively. Since the forces P again neutralize each other, there is left the resultant force which is parallel to f_1 and f_2, is of magnitude $f_1 + f_2$ and acts through A. The position of the line of action of the resultant force may easily be deduced. For, by consideration of similar triangles,

$$\frac{CB}{BA} = \frac{P}{f_1}; \frac{BD}{BA} = \frac{P}{f_2}$$

It follows that

$$\frac{CB}{BD} = \frac{f_2}{f_1}$$

If the parallel forces f_1 and f_2 act in opposite directions, their resultant can be found by similar means. It will act in the plane of f_1 and f_2 and will be found to act outside the space between the lines of f_1 and f_2. The magnitude of the resultant is then the difference between f_1 and f_2, and its direction is that of the greater of the two.

The lines of action of R_1 and R_2 (which fix the point A in fig. 3) are parallel if f_1 and f_2 are equal in magnitude and opposite in direction, so that a single resultant force cannot be found. A pair of forces of this nature is said to constitute a couple, and the plane containing the constituent forces is the plane of the couple.

Suppose that a body could be found with all the forces acting on it in a single plane, say the plane XOY. If a couple acts on the body (fig. 4), then clearly the forces f make no net contribution to either of the relevant force equations of (1), viz.,

$$F_x = F_y = 0$$

The forces do, however, contribute to the third relevant equation, namely,

$$M_z = 0$$

since they have a net moment about the axis OZ (perpendicular to the plane XOY). Their contribution is

$$[(f \times OD) - (f \times OC)] \text{ or } fl$$

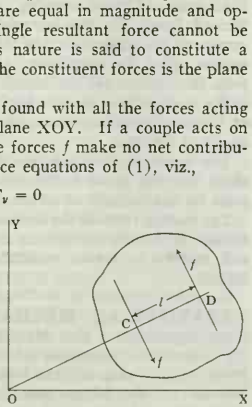


FIG. 4

where OCD is perpendicular to the lines of action of the two forces f, and l is the distance between these lines.

The moment of the couple in this case is of magnitude fl. Evidently the two forces could be moved to act in the plane at any points of the body and in any direction in this plane without affecting the conditions of equilibrium. It is necessary only that the product of the magnitude of the forces and the distance between their lines of action should be kept equal to fl. In this problem, in which all forces act in one plane, a couple is completely defined by the magnitude and direction (clockwise or counterclockwise) of its moment.

Vectorial Nature of Couples.—It has been shown that the effect of a couple on the equilibrium of a body is not altered if the couple is moved in its own plane. It is next shown that if the equilibrium exists, it is not disturbed if the plane of action of the couple is moved parallel to itself.

Let a couple of moment fl act on a body in a plane A (fig. 5[A]). The equilibrium of the body will be unaffected if two pairs of equal forces f are applied with their lines of action distant l apart on a parallel plane B. Now the forces f' may be added, giving the resultant R', while the forces f'' give the resultant R''. And since $R' = R''$, these resultants neutralize each other and may therefore be removed without influencing the equilibrium. There remain only the two forces in plane B, which form a couple of moment fl.

The previous conclusions can now be extended to cover couples in space. It may be seen that a couple is completely defined by (1) the magnitude of its moment; (2) the orientation of the plane in which it acts; and (3) the sense in which it acts (thus the couple in fig. 5[A] acts in the clockwise sense when looking from plane A to plane B). It is evident that a couple can be represented by a line whose length represents its moment, whose direction is normal to the plane of the couple and which points in a direction whose relation to the sense of the couple is that borne by the advancement to the rotation of a right-hand corkscrew. Under this convention, the couple of fig. 5[A] may be represented by the arrow in fig. 5[B].

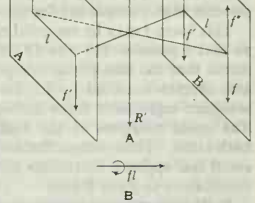


FIG. 5

Lines such as this are, in fact, vectors, since they permit couples to be added according to the parallelogram law. Suppose that two couples act on a body in different planes which intersect in a line containing two points C and D (fig. 6[A]). By movement in their own planes and movement of the type just described, the couples can be replaced by forces f_1 and forces f_2, whose lines of action are separated by the distance l between C and D; it is merely necessary, when finding f_1 and f_2, that the moments of the couple be $f_1 l$ and $f_2 l$.

The forces at C and D can be compounded in the usual way, giving the resultant forces R. Thus the sum of the two original couples is a couple of magnitude Rl whose plane is fixed by the vectors R. The planes containing the pairs of forces f_1 and f_2 and R are indicated in the figure.

Now the original couples can be represented by the lines of length $f_1 l$ and $f_2 l$ (fig. 6[B]) in accordance with the corkscrew rule, these lines being perpendicular to the planes containing the pairs of forces f_1 and f_2, respectively. If these lines are now added by the parallelogram rule as shown, it is easily seen that the construction is geometrically similar to those made before at C and D; the parallelogram is rotated through a right angle, and its sides are l times as long as the previous ones. The diagonal, therefore, correctly represents the resultant couple Rl.

If a body is acted upon by several couples, then these may be compounded by repeated use of the parallelogram rule or, just as with the forces acting on a particle, by means of a space polygon.

If the polygon closes, there is no resultant couple acting. Fur-

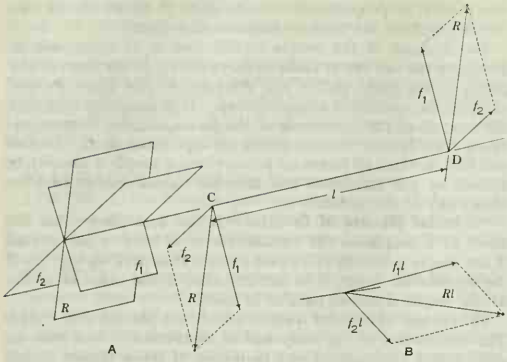


FIG. 6

ther, any couple may be resolved into components by the parallelogram rule.

Static Equilibrium of a Rigid Body.—The analytical conditions for equilibrium of a rigid body are those of equations (1) and (2), which will now be interpreted in terms of vectors.

Suppose that a force f is applied at some point A (x,y,z) of a rigid body and consider its component f_x (fig. 7). This component of the force enters equations (1) and (2) through: (1) its contribution to F_x; and (2) its contributions $-yf_x$ to M_z and zf_x to M_y. If, therefore, the component f_x is replaced by a force of equal magnitude acting at O, together with the couples shown in the figure, the equilibrium of the body would not be influenced. If O lies outside the body, the force and couples may be imagined as being applied to an extension of the body which extends to the origin.

A similar argument may now be applied to the remaining components f_y and f_z of f. The force acting at A is therefore statically equivalent to a force passing through O (with components f_x, f_y and f_z) and a couple whose components in the directions OX, OY and OZ, are $(yf_z - zf_y)$, $(zf_x - xf_z)$ and $(xf_y - yf_x)$. The components of this couple are the contributions of f to M_x, M_y and M_z.

This force f may be regarded as typical of all the applied forces so that equations (1) and (2) can be regarded merely as statements that the sum of the components of these equivalent forces and couples must all be 0. In other words, there must be no resultant equivalent force acting at O and no resultant couple. The space polygons of the equivalent forces and couples must both close. Thus, except for no forces or the special case of two equal and opposite forces, the minimum number that can produce equilibrium is three forces.

Furthermore, an unbalanced couple can be balanced only by an equal couple with opposite sense.

Centre of Gravity.—The weight of a body is the external force exerted upon it by the attraction of the earth. It is the resultant of the contributions, to the total body force, of the separate gravity forces exerted upon the particles which may be thought of as forming the body. The force on each particle is proportional to its mass, and the forces are assumed parallel; under these circumstances it may be shown that their resultant always passes through a particular point fixed in the body.

Referring again to fig. 7, let f_x be the component in the direction OX of the gravity force of one of the constituent particles. Now whatever the direction of the gravity forces on the various particles, their components in any direction will be proportional to the masses of the particles. Let these components be

$$(f_x)_1 = km_1 \; ; \; (f_x)_2 = km_2 \; ; \ldots$$

for the particles m_1, m_2, . . .

It will now be seen by reference to fig. 7 that these various components could be replaced by a force

$$f_x = k\,\Sigma m = kM$$

directed along OX together with couples

$$\Sigma(-yf_x) = -k\,\Sigma my \text{ about OZ}$$

and

$$\Sigma zf_x = k\,\Sigma mz \text{ about OY}$$

Consider now a single force F_x, acting parallel to OX. Evidently it will be statically equivalent to all the components $(f_x)_1$, $(f_x)_2$, . . . etc., provided that its equivalent force and couples at O are the same as before. It is necessary that

$$F_x = kM \; ; \; -F_x\bar{y} = -k\,\Sigma my \; ; \; F_x\bar{z} = k\,\Sigma mz$$

where $\bar{y}$ and $\bar{z}$ are the distances of the line of action of F_x from the planes XOZ and XOY, respectively. It follows that

$$\bar{y} = \frac{\Sigma my}{M} \; ; \; \bar{z} = \frac{\Sigma mz}{M}$$

In the same way the components of gravity force in the directions OY and OZ can be dealt with. It follows that a single force F having components F_x, F_y and F_z can represent the sum of the various forces on the particles. Each component of F is equal to the sum of the corresponding components of the particle forces. Further it is necessary that F should act at the point $(\bar{x}, \bar{y}, \bar{z})$ of the body. This point is referred to in MECHANICS as the centre of mass, but it is referred to as the centre of gravity in the present context. *See* also DYNAMICS.

BIBLIOGRAPHY.—A. P. McMullen and E. W. E. Kempson, *Elementary Experimental Statics for Schools* (1936), contains a very simple introduction to the subject; A. S. Ramsey, *Statics* (1946), an elementary textbook with a mathematical flavour; S. P. Timoshenko and D. H. Young, *Statics*, 4th ed. (1956), introduces the subject of statics and emphasizes its engineering applications; E. J. Routh, *A Treatise on Analytical Statics*, vol. 1 and 2, 2nd ed. (1896–1902), deals with more advanced problems in statics. (R. E. D. B.)

STATIONERY, a word that now covers all writing materials and implements, together with the numerous appliances of the desk and of business offices. For the principal articles and operations of the stationery trade, *see* INK; OFFICE MACHINES AND APPLIANCES; PAPER MANUFACTURE; PEN; PENCIL.

STATIONS OF THE CROSS, a series of 14 pictures or images representing the closing scenes in the Passion of Christ, namely: (1) the condemnation by Pilate; (2) the reception of the cross; (3) Christ's first fall; (4) the meeting with His mother; (5) Simon of Cyrene carrying the cross; (6) Veronica wiping the face of Jesus; (7) the second fall; (8) the exhortation to the women of Jerusalem; (9) the third fall; (10) the stripping of the clothes; (11) the crucifixion; (12) the death; (13) the descent from the cross; (14) the burial. Sometimes a 15th—the finding of the cross by Helena—is added; on the other hand, in the diocese of Vienna the stations were, at the end of the 18th century, reduced to eleven.

The representations are usually ranged round the church; sometimes they are found in the open air, especially on the ascent to some elevated church or shrine.

The normal form of the devotion, which began among the Franciscans, is to visit the stations of the cross wherever represented and exercise a devout meditation on passing from station to station.

See article "Stations of the Cross" in the *Catholic Encyclopaedia.*

STATISTICAL MECHANICS, also called statistical thermodynamics, is that branch of theoretical physics whereby the measurable properties of macroscopic substances, whether solids, liquids or gases, are related to or predicted from the attributes of the microscopic systems, atoms or molecules, of which these substances are composed. Since all substances obey

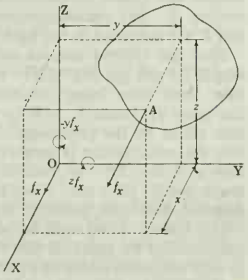


FIG. 7

the laws of thermodynamics, these laws themselves must emerge as a consequence of the molecular structure of matter, and it is by the success of the theory in justifying this supposition that the alternative designation is best vindicated. On the other hand, since the theory itself is basically a combination of mechanics and probability theory, or statistics, the phrase *statistical mechanics*, originating (1901) from the American mathematician J. Willard Gibbs (1839–1903), is perhaps more accurately descriptive of the subject.

Statistical mechanics, indeed, owes to Gibbs much more than its name. Its formal, mathematical structure for computational purposes is still, apart from the inevitable but comparatively minor changes introduced by the new development of quantum mechanics, essentially that presented in Gibbs's *Elementary Principles in Statistical Mechanics* (1902). But much progress has since been made both in the application of this formalism to particular physical situations and in the understanding of its ultimate justification. Satisfactory agreement with experience had to await the creation of the quantum theory, about 25 years later. Moreover, Gibbs himself produced the formal structure as a statistical model which obeyed, or embodied, the laws of thermodynamics without fully justifying this model on the basis of molecular behaviour. Judged pragmatically, Gibbs's formalism certainly provides the correct prescription for calculating equilibrium macroscopic properties from microscopic ones, but a complete justification of the procedure, from the laws of molecular dynamics, presents a difficult and still only partially solved problem. Recent advances in the mathematical theory of probability have, however, done much to clarify the situation. For critical discussion of the foundations of statistical mechanics reference should be made to the books by R. C. Tolman, D. ter Haar and, especially, A. J. Khinchin, listed in the bibliography at the end of this article.

The history of the subject properly starts, however, not with Gibbs, who by largely abandoning the details of molecular mechanics broke with tradition, but three or four decades earlier with the developments by James Clerk Maxwell (1831–79), at Aberdeen and Cambridge, and Ludwig Boltzmann (1844–1906), in Vienna, in the kinetic theory of gases. Kinetic theory itself dates from the 18th century, though it was not until after the work of the English chemist John Dalton and the Italian physicist Amedeo Avogadro, in the early 19th century that the concepts of atoms and molecules could be entertained seriously as physical realities. The laws of thermodynamics, too, are 19th-century discoveries.

Although the atomic hypothesis played its part in the overthrow of the caloric theory of heat, the laws of thermodynamics are purely phenomenological, making no reference to the ultimate constitution of matter. These laws are phrased in terms of concepts, some of which are peculiar to thermodynamics; of these temperature is the most familiar example. Others, such as the work done on a thermodynamic system by external agencies, relate directly to mechanics or to the energetics of electromagnetic theory. It is convenient at this point to summarize these laws, using modern notation and with a minimum of discussion: the account does not pretend to logical completeness and should be regarded solely as a brief recapitulation of phenomenological heat theory.

1. One may speak of thermal equilibrium between two systems, and if two systems are each in thermal equilibrium with a third, then they are in thermal equilibrium with each other. This enables empirical temperature scales to be defined.

2. One may speak of an adiabatically enclosed (thermally insulated) system, and then the work required to change the system from one equilibrium state to another is independent of how this work is supplied. This enables internal energy (U) to be defined apart from an additive constant.

3. No cyclic heat engine can simply extract internal energy from a system and convert it fully into useful mechanical work. This enables a logical introduction of the concepts entropy (S) and absolute temperature (T).· S, like U, is defined apart from an additive constant, and T is arbitrary to the extent of a scale factor.

Statement 1. which underlies the whole science of thermometry, has been accepted as a matter of common experience since at least the time of Galileo, but its recognition as a basic phenomenological law of thermodynamics is comparatively recent. The descriptive phrase, the zeroth law, was coined by Sir Ralph Fowler.

The second statement, 2. comprises the so-called first law of thermodynamics. When a system is not adiabatically enclosed, the extra energy, apart from the work (ΔW) done on the system, required to produce the resulting change in internal energy (ΔU) is known as the heat supplied (ΔQ), so that, automatically,

$$\Delta U = \Delta Q + \Delta W \tag{1}$$

The third statement, 3. comprises the so-called second law of thermodynamics. Alternative statements, such as that without compensating changes heat cannot flow from colder to hotter bodies, can, suitably interpreted, be shown to be equivalent to it. For the purposes of this article, however, it suffices to confine attention to the mathematical consequence of 3. namely, that for any small change from one equilibrium state to another the heat (δQ) absorbed by a thermodynamic system may be written

$$\delta Q \leq T dS \tag{2}$$

where the equality holds only for a reversible, or quasi static, process. Thus for a spontaneous, or natural, adiabatic process, for which $\delta Q = 0$, dS is positive, or the entropy of the system increases. It is important to appreciate that S, like U or T, is a thermodynamic property whose value is fixed when the state of the system is given. It is for this reason that the differential notation dS has been employed on the right-hand side of (2).

The history of both the first and second laws of thermodynamics is complex. Suffice it to note that although the term entropy was first introduced by Rudolf Clausius, of Zürich, as late as 1865, the foundations for the understanding of this law had been laid by Nicolas Carnot 40 years earlier. That heat, to use a modern phrase, is energy in transit was perhaps first convincingly demonstrated by J. P. Joule in the 1840s. Certainly by mid-century both the first and second laws were fairly well established, Clausius and William Thomson, later Lord Kelvin, having played a chief part in clarifying the situation. It was Thomson who introduced the concept of absolute temperature, founding this directly on Carnot's earlier work.

It is to be emphasized, as was clearly appreciated by both Maxwell and Boltzmann, that this thermodynamic story is quite independent of any molecular picture of matter. Moreover, although energy is part of the mechanical description of things, temperature and entropy are concepts foreign to mechanics. Thus if a thermodynamic system, for example a monatomic gas, is regarded as an assembly of enormously many atoms suffering mutual collisions, and perhaps with other forces between them, but essentially moving in accordance with the established laws of mechanics, it is natural to inquire what features of this mechanical situation are reflected, or shown up, in the thermodynamic concepts of temperature and entropy. Landmarks in the development of statistical mechanics are provided by Maxwell's work on the temperature problem (1859) and Boltzmann's on the entropy problem (1877). Although incomplete and, in part, faulty or obscure in their logic, these investigations effectively paved the way to modern statistical theory.

The link between the mechanical aspect of a system and its temperature, developed by Maxwell (and later, 1868, by Boltzmann too), is perhaps the most readily established link between mechanical and thermodynamic properties, though not the strongest. As shall be seen, it did not fully withstand the transition from classical to quantum mechanics; and even had it done so its usefulness would inevitably have been somewhat limited. Nevertheless, its importance in the development of theoretical physics is immense; and, where applicable, it is still a useful working tool.

The connection is most simply established by considering the pressure exerted by a gas on the walls of its containing vessel. Mechanically, this pressure represents the rate of destruction, per unit area of the surface, of momentum directed at right angles

to it. Clausius (1857) showed that if the atoms all had the same velocity, v, and moved in completely random directions, then, there being N atoms each of mass m, the pressure would be given by

$$pV = \tfrac{1}{3}Nmv^2 \tag{3}$$

where V is the volume of the container. Under the conditions of the derivation, the gas is regarded as ideally dilute, and so experimentally satisfies the equation of state

$$pV = nRT \tag{4}$$

where n is the amount of the gas in moles, and R the so-called gas constant. Comparison of (3) with (4) yields

$$\tfrac{1}{2}mv^2 = \tfrac{3}{2}kT \tag{5}$$

where $k = R/N_0$. Here N_0, known as Avogadro's number, denotes the number of atoms per mole (in a monatomic substance): its numerical value is now known to be $6.02 \ldots \times 10^{23}$.

That absolute temperature is thus a measure of the mean translational kinetic energy of the atoms of a gas was fully appreciated by Clausius (for the above argument barely depends on the assumption of equal velocities for all atoms); and Clausius also surmised that if the atoms, or molecules, were capable of other forms of kinetic energy, *e.g.*, rotational motion, then the kinetic energy of translational motion would simply constitute a definite proportion of the total kinetic energy of the molecular systems. It was, however, Maxwell who first derived the actual distribution of velocities among the systems of a gas. It was Maxwell, too (though the theory also owes much to Boltzmann), who gave precision, and more than plausibility, to this surmise of Clausius.

On the first point, namely, the distribution of velocities among the atoms of a monatomic gas (or the molecules of a molecular gas), Maxwell used probability considerations to produce his celebrated formula

$$n(v) = N\frac{4a^3}{\pi^{\frac{1}{2}}}\, v^2 e^{-av^2} \tag{6}$$

where a is a parameter, and $n(v)dv$ denotes the number of systems whose velocities lie in the interval v to $v + dv$. The factor $4a^3/\pi^{\frac{1}{2}}$ simply ensures that all systems are accounted for; *i.e.*, that

$$\int_0^\infty n(v)dv = N$$

If now we use this distribution law to find the mean translational kinetic energy of a system, we find

$$\tfrac{1}{2}m\overline{v^2} = \frac{m}{2}\int_0^\infty v^2 n(v)dv = 3m/(4a)$$

But, by (5), this mean translational energy is $\frac{3}{2}kT$: consequently $a = m/(2kT)$, and equation (6) may be written

$$n(v) = N\frac{4m^3}{\pi^{\frac{1}{2}}(2kT)^{\frac{3}{2}}}\, v^2 e^{-mv^2/(2kT)} \tag{7}$$

Maxwell's arguments leading to (6) have not been detailed here, partly for brevity and partly because the precise extent to which the laws of classical mechanics and probability theory imply this equation is still not entirely clear. Nevertheless, there can be no doubt that equation (6) or (7) is, in fact, truly descriptive of the molecular state of affairs in a gas. There is, perhaps, no direct experimental verification of this, but there are many indirect ones. For example, since all directions of molecular motion are obviously equally probable, equation (7) may be rewritten

$$n(v_1,v_2,v_3) = \frac{N}{4\pi}\,\frac{4m^3}{\pi^{\frac{1}{2}}(2kT)^{\frac{3}{2}}}\, e^{-m(v_1^2+v_2^2+v_3^2)/(2kT)} \tag{8}$$

where now $n(v_1, v_2, v_3)\, dv_1 dv_2 dv_3$ denotes the number of systems whose velocities in three mutually perpendicular directions lie respectively in the ranges v_1 to $v_1 + dv_1$, v_2 to $v_2 + dv_2$, and v_3 to $v_3 + dv_3$. Thus, if $n_1(v_1)dv_1$ is the number whose velocity component in a given direction lies between v_1 and $v_1 + dv_1$ (regardless of any motion perpendicular to this direction), it is found, on integrating over all values of v_2 and v_3, that

$$n_1(v_1) = \frac{mN}{(2\pi mkT)^{\frac{1}{2}}}\, e^{-mv_1^2/(2kT)} \tag{9}$$

This particular consequence of equation (7) can be directly tested in several ways. One way is through the Doppler effect on the wave length of light emitted by a moving atom. This produces broadening of the spectral line and gives it a characteristic profile which can be observed experimentally and also calculated very simply from equation (9). Indeed, such is the certainty of (9) that this equation was used in the mid-20th century to infer, from observed Doppler broadening, the temperatures obtained in pinched plasmas during attempts to achieve controlled thermonuclear reactions.

Maxwell's second chief contribution to kinetic theory lay in his attempt to deduce from the laws of mechanics, together inevitably with always rather fragile probability arguments, just how, if alternative forms of kinetic energy were available to the systems of a gas, the total energy would be distributed between these forms. His work, and later that of Boltzmann, led to what is now known as the Maxwell-Boltzmann "law of equipartition of energy." This asserts that when a system having s mechanical degrees of freedom is in thermal equilibrium at temperature T, then the expectation value of the kinetic energy associated with each degree of freedom is $\frac{1}{2}\,kT$, so that the total kinetic energy, averaged over many such systems, is $\frac{1}{2}skT$ per system.

A particularly simple illustration of this law is afforded by equation (5): a moving atom has three translational degrees of freedom. In general, the number of degrees of freedom of a system is given by the number of co-ordinates required completely to specify its position in space and its internal configuration. For a diatomic molecule, AB, there are necessarily three positional co-ordinates (those of the mass centre), two angles are required to describe the orientation in space of the line AB, and the distance AB must also be regarded as variable. This would suggest that the internal energy of a diatomic gas is given by $U = 3NkT$, since s is now 6. Consequently, the specific heat at constant volume, C_v, or $(\partial U/\partial T)_v$, should be $3R$ per mole for all diatomic gases, which is not the case. It is much closer to $\frac{5}{2}\,R$ at ordinary temperatures but varies with temperature and is not the same for all such gases. For monatomic gases the specific heat C_v is indeed $\frac{3}{2}\,R$ per mole, in agreement with the equipartition law.

Two other illustrations of the practical failure of the law warrant mention. Although derived from considerations relating to thermal equilibrium among the molecules of a gas, there would seem little reason why the law should not be applied equally to calculate the specific heat of a solid. In a monatomic solid each atom is in thermal equilibrium with the rest of the substance, and each atom has three degrees of freedom (corresponding to the three co-ordinates needed to specify its position in space). So it would be expected that the specific heat of a monatomic crystal, like copper, would be $\frac{3}{2}\,R$ per mole. This is about half the room-temperature value. Now it is clear that in a solid the atoms do not move freely in space but vibrate about certain mean positions. They may, indeed, be supposed to execute simple harmonic motion about these mean positions, under the influence of elastic restoring forces. In simple harmonic motion there is, on the average, as much potential as kinetic energy. Consequently, second thoughts demand twice as much internal energy per mole of crystal as that given by the kinetic energy alone, and a specific heat of $3R$ per mole: in good agreement with many room-temperature values for appropriate solids. But this inclusion of an equal potential-energy term when vibrational motions are involved does not help to resolve, but indeed increases, the discrepancy in the case of diatomic gases. And, equally disastrously, the specific heats of solids are found experimentally to decline to zero at very low temperatures.

The second illustration comes from application of the law to thermal, or black body, radiation. Here we are concerned with thermal equilibrium among all the possible electromagnetic waves that can exist within an enclosure of volume V. The number of

mutually independent excitations with frequency between ν and $\nu + d\nu$ can be shown to equal

$$\frac{8\pi V}{c^3}\nu^2 d\nu \tag{10}$$

where c is the velocity of light. Since these are harmonic vibrations, each contributes kT to the internal energy of the enclosure. Consequently, the total energy of black-body radiation, per unit volume, at temperature T, is

$$\frac{8\pi kT}{c^3}\int \nu^2 d\nu \tag{11}$$

However, all frequencies from $\nu = 0$ to $\nu = \infty$ are possible, and the integral diverges at the upper limit; *i.e.*, at high frequencies or short wave lengths. This, the so-called ultraviolet catastrophe, cannot be dismissed on the grounds that an infinite energy might be physically unobservable, since the formula also predicts an infinite specific heat, and it certainly does not require infinite energy to change the temperature within a finite enclosed vacuum.

The earlier of these discrepancies were familiar to Maxwell and Boltzmann. The ultraviolet catastrophe, discovered by Lord Rayleigh in 1900, only dramatically emphasized the unsatisfactory state of statistical theory at that time. Nevertheless, it was not in statistical mechanics as such, but in classical mechanics itself, that the fault lay. These failures of the theory were directly instrumental in promoting the creation of the quantum theory, or the mechanics which is now known to replace classical mechanics in atomic situations. Where classical mechanics is applicable, the law of equipartition of energy is valid. It can be used profitably, for example, to discuss the amplitude of the thermal torsional oscillations of a suspended galvanometer mirror.

Nevertheless, even were it always valid, the law of equipartition of energy could never provide a through route from mechanics to thermodynamics, and Boltzmann's work relating entropy to mechanical concepts must now be discussed.

The most striking property of entropy is that of increasing, as time proceeds, in spontaneous changes of a thermodynamic system. The clue to the mechanical interpretation of entropy was first afforded by Boltzmann's discovery (1872) of a feature of the mechanical description of a molecular assembly having a like property. For simplicity, we may consider a monatomic gas, not necessarily in an equilibrium state. Suppose that, at a particular instant, $f(x,y,z,v_1,v_2,v_3)dx\ dy\ dz\ dv_1 dv_2 dv_3$ is the number of atoms within the volume element $dx\ dy\ dz$ whose velocity components lie in the ranges v_1 to $v_1 + dv_1$, v_2 to $v_2 + dv_2$, v_3 to $v_3 + dv_3$. Then Boltzmann showed that the quantity H defined by

$$H = \int \ldots \int f \ln f\, dx\, dy\, dz\, dv_1\, dv_2\, dv_3 \tag{12}$$

(where *ln* stands for natural logarithm) never increases as the result of atomic collisions. More precisely, it remains constant if $f(x,y,z,v_1,v_2,v_3) = n(v_1,v_2,v_3)/V$, where $n(v_1,v_2,v_3)$ is given by Maxwell's velocity distribution law, equation (8); otherwise it decreases. This result, Boltzmann's celebrated H-theorem, was certainly not proved rigourously: statistical assumptions in the classification of collisions are certainly made, as Boltzmann came to realize; and these demand more justification than they received at that time. Nevertheless, the historical importance of Boltzmann's H-theorem can hardly be overestimated, and for a critical discussion of its status, reference must be made to the appropriate works, already cited, listed in the bibliography below. More generally, for nonuniform molecular gases, each molecule having s degrees of freedom requires s co-ordinates, $q_1, q_2, \ldots q_s$, to specify its positional state. In the language of analytical mechanics, there are now s momenta, $p_1, p_2, \ldots p_s$, conjugate to these coordinates, and $f(q_1,q_2, \ldots q_s,p_1,p_2, \ldots p_s)\ dq_1 \ldots dq_s\, dp_1 \ldots dp_s$ may be defined as the number of systems, at any instant, whose co-ordinates lie in the ranges q_1 to $q_1 + dq_1$, etc., and whose momenta lie in the ranges p_1 to $p_1 + dp_1$, etc. In place of (12), the quantity which, as a result of molecular collisions, is nonincreasing in time is now given by

$$H = \int \ldots \int f \ln f\, dq_1 \ldots dq_s dp_1 \ldots dp_s \tag{13}$$

Boltzmann realized that H, so defined, behaves essentially as the negative of the entropy of the gas.

Boltzmann, however, went further than this. The conceptual space in which a co-ordinate system is provided by the $2s$ quantities $q_1, \ldots q_s, p_1, \ldots p_s$ is known in analytical mechanics as phase space (more precisely, it is the phase space appropriate to a single molecule of the gas); and Boltzmann (1877) imagined this phase space subdivided into a number of equal cells, labeled $1, 2, \ldots i, \ldots r$, each of magnitude $dq_1 \ldots dq_s dp_1 \ldots dp_s$, equal to τ. If now n_i denotes the number of molecules, at a given instant, such that their co-ordinates and momenta define a point within the ith cell in phase space, then, on replacing the integral by a sum, (13) may be written

$$H = (\sum_i n_i \ln n_i)\tau \tag{14}$$

Because only changes in H concern us, just as only changes in entropy can be determined calorimetrically, and since $\sum_i n_i = N$, a fixed quantity, (14) may be modified to read

$$H = \tau[\sum_i (n_i \ln n_i - n_i) - (N \ln N - N)] \tag{15}$$

or

$$H = -\tau \ln \frac{N!}{n_1! n_2! \ldots n_r!} \tag{16}$$

In deriving (16) from (15) we have used the approximation

$$\ln n! = n \ln n - n$$

which is effectively true, mathematically, for the logarithm of the factorial of a large number: we have therefore assumed that all the distribution numbers $n_1, n_2, \ldots n_r$ are large; *i.e.*, that the cells in phase space provide only a rather crude description of the instantaneous dynamical state of the gas.

Now $N!/(n_1!n_2! \ldots n_r!)$ is simply the number of ways in which N molecules can be so distributed among the cells in phase space that the ith cell contains n_i of them; *i.e.*, n_i of them are in dynamical states corresponding to points in the ith cell. Boltzmann called this the number of complexions, W, pertaining to the set of distribution numbers $n_1, n_2, \ldots n_r$; and (16) reads

$$H = -\tau \ln W$$

If, then, H is proportional to the negative of the entropy S, it follows that, for a change in the state of the gas,

$$S_2 - S_1 : \ln W_2 - \ln W_1 \tag{17}$$

The more famous formula

$$S = k \ln W \tag{18}$$

is due not to Boltzmann but to Max Planck, but the constant k, later to be identified with the k of equation (5) above, is known as Boltzmann's constant. Its numerical value is $1.38 \ldots \times 10^{-16}$ ergs per degree. Unlike (17), equation (18) gives an absolute value to the entropy S and, therefore, necessarily exceeds thermodynamical requirements. The quantity W received the unfortunate name "thermodynamic probability"—it is neither a thermodynamic quantity nor a probability, being a larger integer.

Boltzmann did, however, proceed to identify the state of maximum "thermodynamic probability" with the equilibrium state of the gas. It is known that for a given adiabatically enclosed system the entropy increases until an equilibrium state is reached. Consequently, we expect the equilibrium state of the gas to be describable by distribution numbers, $n_1, n_2, \ldots n_r$, which make W a maximum subject to the two restrictive conditions

$$\sum_i n_i = N \tag{19}$$

$$\sum_i n_i \epsilon_i = U \tag{20}$$

Here ϵ_i is the energy appropriate to the i-th cell in phase space; the first condition prescribes a closed thermodynamic system, and the second further specifies adiabatic enclosure. The mathe-

matical problem of maximizing W subject to these restrictive conditions is straightforward, with the solution (in customary current notation)

$$n_i^* = N\lambda e^{\beta\epsilon_i} \tag{21}$$

The parameters λ and β are fixed by the conditions (19) and (20). λ can be eliminated easily enough by using (19); but β demands a physical interpretation.

Returning to the thermodynamical equations, (1) and (2) above, for two neighbouring equilibrium states of any thermodynamic system

$$T\,dS = dU - \delta W \tag{22}$$

and for a gas (or any homogeneous fluid) we can replace δW by $-p\,dV$. Consequently, (22) may be written

$$dU = TdS - pdV \tag{23}$$

the basic equation of all equilibrium thermodynamics. From (23)

$$\left(\frac{\partial S}{\partial U}\right)_v = \frac{1}{T} \tag{24}$$

On the other hand, if we substitute from (21) into (18) we find

$$S = Nk \ln \sum_i e^{\beta\epsilon_i} - k\beta U \tag{25}$$

whence

$$\left(\frac{\partial S}{\partial U}\right)_v = -k\beta \tag{26}$$

even though β itself depends implicitly on U through equation (20). Comparison of (24) and (26) suggests that β has the physical significance of $-1/(kT)$, so that (21) may be written

$$n_i^* = N\lambda\, e^{-\epsilon_i/kT} \tag{27}$$

Before discussing this further, we may also observe that (25) yields

$$A = -NkT \ln \sum_i e^{-\epsilon_i/kT} \tag{28}$$

where A is the thermodynamic quantity known as Helmholtz free energy and defined as $U - TS$, so that, from (23),

$$dA = -S\,dT - p\,dV \tag{29}$$

The above identification of β with $-1/kT$ is not strictly in accordance with the rules of statistical mechanics, since there has been an appeal to thermodynamic formulae which should themselves follow from statistical theory. The identification is most properly made by considering the work done in slightly changing the volume of the gas and showing that this has the form $dU + dS/k\beta$.

However, the present, shorter, approach is not without its interest. In any case, the identification is what would have been expected from equation (7), since for translational motion $\epsilon = \frac{1}{2}mv^2$.

Equation (27), perhaps written in the form

$$\frac{n_i^*}{n_j^*} = \frac{e^{-\epsilon_i/kT}}{e^{-\epsilon_j/kT}} \tag{30}$$

is known as Boltzmann's distribution law. Maxwell's law is a particular case of it, and, indeed, Maxwell himself came close to this generalization of his velocity distribution formula. Equation (28) is particularly basic to further developments and will be discussed in greater detail shortly.

It is, however, time to say something of the generalizations, and rather different attitude, introduced by Willard Gibbs. Although anachronously, they will be described in the language of the quantum theory; and first Boltzmann's results will be recapitulated in this new language. The rather arbitrary cells in phase space have effectively the magnitude h^s, where h is Planck's constant ($6.62\ldots\times 10^{-27}$ erg sec.). More importantly, summation over cells is replaced by summation over allowed energy levels, so that (28) and (30) become

$$A = -NkT \ln \sum_i \omega_i e^{-\epsilon_i/kT} \tag{31}$$

and

$$\frac{n_i^*}{n_j^*} = \frac{\omega_i e^{-\epsilon_i/kT}}{\omega_j e^{-\epsilon_j/kT}} \tag{32}$$

where ω is the quantum-mechanical degeneracy associated with the energy ϵ; *i.e.*, the number of independent wave functions with this energy. In (32), the distribution numbers now refer to molecules having certain energies, rather than pertaining to certain cells in phase space. Since the energy levels themselves explicitly depend on the volume of the enclosure, it is clear that (31) provides an expression for A, as a function of V and T, from which, by (29), all other equilibrium thermodynamic properties of the molecular assembly can be computed. The route from mechanical to thermodynamic properties is complete, at least for gases.

Boltzmann justified his H-theorem by considering in detail the effects of molecular collisions, and in so doing had necessarily particularized assumptions about molecular interactions. Gibbs wished to free statistical mechanics from any such, conceivably unrealistic, suppositions. And at the same time he sought a formalism equally appropriate to solids or liquids as to gases. Instead of considering a gas as a statistical assembly of many molecules, he considered a statistical ensemble of very many replicas of the molecular assembly, whether the latter be gaseous, liquid or solid. The individual members of this ensemble might have different total energies, E_i, and, in particular, Gibbs focused attention on a so-called canonical ensemble in which the number of assemblies with energy E_i is proportional to

$$\Omega_i e^{\beta E_i}$$

Here β is a parameter, independent of E_i and Ω_i, in quantum-mechanical language, is the number of independent wave functions descriptive of a molecular assembly with energy E_i. He then assumed that the measurable properties of an actual physical molecular assembly would be given by averages over the corresponding canonical ensemble, so that, in particular

$$U = \sum_i E_i \Omega_i e^{\beta E_i} / \sum_i \Omega_i e^{\beta E_i} \tag{33}$$

The different members of a Gibbsian ensemble have, of course, no physical interactions with each other: classical mechanics, or the quantum theory, is invoked only to prove that a canonical ensemble will remain a canonical ensemble as time proceeds, which is essential if such averages are to yield equilibrium thermodynamic properties.

The similarity between (33) and the equation of the Boltzmann theory

$$U = \sum_i \epsilon_i n_i^* = N \sum_i \epsilon_i \omega_i e^{-\epsilon_i/kT} / \sum \omega_i e^{-\epsilon_i/kT} \tag{34}$$

is self-evident, and it is not surprising that Gibbs succeeded in proving that the parameter β in a canonical distribution would reflect the physical features of reciprocal temperature. It can, of course, be identified with $-1/kT$. Gibbs showed, indeed, that quantities calculated as averages over canonical ensembles would obey all the laws of equilibrium thermodynamics, the Helmholtz free energy of a molecular assembly being given by

$$A = -kT \ln \sum_i \Omega_i e^{-\epsilon_i/kT} \tag{35}$$

Applied to a gas, whose individual molecules may have energies $\epsilon_1, \epsilon_2, \ldots$, with degeneracies $\omega_1, \omega_2, \ldots$, (35) leads identically to (31) if

$$\Omega = \sum_{\{\text{allowed } n\text{'s}\}} W(n_1, n_2, \ldots)$$

where

$$\sum_i n_i = N, \quad \sum_i n_i \epsilon_i = E$$

and

$$W = \frac{N!}{n_1!\, n_2! \ldots} \omega_1^{n_1} \omega_2^{n_2} \ldots \tag{36}$$

And this, of course, is just the number of ways in which one would expect to be able to ascribe individual energies to the molecules of a gas whose total energy is given.

Gibbs thus brilliantly succeeded in producing a self-consistent statistical model embodying the laws of thermodynamics, and at the same time in providing a universal recipe for computing measurable, macroscopic, physical properties—equations of state, specific heats, etc.,—from the dynamical description of any molecular assembly. This recipe, or prescription, certainly receives full a posteriori justification. What he does not do is properly to justify, a priori, the identification of physical measurements with ensemble averages. And here, as in the Maxwell-Boltzmann theory, the gaps in the logic have to be filled in by more recent developments in the understanding of the foundations of statistical mechanics, developments which are beyond the scope of this article.

Equation (35) is the basic equation of statistical mechanics for computational purposes. The quantity $\sum \Omega_i e^{-E_i/kT}$ is known as the partition function for the molecular assembly. If all the energy within the assembly may be regarded as shared out between individual molecules—as in a sufficiently dilute gas, when intermolecular forces may be ignored—then, as we have shown

$$\sum_i \Omega_i e^{-E_i/kT} = \left(\sum_i \omega_i e^{-\epsilon_i/kT}\right)^N \tag{37}$$

The quantity $\sum_i \omega_i e^{-\epsilon_i/kT}$ is then known as the partition function for a single molecule. In either case, each separate term in the partition function measures the relative probability, in an actual physical system at temperature T, of the particular mechanical state to which it refers.

Equation (37) is simply a particular case of the important factorization property of partition functions. If, for precision, the molecules of a gas are capable of two completely independent sources of energy, for example, translational and rotational energies, ϵ_i' and ϵ_i'', say, with degeneracies ω_i' and ω_i'', then

$$\sum_i \omega_i e^{-\epsilon_i/kT} = \left(\sum_i \omega'_i e^{-\epsilon'_i/kT}\right)\left(\sum_i \omega''_i e^{-\epsilon''_i/kT}\right)$$

since $\epsilon_i = \epsilon_j' + \epsilon_k''$ and $\omega_i = \omega_j'\omega_k''$. It follows that each independent source of mechanical energy makes its own contribution to the free energy of the assembly, and thence its own contribution to the internal energy, U, given by $-T^2\,\partial\,(F/T)/\partial\,T$, and to the specific heat, C_v.

Returning to the language of classical mechanics, the partition function $\sum_i \omega_i e^{-\epsilon_i/kT}$ has now to be replaced by the phase integral

$$\frac{1}{h^s}\int \ldots \int e^{-\epsilon(q_1, \ldots q_s, p_1, \ldots p_s)/kT}\, dq_1 \ldots dq_s dp_1 \ldots dp_s$$

where the energy is now expressed as a function of position in phase space. The factor $1/h^s$, where h is Planck's constant, ensures that the quantum-mechanical partition function and the classical phase integral have the same numerical value at high temperatures; *i.e.*, when kT is large compared with the effective spacing between quantum-mechanically allowed energy levels.

For the translational motion of atomic or molecular systems in a gas, the statistically important energy levels are always very close together compared with kT, and the partition function for translational motion is given by

$$\frac{1}{h^3}\int \ldots \int e^{-[p_1^2 + p_2^2 + p_3^2]/2mkT}\, dx dy dz dp_1 dp_2 dp_3 \tag{38}$$

i.e.,

$$\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} V$$

Consequently (proceeding from A to U), the contribution to the internal energy from the translational motion is $\frac{3}{2}\,kT$ per molecule, in accordance with the equipartition law. Rotational motion can normally also be considered classically and, because the rotational energy is again a separable quadratic function of the relevant momenta, this motion also normally contributes the equipartition value to the internal energy of the gas. However, vibrational motion can seldom be discussed classically, the energy levels being too far apart; which accounts for the various failures of the equipartition law already instanced.

On substitution from (37) and (38) into (35) there is found for the free energy of a monatomic gas the expression

$$A = -\tfrac{3}{2}NkT \ln(2\pi\, mkT/h^2) - NkT \ln V \tag{39}$$

a result that cannot be correct. For increasing V and N proportionately does not likewise increase A, although thermodynamically A is, of course, a simple extensive property of the gas. Gibbs realized this and suggested that the trouble arose from the indistinguishability of the N atoms of the gas. If only distinguishable complexions of the assembly were to be counted, then the above partition function for the gas ought to be divided by $N!$ or, alternatively, in place of (36) there should be

$$W = \frac{\omega_2^{n_1}\, \omega_2^{n_2} \ldots}{n_1!\, n_2!} \tag{40}$$

Equation (39) now becomes

$$A = -\tfrac{3}{2}\, NkT \ln (2\pi\, mkT/h^2) - NkT \ln V + NkT \ln N - NkT \tag{41}$$

which is dimensionally correct. This is certainly the right procedure, but the indistinguishability of the systems of the gas has more profound quantum-mechanical effects than Gibbs was in a position to appreciate.

For simplicity, let it be supposed that the translational energy levels are nondegenerate. Equation (40) then becomes

$$W = \frac{1}{n_1!\, n_2! \ldots} \tag{42}$$

which, since the n's are integers, can hardly describe the number of ways of assigning energies to the systems of the gas in accordance with given distribution numbers. Indeed, according to quantum theory,

$$W = 1 \tag{43}$$

for when the single particle energies are known the state of the system is fixed.

With particles of atomic or molecular mass, equation (42), though strictly incorrect, often (indeed usually) leads to right answers, such as equation (41). The systems are then said to obey Maxwell-Boltzmann statistics. But (43) is necessary in dealing with substantially lighter particles, such as electrons or photons (which have zero rest mass).

For electrons we must use (43) and also invoke the further conditions, imposed by the Pauli exclusion principle, $n_i = 0$ or 1. In this case we speak of using Fermi-Dirac statistics, Enrico Fermi and Paul Dirac having independently deduced this consequence of quantum theory in 1926. A practical consequence is that the conduction electrons of a metal like copper do not, except at very low temperatures, contribute appreciably to the specific heat.

For photons, or for helium atoms, (43) suffices without further conditions on the n's, and we speak of Bose-Einstein statistics, the appropriate theory having been developed by Satyandra Nath Bose and by Albert Einstein in 1924.

The most important features of these specifically quantal statistics are summarized in the analogues of equation (21). For Fermi-Dirac statistics we have

$$n_i^* = \frac{\lambda e^{-\epsilon_i/kT}}{1 + \lambda e^{-\epsilon_i/kT}} \tag{44}$$

while for Bose-Einstein statistics

$$n_i^* = \frac{\lambda e^{-\epsilon_i/kT}}{1 - \lambda e^{-\epsilon_i/kT}} \tag{45}$$

The left-hand side of (44) or (45), as also of equation (21),

must, of course, be interpreted as an average or expectation value; it is not necessarily an integer. In the particular case of photons, or black-body radiation, there is no restriction $\sum_i n_i = N$, since photons are not numerically conserved. This has as consequence that $\lambda = 1$ in (45). Since a photon of frequency ν has energy $h\nu$, (45), (20) and (10) combine to give

$$U = \int_0^\infty \frac{8\pi V}{c^3} \frac{h\nu e^{-h\nu/kT}}{1 - e^{-h\nu/kT}} \nu^2 d\nu = aVT^4 \tag{46}$$

where $a = 8\pi^5 k^4 15 h^3 c^3$.

Equation (46) was discovered semiempirically by Planck (1900) and marked, historically, the beginning of the quantum theory. Unlike (11), the integral in (46) does not diverge, and the integrand correctly reproduces the energy distribution in the frequency spectrum.

Except in the case of photons, the predictions of Fermi-Dirac and Bose-Einstein statistics coincide with those of Maxwell-Boltzmann statistics at sufficiently high temperatures; in fact, whenever

$$\left(\frac{2\pi\, mkT}{h^2}\right)^{\frac{3}{2}} V \gg N$$

And this condition is certainly satisfied in practice for all atomic or molecular gases, with the possible exception of helium at very low temperatures.

This brief outline of the formal structure of statistical mechanics and of its historical origins could well end here. The calculation of transport coefficients, such as coefficients of viscosity or thermal conductivity, still requires detailed consideration of molecular collisions and is perhaps better regarded as part of the kinetic theory of gases than of statistical mechanics proper (*see* Chapman and Cowling [bibliography]). However, one word more about the mechanistic interpretation of entropy may not be out of place.

Because the number of degrees of freedom of a macroscopic molecular aggregate is so large, of the order of 10^{23}, it can be shown that

$$\ln \sum_i \Omega_i e^{-E_i/kT}$$

can, without loss, be replaced by

$$\ln (\Omega^* e^{-E^*/kT})$$

where $\Omega^* e^{-E^*/kT}$ is the greatest of the terms in the above sum; *i.e.*, in the corresponding partition function. Consequently, by (35),

$$S = k \ln \Omega^* \tag{47}$$

where $E^* = U$. More properly, this argument should invoke not a fixed energy but an infinitesimal energy range. Here, however, concern is not with mathematical rigour, and equation (47) must suffice. This is not precisely the same equation as Planck's formula (18), since all complexions for a given energy are now counted and not simply those corresponding to the most probable distribution numbers. But, again because fluctuations are macroscopically negligible, it is virtually equivalent to Planck's formula.

Now Ω^*, or $ln\Omega^*$, is a measure of what is not known about the dynamical state of an assembly if it is known only that it has total energy U. It is for this reason that entropy is sometimes said to measure our ignorance of, or uncertainty about, the detailed dynamical state of the enormous number of microscopic systems, atoms or molecules, which together comprise the macroscopic systems of which there is direct physical experience.

BIBLIOGRAPHY.—L. Boltzmann, *Vorlesungen über Gastheorie*, 2 vol. (1896–98); J. W. Gibbs, *Elementary Principles in Statistical Mechanics . . .* (1902), *Collected Works*, vol. ii, part 1 (1948); R. H. Fowler, *Statistical Mechanics* (1936); R. C. Tolman, *The Principles of Statistical Mechanics* (1938); S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, 2nd ed. (1952); J. E. and M. G. Mayer, *Statistical Mechanics* (1946); A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Eng. trans. by G. Gamow (1949); D. ter Haar, *Elements of Statistical Mechanics* (1954); A. Münster, *Statistische Thermodynamik* (1956); E. A. Guggenheim, *Thermodynamics*, 3rd ed. (1957); L. D. Landau and E. M. Lifshits, *Statistical Physics*, Eng. trans. by E. and R. F. Peierls (1958).

At a more elementary level: E. A. Guggenheim, *Boltzmann's Distribution Law* (1955); G. S. Rushbrooke, *Introduction to Statistical Mechanics* (1949). (G. S. R.)

STATISTICAL ORGANIZATIONS, INTERNATIONAL. The term international statistics refers to compilations of data on economic and social characteristics of peoples and countries, drawn up in such a way as to permit intercountry comparisons. Such comparisons, to be of substantial validity, depend upon the application in each country of broadly uniform concepts, definitions and classifications to the subjects under investigation.

EARLY HISTORY

The International Statistical Congress.—Efforts to develop uniform standards for international comparisons began in 1853 with the meeting in Brussels of the first International Statistical congress. The moving spirit of the first congress, and indeed of later ones, was Lambert Adolphe Jacques Quételet (1796–1874), a noted Belgian astronomer and one of the founders of scientific statistics. The congress held eight subsequent meetings between 1855 and 1876 in various European cities. The objective of the congresses was to agree upon and to seek to give effect to uniform practices for the compilation of national statistics in the interests of international comparability. Statisticians of more than 20 countries usually participated, including such distinguished persons as William Farr, Ernst Engel, Edward Jarvis and, of course, Quételet.

During this period it became clear that continuity between sessions was needed to advance the work more systematically and to avoid repetitious discussions. Various solutions were proposed, culminating in 1872 with agreement to establish a small permanent commission. In the course of the next few years the commission held four meetings and soon sought to become an authoritative institution to impose recommendations upon governments. This policy proved unacceptable; invitations to send representatives to another meeting of the commission in 1879 and to what would have been the tenth meeting of the International Statistical congress, in 1880, were refused by the German prime minister Bismarck and subsequently by the governments of several other European states. The scheduled meetings were not held, and the work of the congress and its permanent commission thereupon collapsed.

Despite the controversies that marked the end of the congress, its activities had given a strong impetus to the development of official statistics in Europe and to improvements in comparability. The congress had encouraged the compilation of internationally comparable tables, the most important of these being the publication in 1865 of *Statistique internationale de la population*, probably the first statistical publication compiled under international auspices, prepared by Quételet and Xavier Heuschling at the suggestion of the fourth congress (1860).

In addition, the congress was instrumental in encouraging the development of uniform standards. At the 1872 meeting, the congress agreed upon a list of items to be collected in population censuses. In 1853 initial work was done on what was to become the International Statistical Classification of Diseases, Injuries and Causes of Death.

The International Statistical Institute.—The dissolution of the congress by no means disposed of the then recognized need for international collaboration in statistics. International collaboration in a number of economic and social fields intensified the usefulness of statistical comparisons. These new developments, which led some statisticians to fear that statistics would become increasingly dispersed, together with the real need to continue the work of the International Statistical congress, led to the creation of the International Statistical institute (I.S.I.) in 1885.

The occasion was the jubilee meeting of the Statistical Society of London (since 1877 the Royal Statistical society) attended by about 40 foreign statisticians invited by the government of the United Kingdom. Unanimous agreement was reached upon pro-

posals to create an international statistical organization that would avoid the difficulties encountered by the defunct congress. It was to be a private, scientific and unofficial body.

The I.S.I. held its first session in Rome in 1887 and met biennially until 1913 when World War I interrupted its meetings. During the period 1923 to 1938 it held ten sessions, its activities being once more interrupted by a world war. Members were elected in their personal capacities; many of them were official statisticians in their own countries.

The institute restricted its agenda to a smaller number of problems than had engaged the congress—a policy which resulted in more systematic work and in more carefully considered resolutions. The system of biennial meetings, while effective, could hardly keep pace with the rapidly expanding needs. A permanent office was established in 1913 to assist in the development of comparable methods, to assemble information from governments and to make arrangements for periodical publications. Special attention was given to demographic statistics. The first publication, in 1916, *Annuaire international de statistique,* was a compilation of the population statistics of European countries. Subsequent yearbooks covered other areas and other subjects. The *Annuaire* continued through 1921 and was superseded in 1922 by the *Aperçu de la démographie des divers pays du monde,* which appeared at intervals up to 1939, there being six issues in all.

In the meantime other international institutions were established. An international labour bureau was founded in Basel, Switz., in 1901. The International Institute of Agriculture (I.I.A.) was established in Rome in 1905, and in 1914 published the first issue of *Annuaire international de statistique agricole.* L'Office International d'Hygiène was organized in Paris in 1907, and in 1909 began to issue a monthly information bulletin, dealing, in part, with statistics on contagious diseases. An international conference on customs statistics, meeting in Brussels in 1910, laid the basis for an international convention, adopted in 1913, establishing the Bureau International de Statistique Commerciale, which was charged with compiling and publishing statistics on external trade in terms of a uniform commodity classification.

DEVELOPMENTS UP TO WORLD WAR II

Problems of Organization.—The outbreak of World War I naturally resulted in restriction or suspension of many of these developing activities. The end of the war brought new developments: the establishment in 1919 of the League of Nations and the International Labour Organization (ILO). The League and the I.S.I. were charged with broad responsibilities calling implicitly for the preparation of comparable statistics. Many of the then important subjects of official statistics had been taken up by one or another specialized organization, while the I.S.I. occupied a strong position of intellectual leadership, with a program actually or potentially embracing all fields of statistics. The newly created intergovernmental organizations could hardly divest themselves of responsibilities of compiling the statistics pertinent to their operations and of making recommendations to governments concerning these statistics. At the same time, the I.S.I. was loath to relinquish the position of leadership it had acquired over so many years.

During the next few years various schemes of collaboration were proposed and rejected. Agreement was finally reached between the I.S.I. and the League on a system of mixed committees that prepared technical reports for submission to sessions of the I.S.I.: recommendations on external trade and fishery statistics were approved in 1923 at Brussels, on industrial (production) censuses in 1925 at Rome and on indexes of industrial activity in 1927 at Cairo.

The International Convention.—The recommendations of the I.S.I. were transmitted to the economic committee of the League, and when they were submitted for government comment, it appeared that several governments wished steps to be taken to give effect to the proposed standards. Accordingly, in 1927, the council of the League authorized the convening of a conference to prepare an international convention incorporating the recommendations. The International Conference Relating to Economic Statistics, meeting in Geneva in 1928, was attended by representatives of 42 states and drew up the International Convention Relating to Economic Statistics. This marked, up to that time, the high point in the development of international comparability.

The articles of the convention set out the type and frequency of statistics to be compiled on a wide range of economic subjects. It also established the Committee of Statistical Experts, which was charged with improving and amplifying the principles of the convention and advising on other classes of statistics in the interests of international uniformity. Certain classes of statistics—on agriculture, labour, transport and finance—were excluded from consideration because they fell within other jurisdictions. The committee was composed mainly of members of the I.S.I. serving in their personal capacities; arrangements were made for representatives of the ILO and the I.I.A. to participate in the work of the committee.

The Committee of Statistical Experts held eight meetings between 1931 and 1939 and made notable contributions to the improvement of comparability and to the development of statistics generally. The main accomplishments of the committee are contained in the series *Studies and Reports on Statistical Methods,* six of which were published by the League in 1938 and 1939 and the remaining three by the United Nations in 1947. Considerable attention was devoted to external trade statistics, the result of which was, among other things, the preparation of the Minimum List of Commodities for International Trade Statistics, designed to replace the Brussels classification referred to above. Other subjects covered by the methodological studies included statistics of the gainfully occupied population (an industrial classification), timber statistics, housing statistics, indexes of industrial production and statistics relating to capital formation. The committee also studied banking and balance of payments statistics. Reports on these latter subjects, and upon national income, were completed after the committee had ceased to function as a result of World War II.

Other Statistical Organizations.—During the period when the I.S.I. and the League were elaborating the principles that were to be contained in the convention, the ILO was active in preparing standards for labour statistics. This work was carried on principally by the International Conference of Labour Statisticians, the first meeting, in 1923, taking up the classification of occupation and industry, and statistics of wages, hours of work and industrial accidents. Later conferences discussed cost of living index numbers, unemployment statistics and real wages (1925); methods of family budget studies and statistics of collective agreements and industrial disputes (1926); and international comparison of real wages (1931). In 1933 the ILO established a committee of statistical experts to undertake preparatory work for the international conferences of labour statisticians and to advise upon problems of international comparisons of wages and the cost of living. The fifth conference (1937) prepared an international convention on statistics of wages and hours of work, signed in June 1938. The I.I.A. continued its work, especially the publication of a monthly bulletin and the *International Yearbook of Agricultural Statistics,* usually published biennially. While some hopes for more sophisticated agricultural statistics were never realized, much useful work was done, including the elaboration of the standards for agricultural and related statistics which were contained in the 1928 convention.

While the League and the I.S.I. were working out harmonious ways of collaboration, a somewhat different situation prevailed between l'Office International d'Hygiène Publique and the Health organization of the League. Agreement was finally reached whereby l'Office retained its independence, the Health organization was put on a permanent basis and the permanent committee of l'Office became the General Advisory Health council to the League although the League had its own health committee. The Health organization, among other subjects, took a strong interest in the International Statistical Classification of Diseases, Injuries and Causes of Death, a classification initiated at the first International Statistical congress in 1853 and finally adopted by the I.S.I. in 1893.

Statistical Publications.—Because of the controversy in the early 1920s between the I.S.I. and intergovernmental organizations, the statistical secretariat in the League was restricted to modest proportions. Nevertheless, in July 1921 the economic and financial section of the League assumed responsibility for the publication of the *Monthly Bulletin of Statistics,* which had been initiated in 1919 by the Supreme Economic Council of the Allied and Associated Powers. By 1927 it was possible to issue the *International Statistical Year-Book,* as a companion volume to the *Bulletin,* containing more subjects and more details. Subsequently, a number of specialized periodical publications were issued, many of them representing new departures in international statistics. The most important of these were: *Memorandum on Balances of Payments and Foreign Trade Balances* (1924) (later *International Trade Statistics*); *Memorandum on Production and Trade* (1926) (later *World Production and Prices*); and *International Trade in Certain Raw Materials and Foodstuffs by Countries of Origin and Destination* (1936–38). The *Network of World Trade,* published in 1942, became something of a classic in its analysis of trade relationships.

The ILO also began the publication of an increased quantity of labour statistics. From 1923 on, special articles appeared in the *International Labour Review,* and in 1936 the *Year-Book of Labour Statistics* appeared as an independent publication. The I.I.A. continued to publish a yearbook on agricultural statistics. Demographic statistics were published in the *Aperçu,* and health statistics mainly in the *Epidemiological Report* of the League's Health organization.

WORLD WAR II AND AFTER

International statistical collaboration was virtually suspended during World War II. In June 1946, following the preparatory conference establishing the United Nations and the report of a temporary statistical commission, the United Nations Economic and Social council (ESC) established a permanent Statistical commission of representatives of 12 (later 15) states to promote the development of national statistics and their comparability and to co-ordinate international statistical work generally. The Statistical office of the United Nations, as a central organ, was also established to collect, compile, analyze and publish statistics, to assist in statistical development and to maintain an international centre for statistics.

Two sessions of the commission were held in 1947 as it was considered necessary to organize the work as rapidly as possible. The commission met annually from 1948 to 1954, thereafter meeting biennially. The League's International convention, while remaining in force, was transferred by protocol to the ESC, which, advised by the Statistical commission, in effect replaced the Committee of Statistical Experts.

The fear, expressed by some observers, that a commission composed of governmental representatives instead of experts acting in their personal capacities would necessarily be ineffective proved groundless. With very rare exceptions, the members nominated by governments were the heads of their national statistical services and commanded international prestige. The fact that these representatives can assist directly to give effect to the recommendations of the commission in their own countries is an important advantage. The technical recommendations of the commission are not mandatory; they depend for their acceptance upon their cogency and applicability.

The problem of co-ordination also required immediate attention. The commission drew up the terms of a standard agreement covering statistical relationships between the UN and each of the specialized agencies. In general, the statistical activities appropriate to the special sphere of each agency were reserved to it.

Secretariat Activities.—In the meantime the normal secretariat duties of collection and publication of statistics were resumed or expanded. The Statistical office of the UN took over the publication of the *Monthly Bulletin of Statistics* with the Jan. 1947 issue, thus continuing without interruption the publication which during the latter part of the war had been published by the League both in Geneva and in Princeton, N.J. Arrangements were made to compile the *Statistical Yearbook,* which resumed publication in 1949. A new publication, the *Demographic Yearbook,* also appeared in 1949. The statistical services of other international agencies commenced publication of current and annual statistics: the International Monetary fund (IMF) initiated the monthly *International Financial Statistics* (1948) and the *Balance of Payments Yearbook* (1949). The Food and Agriculture organization (FAO) continued the *Yearbook of Food and Agricultural Statistics,* issued additional yearbooks on forest products and fisheries (1947) and began publication of the *Monthly Bulletin of Agricultural Economics and Statistics* (1948). The ILO maintained and expanded the *Year Book of Labour Statistics* and the *Statistical Supplement* of the *International Labour Review.* In accordance with the agreements, each agency compiles the statistics relevant to its field and supplies them to any other agency upon request. Thus, labour statistics contained in UN publications are provided by ILO, and the UN provides statistics on population, trade, national income and the other subjects within its sphere.

The *List of Statistical Series Collected by International Organizations,* issued from time to time by the UN Statistical office, gives the sources of the series together with the titles of the international statistical publications.

Taken together, these developments resulted in a world-wide statistical system characterized by the full exchange of statistics and information among international agencies, by extensive consultation on technical questions and by active collaboration of national statistical services upon which international agencies must depend for statistics and for advice and assistance in the formulation of statistical standards. This latter function continues to be one of the most important and takes up a major part of the time of the various statistical services and of the Statistical commission itself.

Among the standards that have been developed by the UN and the specialized agencies through international consultations are: classification of international trade, classification of economic activities, occupational classification and various classifications and definitions applicable to censuses of population, housing, agriculture and industrial production and distribution. Principles have been established for the compilation of vital statistics, index numbers of industrial production, balance of payments, statistical sampling and other important economic and social subjects.

It has long been recognized that the divergent economic, social and political institutions in different countries and regions of the world require differentiated standards, so that absolutely uniform systems are not necessarily applicable from one country to another. Nevertheless, there are many common elements that find broad application in all countries and additional elements that may be applied on a regional or a community of interest basis. This situation has led to the development of regional groups for discussion of common statistical principles. The UN has established arrangements in Europe, Asia and Africa for periodic conferences and working parties to develop standard methods that are based on international standards but that are adapted to regional conditions.

The Inter-American Statistical institute (1940), in co-operation with the UN, serves as a forum for the Americas. Other arrangements of this kind are of long standing: the Scandinavian countries met to discuss statistical matters as early as 1888, and the first conference of statisticians of the British Commonwealth and empire was held in 1920.

Complete lists of applicable international standards are issued by the Statistical office from time to time in the *Directory of International Standards for Statistics.* Standard definitions and methods for many of the important economic and social subjects have been prepared, and work continues under way in many other subjects. Existing standards must be reviewed periodically to evaluate their continued applicability; new problems and concepts constantly emerge requiring statistical treatment. For example, in the late 1950s increasing attention was given to input-output studies, to money-flow analysis, to conditions and levels of living

and to the refinement of national accounting and price-index techniques.

BIBLIOGRAPHY.—A. Quetelet, *Congrès international de statistique* (1873); W. F. Wilcox, "The International Statistical Congress, 1853–1878," *Estadistica* (June 1943); *Bulletin de l'institut international de statistique*, tome I (1886); Friedrich Zahn, *50 Années de l'institut international de statistique* (1934); League of Nations, *Report with Annexes, International Statistical Commission*, A 10 (1921); Asher Hobson, *The International Institute of Agriculture* (1931); International Labour Office, *The International Standardization of Labour Statistics Studies and Reports*, "N Series" (Statistics) no. 25 (1943); Neville Goodman, *International Health Organizations and Their Work* (1952); W. Martin Hill, *The Economic and Financial Organization of the League of Nations* (1946); H. Campion, "International Statistics," *Journal of the Royal Statistical Society*, "A Series" (General), vol. cxii, part ii (1949). (W. R. Lb.)

STATISTICS was a term first applied to such data of importance to the political state as population, yield of taxation, value of trade, and mortality (*see* CENSUS); and to the study and interpretation of such data. As commonly understood nowadays, statistics is a mathematical discipline concerned with the study of masses of numerical data of any kind. (*See* also ECONOMETRICS; INDEX NUMBERS; BIOMETRY; STATISTICAL MECHANICS.)

DESCRIPTIVE STATISTICS

Statistical Tables.—A simple way of arranging numerical data is to set them in tabular form. With knowledge of the amount of sugar consumed in a given country in each year of a given period, for example, a table may be drawn up as indicated by the headings below:

Consumption of Sugar

(1) Year	(2) Total consumption thousand lb.	(3) Population thousands	(4) Consumption per head, lb.

Tabular arrangements of statistical data are found throughout these volumes; for example, *see* OLYMPIC GAMES; DEBT, PUBLIC. A detailed discussion of tabular material is given under MATHEMATICAL TABLES.

Graphs.—It is frequently useful to represent tabular data graphically. A number of examples of such graphs may be found in the articles PRICES, STATISTICS OF; LIFE EXPECTANCY.

Even though graphs fall short of numerical tables in precision, they have the advantage of allowing a large mass of figures to be grasped as a whole much more readily than is possible when those figures are presented in one or more tables (*see* GRAPH).

Averages and Dispersion.—An illustration of a type of statistical table known as a frequency distribution is seen in the following summary of the heights of a number of men:

Height in inches	61	62	63	64	65	66	67
Frequency	2	10	11	38	57	73	106
Height in inches	68	69	70	71	72	73	74
Frequency	126	109	87	75	23	9	4

The table indicates that 126 men were found to be between 5 ft. 7½ in. and 5 ft. 8½ in. and similarly for other heights. In tables such as this some method of dealing with cases falling exactly on the dividing line between two adjacent intervals must be laid down, and a common method is to assign half the number found on the dividing line to each of the classes of which it forms the limit. This procedure may, of course, result in frequencies in some of the intervals which are not integers. The accurate average height of the 730 men covered by the table can be ascertained only by reference to a more detailed statement showing the height of each man exactly, instead of in a number of groups. It will be observed that 297 were not taller than 5 ft. 7½ in., while 307 were 5 ft. 8½ in. or more, the distribution being not quite symmetrical on both sides of the numerically largest group. The approximate average height was about $\frac{1}{20}$ in. less than 5 ft. 8 in. (*see* MEAN).

A table of this kind indicates more than the average height of the individuals represented. Note that 191 were not more than 5 ft. 6½ in. in height and 198 were not less than 5 ft. 9½ in. in height. By calculation from the figures shown, assuming that the individual heights were distributed with approximate regularity along the intervals from inch to inch, the points representing 5 ft. 6.3 in. and 5 ft. 9.7 in. would divide the series of heights so that one-quarter of the whole number fell below the former, and one-quarter of the whole number above the latter. A similar calculation gives, as the point dividing the series into two equally numerous groups, 5 ft. 8.04 in. This last point is called the median (*q.v.*) of the series, and the two others are the lower and upper quartiles, the three points serving to divide the whole number examined into four groups of equal numbers of cases. The distance between the upper and lower quartiles, in the case in question 3.4 in., gives the range within which the middle half of the instances recorded lay. This distance (a measure of dispersion called the interquartile range) expresses much more definitely the degree of concentration of the individuals in the neighbourhood of the median height than, for example, the whole range (14 in.) within which all the measurements lie. If a more exact description of the nature of the distribution than is afforded by the specification of the median and quartiles, but of the same general character, is desired, the group may be divided into a number of parts, *e.g.*, into ten equally numerous parts, the points of division being then known as deciles. Points of division called centiles (percentiles) divide the group into 100 equal parts (*see* PERCENTAGE). For the most commonly used measure of dispersion (*i.e.*, variability or spread within a frequency distribution) *see* STANDARD DEVIATION.

In many statistical problems the form of the distribution is of a recognizably definite character. It is found that, in numerous cases, the manner in which the individual observations are distributed is in close accordance with that of events dependent on chance. Such a case is presented by the following. A number of balls, indistinguishable in size, weight, or form, are placed in a bag, half of the balls being white and half black. If one ball be drawn from the bag, its colour, whether white or black, may be noted. The ball being replaced and the bag shaken, another drawing will give a result wholly independent of the first.

If a large number of sequences of x_r white (or of x_r black) balls are recorded on Y_r occasions, the total number n of drawings will be expressed by $n = x_1Y_1 + x_2Y_2 + x_3Y_3 + \ldots + x_rY_r + \ldots$. The relative frequencies of the different series are y_1, y_2, etc., where $n \cdot y_1 = Y_1$, $n \cdot y_2 = Y_2$, etc., and, when n is indefinitely increased, the relation between the x and y values is represented by what is known as the normal probability curve, sometimes called the Gaussian curve or normal distribution (*see* PROBABILITY, MATHEMATICAL).

In any actual series of observations (*e.g.*, of the colours of balls drawn from a bag) the observed numbers expressing the frequency of different events will be found to differ more or less from those expressed by the normal curve. Prolonged trials would give results approximating to those derived from theoretical calculations and expressed in the shape of the probability curve. Thus, in the case of the heights of 730 men, it is probable that records covering larger numbers of men, of the same race and social condition and within the same limits of age as those from which the 730 were selected, would give a distribution of heights more closely corresponding to that shown by a typical probability curve than the numbers cited. This Gaussian (normal) distribution is illustrated and discussed in greater detail under ERRORS, THEORY OF.

Skew Distributions.—Particular statistical problems yield distributions that are not symmetrical on either side of their mean (as compared with the symmetrical normal probability curve). These are said to be skewed: *i.e.*, grouped more closely on one side of the mean than on the other. The following distribution of weights of men shows marked skewness.

Weight in lb.	105	120	135	150	165	180	195	210
Frequency	17	722	2,175	1,346	485	155	33	3

In this case the average (or arithmetic mean) of the weights of the 4,936 men is, on the assumption of continuous distribution, approximately 141.6 lb. The median is at about 140.5 lb. In the previous illustration of heights, the median was greater than the

mean, so that the skewness shown is of opposite direction in the two cases. If the mean is the greater of the two measures, the distribution is said to be positively skewed, or skewed to the right. When the median is the greater of these two measures of central tendency, the distribution is negatively skewed, or skewed to the left.

This illustration serves also to show a third measure of central tendency called the mode. The group of which 135 lb. is the central point is notably more numerous than any other. The weight 135 lb. may be called, as far as the figures furnish a ready indication, the most generally occurring weight or the modal weight. The point in the distribution thus determined is called the mode. A table like this, however, gives only a crude estimate of the most prevalent (modal) weight.

If the data in the table of weights had been less roughly grouped, the mode could be more accurately estimated. In a skew distribution the mode is related to the other two measures of central tendency as follows: it is lower than the arithmetic mean and the median in positive skewness; and is higher than these two in negative skewness.

Three Variable Elements.—In what precedes, the statistical material considered has consisted of series of pairs of quantities, the values of the two quantities in each pair being related; for example, where one of the quantities is the height of a man, the other the relative frequency with which men of that height were observed (*see* FUNCTION). Now consider series of three quantities, for example, where two of the quantities express the measures of different phenomena (which may be connected in some way, or may be independent) and the third expresses the relative frequency with which any combination of the other two occurs. Such a series is represented in the table given below, which shows how 1,000 marriages were distributed according to the ages of brides and bridegrooms.

Ages of wives at date of marriage	Ages of husbands at date of marriage						Total wives
	Under 25	25–34	35–44	45–54	55–64	65 and over	
Under 25	10	75	46	13	3	..	147
25–34	20	175	167	65	17	3	447
35–44	5	54	104	79	29	7	278
45–54	..	5	18	36	33	9	101
55–64	..	..	1	5	13	5	24
65 and over	..	..	..	..	1	2	3
Total husbands	35	309	336	198	96	26	1,000

It is clear, from the clustering of frequencies in a band crossing the table from its upper left side toward its lower right side, that there is some association (correlation) between the ages of the wives and husbands, old women tending to marry old men, and young women tending to marry young men. (A. W. F.)

Correlation.—Variables that tend to change together systematically are said to be correlated. If their values tend to rise and fall together the variables are positively correlated. Thus, from the table, age of husband is positively correlated with age of wife. If the value of one variable tends to fall as the other rises, the two are negatively correlated. For example, consider a boy at one end of a balanced plank (seesaw), and a girl at the other end. As the boy rises, the girl falls; and vice versa. Thus the elevation of the boy is negatively correlated with the elevation of the girl. Variables that change together in a completely random fashion are said to exhibit zero correlation. The weights of 100 pairs of soldiers selected at random (*see* RANDOMIZATION) from an infantry regiment would be expected to show zero correlation.

If each value of one variable exactly specifies one associated value for its correlated variable, the two variables are said to be perfectly correlated. In the example of the seesaw, the two elevation variables are perfectly correlated since the position of one child always specifies that of the other. The age correlation among marital partners is less than perfect, since any group of men of the same age tend to have wives who span a range of ages. Note from the table that husbands 65 and over have wives whose ages range all the way down to the 25–34 interval. This is an example of what is called regression: extreme values on one variable X tend to be associated with a range of values on the other, Y. The arithmetic mean of this Y range tends to be less extreme (to regress, or fall toward the Y average). In perfect correlation there is zero regression, since each X value exactly specifies one Y value; there is complete regression in zero correlation, each X value being associated with a range of Y values whose arithmetic mean is the same as the mean for all Y values combined.

Correlation can be quantified with a number (correlation coefficient) that ranges from +1.00 (perfect positive), through 0.00 (completely random), to −1.00 (perfect negative). The algebraic sign indicates only the direction (positive or negative) of correlation. Degree of relationship is shown by the absolute size of the coefficient. Thus a correlation coefficient of +.50 shows the same degree of relationship as does one of −.50.

Suppose five people A, B, C, D, E are measured on variables X,Y to yield the following:

Individual	X Score	Y Score
A	60	83
B	55	109
C	40	121
D	20	126
E	10	130

Note that these data represent a perfect negative correlation (−1.00), since A scored highest on X, lowest on Y; B next highest on X, next lowest on Y; and so on.

One way to express this quantitatively is to calculate a Spearman rank-order correlation coefficient ρ_{XY} based on

$$\rho_{XY} = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

where D is the difference between each pair of scores (each score expressed in rank form), N is the number of individuals (in this case 5), and ΣD^2 is the sum of squared differences. Thus, assigning rank 1 to the highest score in each array, rank 2 to the next highest, and so on, the initial calculations are

Individual	Rank X	Rank Y	D	D^2
A	1	5	−4	16
B	2	4	−2	4
C	3	3	0	0
D	4	2	2	4
E	5	1	4	16
				$40 = \Sigma D^2$

Substitute appropriately in the equation and the correlation between X and Y is $\rho_{XY} = 1 - [6 \times 40/5(25 - 1)] = -1.00$; a perfect negative correlation as expected.

Alternately the degree of relationship may be quantitatively estimated as a Pearson product-moment correlation coefficient r_{XY} calculated from

$$r_{XY} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

where x is the difference between each value of X and the arithmetic mean M_X of all X values (*i.e.*, $x = X - M_X$); and $y = Y - M_Y$. Thus r_{XY} considers the actual values X,Y while ρ_{XY} considers only their ranks.

Both ρ and r can be used only with continuous variables (*e.g.*, height *v.* weight, gas pressure *v.* temperature); for such discontinuous (discrete) variables as sex (male *v.* female), marital status (married *v.* not married), employment status (employed *v.* unemployed), other methods for calculating correlation coefficients must be sought. Among these special methods (only mentioned here) are the biserial, point-biserial, tetrachoric, and phi correlation coefficients.

The Pearson r will underestimate the degree of relationship when X and Y fail to change together in a linear (straight-line) fashion; *i.e.*, when they fail to show linear regression. This may be observed by constructing what is called a scatterplot, representing each pair of values X,Y as a point in rectangular Cartesian coordinates. If the scatterplot shows a curvilinear relationship between X and Y, the Spearman ρ may be substituted. However, if the curvilinearity is so marked that the slope of the scatterplot reverses direction (tends to become parabolic), ρ will also under-

estimate the degree of relationship, and the so-called eta coefficient (only mentioned here) should be used. Such parabolic scatterplots are sometimes encountered in industry, where employees on simple routines tend to produce best if they are of middle-range intelligence, those who are extremely dull or bright tending to do poorly.

To interpret the meaning of correlation coefficients it is helpful to think of them as expressing the degree to which variables X,Y exhibit variation in common. Such variation is usually given as the variance (*i.e.*, the square of the standard deviation) for each variable. If all assumptions underlying the accurate use of r are met, then an r_{XY} of either $+1.00$ or -1.00 means that X accounts for all of the Y variance; it is then said that X and Y show 100% variance overlap, or that Y explains 100% of the X variance. However, an r of $\pm.90$ does not represent 90% overlap; instead, it indicates 81% common variance. In other words, variance overlap is a function of the square of the correlation coefficient: $r = \pm.80$ signifies 64% common variance; $\pm.70$ represents 49% overlap; and so on.

For example, in estimating the effectiveness (validity) of intelligence test scores in predicting academic grades, it is common to compute the correlation coefficient between test scores and grades for different samples of students (*see* PSYCHOLOGICAL TESTS AND MEASUREMENTS). A typical validity coefficient from such studies is $+.60$. In terms of overlap the test is then said to explain 36% of the variance in academic grades, 64% of the variance remaining unexplained. (C. CE.)

STATISTICAL INFERENCE

Thus far the discussion has been concerned with descriptive statistics: methods for succinctly describing masses of numerical data. Statistical inference deals with techniques used to assess the probable characteristics of masses of data when these data are incompletely known. For example, such methods are used to infer the probable average height of a large human population (or universe) on the basis of measurements taken from a sample of people derived from the universe. An understanding of STATISTICS, MATHEMATICAL should be helpful in understanding what follows.

Statistical Tests.—A statistical test is a procedure for testing a given hypothesis about the distribution law of a population from which a sample is assumed to have been drawn. The hypothesis in its simplest form is whether the cumulative distribution function of the population is some specific function $F_1(x)$ alternative to a specified function $F_0(x)$. It is assumed that the sample comes from one of these populations. To test the hypothesis, a function $u(x_1, x_2, \ldots, x_n)$ of the sample x values is constructed, and a range R of critical values of this function is found so that the probability is a given value δ (usually .01 or .05) of $u(x_1, x_2, \ldots, x_n)$ lying in R if the population cumulative distribution function is $F_0(x)$, and also such that the probability of $u(x_1,x_2, \ldots, x_n)$ falling in R when the population distribution $F_1(x)$ is made as large as possible—certainly larger than δ. Then the claim can be made that the sample has been drawn from the population having distribution $F_1(x)$ with a risk of probability δ of making a type-I error. By a type-I error is meant the error committed by saying that the sample is not from the population having distribution $F_0(x)$, when, in fact, it is. In other words, when the value of $u(x_1, x_2, \ldots, x_n)$ falls in R there is a risk of erroneously claiming the sample not to have come from the population with distribution $F_0(x)$ when it actually does come from such a population. If the value of $u(x_1, x_2, \ldots, x_n)$ does not fall in range R the test calls for the assertion that the sample comes from the population having distribution $F_0(x)$, which is considered as the only alternative to $F_1(x)$ at present. There is a risk of being in error in this assertion, too, because the sample could have come from the population with distribution $F_1(x)$. The probability ϵ of this error is obtained by finding the probability of $u(x_1, x_2, \ldots, x_n)$ not falling in range R, when the sample has actually been drawn from the population having distribution $F_1(x)$. This error is called a type-II error. The probability of $u(x_1, x_2, \ldots, x_n)$ falling in R when the sample is from the population having distribution $F_1(x)$ is $1 - \epsilon$ and is called the power of the test based on $u(x_1, x_2, \ldots, x_n)$. When $u(x_1, x_2, \ldots, x_n)$ falls in R it is often said to be statistically significant at the δ probability level. Usually, $F_1(x)$ is a member of a family of functions $F(x, \theta)$, when θ is a parameter, which would be a specific member of the family, say $F(x, \theta_0)$. In this case the power $1 - \epsilon$ for the various alternatives corresponding to all values of θ (except θ_0) would depend on θ and is called the power function of the test based on $u(x_1, x_2, \ldots, x_n)$. The essential idea underlying type-I and type-II errors was originally introduced by Bell Telephone Laboratories' quality control engineers who first referred to them in their own applications as producer's risk and consumer's risk, respectively. The producer's risk is the probability that a given lot of products of satisfactory quality offered by a producer will be rejected on basis of the results of an inspected sample from the lot, whereas the consumer's risk is the probability of the consumer's accepting a lot of products of unsatisfactory quality on basis of the results of the inspected sample. J. Neyman and E. S. Pearson (1928) developed the idea that underlies these two risks into the present theory of testing statistical hypotheses. Contributions to this theory and its application to various problems are numerous. Two important examples will be briefly discussed.

A. Wald (1950) further generalized the ideas by developing a theory of statistical decision functions in which decisions regarding a statistical hypothesis may be more numerous than simply acceptance or rejection and where the risks associated with such decisions are measured.

The Student t-Test.—Consider a sample of n items taken from a normally distributed population to test whether this population has a given mean μ_0. Each measurement based on one case within the sample is signified by x_i. The criterion for this t-ratio test is

$$\frac{\sqrt{n}\,(\bar{x} - \mu_0)}{s} = t$$

where $\bar{x}$ is the sample mean, and s^2 is the sample variance

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

The sampling distribution of t when the sample has been drawn from a normal population with mean μ_0 is

$$dF(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi(n-1)}}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt \qquad (1)$$

where $\Gamma(m)$ is the gamma function which equals $(m-1)!$, if m is a positive integer and provides a good interpolation function for $(m-1)!$ if m is real but not a positive integer. In either case, $\Gamma(m)$ can be expressed as a definite integral. The distribution law expressed by equation (1) is known as the Student (1908) distribution law with $n - 1$ degrees of freedom. The range R of critical values of the function t is taken as the set of values of t outside the interval $(-t_\delta, t_\delta)$ where t_δ is given by the relation

$$\int_{-t_\delta}^{+t_\delta} dF(t) = 1 - \delta$$

This means that the probability is $1 - \delta$ that the sample will be such that

$$-t_\delta < \frac{\sqrt{n}\,(\bar{x} - \mu_0)}{s} < +t_\delta \qquad (2)$$

when the sample is actually from a normal population with mean μ_0. If the sample yields values of $\bar{x}$ and s so that the inequality (2) is not satisfied the hypothesis is rejected that the sample is from a normal population with mean μ_0. The risk of rejecting this hypothesis falsely is measured as the type-I error δ. Tabulations of values of t_δ have been presented in various statistics books for $\delta = .01, .05$, and other values and for different values

of $n - 1$. There are many applications of the Student t-test.

The Chi-Square Test.—As a second example, consider an infinite population in which each element belongs to one and only one of the mutually exclusive classes $A_1, A_2, \ldots, A_k$ having probabilities $p_1, p_2, \ldots, p_k$, where

$$\sum_{i=1}^{k} p_i = 1$$

In a sample of n elements from this population, the probability of getting x_1 A_1's, x_2 A_2's, $\ldots$, x_k A_k's is given by the multinomial law

$$m(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

It is important in many statistical problems to test whether the observed frequencies $x_1, x_2, \ldots, x_k$ could have reasonably come from a population in which $p_1, p_2, \ldots, p_k$ have specified values $p'_1, p'_2, \ldots, p'_k$; *i.e.*, whether the x values depart significantly from their expected values $np'_1, np'_2, \ldots, np'_k$ under the hypothesis that the sample came from a population with these specified p values. The criterion for testing this hypothesis which was originally proposed by K. Pearson (1900) is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - np'_i)^2}{np'_i} \tag{3}$$

known as the chi-square test criterion. It will be seen that it satisfies the intuitive requirement that it should somehow measure the discrepancy between each x_i and its expected value. The more the x values depart from their mean or expected values, the larger the value of χ^2 (chi-square). Thus, the range R of critical values which suggests itself consists of all values of χ^2 greater than a suitable chosen value χ^2_δ. For larger values of n, and assuming that none of the p values are very small, the sampling distribution of χ^2, if the hypothesis is true that the sample came from a population in which the p's have the values specified, is approximately

$$dF(\chi^2) = \frac{\left(\frac{\chi^2}{2}\right)^{\frac{k-3}{2}}}{2\Gamma\left(\frac{k-1}{2}\right)} e^{-\frac{1}{2}\chi^2} d(\chi^2) \tag{4}$$

This is an important sampling distribution, known as the chi-square distribution with $k - 1$ degrees of freedom. Values of χ^2_δ have been tabulated for which

$$\int_{\chi^2_\delta}^{\infty} dF(\chi^2) = \delta$$

has various values from .01 to .99 and for values of k from 2 to 30. There are many statistical functions having sampling distributions that are of the form given as equation (4).

As a simple example of the application of equation (3), suppose a bridge player has kept a record of the number of aces he has obtained in each of 500 bridge hands. Suppose he received no aces in 165 hands, one ace in 210 hands, and more than one ace in 125 hands. Is this distribution of aces consistent with the hypothesis of thorough shuffling and fair dealing? Under the hypothesis of thorough shuffling and fair dealing one finds by applying the theory of combinations (*see* Combinatorial Analysis) that the probabilities of no aces, one ace, and more than one ace are approximately 0.30, 0.44, 0.26, respectively. The expected numbers of hands containing no ace, one ace, and no aces in 500 hands are, therefore, 150, 220, 130, respectively. The value of χ^2 computed from equation (3) is

$$\chi^2 = \frac{(165 - 150)^2}{150} + \frac{(210 - 220)^2}{220} + \frac{(125 - 130)^2}{130} = 2.147$$

This is not a significantly large value of χ^2. For example, in a large number of sets of 500 bridge hands more than 30% of the sets would yield hands deviating more widely from the expected distribution of aces than in the present case. To have significant discrepancies at the 0.05 probability level the observed distribution of aces would have to yield a value of χ^2 equal to about 6 or larger.

Analysis of Variance.—With the rapid development of statistical theory, an increasing amount of attention is being paid to the original design or setting up of any type of an experiment or an investigation which depends heavily upon statistical analysis of the results for interpretation. Consideration of experimental design led to the development of theory and procedures for determining what combinations of factors important to the entire experiment should be introduced to simplify statistical analysis of the data and its interpretation, and also to determine how many repetitions or replications of the experiment should be performed to detect important differences in experimental results attributable to the various factors or combinations of factors. R. A. Fisher was the leading exponent of the importance of carefully designed experiments. Fisher and his school developed experimental designs together with an appropriate statistical procedure known as analysis of variance for analyzing the data from such experiments. The principal idea in this work is to arrange the entire experiment in an orderly manner as a set of subexperiments (which may consist of as few as one trial each), these subexperiments differing from each other by varying one or more factors, so as to permit the effect of the variation of the factor (or factors) on the outcome of the experiment to assert itself in case there is such an effect. A statistical test function is designed to detect whether the outcome of the experiment is such that the differences in results from one subexperiment to another are attributable to the variation of the factors being tested, or whether they are simply chance differences; and to measure such differences if they are statistically significant (*i.e.*, unlikely to have occurred by chance).

Suppose samples are drawn from each of two populations that have normal distribution laws with equal variances. The question is this: could the two samples have reasonably come from normal populations with the same mean? If the numbers of items, means, and variances of the two samples are n', n''; $\bar{x}'$, $\bar{x}''$; s'^2, s''^2, respectively, a suitable criterion for making the test is $\sqrt{n}(\bar{x}' - \bar{x}'')/s$ where $n = n' + n''$, and

$$s^2 = \frac{(n' - 1)s'^2 + (n'' - 1)s''^2}{n' + n'' - 2}$$

Thus, the greater the difference between the sample means in relation to s, the greater the value of the criterion and the more unfavourable it is to the hypothesis of equality of population means. This criterion computed from pairs of samples from normal populations having the same means has a distribution law given by equation (1) with $n' + n'' - 2$ degrees of freedom; *i.e.*, with $n - 1$ replaced by $n' + n'' - 2$. Therefore, if the absolute value of the criterion exceeds t_δ, the samples' means are said to be significantly different at the δ probability level, which means that the samples are likely to have come from normal populations with the same means with a risk of probability δ of making a type-I error. This test is useful in testing whether the effects of two treatments A and B on a given measurable characteristic are significantly different; *e.g.*, whether two rations A and B are different in their effects in increasing the weight of pigs. The experiment might be set up with 30 pigs of a given age to receive ration A and 30 more pigs matched with those in the first group on age and weight to receive ration B. Thus, there are two subexperiments. After a given time each of the 30 pigs in each sample would be weighed, and these two sets of weights would be the two samples of x values to be used in the criterion.

Note that in using the two-sample criterion described above it has been assumed that the variances of the populations are equal. Actually, it may be wished to test this assumption, assuming only that the two population distributions are normal. A criterion (sometimes called an F-ratio; sometimes written $\mathcal{J}$-ratio) is available for making this test, namely

$$\mathcal{J} = s'^2/s''^2$$

which has the following distribution law when the hypothesis of equal population variances is true:

$$dF(\mathcal{J}) = \frac{\Gamma\left(\frac{n'+n''-2}{2}\right)}{\Gamma\left(\frac{n'-1}{2}\right)\Gamma\left(\frac{n''-1}{2}\right)}\left(\frac{n'-1}{n''-1}\right)^{\frac{n'-1}{2}} \mathcal{J}^{\frac{n'-3}{2}} \cdot \left(1+\frac{n'-1}{n''-1}\mathcal{J}\right)^{-\frac{n'+n''-2}{2}} d\mathcal{J} \tag{5}$$

which is known as Snedecor's $\mathcal{J}$ distribution and defined on the interval $(0, \infty)$. The $\mathcal{J}$ criterion is intuitively reasonable since small or large values of it are unfavourable to the hypothesis of equal population variances. The critical range R of values of $\mathcal{J}$, for a given level of significance δ, are found by cutting off the lower tail of the distribution at $1/\mathcal{J}_\delta$ and the upper tail at $\mathcal{J}_\delta$ so that the probability of $\mathcal{J}$ falling between $1/\mathcal{J}_\delta$ and $\mathcal{J}_\delta$ is $1 - \delta$. Equation (5) is the basic probability law involved in making statistical tests in analysis of variance.

Next, consider a slightly more elaborate experiment. Suppose it is desired to test whether the quality of screws of a given type as measured by number of defectives per 1,000 varies significantly from machine to machine or from operator to operator. An experiment might be designed along the following lines: consider eight machines and eight operators; let each operator make 1,000 screws on each machine, and let x_{ij} be the number of defectives in the 1,000 screws produced by the ith operator on the jth machine. Assume that x_{ij} is made up of the sum of four parts: m which is the same for all combinations of operators and machines, r_i an effect attributable to the ith operator, c_j an effect attributable to the jth machine, and a random error e_{ij}, which is normally and independently distributed for all combinations of i and j with mean zero and variance σ^2. Then $x_{ij} = m + r_i + c_j + e_{ij}$. The r_i and c_j may be thought of as corrections to m, which may be positive or negative; r_i a correction to be made for the ith operator; and c_j a correction of the jth machine. They are such that

$$\sum_{i=1}^{8} r_i = \sum_{j=1}^{8} c_j = 0$$

The x_{ij} are sample values and what is wished is to set up a criterion for testing whether the operator corrections r_i are all zero (*i.e.*, no average change from operator to operator), and similarly whether the machine corrections c_j are all zero. The likelihood ratio or least squares criterion for testing for no significant variation from operator to operator is

$$7 \sum_{i=1}^{8} (\bar{x}_{i\cdot} - \bar{x})^2 \Big/ \sum_{i,j=1}^{8} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$$

where $\bar{x}_{i\cdot}$ is the average of $x_{i1}, x_{i2}, \ldots, x_{i8}$, with a similar meaning for $\bar{x}_{\cdot j}$; $\bar{x}$ is the average of all x values. The denominator of this expression is rather insensitive to variations due to operators or machines although the numerator tends to be sensitive to operator variations, and the ratio tends to have large values if there is significant variation from operator to operator. The criterion has as its sampling distribution the $\mathcal{J}$ distribution of equation (5) with $n' - 1 = 7$, and $n'' - 1 = 49$ if there is no significant operator to operator variation, which provides a method of obtaining a critical value of the ratio beyond which it would be declared that operator variation is significant. A similar analysis would be made for testing significance of machine to machine variation.

See also references under "Statistics" in the Index.

(S. S. W.)

BIBLIOGRAPHY.—L. G. Gotkin and L. S. Goldstein, *Descriptive Statistics*, 2 vol. (1964); C. R. Li, *Statistical Inference*, 2 vol. (1964); N. L. Johnson and F. C. Leone, *Statistics and Experimental Design in Engineering and the Physical Sciences*, 2 vol. (1964); W. C. Guenther, *Analysis of Variance* (1964); G. M. Smith, *A Simplified Guide to Statistics, for Psychology and Education*, 3rd ed. (1962); F. E. Croxton and D. J. Cowden, *Practical Business Statistics*, 3rd ed. (1960); R. A. Fisher, *Statistical Methods for Research Workers*, 13th ed. rev. (1958); H. M. Walker and J. Lev, *Elementary Statistical Methods*, rev. ed. (1958); G. U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, 14th rev. ed. (1958); D. A. S. Fraser, *Nonparametric Methods in Statistics* (1957).

STATISTICS, MATHEMATICAL. In traditional usage mathematical statistics denotes a conglomeration of mathematical discussions connected with efforts to most efficiently collect and use numerical data subject to random variation (*see* STATISTICS). These efforts can be classified as behaviouristic and inductive.

BEHAVIOURISTIC APPROACH

Behaviouristically, mathematical statistics is concerned with methods people use in guiding their actions to minimize the chances of committing errors. More exactly, it may be defined as dealing with statistical decision functions (*see* below). The definition originates from the human habit (so-called inductive behaviour) of adjusting to pertinent frequency relations. For example, since accumulations of clouds are frequently followed by rain, people who see clouds when embarking on a trip may think of taking an umbrella. Vaccinated persons contract smallpox much less frequently than others; thus smallpox vaccination is almost universal. In these examples the decisions (to take an umbrella or not, to be vaccinated or not) are based on the relative frequency of specific occurrences and the possible consequences of the different decisions.

Thus mass vaccination is desirable if the vaccine frequently prevents disease and if it causes dangerous reactions only infrequently. Obviously, people prefer to speak in terms of certainty instead of in terms of relative frequency. However, either the nature of the phenomena or ignorance of their causes compel an indeterministic approach based on how frequently this or that happens.

Questions of relative frequency are the concern of the theory of probability (*see* PROBABILITY, MATHEMATICAL). It deals with so-called chance mechanisms that combine the unpredictability of a single trial with the stability of relative frequencies in a long series of trials. The probability is an idealization of the long-run relative frequency of an outcome of a chance mechanism. When the probability that a given coin will fall heads is said to equal $\frac{1}{3}$, it means that this unusual coin tends to fall heads one-third of the time in a long series of tosses.

In the theory of probability the term random variable denotes a variable quantity determined by a chance mechanism. If X is a random variable and x any specified number, then the relevant chance mechanism determines the probability $P[X \leq x]$ that X will assume a value not exceeding x. This same mechanism determines the probability $P[X = x]$ that X will equal x. The probability $P[X \leq x]$, considered as a function of x, determines the properties of the random variable X and is called its distribution function, or simply its distribution.

Solution of typical probability problems consists in determining the distributions of some random variables from the given distributions of other random variables. A typical result of elementary probability theory is the distribution of the so-called binomial variable X (*see* BINOMIAL THEOREM), defined as the number of successes in n independent trials, where the probability of success has the same value p. Here n is a positive integer and p an arbitrary number between 0 and 1. The distribution function of X so defined may be obtained from

$$P[X = x] = \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x} \tag{1}$$

for $0 \leq x \leq n$. In particular, this formula may serve to compute the probability that in $n = 10$ independent tosses, with a coin falling heads with a long-run frequency p of $\frac{1}{3}$, heads will appear not more than once, twice, and so on.

This example deals with a chance mechanism, the throw of a coin, repeated independently $n = 10$ times; that is, the outcome of any given throw is assumed to have no effect on the outcomes of the others. If the order in which heads may occur is ignored, there are $n + 1 = 11$ possible outcomes: the number X of heads thrown may be 0, 1, 2, . . . , 10.

The unpredictably varying outcomes of scientific experiments are treated as manifestations of a chance mechanism and the quantitative characteristics of these outcomes are treated as random variables. Simplifying severely, the results of the use of a vac-

cine on a particular subject may be expected to vary according to his individual susceptibility to the disease, to aftereffects of vaccination, and according to the degree of his exposure to infection with the disease.

The decision for or against mass vaccination may be based on an experiment using a random selection out of a large parent population (universe) of n (experimental) individuals to be vaccinated and an equal number of (control) individuals to be left untreated. Then all $2n$ individuals are exposed to infection and observed. The outcome may be described by the number X_1 of "good" results among the vaccinated and the number X_2 of such results among the controls. The two variables X_1 and X_2 have the same set of possible values: $0, 1, 2, \ldots, n$. It is assumed that X_1 and X_2 are random variables with approximate distributions represented by equation (1), with $p = p_v$ for the vaccinated group and $p = p_0$ for the nonvaccinated. The symbol p_v represents the proportion of individuals in the parent population for whom the vaccine would give "good" results: p_0 stands for the proportion of the parent population for whom such results would be observed without vaccination. The desirability of mass vaccination depends on the values of p_v and p_0. Unfortunately these values are unknown and the decision for or against mass vaccination must be taken on the values of the random variables X_1 and X_2 found empirically from the sample of subjects in the experiment.

In practice, the outcomes of trials with a vaccine are not classified as just "good" and "bad," and many different trials are performed. In many situations there is a set A of possible actions $a_1, a_2, \ldots$ from which one must be selected. Also, the desirability of this or that action depends on the distribution F of some observable random variables $X = (X_1, X_2, \ldots, X_n)$. The distribution F is unknown, at least in all details, and the choice of action has to be based on the observed values $x = (x_1, x_2, \ldots, x_n)$ of the random variables X. Each possible combination x of values of X is described as a possible sample point and the set of all such combinations is the sample space.

In the vaccine example there are only two observable random variables X_1 and X_2. The sample space W consists of $(n+1)^2$ different points with coordinates $(0,0), (0,1), (1,0), \ldots, (n,n)$. The datum of the general problem also includes a family or set Ω of possible distributions of the observable random variables X, described as the set of admissible hypotheses or the set of possible states of the universe. To each distribution F of the set Ω there corresponds a "preferred" action, and for each F in Ω and each possible action a in A there is a more or less explicitly ascertainable degree of desirability. The datum of the problem may (but usually does not) give the so-called a priori probabilities of the various distributions F of the set Ω (*see* PROBABILITY: *Inverse Probability*).

In the vaccine example, when the population from which subjects are selected is large, the family Ω of possible distributions of the two observable variables X_1 and X_2 is determined approximately by all combinations of two equations of type (1). If p_0 (the potential proportion of "good" results in the parent population without vaccination) is greater or equal to p_v, then the preferred action is a_1 (abstain from mass vaccination). On the other hand if p_v is greater then the preferred action is a_2 (embark on mass vaccination) with the desirability of a_2 increasing with the difference $p_v - p_0$.

Any systematically defined rule for selecting actions can have the following form: if the observations yield $X = x'$, then take action $a(x')$; if the observations yield $X = x''$, then take action $a(x'')$, and so on, for all possible sample points. In effect a rule for making decisions must establish the action to be taken as a function of the sample point to be observed. This function $a(x)$ is called the statistical decision function. Since X is a random variable, so is the degree of desirability of the action taken using a given decision function. The datum of the problem allows calculation of the probability (at least the limits within which the probability is inferred to lie) that a given decision function will lead to the adoption of a very undesirable action. Thus mathematical statistics may be defined as the mathematical discipline dealing with properties of statistical decision functions. The general problem of mathematical statistics is to determine (from the data of the given problem) the particular decision function that has specified desirable properties; *e.g.*, the decision function (if it exists) that minimizes the probability of a wrong decision.

Probabilities are involved in probability theory and in mathematical statistics; if the unknown is a probability, the problem is one of probability theory; if the unknown is a decision function, the problem belongs to mathematical statistics. The viewpoint and terminology used here are generalizations due to Abraham Wald (1902–50) of two earlier theories described below. While many mathematical statisticians would agree with this characterization of their field, others would feel it too restrictive.

Theory of Testing Statistical Hypotheses.—The theory of testing statistical hypotheses, developed in the 1930s by Jerzy Neyman and Egon S. Pearson, is a chapter of mathematical statistics referring to the case where the set A of possible decisions contains only two elements: action a_1 and action a_2 considered as the negation of a_1 (*e.g.*, a_1 = arrange a mass vaccination campaign, and a_2 = do not). The terminology refers to statistical hypotheses and their rejection. A statistical hypothesis means any assumption regarding the distribution of the observable random variables. The fact that only two possible actions are contemplated implies the existence of two contradictory hypotheses H_1 and H_2 such that if H_i happens to be true then the preferred action is a_i for $i = 1,2$. This indicates symmetry between H_1 and H_2 that usually disappears when the consequences of selecting a wrong action are considered. There are two different situations in which a wrong action may be selected; *i.e.*, there are two different kinds of error possible in testing a hypothesis:

(H_1,a_2) = the true hypothesis is H_1 and the action taken is a_2 which is reject H_1

(H_2,a_1) = the true hypothesis is H_2 and the action taken is a_1 which is accept H_1.

Alternatively, a_2 may be described as acceptance of H_2, and a_1 as rejection of H_2. Usually error (H_1,a_2) is qualitatively different from error (H_2,a_1). According to the individual's subjective evaluation one of the errors may be judged more important to avoid than the other. The one held more important to avoid is called the error of the first kind; the other is called the error of the second kind. One of the hypotheses H_1 and H_2 is selected to be the hypothesis tested or the null hypothesis H so that unjust rejection of H represents the error of the first kind. Thus if (H_1,a_2) is more important to avoid than (H_2,a_1) the hypothesis tested is H_1. For two contemplated actions the definition of any statistical decision function is now equivalent to selecting a subset w of the sample space W, called the critical region or region of rejection. Then a rule is made to reject the tested hypothesis H whenever the observations determine a sample point falling within w, and to abstain from this action in all other cases. Thus to define a test means to select a critical region.

The general character of the problems of testing hypotheses may be illustrated in the simplest case where the set Ω of admissible hypotheses contains only two hypotheses (H and its negation $\bar{H}$), each specifying the distribution of the observable variables in all details. It is assumed that there is a clear distinction in the importance of avoiding the two kinds of error.

Noting the importance of avoiding the error of the first kind, arbitrarily select a small number α called (somewhat inappropriately) the level of significance, and seek the critical regions $w(\alpha)$ such that their use will insure that the probability of committing an error of the first kind cannot exceed α. Ordinarily there are many such regions $w(\alpha)$; select for use the one $w_0(\alpha)$ that insures the greatest probability of avoiding the error of the second kind. The region $w_0(\alpha)$ that satisfies these two conditions is called the most powerful critical region for testing H against $\bar{H}$ at the level of significance α. Its use insures that the probability (i) of rejecting H when H is true is no more than the preassigned level α and, subject to restriction (i), the probability (ii) of rejecting H when H is false is greatest.

After a region $w_0(\alpha)$ is found (on occasion there are many) it is appropriate to compute the probability $\beta\ [\omega_0(\alpha)]$ (termed power) that the test will reject H when this hypothesis is false.

When the selected level of significance α is very low then the power of the test is very low too, and application of the test does not hold much promise in "detecting the falsehood of H." It may then be decided to take a greater risk in controlling the error of the first kind (*i.e.*, a larger α) to increase chances of rejecting the hypothesis tested when it is false.

In summary, the basic concepts of the Neyman-Pearson theory are those of critical region, two kinds of error, and power of a test.

Theory of Estimation.—The problem of statistical estimation has two different forms, known as point estimation and interval estimation (more generally, estimation by a set). Both are particular cases of the general problem of mathematical statistics and refer to situations in which the set of contemplated actions contains (roughly speaking) a large number of actions, comparable with the number of distributions of the observable random variables X. More specifically, to each distribution F of the set Ω there corresponds one particular preferred action $a(F)$ belonging to A. In these circumstances the decision to take a particular action $a(F)$ in a sense is equivalent to the assertion that the function F represents the distribution of X. Such assertions based on observed values of random variables represent the process of statistical estimation.

In many cases the possible distributions of observable random variables X are identified by values of one or more parameters $\theta_1, \theta_2, \ldots, \theta_s$. Thus the distributions of variables X_1 and X_2 in the vaccine example are identified by values of two parameters p_v and p_0. In such situations the problem of estimation is reduced to that of making assertions regarding the values of all or of some of the parameters involved. Consider one particular parameter θ with possible numerical values contained in interval (a,b).

It is called point estimation of the parameter θ when the statistical decision function contemplated has the following form: if the observations yield the sample point $X = x$, then θ has the value $\hat{\theta}(x)$. Here $\hat{\theta}$ stands for a function defined for all possible sample points and has its values in the interval (a,b). The function $\hat{\theta}$ so used is called a point estimator of θ.

It is called interval estimation of θ when the decision function contemplated depends on functions $\hat{\theta}_1$ and $\hat{\theta}_2$, both defined over the sample space W, both having their possible values in the interval (a,b), and when for every possible sample point $\hat{\theta}_1 < \hat{\theta}_2(x)$. Then the statistical decision function is defined as follows: if the observations yield $X = x$, then assert that $\hat{\theta}_1(x) \leqq \theta \leqq \hat{\theta}_2(x)$. In other words instead of trying to predict the exact value of θ, interval estimation involves an effort to predict the limits $\hat{\theta}_1(x)$ and $\hat{\theta}_2(x)$ within which this value presumably lies.

Suppose functions $\hat{\theta}_1(x)$ and $\hat{\theta}_2(x)$ are defined. Substituting random variable X for its particular value x produces a random interval $I(X) = [\hat{\theta}_1(X), \hat{\theta}_2(X)]$. This interval is called an interval estimator of θ.

Explicit consideration of the problems of point estimation goes back at least to Galileo, but the major early figures in the development discussed here were Pierre Simon Laplace (1749–1827) and Carl Friedrich Gauss (1777–1855). Their view, adopted by A. A. Markov (1856–1922) and subsequently generalized by Wald, may be described by comparison with a game of chance in which the statistician can never win, but can lose an amount that increases with the size of the error of estimation. For every possible value ϑ_1 of the estimated parameter θ, and for every possible value ϑ_2 of the point estimator, there is assumed a loss function $L(\vartheta_1,\vartheta_2) \geqq 0$ representing the amount to be paid by the statistician if $\theta = \vartheta_1$ and he asserts that $\theta = \vartheta_2$. The exact nature of the loss function must be specified by practical considerations; Laplace considered $L(\vartheta_1,\vartheta_2) = |\vartheta_1 - \vartheta_2|$; Gauss preferred $L(\vartheta_1,\vartheta_2) = (\vartheta_1 - \vartheta_2)^2$ and subsequently other loss functions were considered. The various aspects of point estimation are based on different limitations imposed on the estimator, on different loss functions adopted, and on different definitions of optimum properties of an estimator. Here is an example:

(a) The estimator $\hat{\theta}(X)$ shall be a linear function of the observable random variables $\hat{\theta}(X) = a_1X_1 + a_2X_2 + \ldots + a_n X_n$. This restriction insures that the estimator will be easy to compute.

(b) The estimator $\hat{\theta}(X)$ shall be unbiased; *i.e.*, its mathematical expectation shall be equal to the parameter θ.

(c) The loss function $L(\vartheta_1,\vartheta_2)$ shall be equal to the square of the error of estimation (*i.e.*, shall be Gaussian).

(d) Under conditions (*a*), (*b*), and (*c*), the estimator $\hat{\theta}(X)$ shall minimize the mathematical expectation of the loss.

This particular problem was first formulated by Gauss who made a very elegant broadly applicable solution that became known as the method of least squares (*see* Least Squares, Method of). Developments in this direction include the study of estimators subject to conditions (c) and (d), but not necessarily (a) and (b); loss functions other than the Gaussian also are considered.

Although vague ideas underlying interval estimation can be traced to the end of the 18th century, the first clear statement was formulated in 1931 by Harold Hotelling. In the 1930s a comprehensive theory was developed by Neyman. The basic concept here is that of the confidence interval. It is said that the interval estimator $I(X)$ is a confidence interval for estimating θ at the confidence coefficient α if, whatever be the value ϑ of the estimated parameter θ, the probability that the interval $I(X)$ will cover ϑ is at least equal to α.

It is shown that (in many cases) more than one confidence interval for a given parameter can be found, all corresponding to the same confidence coefficient α. In such cases the problem of optimality arises, an important section of the theory. A typical problem is as follows:

(a) For a given confidence coefficient α (say $\alpha = .99$), and whatever be the true value ϑ of the parameter θ, the interval estimator $I_0(X)$ shall cover ϑ with probability equal to α.

(b) Whatever be the true value ϑ_1 of θ, whatever be the false value ϑ_2 of the same parameter, and whatever be the interval estimator $I(X)$ satisfying (a), the interval estimator $I_0(X)$ shall cover the false value ϑ_2 less frequently (more precisely, not more frequently) than $I(X)$.

If an estimator $I_0(X)$ satisfying (a) and (b) exists, it is called the shortest confidence interval for estimating θ. The systematic use of shortest confidence intervals (to estimate the same parameter or different parameters) insures the prescribed long-run frequency (*e.g.*, 99%) of correct assertions combined with the greatest precision of such assertions understood in the sense of condition (b).

Theory of Experiments.—Statistical considerations initiated an important scientific discipline called theory of experiments or design of experiments. The first attempt to build a unified theory was by Ronald A. Fisher (1930s) who sought ways of designing experiments to provide more relevant information than others requiring the same investment in funds and effort. Thus originated the systematic study of optimal designs.

Fisher and his school were chiefly concerned with designs of single experiments, with what might be called experimental tactics. Developments concerned with experimental strategy are due to Wald (1940s), and deal with a sequence of experiments that in principle may be continued indefinitely.

Fisher originally was concerned with agricultural experimentation. Later it was found that the same kinds of problems are encountered in almost all experimental work. One of these problems is that of biased selection of material used in experiments. To avoid such bias he introduced the principle of randomization (*see* Randomization), an outstanding achievement.

Many experimental tactics introduced by Fisher and developed by his school are meant to diminish the effects of variability in experimental material, and include such specific designs as randomized blocks and Latin squares (*see* Magic Square). In testing a drug on mice, the randomized blocks design consists in using N litters of mice and in selecting at random from each litter; for example, three mice to be treated and another three to be left untreated. The six mice of each litter so selected constitute one randomized block. Considering substantial hereditary variation from one litter to the next, intuition and theory agree that such an experiment is less subject to random error than one based on unrestricted randomization of $6N$ mice.

Fisher was also concerned with the possibility of experiments answering several questions at once. This may be illustrated by an elegant example due to Frank Yates. Consider an effort to weigh three objects A, B, and C with a balance that has an undetermined (changeable) zero point so that at each weighing the indicator has to be read while there is no load. Let the symbol (1) denote such a reading; also let (A), (B), . . . , (BC), and (ABC) stand for readings when the scale is loaded with A alone, with B alone, . . . , with B and C combined, and with A, B, and C combined.

An estimate of weight A will be provided by the difference (A) − (1); *i.e.*, weight A differs from (A) − (1) only by an unavoidable random experimental error. One method of weighing A would be to make some (say $4n$) independent readings (1) and the same number of (A), to compute $4n$ differences (A) − (1) and to average them. Done separately for all three objects A, B, and C, the total number of readings would be $24n$. With the so-called factorial design due to Fisher, estimates of these three weights of exactly the same precision may be made from only $8n$ readings. This is done by making n independent readings each of (1), (A), (B), (C), (AB), (AC), (BC), and (ABC). To estimate weight A arrange the $4n$ differences as follows:

(A) − (1)
(AB) − (B)
(AC) − (C)
(ABC) − (BC)

It may be seen that each of these differences is an independent estimate of weight A and that the average of $4n$ of them is just as precise an estimate as the average of $4n$ differences (A) − (1). The same readings provide an estimate of weight B as follows:

(B) − (1)
(AB) − (A)
(BC) − (C)
(ABC) − (AC)

A similar procedure is used for weight C. Thus one factorial experiment of $8n$ readings can yield the precision requiring three times as many readings arranged conventionally.

A limitation on factorial designs is imposed by the requirement that the treatments studied must be additive. In the example just given (assuming a robust balance is used) this condition will be satisfied so that, apart from random error, (A) − (1) is equal to (AB) − (B), and so on. However, if the phenomena studied are not weights but the effects of chemical elements in fertilizers (*e.g.*, nitrogen, potassium, and phosphorus) on the yield of wheat per acre, then the effect of nitrogen alone (N) − (1) need not (and usually does not) equal the effect of the same nitrogen in the presence of potassium (NK) − (K), and so on. Thus in such a case the use of fractional design may lead to unexpected complications.

In problems of experimental strategy concerned with sequences of tactically correct experiments, after n experiments have been completed there are different ways in which the $(n + 1)$th experiment may be designed. There is also the possibility of discontinuing experimentation and reaching a so-called terminal decision. The problem of experimental strategy is to develop rules for using the results of n completed experiments to decide what to do next. The statement of the general problem and some remarkable mathematical developments were published by Wald (1950). One strategic problem of experimentation has attracted considerable attention and created a substantial literature. It is known as the problem of the Robbins-Monroe procedure and consists in the following.

For every number x in the interval (a,b) an experiment can be made yielding a value of a random variable $Y(x)$. Only a limited amount of information is available about $Y(x)$. However, this random variable is known to have an expectation $M(x)$. The exact nature of the function $M(x)$ is not known except that it is continuous and increasing and that for some value x_0 of x it is equal to zero. The experimental problem consists in determining the approximate value of x_0. The Robbins-Monroe procedure for attaining this goal begins by selecting in (a,b) an arbitrary number x_1, perhaps the midpoint of (a,b), and in observing $Y(x_1)$. Let the observed value be y_1. Then using the values x_1 and y_1 another choice is made of a number x_2 within (a,b) and an experiment is made to observe $Y(x_2)$. The result of this observation is denoted by y_2. By continuing this procedure it is hoped to obtain a sequence of numbers $x_1, x_2, \ldots x_n, \ldots$ approaching the unknown x_0 as a limit. The problem of experimental strategy is to determine how best to use the observed values of x and y to choose x_{n+1} so that the sequence of x values converges rapidly to x_0.

The characteristic feature that identifies this as a problem of experimental strategy is the chainlike dependence of the arrangements for the $(n + 1)$th experiment on the designs and on the outcomes of the earlier experiments.

INDUCTIVE APPROACH

A view that is sharply antagonistic to the behaviouristic approach is represented by several inductive schools of thought symbolized by the names of Fisher and Harold Jeffreys. As in the behaviouristic approach, all these schools are concerned with the best utilization of data subject to random variation. Hence their ultimate results (unavoidably) are particular statistical decision functions; that is, the conclusion is based on the observational data. However, the statistical decision functions are not considered explicitly; when a particular function is proposed, arguments in its favour are formulated not in terms of the consequences of its use but in terms of degrees of confidence in the underlying hypotheses. Essentially each inductive school begins by laying down one or more basic formulas (posterior probability of a hypothesis in some cases, likelihood and fiducial probability in others) represented as norms of human beliefs and actions. The handling of these formulas is then labeled by such terms as inductive reasoning or scientific inference. More radical inductive arguments in favour of a particular point estimator (called maximum likelihood estimator) ignore the consequences of its systematic use but refer to the fact that the estimator maximizes the likelihood described (Fisher) as "the mathematical quantity . . . appropriate for measuring our order of preference. . . ." In relatively simple problems the decision functions developed by inductive schools frequently coincide with those of behaviouristic theory. In particular, as shown by Hotelling, Joseph L. Doob, and Wald, the maximum likelihood estimators have desirable properties in a broad category of problems. However, in the 1960s these inductive systems represented major areas of controversy in mathematical statistics.

HISTORICAL BACKGROUND

Development of the conceptual content of mathematical statistics was gradual, extending over a century and a half. As noted earlier the behaviouristic approach to problems of estimation is found in the work of Laplace and Gauss. Laplace seems one of the first to test statistical hypotheses; but at that time there was no question of mathematical statistics as an independent discipline. Laplace's criteria for testing statistical hypotheses were intuitive rather than mathematical and in no way similar to his approach to the problem of point estimators. The behaviouristic approach to estimation was soon forgotten.

Mathematical statistics as an independent discipline seems to have appeared in the latter part of the 19th century in the work of A. Quételet, G. T. Fechner, K. Pearson (*qq.v.*), E. Czuber, G. U. Yule, and L. von Bertkievicz. The then existing branches of applied mathematics, (*e.g.*, analytical mechanics) were, so to speak, individualistic. Many scientific problems required a new mathematical discipline concerned with aggregates of objects, all of which satisfy a specific definition but differ in individual characteristics. Such an aggregate (population) might be all men of military age in a country; these men differ in height. Another population would be the set of gas molecules in a container; each molecule has a unique velocity as its individual characteristic.

However, any population has its own characteristics, conditioned by those of the individual members, but essentially different. Gas pressure is a characteristic of the population of molecules con-

ditioned by their velocities; temperature is another such characteristic. These are examples of what Fechner called *Kollektivmass*. The average height of recruits is also a *Kollektivmass*, as is the frequency with which they are found with height between 5 ft. 10 in. and 5 ft. 11 in. The set of such frequencies (the frequency distribution) is also a *Kollektivmass*.

In the late 19th and early 20th centuries mathematical statistics emphasized the *Kollektivmasslehre*, a mathematical discipline concerned with collective characteristics of populations. The term descriptive statistics was introduced to mean the use of a variety of methods (including graphical and tabular methods) for describing such characteristics of populations. An important method of descriptive statistics is to consider a family of flexible curves or surfaces (depending on a moderate number of parameters) that can be used to approximate the empirical frequency distribution. A number of such families was developed, all representing interpolation formulas. The most successful system seems to be that due to Karl Pearson.

A conceptual problem studied by Wilhelm Lexis (1837–1914) developed to deprive *Kollektivmasslehre* of its independent status. Lexis was concerned with the year-to-year fluctuation of sex ratio among children born in a city, and similar ratios arising in the study of vital statistics. The question was whether an empirical index of dispersion is consistent with the assumption that determination of sex is governed by a simple chance mechanism such as a toss of a coin. This renewed a continuing effort concerned with chance mechanisms acting on individuals to produce the observed collective characteristics.

Interest in flexible formulas specifically developed to fit empirical distributions gradually decreased, except for those that could be deduced from clearly defined chance mechanisms. In the 1920s George Pólya constructed a system of chance mechanisms that can generate almost all the distributions of Karl Pearson's system. Thus mathematical statistics shifted from *Kollektivmasslehre* to the construction of chance mechanisms or (as they are called) stochastic models of phenomena. This idea was explicitly stated by Émile Borel (1871–1956): "The basic problem of mathematical statistics is to invent a system of simple chance mechanisms, such as throws of a coin, so that the probabilities determined by this system agree with the observed relative frequencies of the various details of the phenomena studied."

Although Borel's definition was appealing, developments following its formulation indicated that it leaves mathematical statistics without a field of its own. Depending on the attitudes of given research workers, stochastic models belong to the relevant substantive fields or to the theory of probability. Thus stochastic models of hereditary phenomena are an integral part of genetics; stochastic models of epidemics are an integral part of epidemiology; and so on. However, considerable literature on stochastic processes developing in time (*e.g.*, birth and death stochastic process constructed to represent the growth of biological populations) is one of the most active sections of the theory of probability. Each problem of the theory of stochastic processes is reducible to that of determining probabilities (*e.g.*, that a biological population will eventually die out). While the field indicated in Borel's definition continued to grow in importance, it was claimed by other disciplines, leaving nothing to mathematical statistics.

While these ideas were in the process of dissolution, increasing effort was given to problems of estimation and of testing hypotheses. In connection with his system of curves, Karl Pearson developed a method of point estimation known as the method of moments; in 1900 he published his famous χ^2 (chi-square) test for goodness of fit. In 1908 F. Y. Edgeworth (1845–1926) found that consistent use of the method of moments must yield an excessive frequency of large errors of estimation (a distinctly behaviouristic approach) and proposed a new method of estimation, conjectured to be much better. Fourteen years later, Fisher discussed the same ground more rigorously and intensively. He introduced the term method of maximum likelihood. Student (pseudonym of W. S. Gosset) and Fisher developed a number of tests of particular hypotheses. However, studies of this kind were not yet connected by any unifying idea, and were limited to particular decision functions. Problems of testing hypotheses and of estimation remained unrecognized as an independent field for systematic study.

Such recognition came when Fisher (1922) gave a new definition of statistics. Stating that the object of statistical methods is the reduction of bulky data, Fisher distinguished three basic problems: those of specification of the kind of population from which the data come (*i.e.*, of the *Kollektivmass*), of estimation, and of distribution (probabilistic problems connected with point estimation). This was followed by a paper concerned specifically with the theory of estimation; the ideas underlying these papers are distinct precursors of the behaviouristic approach to mathematical statistics described earlier.

See also GAMES, THEORY OF.

BIBLIOGRAPHY.—R. B. Braithwaite, *Scientific Explanation* (1953); W. G. Cochran and G. M. Cox, *Experimental Designs*, 2nd ed. (1957); H. Cramér, *Mathematical Methods of Statistics* (1946); R. A. Fisher, *The Design of Experiments*, 7th ed. (1960), *Statistical Methods and Scientific Inference* (1956); M. Fisz, *Probability Theory and Mathematical Statistics* (1963); H. Jeffreys, *Scientific Inference*, 2nd ed. (1957); E. L. Lehmann, *Testing Statistical Hypotheses* (1959); A. M. Mood and F. A. Graybill, *Introduction to the Theory of Statistics*, 2nd ed. (1963); J. Neyman, *First Course in Probability and Statistics* (1950), *Lectures and Conferences on Mathematical Statistics*, U.S.D.A. Graduate School (1953); L. J. Savage, *The Foundations of Statistics* (1954); A. Wald, *Sequential Analysis* (1947), *Statistical Decision Functions* (1950); S. S. Wilks, *Elementary Statistical Analysis* (1948).

STATIUS, PUBLIUS PAPINIUS (*c.* A.D. 45–96), one of the principal epic and lyric poets of the silver age of Latin literature, was born in Naples. His father was also a poet, and Statius seems to have been trained as one from childhood. Little is known of his life. He lived in Rome, and was a court poet under the emperor Domitian, who awarded him a prize at the poetic festival at Alba, in A.D. 89 or 90. He was, however, unsuccessful in the Capitoline competition, probably on its third celebration in 94, and shortly afterward he returned to Naples.

The role of court poet seems to have suited Statius. His flattery is as different from the gentle propitiatory tone of Quintilian as it is from the exaggerated humility of Martial. Statius adopted without scruple the flattery that was inevitable under Domitian, exploiting it in a way which suited his own nature. He was exceptionally talented, and his poetic expression is, with all its faults, richer, as well as more buoyant and felicitous than is generally to be found in silver Latin.

Statius is at his best in his occasional verses, the *Silvae*, in five books containing 32 poems altogether, of which all but 6 are in hexameters. Five poems are devoted to flattery of the emperor and his favourites. Another group gives picturesque descriptions of the villas and gardens of the poet's friends. The *Kalendae Decembres* gives a striking description of the gifts and amusements provided by the emperor for the Roman populace at the Saturnalia. In his attempt at an epithalamium Statius is forced and unhappy, but his birthday ode in Lucan's honour has, along with the accustomed exaggeration, some good lines, and shows appreciation of earlier Latin poets.

Those poems, too, which deal with family affection or personal loss are moving, and the short poem to Sleep (v, 4) is deservedly famous.

Statius also wrote two epics, the *Thebais* (in 12 books) and the *Achilleis* (in two, unfinished). The latter gives a charming and accomplished account of the early education of Achilles by Chiron the centaur, and of his being hidden in girl's clothing among the daughters of Lycomedes by his mother, Thetis, in order to save him from the Trojan war. He is discovered by Odysseus and taken off to Troy; but at this point the poem was evidently interrupted by the poet's death.

The *Thebais* is a more ambitious work, describing the struggle of the brothers Polynices and Eteocles for the throne of Thebes. It has been called an "episodic" epic; but apart from one long digression it keeps to its theme well enough. It has many features borrowed from Virgil, such as the catalogue of warriors, the games and the battle scenes. It often suffers from overstatement and exaggeration, and some of the scenes of horror are too gruesome,

but many more are admirable and original, such as the fight between Tydeus and his pursuers in book ii and the martyrdom of Menoeceus in book x. The work begins and ends with two passages, the fight in a storm between the two young warriors on the threshold of the king's palace at Argos and the meeting of Polynices' mother and sister at night on the battlefield, which convey an atmosphere of dramatic tension and reveal considerable tragic power.

BIBLIOGRAPHY.—Editions in the "Teubner Series" by A. Klotz, 3 vol. (1902–07); in the Oxford classical texts by H. W. Garrod (*Thebais* and *Achilleis*, 1906) and J. S. Phillimore (*Silvae*, 2nd ed., 1918); in the "Loeb Series" with Eng. trans. by J. H. Mozley, 2 vol. (1928); *Silvae* with commentary by F. Vollmer (1898); *Thebais* i–vi with commentary by O. Müller (1870); *Achilleis* with English commentary by O. A. W. Dilke (1954). Eng. trans. of the *Silvae* by D. A. Slater (1908). (AN. K.)

STATUE OF LIBERTY. *See* LIBERTY, STATUE OF.

STATUTE LAW. A statute was defined by Sir Edward Coke as "an Act of Parliament made by the King, the Lords and the Commons." This definition is still apt in English law, save that in certain cases, provided for by the Parliament act, 1911, a bill may become an act even though the house of lords has not consented to it. The procedure for the enactment of statutes is regulated mainly by usage and the standing orders of parliament. Church assembly measures passed under the Church of England Assembly (Powers) act of 1919 and presented for the royal assent after a favourable resolution of both houses of parliament have the power and effect of acts of parliament.

The earliest English statute is the Statute of Merton, 1235. Acts were engrossed upon the statute roll from the time of the Statute of Gloucester (6 Edw. I) but this practice has long been discontinued. Official copies of acts are now made on vellum and lodged at the public record office and in the house of lords. In the case of Church assembly measures, a third vellum copy is printed and sent to the Church assembly.

Classification of Statutes.—Acts of parliament are classified as (1) public general acts; (2) local (or "local and personal") acts, including public acts of a local character; (3) private personal acts; (4) Church assembly measures (*see* above).

With the exception of certain classes of private personal acts (such as naturalization and divorce acts), which are virtually obsolete, all acts of parliament are printed by the queen's printer. There are two separate series of volumes, one containing the public general acts and Church assembly measures, the other containing local and personal acts together with certain statutory orders (*see* below).

For the purposes of the law of evidence, however, a different classification applies, and by section 9 of the Interpretation act, 1889, every act passed after 1850 is a public act unless the act expressly provides to the contrary. Such acts are judicially noticed and no proof of them is required. Acts which are stated not to be public acts are usually provable by the production of a queen's printer's copy.

Acts may also be classified according to their object by such terms as penal, beneficial, imperative, directory, enabling, disabling, taxing, codifying, consolidating, perpetual, temporary and so forth. This last classification is important for the purposes of interpretation (*see* below).

Penal acts are those which impose a new disability; beneficial, those which confer a new favour. An imperative statute (often negative or prohibitory in its terms) makes a certain act or omission absolutely necessary, and subjects a contravention of its provisions to a penalty. A directory statute (generally affirmative in its terms) recommends a certain act or omission, but imposes no penalty on nonobservance of its provisions. To determine whether an act is imperative or directory the act itself must be looked at, and many nice questions have arisen on the application of this rule of law to a particular case. Enabling statutes are those which enlarge the common law, while disabling statutes restrict it. Declaratory statutes, or those simply affirming the common law, once not uncommon, are now practically unknown; the Treason act is an example. Statutes are sometimes passed in order to overrule specific decisions of the courts. Examples are the Territorial Waters Jurisdiction act, 1878, the Married Women's Property act, 1893, and the Trade Disputes act, 1906.

Every act has a short title by which it may be cited; *e.g.*, the Transport act, 1953.

Interpretation of Statutes.—The Interpretation act, 1889, provides an authentic interpretation for numerous words and phrases of frequent occurrence in statutes. Moreover, most modern acts contain an interpretation section which explains particular expressions in that act. The main body of the law on this subject is, however, to be found in judicial decisions; *i.e.*, the common law. The main rules are as follows: (1) The meaning of an act is primarily to be sought in the act itself. (2) The words of an act are to be construed in their ordinary and natural sense, unless in the case of particular expressions there is something to the contrary in the context or in the scheme of the act. (3) The act is to be construed as a whole and acts *in pari materia* are to be construed together as forming one system. (4) Contemporaneous circumstances may be taken into account where the intention of the legislature cannot otherwise be ascertained. What is known as the rule in Heydon's case (1584) (3 Coke Rep., 7a) provides that the factors to be ascertained in this connection are (*a*) what was the common law before the act; (*b*) what was the mischief or defect for which the common law did not provide; (*c*) what remedy parliament has resolved and appointed to cure the disease; and (*d*) the true reason for the remedy. (5) Where words can be interpreted in two different ways the inconvenient, unjust or absurd consequences of adopting the one way may guide the court to adopt the other. (6) The full title and the preamble but not the short title, heading or marginal note may be used for the purpose of interpreting the act as a whole, or in order to resolve any ambiguity. (7) Acts which purport to interfere with existing rights must be clearly expressed. (8) Statutes are to be so interpreted as far as their language admits as not to be inconsistent with the comity of nations or with the established rules of international law. (9) Some acts are to be liberally or benevolently construed, others strictly. Instances of those which fall into the former category are acts which are remedial; amending; explanatory; for the maintenance of religion, the advancement of learning and the relief of the poor; directory as distinguished from mandatory or imperative enactments; and those in favour of the liberty of the subject. The acts to be strictly construed include those restricting the liberty of the subject; penal acts; fiscal and revenue acts; acts creating statutory powers and duties; private acts; acts affecting the prerogative of the crown, existing private rights or the jurisdiction of the courts. In the case of penal statutes, for instance, strict construction means that the thing charged as an offense must be within the plain meaning of the words used, and must not strain the words on any notion that there has been a slip, that there has been a *casus omissus*, or that the thing is so clearly within the mischief in the mind of the legislature that it must have been intended to be included and would have been included if thought of. Great care must be taken not to extend penal laws beyond what the legislature has clearly meant and adequately expressed. Similarly in the case of taxation statutes the court is not to be guided so much by the objects which it thinks such acts are to achieve, as by considering whether the words of the act have reached the alleged subject of taxation. There is no presumption as to a tax; nothing is to be read in; nothing is to be implied. One can only look fairly at the language used. (10) Codifying and consolidating acts give rise to an inference that it is intended to reproduce the existing position and not to alter the law. A codifying act is, however, in the first instance, to be looked at without reference to the previous law if it is unambiguous, whereas reference is to be made to the previous law to ascertain the intention of the legislature in the case of an amending or consolidating act. (11) In construing acts, such as taxing acts, which apply to both England and Scotland, it is well settled that a construction should, if possible, be adopted which will make the incidence of taxation the same in both countries. (12) A special act which imposes mutual obligations on several parties so that it has a contractual operation should as between parties be construed in the same way as a contract to the same effect.

Operation of Statutes.—The date of the "passing" of an act is the date when it becomes law. This is the date when it receives the royal assent. This date is also, in the absence of a provision to the contrary, the date of the "commencement" of the act; *i.e.*, the date when its provisions come into effect.

An act of the United Kingdom parliament is to be construed *prima facie* as applying to the whole of the United Kingdom and not to any place outside. The United Kingdom does not include the Channel Islands or the Isle of Man.

An "extent" or "application" clause will, however, vary the effect of this rule. Acts are extended to apply to waters adjacent to the United Kingdom and other parts of the British Commonwealth (*see* below) for certain purposes.

An act generally applies only to those within the allegiance of the crown and does not extend to foreign subjects out of the queen's dominions. An act applicable to the queen's dominions is, if the context permits, to be construed as applying to all British subjects wherever they may be.

It is a clear rule of law that the crown is not bound by an act of parliament, unless it is expressly named or clear words of implication are used. The crown for this purpose includes the great departments of state, but bodies such as the British Transport commission and other nationalized industries have been held not to be the servant or agent of the crown for this purpose. The crown may, however, be held to be bound by implication if the beneficent purpose of the act would be otherwise wholly frustrated. The crown is, however, entitled to the benefit of an act although not named.

Generally speaking the courts will hold that statutes are not retrospective, if it is possible so to construe them. This fundamental rule of English law involves another to the effect that a statute is not to be construed so as to have a greater retrospective operation than its language renders necessary. Where an act is repealed the repeal does not in general affect its previous operation.

The provisions of a public act can be overridden or varied only by express enactment or clear implication. Where the common law is to be changed, and especially the common law which a statutory provision has recognized and enforced, the intention of any new enactment to abrogate it must be plain to exclude a construction by which both may stand together. Similarly, existing rights are not destroyed by a statute unless there are express words or the plainest implication to that effect. The principle applies whether the rights be common law rights, rights of action, rights of ownership or rights arising under a contract. Again, the jurisdiction of the superior courts over matters originally cognizable by them cannot be taken away except by express words or necessary implication. On the other hand, it has been settled for a long time that the creation of a right of appeal is an act requiring legislative authority.

Temporary acts may be made perpetual or may be continued from time to time. The latter is now achieved largely by periodic Expiring Laws Continuance acts. The Army and Air Force acts are instances of temporary acts specifically re-enacted annually.

Enforcement of Statutes.—Crimes created by statute are tried either summarily or by indictment and there are some which are tried by courts, such as courts-martial, having jurisdiction over a limited class of persons. Again, a breach of a statutory command in a matter of public concern is indictable at common law, and in the absence of a special statutory remedy the performance of a public duty may be enforced by indictment or by information filed by the attorney general. An injunction against a continuance of the breach may be granted in appropriate cases. An order of *mandamus* lies for the enforcement of a public duty provided there is no more appropriate remedy.

In some cases a person damnified by a breach of statutory duty may bring proceedings to recover his loss. This holds only where no remedy for the breach by way of penalty or otherwise is prescribed, or where it appears that the duty is imposed for the benefit of particular persons.

The operation of an act may be suspended by a subsequent statute. Many instances of this occurred during World War II by means of emergency legislation. Similarly, an act may be repealed or amended by a later act. Repeal by implication may occur where the provisions of a later enactment are repugnant to the earlier, but such repeal is never to be favoured. An act in its application to England cannot be repealed by non-user (*see* below as to Scotland). There is a presumption that, where special statutory provision has been made for a specific class of cases, a subsequent general enactment is not intended to interfere with such provision. The Interpretation act, 1889, contains many provisions relating to repeal.

Delegated Legislation.—Modern statutes tend increasingly to provide that a particular authority shall have power to make orders, rules and regulations in relation to particular matters arising under the act. Power may also be conferred to make statutory orders in council. These were previously termed statutory rules and orders, but by the provisions of the Statutory Instruments act, 1946, they were named statutory instruments, the procedure in relation to them being governed by that act. Copies of such instruments printed under the authority of her majesty's stationery office are receivable in evidence. Some hold that legislation by statutory instrument has gone too far and that parliament itself should consider substantive legislation even if this requires the establishment of Scots and Welsh and possibly other regional legislatures in order to lighten the burden of the United Kingdom parliament.

Scotland.—Much of what has already been said applies to Scotland although it must be emphasized that Scotland has its own courts whose decisions are often quite different from those of the English courts. A public general statute applies generally to Great Britain and Northern Ireland, which includes Scotland, unless otherwise stated. By the Private Legislation Procedure (Scotland) act, 1936, which applies to Scotland only, where any public authority or persons desire a private act they may apply to the secretary of state for Scotland for a provisional order. In the absence of any opposition the secretary of state for Scotland is directed to introduce into parliament a confirmation bill, and any resulting act is a public act of parliament.

In the 17th and 18th centuries the court of session promulgated acts of *sederunt* amounting to general legislation. In modern times these acts relate exclusively to procedure and are usually passed in virtue of a specific authority in some particular statute. In 1948 an act of *sederunt* approved new consolidated rules and repealed the provisions of any previous act inconsistent with them. An act of adjournal is an act passed by the high court of justiciary for regulating procedure in that court and in inferior criminal courts.

A statute may be impliedly repealed by falling into desuetude. This rule is not applicable except to Scots acts; *i.e.*, those passed before 1707. There is no analogous principal in England, and even in Scotland no statute passed after the union in 1707 has been held by the Scottish courts to have fallen into desuetude.

Scots acts, which are usually expressed briefly and without detailed provisions, have been interpreted in a very liberal spirit. In their construction it is a recognized rule that, whatever the literal meaning of the statutory words, they are to be read as interpreted by decisions pronounced shortly after the act was passed. Apart from this important difference, the rules of interpretation in Scotland are similar to those discussed above.

Northern Ireland.—The parliament of Northern Ireland consists of the queen, the senate of Northern Ireland and the house of commons in Northern Ireland. The Government of Ireland act, 1920, deals with legislation in Northern Ireland and gives power to make laws for the peace, order and good government of Northern Ireland with certain matters reserved to the United Kingdom parliament. Parliament at Westminster does not as a matter of practice legislate for Northern Ireland in matters within the jurisdiction of the Northern Ireland parliament although it has power to do so.

British Commonwealth.—Acts of the imperial parliament do not extend to self-governing commonwealth countries, colonies or dependencies unless these are specially named therein. Moreover, the United Kingdom parliament cannot legislate for the self-governing countries without their consent. Colonies are either self-governing or crown colonies. By the Colonial Laws Validity

act, 1865, the "charter of colonial independence," any colonial law repugnant to the provisions of any act of parliament extending to the colony is void to the extent of such repugnancy, and no colonial law is to be void by repugnancy to the law of England unless it is repugnant to such an act of parliament. For colonies without representative legislatures the crown usually legislates. The relationship between the U.K. parliament and the self-governing countries' legislatures is dealt with by the Statute of Westminster, 1931.

Other Countries.—In most other countries, with the exception of the United States (*see* LEGISLATION), there is a code of law. The assent of two chambers and of the head of the state is generally necessary to legislation.

The term "statute" is used by international jurists to denote the whole body of the municipal law of the state. In this sense, statutes are either real, personal or mixed. A real statute is that part of the law which deals directly with property, whether movable or immovable. A personal statute has for its object a person and deals with questions of status such as marriage, legitimacy and infancy. A mixed statute affects both property and persons and it is said to deal with acts and obligations. Personal statutes are of universal validity; real statutes have no extraterritorial authority. *See* also PARLIAMENT. (P. E. L.)

STAUDINGER, HERMANN (1881–), German chemist, awarded the 1953 Nobel prize for chemistry, for his work on the macromolecule. He was born on March 23, 1881, in Worms, Ger. He was educated in the universities of Halle, Munich and Darmstadt. At Halle he became doctor of philosophy in 1903. He taught in Strasbourg, in Karlsruhe as professor of organic chemistry at the Technische Hochschule and in Zürich as professor of general chemistry (1912–26). In 1926 he was elected to the chair of chemistry in Freiburg im Breisgau, where in 1940 he became director of the chemical laboratory and research institute for macromolecular chemistry. His main contributions to chemistry have been in the important developing field of plastics. He showed that small molecules polymerize by chemical interaction, not simply by physical aggregation, and that linear molecules could be built up synthetically in this way by a variety of processes and maintain their individuality even when subject to chemical modification.

He further showed how it is possible to build up high-polymer networks and that these expanded only to a finite extent when immersed in suitable solvents. He established the relationship between the molecular weight of a high polymer and its viscosity in solution.

See PLASTICS; POLYMERIZATION. (D. MCK.)

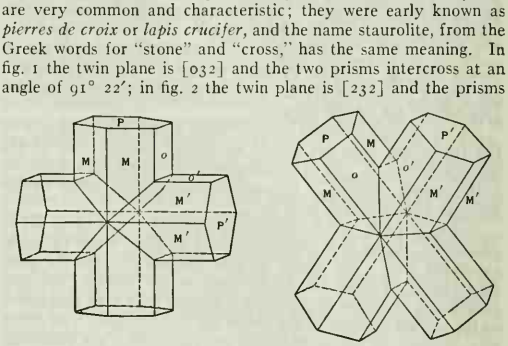
STAUFEN: *see* HOHENSTAUFEN.

STAUNTON, a city of Virginia, U.S., is located 110 mi. N.W. of Richmond, near the southern end of the Shenandoah valley; the seat of Augusta county. This city of narrow streets and steep hills was founded by John Lewis on land granted to William Beverley by King George II in 1736 and was named for Lady Rebecca Staunton, wife of Sir William Gooch, a colonial governor. Settled by Scotch-Irish and English, the town in 1738 became the seat of the newly organized county, which then stretched to the Mississippi river. It was an important junction point for travelers. During the American Revolution the Virginia general assembly met for a brief period in Trinity church there. The American Civil War twice brought Federal occupation.

Staunton is the heart of a general farming area which produces sheep, cattle and apples. The manufacture of flour, furniture, clothing, air-conditioning units and razor blades has brought in new workers from the north and east. Mary Baldwin college, a Presbyterian school for women, was established in 1842. Stuart Hall, an Episcopal preparatory school for girls, was founded in 1843. Staunton Military academy, a private school for boys, was founded in 1860. The Virginia School for the Deaf and Blind and two state hospitals are located in the city. Public schools have kept pace with population growth through new construction and modernization. Woodrow Wilson was born in the manse of the First Presbyterian Church and in 1941 the building was made a national shrine.

Incorporated as a city in 1871, Staunton in 1908 became the first city in the U.S. to adopt the city-manager form of government. For comparative population figures *see* table in VIRGINIA: *Population.* (G. M. BE.)

STAUROLITE, a mineral consisting of basic aluminum and ferrous iron silicate. Crystals are orthorhombic and have the form of six-sided prisms. Interpenetrating cruciform twinned crystals are very common and characteristic; they were early known as *pierres de croix* or *lapis crucifer,* and the name staurolite, from the Greek words for "stone" and "cross," has the same meaning. In fig. 1 the twin plane is [032] and the two prisms intercross at an angle of 91° 22′; in fig. 2 the twin plane is [232] and the prisms



FIGS. 1 AND 2.—TWIN CRYSTALS OF STAUROLITE

intercross at nearly 60°. The mineral is translucent to opaque and dark reddish-brown in colour. Waterworn pebbles of material sufficiently transparent for cutting as gem stones are occasionally found in the diamantiferous sands of Brazil. The formula is $FeAl_4(SiO_4)_2(OH)O$. The hardness is 7.5; specific gravity 3.75. Staurolite is a characteristic mineral of crystalline schists, and it is also a product of contact metamorphism. It is found at various points in Virginia, North Carolina and Georgia embedded in schist or weathered loose in the soil, often as fine twinned crystals that are prized as good luck charms (fairy stones). Lustrous brown crystals occur with kyanite on Monte Campione, Ticino, Switz., and large twinned crystals are found in the mica schists of Brittany. (CL. F.)

STAVANGER, a seaport of Norway, capital of Rogaland *fylke* (county), on the west coast on the Bukken fjord. Pop. (1959 est.) 52,787. The town is one of the oldest in Norway, founded in the 8th or 9th century, but the greater part of the present town is modern and built of stone. It became the seat of a bishopric in the 13th century, and though the see was removed to Christiansand in 1685, it was restored in 1925. The cathedral church of St. Swithun, founded by the English bishop Reinald at the end of the 11th century, and rebuilt after being burned in 1272, is, next to the cathedral of Trondheim, the most interesting stone church in Norway. The old episcopal palace of Kongsgaard is now a Latin school. The fisheries are important—for herring, mackerel, sprats, cod, salmon, lobsters and anchovies. The extraordinary growth of Stavanger was mainly due to its fish-canning industry. Other industries are the making of iodine from seaweed, and shipbuilding.

STAVELEY, a town and urban district in the Chesterfield parliamentary division of Derbyshire, Eng., 29 mi. N.N.E. of Derby by road. Pop. (1961) 18,071. Area 10.2 sq.mi. The main industries are coal mining, the production of iron and chemicals, etc., and agriculture. There are locomotive sheds. The church contains a 12th-century font and monuments to the Frecheville family who lived at Staveley hall, built in 1604 and now the rectory.

STAVELOT, town on the Amblève, province of Liège, Belgium. Pop. (1955 est.) 4,766. There Charles Martel gained a signal victory over Neustria in 719. The prince-abbot of the monastery of Stavelot (established 7th century) exercised secular authority over many towns in the Amblève and Warche valleys, including Malmédy, and had a seat in the old German Diet.

STAVROPOL, chief town of Stavropol *kray* (territory), Russian Soviet Federated Socialist Republic of the U.S.S.R., was

founded in 1777 as a fortress along the Azov-Mozdok defense line. In 1847 it became chief town of Stavropol province (*gubernya*). Its population rose between 1897 and 1956 from 41,600 to an estimated 123,000. After the revolution of 1917, from July 1918 to March 1920, the town was held by the "White" Russian forces under Gen. A. I. Denikin. Under the Soviet rule the city was for a time renamed Voroshilovsk but after World War II recovered its historic name. It was occupied by the Germans in Aug. 1942 and was liberated on Jan. 21, 1943. At the beginning of the 20th century Stavropol was exporting about 500,000 metric tons of grain annually. It is now a centre of agricultural and subsidiary industries. In 1956 it had four institutions of higher education. Stavropol territory lies north of the Caucasus and is bounded by the Rostov, Astrakhan and Grozny regions and by North Ossetia, Kabarda and Georgia and by Krasnodar territory. Its area is 29,575 sq.mi. (incl. Cherkess autonomous region, 1,544 sq.mi.).

STAWELL, a mining, industrial and agricultural town in the shire of the same name in Victoria, Austr., 150 mi. N.W. by rail from Melbourne. Pop. (1954) 5,463. Gold was first mined in the area in 1858, and building stone is also obtained from the Grampian mountains near by. The surrounding district produces merino wool, fruit and tobacco.

STEAD, WILLIAM THOMAS (1849–1912), English journalist, was born at Embleton, Northumberland, on July 5, 1849, the son of a Congregational minister. He was early apprenticed in a merchant's office at Newcastle-on-Tyne; he soon gravitated, however, into journalism, and in 1871 became editor of the Darlington *Northern Echo*. In 1880 he went to London to be assistant editor of the *Pall Mall Gazette* under John Morley, and when the latter retired he became editor (1883–1889). Up to 1885 he had distinguished himself for his handling of public affairs, and his brilliant modernity in the presentation of news. He introduced the "interview," made a feature of the *Pall Mall Gazette* "extras," and his enterprise and originality exercised a potent influence on contemporary journalism and politics.

In 1885 he entered upon a crusade against vice by publishing a series of articles on the "Maiden Tribute of Modern Babylon." Though his action undoubtedly furthered the passing of the Criminal Law Amendment act, it made his position on the paper impossible; and he was imprisoned in Holloway jail for three months on a charge arising out of his crusade. On leaving the *Pall Mall Gazette* he founded the monthly *Review of Reviews* (1890).

He started cheap reprints (*Penny Poets* and *Prose Classics*, etc.), conducted a spiritualistic organ, called *Borderland* (1893–1897), in which he gave full play to his interest in psychical research; and became an enthusiastic supporter of the peace movement, and of many other movements, popular and unpopular. He wrote with facility and sensational fervour on all sorts of subjects, from *The Truth about Russia* (1888) to *If Christ Came to Chicago* (1893), and from *Mrs. Booth* (1900) to *The Americanization of the World* (1902). In private life his keen sense of merit and kindly interest influenced many aspirants to journalism and literature.

Stead went down in the "Titanic" on April 15, 1912.

See *Life of W. T. Stead* (1913), by his daughter; and F. Whyte, *Life of W. T. Stead* (1925).

STEAM. Steam is the term commonly applied to vaporized water. In the pure state steam is an invisible gas, but when interspersed with minute droplets of water, it has the white cloudy appearance familiar to anyone who has seen it issuing from the spout of a teakettle.

The article which follows is divided into the following sections:

I. Steam Engine
  A. Historical
  B. Action of Reciprocating Engines
  C. Types of Reciprocating Engines
II. Properties of Steam
III. Steam Generation
IV. Steam Power Plants

I. STEAM ENGINE

A steam engine is a heat engine in which the working substance is steam. By a heat engine is meant a machine for doing mechanical work through the agency of heat: it does this by taking in heat at comparatively high temperature, converting part of the heat into another form of energy and rejecting the remainder of the heat at a lower temperature. The working substance is the vehicle by which heat is taken in and rejected. In a steam engine, the substance takes in heat mainly in a separate vessel—the boiler—in the process of being vaporized; it does work by expanding under pressure and thereby converts part of the heat that is taken in; finally the remainder of the heat is rejected, either by allowing the steam to escape into the atmosphere, as in locomotives, or by condensing it at a comparatively low temperature and pressure, as in marine and many other engines. This gives a broad distinction between noncondensing and condensing steam engines. The latter have the great advantage that they allow the effective expansion of the steam to be carried much farther; the substance rejects heat at a lower limit of temperature, and this enables the engine to convert into work a larger fraction of the heat which it has received. That fraction expresses what is called the efficiency of the engine as a contrivance for converting heat into work. The addition of a condenser, while it increases the efficiency, of course complicates the mechanism: it requires a supply of cooling water or some equivalent means of keeping down the temperature by absorbing the heat which the steam gives up in the act of being condensed, and also a pump or other means of removing the condensed substance together with any air that may be present. But the advantage which it brings about in respect of efficiency is so great that all engines of large power, where economy of fuel is an important factor, and where the use of a condenser is practicable, are of the condensing kind.

Given the upper limit of temperature, at which heat is taken in, the efficiency which the engine may attain is determined by the lowness of the temperature at which heat is rejected. Similarly, when the lower limit of temperature at which heat is rejected is assigned, the efficiency which the engine may attain is increased by raising the temperature at which the working substance takes in heat. To secure high efficiency there must be a wide range through which the temperature of the working substance falls, as a consequence of expansion within the engine, from the level of temperature at which heat is received to the level at which heat is rejected. Thus in the steam engine the most efficient performance, that is to say the greatest output of work in relation to the heat supplied, is secured by keeping the condenser as cold as the available cooling water will allow, and at the same time using a high boiler pressure, so that the working substance is very hot while it receives heat in the act of changing from water into steam. For this reason, modern practice tends toward higher and higher boiler pressures, as the mechanical difficulties of boiler construction and high pressure working are overcome.

After conversion into steam the working substance may take in a supplementary supply of heat on its way from the boiler to the engine, by passing through a superheater, in which its temperature is raised above that of the boiler. A common form of superheater is a group of parallel pipes with their surfaces exposed to hot gases of the boiler furnace.

Steam engines are classified into two general types according to the manner in which the steam does work during its expansion. In the first, or piston-and-cylinder type, the steam, in a confined space, namely the part of the cylinder behind the piston, enlarges the volume of that space by pushing the piston forward. It does work by exerting a static pressure on the moving piston: the movement of the steam itself is of no significance. In the second class, to which belong all kinds of steam turbines, the action is less direct. The pressure of the steam is first employed to set the steam itself into motion, forming a jet or group of jets, the momentum of which causes work to be done on a moving part of the machine, either by the impulsive action of the jet or jets on revolving vanes, or by the reaction on revolving guide blades during the formation of the jets, or, as in Parsons' turbine, by a combination of impulse and reaction. In any turbine the action of the steam is kinetic, in contrast with its static action in an engine

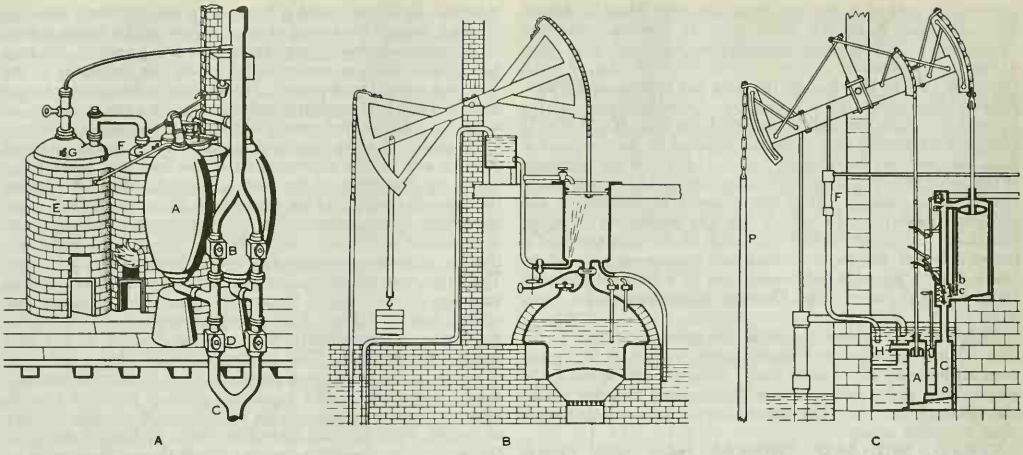


FIG. 1.—EARLY STEAM ENGINES: (A) SAVERY'S ENGINE OF 1698; (B) NEWCOMEN AND CAWLEY ENGINE OF 1705; (C) WATT'S PUMPING ENGINE, 1769

of the piston type. In both types there is progressive expansion of the steam from the high pressure and relatively small volume at which it is admitted, to the low pressure and relatively very great volume at which it is discharged. The principle already stated, that a large range of temperature and pressure, between admission and exhaust, is essential to efficiency applies equally to both types. In practice, the turbine has a notable advantage over the piston-and-cylinder engine in this respect, that it allows the last stages of the expansion, when the volume of the steam has become very great, to be effectively utilized, to a degree which is impracticable in the other type, because of the enormous size which the cylinder would have to assume and the waste of work that would be caused by piston friction, if the steam were expanded down to a very low pressure by the piston method. The turbine method escapes these difficulties; mainly for that reason it has become the most efficient way of converting heat into work, on a large scale, through the agency of steam. It has further advantages in compactness, in simplicity of working, and in the facility with which it can be adapted to take steam of exceptionally high initial pressure and high superheat.

A. Historical

In any historical sketch of the steam engine, stress must be laid on the work of James Watt. But a process of evolution had been going on before his time which prepared a crude and primitive device for the immense improvements it received at his hands. The labours of Watt stand in natural sequence to those of Newcomen, and Newcomen's to those of Papin and Savery. Savery's engine, in its turn, was the reduction to practical form of an ancient scientific toy. Along another line of development, the modern steam turbine, which we owe to the genius of Charles Parsons, can be traced back to an early prototype.

In the *Pneumatica* of Hero of Alexandria (*c.* 130 B.C.) there is described the aeolipile, which may be called a primitive steam reaction turbine. It consists of a hollow globe pivoted so that it can turn on a pair of central trunnions, and supplied with steam through one of them, which is hollow. The steam escapes from the globe to the outside air through two bent pipes facing tangentially in opposite directions at the ends of a diameter perpendicular to the axis. The globe revolves by reaction from the escaping steam. Hero's volume also mentions another device which may be described as the prototype of the pressure engine. A hollow altar containing air is heated by kindling a fire on it; the air expands and by its pressure drives some of the water in a vessel below into a hanging bucket, which then descends, opening the doors of a shrine. When the fire is extinguished the air contracts, the bucket empties, and the doors close.

In a treatise on pneumatics (1601) by Giambattista della Porta there is shown a somewhat similar apparatus, but with steam for working substance. Its pressure forces up water from a separate vessel. He also points out that the condensation of the steam may be used to produce a vacuum and thereby suck up water from a lower vessel. His suggestions go far to anticipate the engine which, a century later, in the hands of Savery, became the first commercially successful steam engine.

Meanwhile Edward Somerset, 2nd marquis of Worcester, described in his *Century of Inventions* (1663) a method of raising water by the agency of steam. His description is obscure, and no drawings of the device are extant. It appears to have consisted of a pair of displacement chambers, from each of which alternately water was forced, probably by admitting steam from an independent boiler, while the other vessel was allowed to refill.

Savery, 1698.—The earliest steam engine to take a practical form and find employment in industry was that of Thomas Savery, who, in 1698, obtained a patent for a water-raising engine, shown in fig. 1(A). Steam is admitted to one of the oval vessels A, displacing water, which it drives up through the check valve B. When the vessel A is emptied of water the supply of steam is stopped, and the steam already there is condensed by allowing a jet of cold water from a cistern above to stream over the outer surface of the vessel. This produces a vacuum and causes water to be sucked up through the pipe C and the valve D. Meanwhile steam has been displacing water from the other vessel, and is ready to be condensed there. The valves B and D open only upward. The supplementary boiler and furnace E are for feeding water to the main boiler; E is filled while cold and a fire is lighted under it; it then forces a supply of feeding water into the main boiler F. The gauge cocks G, G are an interesting detail. Another form of Savery's engine had only one displacement chamber and worked intermittently. In the use of artificial means to condense the steam, and in the application of the vacuum so formed to raise water by suction from a level lower than that of the engine, as well as in practical features, Savery's engine marked an inventive advance. It found considerable employment in pumping mines and in raising water to supply houses and towns, and even to drive water wheels. A serious difficulty which prevented its general use in mines was the fact that the height through which it would lift water was limited by the pressure the boiler and vessels could bear. Pressures as high as eight or ten atmospheres were employed—and that, too, without a safety valve—but Savery found it no easy matter to deal with high-pressure steam. Apart from this drawback, the waste of fuel was enormous, from the condensation of steam which took place on the surface of the water and on the sides of the displacement chamber at

each operation; the consumption of coal was, in proportion to the work done, about 20 times greater than in a modern engine.

Before Savery's engine was displaced by its successor, Newcomen's, it was improved by J. T. Desaguliers who applied to the safety valve (invented by Papin), and substituted condensation by a jet of cold water within the vessel for the surface condensation used by Savery. To Savery is ascribed the first use of the term horsepower as a measure of performance.

In 1690 Dènis Papin suggested that the condensation of steam should be employed to make a vacuum under a piston previously raised by the expansion of the steam. Papin's was the earliest cylinder-and-piston steam engine, and his plan of using steam was that which afterward took practical shape in the atmospheric engine of Newcomen. But his scheme was made unworkable by the fact that he proposed to use but one vessel as both boiler and cylinder. A small quantity of water was placed at the bottom of a cylinder and heat was applied. When the piston had risen the fire was removed, the steam was allowed to cool, and the piston did work in its downstroke under the pressure of the atmosphere. After hearing of Savery's engine in 1705 Papin turned his attention to improving it, and devised a modified form, with a floating diaphragm or piston on the top of the water to keep the water and steam from direct contact with one another. Papin's engine may be described as a noncondensing single-acting steam pump, with steam cylinder and pump cylinder in one.

Newcomen's Atmospheric Engine, 1705.—While Papin was thus going back from his first notion of a piston engine to Savery's cruder type, a new inventor had appeared who made the piston engine a practical success by separating the boiler from the cylinder and by using (as Savery had done) artificial means to condense the steam. This was Thomas Newcomen, who in 1705, with his assistant, John Cawley, gave the steam engine the form shown in fig. 1(B). Steam admitted from the boiler to the cylinder by the opening of a valve allowed the piston to be raised by a heavy counterpoise on the other side of the beam. Then the steam valve was shut and a jet of cold water entered the cylinder and condensed the steam. The piston was consequently forced down by the pressure of the atmosphere and did work on the pump. The next entry of steam expelled the condensed water from the cylinder through an escape valve. The piston was kept tight by a layer of water on its upper surface. Condensation was at first effected by cooling the outside of the cylinder, but an accidental leakage of the packing water past the piston showed the advantage of condensing by a jet of injection water, and this plan took the place of surface condensation. The engine used steam whose pressure was little if at all greater than that of the atmosphere.

About 1711 Newcomen's engine began to be introduced for pumping mines. It is doubtful whether the action was originally automatic, or depended on the periodical turning of taps by an attendant. The common story is that in 1713 a boy named Humphrey Potter, whose duty it was to open and shut the valves of an engine he attended, made the engine self-acting by causing the beam itself to open and close the valves by suitable cords and catches. This device was simplified in 1718 by Henry Beighton, who suspended from the beam a rod called the plug tree, which worked the valves by means of tappets. By 1725 the engine was in common use in collieries, and it held its place without material change for about three-quarters of a century in all. Near the close of its career Newcomen's engine was much improved in its mechanical details by John Smeaton, who built many large engines of this type about the year 1770, just after the great step which was to make Newcomen's engine obsolete had been taken by James Watt.

Compared with Savery's engine, Newcomen's had (as a pumping engine) the great advantage that the intensity of pressure in the pumps was not in any way limited by the pressure of the steam. It shared with Savery's, in a scarcely less degree, the defect already pointed out, that steam was wasted by the alternate heating and cooling of the vessel into which it was led. Though obviously capable of more extended uses, it was in fact almost exclusively employed to raise water—in some instances for the purpose of turning water wheels to drive other machinery. Even contemporary writers complain of its great consumption of fuel.

James Watt, 1763, 1769, 1781, etc.—In 1763 James Watt, an instrument maker in Glasgow, while engaged by the university in repairing a model of Newcomen's engine, was struck with the waste of steam to which the alternate chilling and heating of the cylinder gave rise. He saw that the remedy, in his own words, would lie in keeping the cylinder as hot as the steam that entered it. With this view he added to the engine a new organ—namely the "separate condenser"—an empty vessel separate from the cylinder, into which the steam should be allowed to escape from the cylinder, to be condensed there by the application of cold water either outside or as a jet. To preserve the vacuum in his condenser he added a pump called the air pump, whose function was to extract from it the condensed steam and water of condensation, as well as any air which might come in by leakage or by solution in the steam or the injection water. Then, as the cylinder was no longer used as a condenser, he was able to keep it hot by clothing it with nonconducting bodies, and in particular by the use of what is called a steam jacket—a layer of hot steam between the cylinder and an external casing. Further and still with the same object, he covered in the top of the cylinder, taking the piston rod out through a steamtight gland or stuffing box, and allowed steam instead of air to press upon the piston's upper surface. After much experiment Watt patented his improvements in 1769; they are described in a specification from which the following extracts are taken:—

> My method of lessening the consumption of steam, and consequently fuel, in fire-engines, consists of the following principles:—
>
> *First*, That vessel in which the powers of steam are to be employed to work the engine, which is called the cylinder in common fire-engines, and which I call the steam-vessel, must, during the whole time the engine is at work, be kept as hot as the steam that enters it; first by enclosing it in a case of wood, or any other materials that transmit heat slowly; secondly, by surrounding it with steam or other heated bodies; and, thirdly, by suffering neither water nor any other substance colder than the steam to enter or touch it during that time.
>
> *Secondly*, In engines that are to be worked wholly or partially by condensation of steam, the steam is to be condensed in vessels distinct from the steam-vessels or cylinders, although occasionally communicating with them; these vessels I call condensers; and, whilst the engines are working, these condensers ought at least to be kept as cold as the air in the neighbourhood of the engines, by application of water or other cold bodies.
>
> *Thirdly*, Whatever air or other elastic vapour is not condensed by the cold of the condenser, and may impede the working of the engine, is to be drawn out of the steam-vessels or condensers by means of pumps, wrought by the engines themselves, or otherwise.
>
> *Fourthly*, I intend in many cases to employ the expansive force of steam to press on the pistons, or whatever may be used instead of them, in the same manner in which the pressure of the atmosphere is now employed in common fire-engines. In cases where cold water cannot be had in plenty, the engines may be wrought by this force of steam only, by discharging the steam into the air after it has done its office.

The common fire engine alluded to was the steam engine of Newcomen. Highly important as Watt's first inventions were, they resulted for a time in the production of nothing more than a greatly improved engine of the Newcomen type, much less wasteful of fuel, able to make faster strokes, but still only suitable for pumping, still single-acting, with steam admitted during the whole stroke, the piston, as before, pulling the beam by a chain working on a circular arc. The condenser was generally worked by injection, but Watt has left a model of a surface condenser made up of small tubes, in every essential respect like the condensers now used.

Fig. 1(C) is an example of the Watt pumping engine of this period. It should be noticed that, although the top of the cylinder is closed, and steam has access to the upper side of the piston, this is done only to keep the cylinder and piston warm. The engine is still single-acting; the steam on the upper side merely plays the part which was played in Newcomen's engine by the atmosphere; and it is the lower end of the cylinder alone that is ever put in communication with the condenser. There are three valves: the steam valve a, the equilibrium valve b, and the exhaust valve c. At the beginning of the downstroke c is opened to produce a

vacuum below the piston and a is opened to admit steam above it. At the end of the downstroke a and c are shut and b is opened. This puts the two sides in equilibrium and allows the piston to be pulled up by the pump rod P, which is heavy enough to serve as a counterpoise. C is the condenser, and A is the air pump, which discharges into the hot well H, whence the supply of the feed pump F is drawn.

In a second patent (1781) Watt describes the sun-and-planet wheels and other methods of making the engine give continuous revolving motion to a shaft provided with a flywheel. The crank and connecting rod—already a familiar mechanical device from its use on the treadle of a lathe—would have been the natural means of doing this, but its application to the steam engine in a particular manner had been made the subject of a patent by James Pickard, and Watt, rather than make terms with Pickard, whom he regarded as a plagiarist of his own ideas, made use of his sun-and-planet motion until the patent on the crank expired. The reciprocating motion of earlier forms had served only for pumping; by making the steam engine drive a revolving shaft Watt opened up for it many other channels of usefulness. The engine was still single-acting; the connecting rod was attached to the far end of the beam, and that carried a counterpoise which served to raise the piston when steam was admitted below it.

In 1782 Watt patented two further improvements of the first importance, both of which he had invented some years before. One was the use of double action, that is to say, the application of steam and vacuum to each side of the piston alternately. The other (invented as early as 1769) was the use of steam expansively, in other words the plan, essential to economy of fuel, of stopping the admission of steam when the piston had made only a part of its stroke, and allowing the rest of the stroke to be performed by the expansion of the steam already in the cylinder. To let the piston push as well as pull the end of the beam Watt devised his so-called parallel motion, an arrangement of links connecting the piston-rod head with the beam in such a way as to guide the rod to move in a very nearly straight line. He further added a throttle valve, for regulating the rate of admission of steam, and a centrifugal governor, in the form of a double conical pendulum, which controlled the speed by a throttle valve.

Among other important devices associated with Watt was the indicator, by which diagrams showing the relation of the steam pressure in the cylinder to the movement of the piston are automatically drawn: its invention seems to have been mainly due to his assistant John Southern.

In partnership with Matthew Boulton, Watt carried on in Birmingham the manufacture and sale of his engines with great success, and held the field against all rivals in spite of severe assaults on the validity of his patents. Notwithstanding his knowledge of the advantage to be gained by using steam expansively, he continued to employ only low pressures—seldom more than 7 lb. per square inch over that of the atmosphere. His boilers were fed, as Newcomen's had been, through an open pipe which rose high enough to let the column of water in it balance the pressure of the steam. He gave a definite numerical significance to the term horsepower, defining it as the rate at which work is done when 33,000 lb. are raised one foot in one minute.

In the fourth claim of Watt's first patent, quoted above, the second sentence describes a noncondensing engine, which would have required steam of a higher pressure. This, however, was a line of invention which Watt did not follow up, perhaps because so early as 1725 a noncondensing engine had been described by Jacob Leupold in his *Theatrum machinarum*.

It was not until much later that the thermodynamic principles underlying the action of the steam engine came to be understood. Engineers were consequently slow to appreciate the fact that to obtain economy of fuel it was advantageous to employ a high initial pressure, in combination with much expansion in the cylinder and with the separate condenser of Watt.

Trevithick, Bull and Evans.—The introduction of the noncondensing and, at that time, relatively high-pressure engine was effected in England by Richard Trevithick and in America by Oliver Evans about 1800. Both Evans and Trevithick applied their engines to propel carriages on roads, and both used for boiler a cylindrical vessel with a cylindrical flue inside containing the fire—the construction now known as the Cornish boiler. In association with Edward Bull, Trevithick had previously made direct-acting pumping engines, with an inverted cylinder set over and in line with the pump rod, thus dispensing with the beam that had been a feature in all earlier forms. But in these Bull engines, as they were called, the steam was condensed by a jet of cold water in the exhaust pipe, and Boulton and Watt successfully opposed them as infringing Watt's patent. To Trevithick belongs the honour of being the first to use a steam carriage on a railway; in 1804 he built a locomotive in the modern sense, to run on what had formerly been a horsecar route, in Wales. In this connection it may be added that as early as 1769 a steam carriage for roads had been built in France by Nicolas Joseph Cugnot, who used a pair of single-acting high-pressure cylinders to turn a driving axle step by step by means of pawls and ratchet wheels. To the initiative of Evans may be ascribed the early general use of high-pressure steam in the United States, a feature which for many years distinguished American from English practice. (*See* RAILWAY: *Locomotives.*)

Compound Engine and Cornish Engine.—Among contemporaries of Watt the name of Jonathan Hornblower deserves special mention. In 1781 he constructed and patented what would now be called a compound engine, with two cylinders of different sizes. Steam was first admitted into the smaller cylinder, and then passed over into the larger, doing work against a piston in each. In Hornblower's engine the two cylinders were placed side by side, and both pistons worked on the same end of a beam overhead. This was an instance of the use of steam expansively, and as such was earlier than the patent, though not earlier than the invention, of expansive working by Watt. Hornblower was crushed by the Birmingham firm for infringing their patent in the use of a separate condenser and air pump. The compound engine was revived in 1804 by Arthur Woolf, with whose name it is often associated. Using steam of fairly high pressure, and cutting off the supply before the end of the stroke in the small cylinder, Woolf expanded the steam to several times its original volume. Mechanically the two-cylinder compound engine has some advantage over a one-cylinder engine with the same amount of expansion, in exerting a more uniform driving effort. But another and more important merit of the system lies in the fact that by dividing the whole range of expansion into two parts the cylinders in which these are separately performed are subject to a reduced range of fluctuation in their temperature. This helps to limit a source of waste which is present in all piston engines, namely the waste which results from the heating and cooling of the metal by its alternate contact with hot and cooler steam. The introduction of compound expansion forms the most outstanding improvement which steam engines of the piston and cylinder type have undergone since the time of Watt.

Woolf introduced the compound engine somewhat widely about 1814 as a pumping engine in the mines of Cornwall. But here it met a strong competitor in the high-pressure single-cylinder engine of Trevithick, which had the advantage of greater simplicity in construction. Woolf's engine fell into comparative disuse, and the single-cylinder type took a form which, under the name of the Cornish pumping engine, was for many years famous for its great economy of fuel. In this engine the cylinder was set under one end of a beam, from the other end of which hung a heavy rod which operated a pump at the foot of the shaft. Steam was admitted above the piston for a short portion of the stroke, thereby raising the pump rod, and was allowed to expand for the remainder. Then an equilibrium valve, connecting the space above and below the piston, as in fig. 1(C), was opened, and the pump rod descended, doing work in the pump and raising the engine piston. The large mass which had to be started and stopped at each stroke served by its inertia to counterbalance the unequal pressure of the steam, for the ascending rods stored up energy of motion in the early part of the stroke, when the steam pressure was greatest, and gave out energy in the later part, when the pressure was much lowered by expansion. The frequency of the stroke was

controlled by a device called a cataract, consisting of a small plunger pump, in which the plunger, raised at each stroke by the engine, was allowed to descend more or less slowly by the escape of fluid below it through an adjustable orifice, and in its descent liberated catches which held the steam and exhaust valves from opening. A similar device controlled the equilibrium valve, and could be set to give a pause at the end of the piston's downstroke, so that the pump cylinder might have time to become filled.

The final revival of the compound engine did not occur until about the middle of the 19th century, and then several agencies combined to effect it. In 1845 John M'Naught introduced a plan of improving beam engines of the original Watt type, by adding a high-pressure cylinder whose piston acted on the beam between the centre and the flywheel end. Steam of higher pressure than had formerly been used, after doing work in the new cylinder, passed into the old or low-pressure cylinder, where it was further expanded. Many engines whose power was proving insufficient for the extended machinery they had to drive were "M'Naughted" in this way, and after conversion were found not only to yield more power but to show a marked economy of fuel. The compound form was selected by William Pole for the pumping engines of Lambeth and other waterworks about 1850; in 1854 John Elder began to use it in marine engines; in 1857 E. A. Cowper added a steam-jacketed intermediate reservoir for steam between the high- and low-pressure cylinders, which made it unnecessary for the stroke of the low-pressure piston to be just beginning when that of the other piston was just ending. As facilities increased for the use of high-pressure steam, compound expansion came into more general use, its advantage becoming more conspicuous with every increase in boiler pressure. In marine practice, where economy of fuel was from the first an obviously important factor in design, the principle of compound expansion was extended by the introduction of triple- and even quadruple-expansion engines.

Application to Locomotives and Steamboats.—The adaptation of the steam engine to railways, begun by Trevithick, became a success in the hands of George Stephenson, whose engine, the "Rocket," when tried along with others in 1829, distanced its competitors. The principal features of the "Rocket" were an improved steam blast for urging the combustion of coal and a boiler (suggested by Henry Booth) in which a large heating surface was given by the use of many small tubes through which the hot gases passed. Further, the cylinders, instead of being vertical as in earlier locomotives, were set at a slope, which was afterward altered to a position more nearly horizontal. To these features there was added later the link motion, a contrivance which enabled the engine to be easily reversed, and the amount of expansion to be readily varied. In the hands of George Stephenson and his son Robert the locomotive took a form which in the main was retained by the far heavier locomotives later in use.

The first practical steamboat was the tug "Charlotte Dundas," built by William Symington, and tried in the Forth and Clyde canal in 1802. A Watt double-acting condensing engine, placed horizontally, acted directly by a connecting rod on the crank of a shaft at the stern, which carried a revolving paddle wheel. The trial was successful, but steam towing was abandoned for fear of injuring the banks of the canal. Ten years later Henry Bell built the "Comet," with side paddle wheels, which ran as a passenger steamer on the Clyde; but an earlier inventor to follow up Symington's success was the American, Robert Fulton, who, after unsuccessful experiments on the Seine, fitted a steamer on the Hudson in 1807 with engines made to his designs by Boulton and Watt, and brought steam navigation for the first time to commercial success.

Rise in Steam Pressure and in Piston Speed.—With improvements in the details of design and construction it gradually became practicable to use higher steam pressures and higher piston speeds, and consequently to obtain not only greater efficiency, but also a greater amount of power from engines of given bulk. The triple-expansion engine, introduced by A. C. Kirk in 1874, did not come into general use until after 1881. It became the normal type of marine engine, with pressures ranging, as a rule, from 150 to 200 lb. per square inch, piston speeds generally of 500 or 600 ft. per minute, but sometimes as high as 900 or 1,000, and coal consumption of about 1½ lb. per hour per indicated horsepower. It continues to be very largely used in steamships which are not driven by turbines. In some instances quadruple expansion has been preferred, with somewhat higher pressures, but when the pressure is much raised the tendency is to abandon the piston type in favour of the steam turbine. This is true both in marine and in land practice. The gigantic concentration of steam power that is found in a great steamship or in a power station has been made practicable by the turbine. The selection of Parsons turbines in 1907 as the motive engines of the Cunarders "Lusitania" and "Mauretania" constituted a new departure in steam engineering. It was then a novelty to develop some 70,000 horsepower in the engine room of a single ship. Many of the turbines in power stations now exceed this figure.

Introduction of the Steam Turbine.—The invention of the steam turbine revolutionized marine engine practice, in respect especially of the largest and fastest vessels. For the generation of electricity the turbine has a notable advantage in directly developing the high speed of rotation which a dynamo requires, and, when designed on a large scale, its efficiency is unrivaled by steam engines of other kinds. To Sir Charles Parsons we owe not only the main idea of the modern steam turbine, but also the invention of many mechanical features and details essential to its practical success and general adaptation.

In the steam turbine, as we have seen, pressure, instead of being exerted on a piston, is employed in the first instance to set the fluid itself in motion. There is a conversion of pressure energy into velocity energy as a preliminary step toward obtaining the effective work of the machine. If this were done in a single step it would involve immensely high velocities in the steam jet and in the vanes on which the jet acts. Attempts to design a steam turbine were made by numerous inventors, but fell short of practical success mainly because of the difficulty of arranging for high enough velocity in the working parts to utilize a reasonably large fraction of the kinetic energy of the steam. There was a further difficulty in getting the energy of the steam into a suitable kinetic form, namely, to get the stream of issuing particles to take a single direction, without undue dispersion, when steam was allowed to expand through an orifice from a chamber at high pressure into a space where the pressure was greatly less.

In 1889 Gustaf de Laval introduced a form of steam turbine in which both of these difficulties were, to a considerable degree, overcome, partly by the special form of the nozzle used to produce the steam jet and partly by features of design which allowed an exceptionally high speed to be reached in the wheel carrying the vanes against which the steam impinged.

Parsons attacked the problem at an earlier date and in a different way, by his invention of the compound turbine. He divided the whole expansion of the steam into a great number of successive and separate steps and thereby limited the velocity acquired at each step to such an extent as to make it comparatively easy to extract the greater part of the kinetic energy as work done upon the moving blades, without making the velocity of these blades inconveniently high. Moreover, in Parsons's compound turbine the range of pressure through which the steam expands in each separate step is too small to cause any difficulty in the formation of the jets. The guide blades, which form the jets, are distributed around the circumference of the revolving wheel, and all the revolving blades are consequently in action at once. The steam streams from end to end of the turbine through an annular space between a revolving drum and the casing which surrounds it. Parallel rings of fixed guide blades project inward from the casing at suitable distances, and between these are rings of moving blades which project outward from the drum and revolve with it. At each step in the expansion the steam streams through a ring of fixed guide blades, and the streams so formed impinge on the next ring of moving blades, and so on. The construction, which is of great simplicity, is described, along with others, in the article TURBINE: STEAM. It lends itself well to the generation of power on a large scale, especially where a fairly high speed of rotation is wanted.

Parsons introduced his compound steam turbine in 1884. For some years it was made in small sizes only, and the steam was discharged to the atmosphere without condensation. So long, however, as this was done the steam turbine was sacrificing one of its most important advantages, namely, its exceptional capacity for utilizing the energy of low-pressure steam down to the lowest vacuum obtainable in a condenser. In 1891 it was first fitted with a condenser, and it then began to be used in electric supply stations. The first application to marine propulsion was in the "Turbinia," in 1897. The success of this little experimental vessel of 100 tons, which, with its horsepower of 2,100, made what was then a record in speed for any ship, was soon followed by the application of the turbine to warships and other steamers. In merchant vessels its use was at first limited to those of the highest speed, for the turbine shaft was directly coupled to the shaft of the screw propeller; but in 1910 Parsons introduced a mechanical reducing gear between the two, which allowed the turbine shaft to run much faster than the propeller shaft, to the great advantage in efficiency of both turbine and propeller. Later he followed this up by a double-reduction gearing which admitted of a still greater difference in speed of rotation between the propeller and the turbine. In most ships the single-reduction system is sufficient for the purpose: its introduction greatly extended the range within which the turbine could be advantageously substituted for the three-cylinder or four-cylinder compound engine of the piston type in ocean-going steamships.

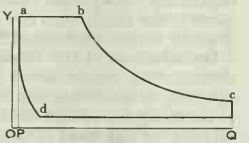


FIG. 2.—INDICATOR DIAGRAM SHOWING DISTRIBUTION OF STEAM (SEE TEXT FOR FURTHER INFORMATION)

Enough has been said to show that the invention of the steam turbine is by far the most important step in steam engineering since the time of Watt. It solved the problem of using steam efficiently in an engine without reciprocating parts. (For a more complete discussion of the steam turbine *see* TURBINE: STEAM.)

Early Theory.—In the early development of the steam engine inventors had little in the way of theory to guide them. Watt had the advantage of a knowledge of Joseph Black's doctrine of latent heat; but there was no philosophy of the relation of work to heat until long after the inventions of Watt were complete. The theory of the steam engine as a heat engine may be said to date from 1824, when N. L. Sadi Carnot published his *Réflexions sur la puissance motrice du feu*, a remarkable essay in which he showed that heat does work only by being let down from a higher to a lower temperature. But Carnot was not then aware that any of the heat disappears in the process, and it was not until the doctrine of the conservation of energy was established in 1843 by the experiments of James Prescott Joule that the theory of heat engines began a vigorous growth. From 1849 onward the science of thermodynamics was developed with extraordinary rapidity by Rudolf Clausius, W. J. Rankine and William Thomson (Lord Kelvin) and was applied, especially by Rankine, to practical problems in the use of steam. Rankine's *Manual of the Steam Engine*, published in 1859, was the first attempt at a systematic treatment of steam-engine theory. It involved the simplifying assumption that the cylinder and piston might be treated as behaving to the steam like nonconducting bodies, in other words, that the transfer of heat between the steam and the metal might generally be disregarded. One effect of this was to treat the volume of steam consumed per stroke as corresponding to the volume of the cylinder up to the point of cutoff. When steam enters the engine cylinder it finds the metal chilled by the previous exhaust, and a portion of it is at once condensed. This has the effect of increasing, often very largely, the volume of boiler steam required per stroke.

As expansion goes on, the water that was condensed during admission begins to be re-evaporated, and this action is often prolonged into the exhaust. It is now recognized that exchanges of heat between the steam and its metal envelope cannot be ignored. They cause the actual performance to fall short, in some cases very much short, of the ideal limit. They may be reduced by suitable design.

B. Action of Reciprocating Engines

Most, though not all, reciprocating engines are double-acting: that is to say, steam from the boiler is alternately admitted to each side of the piston. In each double stroke, or revolution, there are four events for each end of the cylinder:—(1) admission, which begins by the opening of a steam valve when the piston is at or very near the limit of its travel; (2) cutoff, at which the steam valve is closed and admission ceases. This may take place early in the stroke. The steam which is enclosed behind the piston then expands, with falling pressure, while the stroke continues, until (3) release occurs; that is to say, an exhaust valve opens, allowing the steam to escape from the cylinder. Its discharge generally continues for a large part of the backstroke, until event (4), at which the exhaust valve closes, and compression begins: from there to the end of the backstroke, the steam remaining in the cylinder is compressed into the clearance space behind the piston. This compression of the residue of steam, which is called cushioning, assists smoothness of working as the piston passes what is called the dead point, at the limit of its stroke; and the cushioning effect is often augmented by giving the steam valve what is called lead,—that is, causing it to open a little before the piston reaches the dead point.

Distribution of the Steam: Indicator Diagram.—Together, these various events of the stroke constitute what is termed the distribution of the steam. They are conveniently exhibited by drawing an indicator diagram of the action, such as that illustrated in fig. 2. There, on a base PQ which represents the stroke, lines are drawn to show the continuous changes of pressure that go on within the cylinder during the whole action. Starting from the beginning of the stroke at a, for one side of the piston, steam is admitted up to the point b which shows the cutoff. At (or near) the end of the forward stroke, at c, the steam is allowed to escape. Compression, in the return stroke, begins at d. From b to c the steam confined in the cylinder is expanding, with falling pressure, doing work on the piston. The whole work done in the revolution, namely the integral of the force acting on the piston and the distance through which it moves, is represented by the enclosed area of the indicator diagram, a b c d.

The diagram is idealized in the sense that the events are shown as if they happened suddenly, giving sharply defined changes from

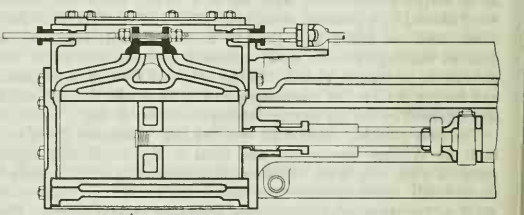


FIG. 3.—CROSS SECTION OF A SLIDE VALVE SHOWING THE EXHAUST AND INLET PORTS BELOW IT, AND CYLINDER AND PISTON

admission to expansion and so on. In any real engine the events are necessarily gradual, for no valve—whatever its type—can close or open quite instantaneously. A real indicator diagram accordingly has a smoother outline with rounded corners to mark the places of cutoff, release, compression and admission. In practice release always occurs before the piston has quite completed its stroke.

The distance PQ which represents the stroke may (by the use of an appropriate scale) be interpreted as a volume, namely the volume swept through by the piston, and a point O behind P may be so taken that OP represents, on the same scale, the volume of the clearance. The distance measured horizontally from the line OY to any point of the curve represents the whole volume of steam behind the piston at that point, and the curve between admission and release exhibits the relation of the pressure to the total volume of the steam enclosed behind the piston throughout

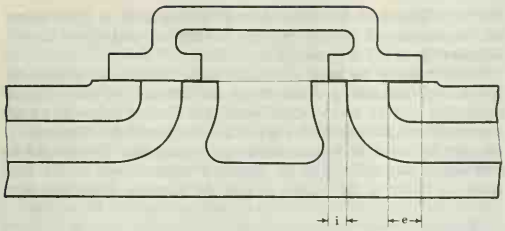


FIG. 4.—CROSS SECTION OF SLIDE VALVE IN MIDDLE POSITION: (i) INTERNAL LAP. (e) EXTERNAL LAP

its process of expansion.

Actual indicator diagrams may be automatically drawn by attaching to the cylinder of the engine a device in which a pencil is made to rise or fall proportionally to the variations of steam pressure, while a paper, on which it inscribes a record, moves back and forth horizontally through distances proportional to the movement of the engine piston. Many forms of indicator have been designed to give, in some such way, a diagram in which the coordinates represent respectively the displacement of the piston and the pressure of the steam against it. From such diagrams it is easy to infer the mean effective pressure throughout the stroke, and also to observe how the valves are working. By connecting the instrument successively to each of the two ends of the cylinder, in a double-acting engine, a complete record is obtained for the two strokes which make up a revolution. The indicated horsepower is determined by the formula

$$\text{i.h.p.} = \frac{nL(p_1A_1 + p_2A_2)}{33{,}000}$$

where A_1 and A_2 are the areas of the two sides of the piston, in square inches; p_1 and p_2 are the mean effective pressures on the two sides, in pounds per square inch, as determined from the diagrams; n is the number of revolutions per minute, and L is the stroke in feet. In trials of a steam engine such diagrams are taken during a considerable period. The amount and condition of the steam passing through the engine is observed, by measuring either the quantity of water discharged from the condenser (if of the surface type), or the quantity of feed water required to be supplied to the boiler to keep the water level constant. From these data it becomes possible to compare the power developed with the heat supplied. Measurements of the heat rejected may also be made by observing the quantity of condensing water used and its rise in temperature; and together these figures give material for drawing up a balance sheet of the disposal of thermal energy.

The Slide Valve.—The admission and exhaust of steam, for each end of the cylinder, may be controlled by separate valves, but very commonly a single moving piece, called a slide valve, serves to control the events for both ends. The slide valve was invented by William Murdock, an assistant of Watt. A common form of it is illustrated in section in fig. 3, which also shows the cylinder, with the piston, and the steam ports and passages leading to each end of the cylinder. The face on which the valve slides is a plane surface on one side of the cylinder with three ports or openings which extend across the greater part of the cylinder's width. The central opening is the exhaust port, through which the steam escapes after doing its duty in the cylinder. The others, which are narrower, lead to the two ends. The valve itself is a box-shaped cover, sliding on the face, and contained in the valve chest, or chamber, to which steam is admitted from the boiler. When the valve moves a sufficient distance to either side of the middle position steam is admitted to one end of the cylinder, past the outer edge. Similarly steam escapes from the other end of the cylinder, through the cavity of the valve, to the exhaust port. The valve takes its motion from an eccentric on the engine shaft, which is set more than 90° ahead of the crank, so that the valve has already begun to uncover the port on one side when the piston is at the corresponding dead point, with the result that steam passes from the steam chest above the valve into the space behind the piston. Fig. 4 shows a slide valve in its middle position, and illustrates the internal lap i and the external lap e which the valve must have to make expansive working possible. At the beginning of the stroke it has already passed its middle position by an amount at least equal to e, so that steam may enter. It continues to move farther, opening the port more widely, and then begins to return, while the piston is still advancing in the cylinder. At a particular point in its return the outer edge of the valve closes the port; this determines the instant of cutoff. The piston continues to advance, with expansion of the steam confined in the cylinder, while the valve continues to move back, until the inner edge of the valve begins to uncover the port, allowing steam from the cylinder to escape to the exhaust channel. This point, which depends on the internal lap i, determines the instant of release. The port then opens more widely to exhaust, and remains open during most of the backstroke of the piston, until the valve, again moving in the original direction, brings its inner edge again over the port and compression begins. Similar events for the other end of the cylinder are determined by the other side of the valve. The positions of the piston when the several events occur depend on the amounts of the laps e and i, and of the angular advance which is the excess over 90° in the angle by which the eccentric stands ahead of the crank. If there were no laps and no angular advance, steam would be admitted during the whole forward stroke and exhausted during the whole backward stroke. By giving the valve laps and angular advance expansive working becomes possible. But the slide valve is not well adapted to effect a cutoff early in the stroke, and when much expansion is desired, a device called a separate expansion valve is added, or valves of a different type are used to control the distribution of the steam.

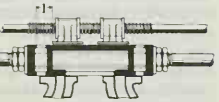


FIG. 5.—SEPARATE EXPANSION VALVE

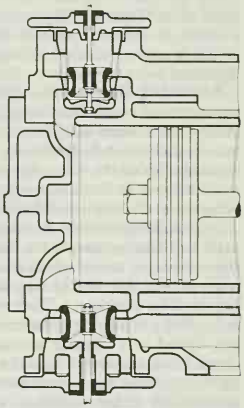


FIG. 6.—CROSS SECTION OF ONE END OF A CYLINDER SHOWING DROP VALVES

Separate Expansion Valves.—Fig. 5 shows Meyer's expansion valve, which consists of two blocks sliding on the back of a slide valve. It is caused to reciprocate by an eccentric, generally set directly opposite the crank, so that it moves to the left while the piston moves to the right, and vice versa. The admission of steam ceases when the relative displacement of the expansion blocks over the slide valve amounts to the distance l, and this can be arranged to happen early in the piston's stroke without affecting the release or the compression.

In this example the rod which carries the two blocks is fitted with right- and left-handed screws, so that by turning the rod the distance between the blocks may be adjusted, and the distance l consequently altered. This allows the valve to be set to give an earlier or later cutoff: when l is increased by bringing the blocks nearer together, the cutoff comes later in the stroke. The adjustment may be made while the engine is running. Another way to change the cutoff with a valve of the Meyer type is to alter its travel. Increasing the travel of the expansion valve, while l remains constant, makes the cutoff come later.

In many engines the events are controlled by using, at each end of the cylinder, separate valves for admission and exhaust (fig. 6). Sometimes these are drop valves which are double-beat in the sense that the valve when it drops into its seat closes simultaneously two passages. These are so arranged that the steam pressure tending to close or open the valve remains balanced, and no large force has to be exerted. Such valves are often

worked by cams and levers from a lay shaft alongside the cylinder, driven from the main shaft so as to turn in unison. In many examples the valves giving admission are caused to close by a tripping device; that is to say they spring suddenly into the closed position, at the instant when cutoff is desired, through the action of a trigger which brings a closing spring into play; and the tripping device is connected with the governor of the engine in such a manner that the cutoff comes earlier or later when the speed exceeds or falls short of the normal. In this way the engine automatically adjusts the amount of power it develops to suit changes in the demand made upon it by the mechanism which it drives.

Balanced Valves.—Fig. 3 shows a slide valve form usual in locomotives and small engines but in large sizes the objection is that the unbalanced pressure of steam on the back of the valve would cause much friction and wear in the movement of the valve over the working face on which it slides. In large engines, such as those of steamships, it is usual to relieve the pressure of the valve on the working face by fitting the back of the valve with what is called a relief frame, extending between the valve and the steam-chest cover, the effect of which is to prevent the steam of the steam chest from having access to the whole back surface of the valve.

Another device, often resorted to, is to substitute for the flat form of slide valve a piston form in which the effective sliding portions of the valve extend all around the surface of a cylinder and slide over cylindrical fixtures in the valve chest, around which the steam ports extend.

Reversing Gears.—The slide valve is a particularly convenient means of controlling the admission of steam when, as in locomotives or marine engines, etc., the engine's direction of motion has frequently to be reversed. It is only necessary to provide means by which the valve may be actuated by either of two eccentrics, one set at the angle suitable for forward running and the other set at the angle suitable for backward running.

In Stephenson's link motion, shown in fig. 7(A), which is an early and familiar device for this purpose, the engine shaft carries two eccentrics, one of which is set ahead of the crank at a suitable angle for one direction of running, and the other at a suitable angle for the other direction. Their rods are connected to the ends of a link which gives its name to the contrivance. The link is a slotted bar curved to a circular arc, and capable of being moved up or down, being suspended from a rod which can be raised or lowered by means of the hand lever above. The valve rod ends in a block which slides within the link. This device allows either the forward eccentric or the backward eccentric to determine the motion of the valve, and so allows the engine to be reversed. But it does more, for if the link be set to an intermediate position, in which it takes its motion chiefly from one eccentric but partly from the other, the valve receives a motion virtually the same as that which it would get from a single eccentric of shorter travel and greater angular advance. The effect is to give a distribution of steam in which the cutoff is hastened and the expansion and compression are increased.

Thus beside allowing of full forward and full backward gear the link may be notched up to give a much shorter admission of steam along with more compression, a matter of practical value in the running of a locomotive, where frequent large variations in the demand for driving effort have to be met.

Various other forms of link motion have been devised which share this advantage. It is also shared by a group of equivalent mechanical devices, called by the general name of radial valve gears, in which only one eccentric is employed but the valve is caused to take a motion capable of being varied in the same way. Of these the most widely used is the Walschaert's gear, very common in locomotives, a skeleton diagram of which is shown in fig. 7(B). There A is a projection moving with the engine piston. The valve V takes motion partly from it, through the rods AB and BV; but BV has a fulcrum at D, the position of which varies through the movement of the rod DG. That rod is supported by a suspension rod J, and its end at G can slide in a link HI rocking about a fixed fulcrum at O. There is one eccentric, CF, set at right angles to the crank CK, and giving motion to the rocking link HI by the rod FH. The block G can be shifted from near H to I by raising the suspension rod J. This reverses the motion, and allows also of notching up to any intermediate position.

Governing.—To keep a steam engine running at a uniform speed, notwithstanding variations in the demand for power, hand regulation may to some extent serve, but in very many cases an automatic regulator called a governor is provided. Regulation, whether by hand or by automatic governor, may be effected in two ways: the steam may be throttled more or less on its way from the boiler to the engine, so that the pressure of admission is

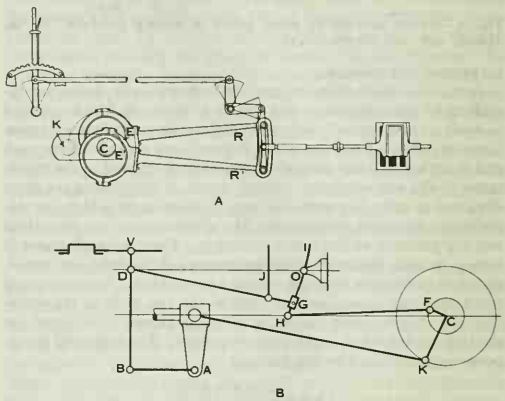


FIG. 7.—(A) STEPHENSON'S LINK MOTION: (E, E') ECCENTRICS, (R, R') RODS CONNECTED TO ECCENTRICS; (B) WALSCHAERT'S GEAR

reduced when the speed tends to rise, and increased when the speed tends to fall; or the point of cutoff may be varied, with the result of admitting a greater or less volume of steam in each stroke. In small engines the governor often acts on a throttle valve, but for engines which aim at efficiency it is preferable to govern by altering the cutoff. In such engines, to secure in any case a proper amount of expansion, there is an early cutoff, and a small change in its position largely affects the amount of steam admitted and consequently the amount of work done.

Watt's original governor was a pair of balls revolving about a vertical spindle as conical pendulums under the combined effect of their own weight and the centrifugal force due to their speed of revolution. Any increase of speed sufficient to overcome the slight friction of the attached gearing made them take up a higher position, and in so doing they partially closed a throttle valve. This went on until, at a slightly greater speed than before, equilibrium was established between the supply of steam and the demand for power. Here, as in all governors, the effect is not to maintain a strictly constant speed, but to prevent the variation, either way, from exceeding narrow limits.

In modern governors the centrifugal force of the balls generally acts not simply against their own weight, but against an extra load or against the force of a spring.

In many factory and other engines the governor determines the position of cutoff by means of a trigger action of the type already mentioned: the admission valve, which has been opened and held open by a cam turning with the engine shaft, is suddenly disconnected by the release of a trigger, at a point depending on the exact height of the governor balls.

Function of the Flywheel.—The function of the governor is, generally, to keep the engine running at a nearly uniform rate, measured in turns per minute; the function of the flywheel is to keep the rate nearly uniform within the limits of any one revolution. For this purpose it must act as a reservoir of energy, alternately storing and restoring the excess amounts which result from periodic variations of turning moment in the operation of the piston on the crank. At the dead points the piston's rate of doing work is zero, and it rises to a maximum at an intermediate point

the position of which may be much affected by an early cutoff. The flywheel is drawn upon for energy during those parts of the revolution in which the work done by the piston (or pistons) on the engine shaft is less than the work done by the shaft on the mechanism which the engine drives; and it takes up the surplus during the other parts of the revolution.

To effect this alternate give and take of energy the flywheel must undergo small variations of speed, the magnitude of which may be kept down to any assigned amount by giving the rim a sufficient mass and speed.

Inertia of Reciprocating Parts.—The effective effort on the crank is greatly influenced, especially in high-speed engines, by the inertia of the piston, pistonrod and connecting rod. At the beginning of the stroke, when the piston is coming toward the crankshaft, these reciprocating masses are acquiring motion; from a point near the middle to the end of the stroke they are losing motion. Hence in the early part the effort on the crank is reduced below what is due to the steam pressure on the piston and in the later part it is increased. In a high-speed engine the forces due to acceleration become so large as seriously to affect questions of design.

In direct-acting steam pumps the inertia of pump rods and buckets forms an additional item in the consideration of reciprocating mass; its effect there is on the whole beneficial in tending to equalize the force which is exerted on the pump throughout the stroke.

Balancing of Engines.—Another aspect of the acceleration of reciprocating parts is its effect on the balance of the engine as a whole. This is especially important in locomotives and marine engines. In a locomotive the forces on the engine frame arising from the inertia of the pistons and rods tend to make the engine sway laterally; this is to some extent prevented by placing balance masses on the wheels. But such masses, although they may be adjusted nearly to balance the horizontal forces, introduce unbalanced vertical forces, causing what is called hammerblow. In extreme cases these vertical forces might even lift the wheels from the rails; short of that, however, they are objectionable in causing periodic variations in the pressure on the rails, which tend to set bridges into oscillation and may add substantially to the stresses for which the railway engineer must provide.

C. Types of Reciprocating Engines

In classifying engines with regard to their general arrangement of parts and mode of working, account has to be taken of a considerable number of independent characteristics. We have first a general division into condensing and noncondensing engines, with a subdivision of the condensing class into those which act by surface condensation and those which use injection. Next there is the division into compound and noncompound, with a further classification of the former as double-, triple- or quadruple-expansion engines. Again, engines may be classed as single- or double-acting, according as the steam acts on one or alternately on both sides of the piston. Again, a few engines—such as steam hammers and certain kinds of steam pumps—are nonrotative, that is to say, the reciprocating motion of the piston does work simply on a reciprocating piece; but generally an engine does work on a continuously revolving shaft. In most cases the crankpin of the revolving shaft is connected directly with the piston rod by a connecting rod, and the engine is then said to be direct-acting; in other cases, of which the beam engine is an historical example, a lever is interposed between the piston and the connecting rod. The same distinction applies to nonrotative pumping engines, in some of which the piston acts directly on the pump rod, while in others it acts through a beam. The position of the cylinder is another element of classification, giving horizontal-, vertical- and inclined-cylinder engines. In most vertical engines the cylinder is above the connecting rod and crank. In oscillating-cylinder engines (a type still found in some river boats) the connecting rod is dispensed with; the piston rod works on the crankpin, and the cylinder oscillates on trunnions to allow the piston rod to follow the crankpin around its circular path. In rotary engines there is no piston in the ordinary sense; the steam does work on a revolving piece, and the necessity is thus avoided of afterward converting reciprocating into rotary motion.

Beam Engines.—In the single-acting atmospheric engine of Newcomen the beam was a necessary feature; the use of water packing for the piston required that the piston should move down in the working stroke, and a beam was needed to let the counterpoise pull the piston up. Watt's improvements made the beam no longer necessary; and in one of the forms he designed it was discarded—namely, in the form of pumping engine known as the Bull engine, in which a vertical inverted cylinder stands over and acts directly on the pump rod. However, the beam formed a convenient driver for pump rods and valve rods; and the parallel motion invented by Watt as a means of guiding the piston rod, which could easily be applied to a beam engine, was, in the early days of engine building, an easier thing to construct than the plane surfaces which are the natural guides of the piston rod in a direct-acting engine. In modern practice the direct-acting type has almost wholly displaced the beam engine.

Direct-acting Engines.—Of direct-acting engines the horizontal arrangement has in general the advantage of greater accessibility, but the vertical economizes floor space. In small forms the engine is generally self-contained, that is to say, a single frame or bedplate carries all the parts including the main bearings in which the crankshaft with its flywheel turns. The frame often takes what is called a girder shape, which brings a portion of it into a favourable position for taking the thrust between the cylinder and the crankshaft bearings and allows two surfaces to be formed on the frame to serve as guides for the crosshead. When a condenser is used with a horizontal engine it is often placed behind the cylinder, and the air pump, which may be within the condenser, has a horizontal plunger or piston on a tail rod or continuation of the main piston rod through the back cover of the cylinder. In large horizontal engines the condenser generally is in a well below, and its pump is driven by a mechanism linked with the crosshead.

Coupled Engines.—When uniformity of driving effort or the absence of dead points is important, two independent cylinders often work on the same shaft by cranks at right angles to each other. Such engines, which are called coupled, can start readily from any position; the ordinary locomotive engine is an example. In some engines three cylinders, set to work on cranks 120° apart, co-operate in giving a still more uniform drive.

Winding engines for mines and collieries, in which ease of starting, stopping and reversing is essential, are very generally made by coupling a pair of cylinders on opposite sides of the winding drum.

Compound Engines, Coupled or Tandem.—Large direct-acting engines are usually compounded either by having a high- and a low-pressure cylinder side by side, with cranks at right angles, or by putting one cylinder behind the other with a common centre line. The latter is called a tandem arrangement. In a tandem engine, since the pistons agree in phase, the steam may expand directly from the small into the large cylinder. But the connecting pipe and steam chest form a receiver of considerable size, and it is preferable to cut off the supply of steam to the large cylinder at an earlier stage of the stroke. For mill engines the compound tandem and compound coupled types of engine are very usual. The high-pressure cylinder is often fitted with some form of trip gear.

The chief advantage of compounding, as was pointed out in the historical section, is that it reduces the condensation which hot steam undergoes in the cylinder through contact with metal which has been chilled by the exhaust of the previous stroke. Compound working acts beneficially by narrowing the range through which the temperature periodically fluctuates in the metallic surfaces exposed to contact with the operative steam. The amount of steam that is at once condensed on admission depends on the extent to which the surface with which it comes in contact has been chilled during the previous exhaust. For this reason there is less loss through initial condensation in a compound engine giving the same total ratio of expansion.

Uniflow Engine.—The periodic give and take of heat between the steam and the metallic surfaces, which is a serious source of

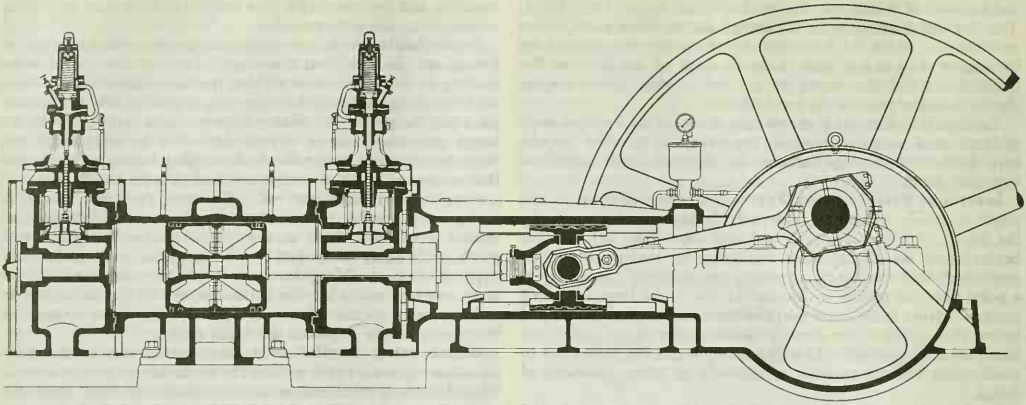


FIG. 8.—UNIFLOW ENGINE IN WHICH THE PISTON ACTS AS EXHAUST VALVE AND THE ADMISSION VALVES ARE IN THE CYLINDER ENDS

loss in any reciprocating engine, can be reduced by having separate valves for admission and exhaust, so that the same port surfaces and valve surfaces are not brought into contact with hot and comparatively cold steam. This idea is carried further in what is called the uniflow engine (fig. 8), by an ingenious arrangement which secures that the ends of the cylinder, where the steam is admitted, shall always be kept hot, the exhaust taking place at the middle of the cylinder through ports in the circumference which are uncovered by the motion of the piston.

The uniflow design is a recent but conspicuously successful departure from earlier practice. It allows high-pressure steam to be expanded in a single cylinder without the drawbacks ordinarily entailed in noncompound working. There is, however, still enough exchange of heat between the metal and the steam to make superheating highly advantageous.

Condensation.—In land engines a jet form of condenser is common, but surface condensation is resorted to when the available water supply is unsuited for boiler feed. When there is no large supply of condensing water a very fair vacuum can be obtained by using an evaporative condenser, consisting of a stack of pipes into which the exhaust steam is admitted and over which a small amount of cooling water is allowed to drip. This water is evaporated by the heat which the condensing steam gives up. Such a condenser is placed in the open, generally on a roof where the air has free access. The amount of water it uses need not exceed the amount of steam that is condensed, and is a small fraction of the amount that would be required in a jet or surface condenser.

High-speed Direct-acting Engines.—Prior to the development of the steam turbine the demand for engines suitable for driving electric generators without the intervention of a belt led to the introduction of various forms of direct-acting engine adapted to run at a high speed. Some of these were single-acting, steam being admitted to one side of the piston only, generally the back, with the result that the rods could be kept in a state of thrust throughout the revolution, and alternations of stress in them and at the joints thereby avoided, together with the knocking and wear of the bearing brasses which it is apt to cause. To assure, however, that the connecting rod should always push and never pull against the crankpin there had to be much cushioning during the outstroke on account of the fact that from about the middle of that stroke to the end the reciprocating mass was being retarded.

In engines of this class designed by P. W. Willans, which were at one time much used, but are now displaced by turbines, the cushioning was provided by means of a supplementary piston which compressed air during the outstroke; the energy which the reciprocating masses had to part with in losing their motion during the second half of the outstroke was stored in this air and was restored in the succeeding downstroke. The engine was entirely enclosed in a casing the bottom of which formed an oil bath in which the cranks splashed to ensure ample lubrication. This feature is retained in many modern high-speed engines.

Pumping Engines.—In engines for pumping or for blowing air it is not essential to drive a revolving shaft, and in many forms the reciprocating motion of the steam piston is applied directly to produce the reciprocating motion of the pump piston or plunger. Often an inverted vertical triple-expansion engine is used, resembling the usual marine form, with a pump plunger under each of the three cylinders. Most of the power is directly transmitted from the piston to the pump, but there is an auxiliary crank for each cylinder, driving a shaft with a flywheel to equalize the effort on the pump.

A familiar example of direct-acting pump is the small donkey pump used for feeding boilers, in which the steam piston and pump plunger are on one and the same rod. In some of these pumps a rotative element is introduced, partly to secure steadiness of running and partly for convenience in working the valves. But many pumps of this class are entirely nonrotative, and in such cases the steam is generally admitted throughout the stroke without expansion. The valve may be actuated by tappets from the piston rod. In some cases a tappet worked by the piston as it reaches each end of its stroke throws over an auxiliary steam valve, which admits steam to one or other side of an auxiliary piston carrying the main slide valve.

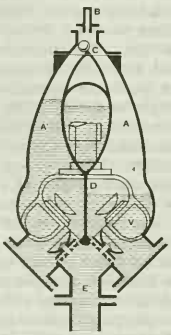


FIG. 9.—HALL'S PULSOMETER

Pulsometer.—Hall's pulsometer is a peculiar pumping engine without cylinder or piston, which may be regarded as the modern representative of the engine of Savery. The sectional view, fig. 9, shows its principal parts. There are two chambers, A, A′, narrowing toward the top, where the steam pipe B enters. A ball valve C allows steam to pass into one of the chambers and closes the other. Steam entering (say) the right-hand chamber forces water out of it past the clack valve V into a delivery passage D, which is connected with an air vessel. When the water level in A sinks so far that steam begins to blow through the delivery passage, the water and steam are disturbed and so brought into intimate contact, the steam in A condenses, and a partial vacuum is formed. This causes the ball valve C to rock over and close the top of A, while water rises from the suction pipe E to fill that chamber. At the same time steam begins to enter the other chamber A′, discharging water from it, and the same series of actions is repeated in each chamber alter-

nately. While the water is being driven out there is comparatively little condensation of steam, partly because the shape of the vessel does not promote the formation of eddies, and partly because there is a cushion of air between the steam and the water. Near the top of each chamber is a small air valve opening inward, which allows a little air to enter each time a vacuum is formed. When any steam is condensed, the air mixed with it remains on the cold surface and forms a nonconducting layer. The pulsometer is, of course, far from efficient as a thermodynamic engine, but its adaptability to situations where other steam pumps cannot easily be applied, and the simplicity of its working parts, sometimes make it practically useful.

Marine Engines.—The early steamers were fitted with paddle wheels, and the engines used to drive them were for the most part modified beam engines. Bell's "Comet" was driven by a species of inverted beam engine, and another form of inverted beam, known as the side-lever engine, was for long a favourite with marine engineers. An old form of direct-acting paddle engine was the steeple engine, in which the cylinder was set vertically below the crank. Two piston rods projected through the top of the cylinder, one on each side of the shaft and of the crank. They were united by a crosshead sliding in vertical guides, and from this a return-connecting rod led to the crank. Most modern paddle-wheel engines are direct-acting engines of the ordinary connecting-rod type, with the cylinders fixed on an inclined bed and the guides sloping up towards the shaft.

When the screw propeller began to take the place of paddle wheels in ocean steamers, the increased speed which it required was at first obtained by using spur-wheel gearing in conjunction with one of the forms of engines then usual in paddle steamers. Types of engine better suited to the screw were introduced later, and were driven fast enough to be connected directly to the screw shaft. After passing through various forms, now obsolete, the marine engine settled into one which is now universal in all screw steamships that are not driven by steam turbine.

This is the inverted vertical direct-acting engine, with two or more cylinders placed side by side directly over the shaft. It has the great advantage that the shaft can readily be put at the low level necessary for screw propulsion. Two, three and four cranks are employed, the arrangement with four cranks being especially suitable, as has already been pointed out, when a balance of the engine at high speeds has to be secured. In vessels of high speed and power the engines are often arranged in twin sets, on two shafts with twin screw propellers.

The marine engine is always furnished with a surface condenser, consisting of a multitude of brass tubes about $\frac{3}{4}$ inch in diameter cooled by sea water which is caused to circulate through the condenser by means of a circulating pump. This pump and the air pump are often driven independently of the main engine.

It is in marine practice that the most powerful reciprocating engines are still to be found, although in the largest and fastest vessels the turbine has taken their place. Another and more recent rival is the internal combustion engine, using the Diesel cycle of action, which by 1928 had found favour in many new liners. The reciprocating engine continues, however, to be much used for sea-going vessels, especially those of the tramp class, most commonly in the triple-expansion form.

Often the third stage of the expansion is performed in two cylinders, making four in all. This avoids the use of a very large cylinder, and secures an advantage in the better balance which can be obtained with four cranks.

Locomotive Engines.—The ordinary steam locomotive consists of a pair of direct-acting horizontal or nearly horizontal engines, fixed in a rigid frame under the front end of the boiler, and coupled to the same shaft by cranks at right angles, each with a single slide valve worked by a link motion, or by a form of radial gear. The engine is noncondensing, except in a very few special cases, and the exhaust steam, delivered at the base of the funnel through a blast pipe, serves to produce a draft of air through the furnace. In some instances a portion of the exhaust steam, amounting to about one-fifth of the whole, is diverted to heat the feed water.

On the shaft are a pair of driving wheels, whose frictional adhesion to the rails furnishes the necessary tractive force. Nearly always a greater tractive force is secured by having two or more driving wheels on each side, connected by a coupling rod between pins on the outside of the wheels.

It is general to have under the front of the engine a group of smaller wheels which do not form part of the driving system. These are carried in a bogie, that is, a small truck upon which the front end of the boiler rests by a swivel pin or plate which allows the bogie to turn, so as to adapt itself to curves in the line, and thus obviate the grinding of tires and danger of derailment which would be caused by using a long rigid wheel base.

In inside-cylinder engines the cylinders are placed side by side within the frame of the engine, and their connecting rods work on cranks in the driving shaft. In outside-cylinder engines the cylinders are spread apart far enough to lie outside the frame of the engine, and work on crankpins on the outside of the driving wheels. Many modern locomotives combine the outside and inside arrangement, in order to get greater power, placing one or in some cases two cylinders within the frame, as well as two outside. The principle of compounding has often been applied to locomotive engines, but without much advantage. (*See* RAILWAY: *Locomotives.*)

BIBLIOGRAPHY.—Robert H. Thurston, *A History of the Growth of the Steam-Engine,* centennial ed. (Ithaca, N.Y., London, 1939); Henry Winram Dickinson, *A Short History of the Steam Engine* (New York, London, 1939); W. E. Dalby, *Valves and Valve Gear Mechanisms* (New York, London, 1906), *Steam Power* (New York, London, 1915); J. A. Ewing, *The Steam-Engine and Other Heat-Engines,* 4th ed. (New York, 1926). (J. A. E.)

II. PROPERTIES OF STEAM

To understand the properties of steam imagine a vertical cylinder containing one pound of water as illustrated in fig. 10(A). Resting on the water is a weighted movable piston exerting a constant pressure on the contents. Imagine the water to be initially at room temperature. In this condition it is called subcooled liquid because it exists at a temperature below its boiling point. It occupies the volume indicated at point A′ in fig. 10(F). Next

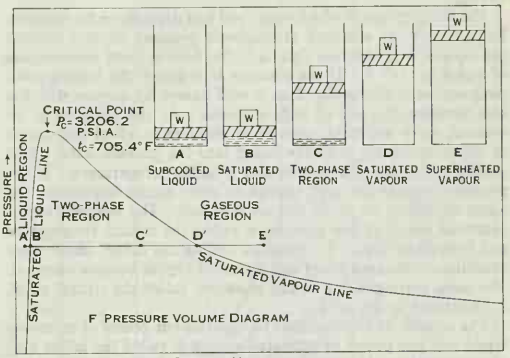


FIG. 10.—DIAGRAM ILLUSTRATING THE BEHAVIOUR OF STEAM. IN THE TWO-PHASE REGION (C) BOTH SATURATED LIQUID AND SATURATED VAPOUR ARE PRESENT

imagine heat to be slowly added to the water until it reaches the boiling point temperature and volume represented by point B′, which is slightly greater than A′. In this condition the water is said to be in the saturated liquid state. The volume is often designated by the symbol v_f meaning specific volume of the saturated fluid (liquid). If still more heat is added, boiling occurs and steam is formed underneath the piston forcing it upward as illustrated in fig. 10(C).

When liquid and steam coexist the contents are said to be in the two-phase region, meaning that the water exists in two physical forms, liquid and gas. As long as this condition exists, the mixture will remain at its boiling-point temperature because additional heat

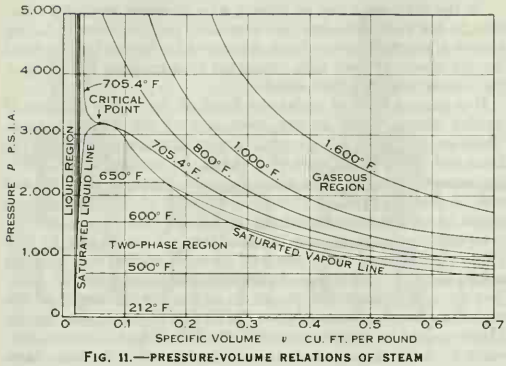


FIG. 11.—PRESSURE-VOLUME RELATIONS OF STEAM

will merely cause more water to boil. Similarly, if heat is withdrawn, some of the steam will condense but the temperature will not change. Between points B′ and D′ in fig. 10(F), therefore, the temperature is constant and will depend only on the pressure exerted by the piston.

If just enough heat is added to vaporize the last droplet of liquid the entire volume will be occupied by steam. Under this condition the steam is said to be dry saturated and will be represented by the point D′ in fig. 10(F). Its volume is usually designated by the symbol v_g, meaning specific volume of the saturated gas (vapour). If still more heat is added there will no longer be any liquid to boil and the temperature will rise. In this condition the steam is said to be superheated, meaning that it is heated above its boiling- or condensation-point temperature. As indicated in fig. 10(E) the volume of a pound of superheated steam is greater than that of a pound of saturated steam at the same pressure. A typical point in the superheated region is that represented by E′ in fig. 10(F).

The temperature at which water will boil depends on its pressure. For example, at standard atmospheric pressure of 14.7 pounds per square inch absolute (p.s.i.a.), the boiling-point temperature of water is 212° F. If the pressure is reduced, the boiling-point temperature is also reduced, as is well known by anyone who has had occasion to cook at high altitudes. As the pressure is increased, more molecular activity is required to effect the escape of water molecules from the liquid into the gaseous state. This greater molecular activity requires a higher temperature. If the pressure is made just high enough to make escape impossible, the water is said to be at its critical pressure. The temperature of saturated liquid at this pressure is called the critical temperature and is equal to 705.4° F. Since no boiling can occur, under these conditions, saturated liquid and saturated vapour become identical. The point corresponding to this situation, called the critical point, is illustrated in fig. 10(F).

The amount of heat required to vaporize one pound of saturated liquid into one pound of saturated vapour is called the latent heat of vaporization. It varies with the pressure, being greatest at low pressures and diminishing to zero at critical pressure.

The importance of steam for the heating of buildings, generation of power and many other applications has resulted in a need to know its physical properties over a wide range of pressure and temperature.

The most important properties of steam are its specific volume, internal energy, enthalpy and entropy.

Specific Volume.—The specific volume is defined as the volume per unit weight. In the English system of units it is measured in cubic feet per pound. The variation of specific volume with pressure and temperature is illustrated by the pressure-volume diagram of fig. 11. The diagram presents the variation of volume with pressure along constant temperature lines or isotherms. The diagram is divided into three regions: liquid, two-phase, and gaseous. Everything to the left of the saturated liquid line and below the critical temperature of 705.4° F. is normally considered to be in the liquid region. Everything to the right of the saturated vapour line or above 705.4° F. is normally considered to be in the gaseous region. At pressures above the critical this designation is quite arbitrary, since there is really no physical distinction between liquid and vapour across the critical temperature line. The two-phase region lies between the saturated liquid and saturated vapour lines. In this region lines of constant temperature are horizontal and therefore correspond to lines of constant pressure.

Examination of one of the constant temperature lines reveals that there is little change of volume with pressure in the liquid region, indicating that liquid water is almost incompressible. During boiling the specific volume increases at constant pressure and temperature until the saturated vapour line is reached, after which a reduction in pressure is required for an increase in volume at constant temperature.

Specific volumes along the saturated liquid line are designated as v_f while those along the saturated vapour line are designated as v_g. In the two-phase region, which is sometimes called the saturated region, the fraction by weight of the mixture which is vapour is called the quality. It is usually designated by the symbol x. Thus, if at a given pressure 90% of the water-vapour mixture is saturated vapour and 10% saturated liquid, the steam would be said to have a quality of $x = 0.90$ or 90%. From a knowledge of the quality the specific volume of the steam can readily be computed since one need only add together the fraction which is liquid times its saturated liquid specific volume and the fraction which is vapour times the specific volume of saturated vapour. Since the fraction which is vapour is x and that which is liquid is $(1-x)$, the specific volume in the two-phase region will be

$$v=(1-x)v_f+xv_g=v_f+x(v_g-v_f)=v_f+xv_{fg} \qquad (1)$$

where v_{fg} represents the change in volume from liquid to vapour (v_g-v_f).

Internal Energy.—The internal energy of steam is that energy possessed by virtue of the motion and configuration of its molecules. It is usually measured above a fixed datum state in which its value is arbitrarily taken as $-0.185\ p_f v_f$ where p_f and v_f are the pressure and specific volume of saturated liquid at the freezing-point temperature of 32° F. The reason for this value will become clear in the following section on enthalpy. The internal energy is most frequently designated by the symbol u and in the English system of units is measured in British thermal units (B.T.U.) per pound. As in the case of volume it also varies with pressure and temperature, and is designated at important states with the same subscripts. Thus

u_f = internal energy of saturated liquid
u_g = internal energy of saturated vapour
$u=(1-x)u_f+xu_g=u_f+xu_{fg}$ = internal energy in the two-phase region.

Enthalpy.—The enthalpy of steam is a property obtained by adding the pressure-volume product to the internal energy. Thus, in consistent units,

$$h=u+\frac{144pv}{778}=u+0.185pv \text{ B.T.U. per pound.} \qquad (2)$$

In this expression p is the pressure in pounds per square inch absolute and v is the specific volume in cubic feet per pound. The numerical constants 144 and 778 are conversion factors representing the number of pounds per square inch in one pound per square foot and the number of foot pounds in a British thermal unit respectively. Thus, the units of enthalpy are the same as those of internal energy, British thermal units per pound.

In the Keenan and Keyes steam tables the value of enthalpy is arbitrarily taken as zero at 32° F. and at 0.08854 p.s.i.a., the saturation pressure corresponding to this temperature. It is for this reason that the internal energy u at this state is $-0.185\ p_f v_f$ as indicated in the previous section. The enthalpy varies with pressure and temperature in much the same manner as internal energy, and at important states is designated with the same subscripts. Thus

h_f = enthalpy of saturated liquid

h_g = enthalpy of saturated vapour
$h = (1-x)h_f + xh_g = h_f + xh_{fg}$ = enthalpy in the two-phase region.

The importance of the enthalpy of steam arises from the fact that it occurs in the steady flow energy equation. If change in velocity and elevation during a flow process are neglected, the steady flow energy equation between any two sections becomes

$$h_1 \pm {}_1Q_2 \pm {}_1W_2 = h_2 \tag{3}$$

where h_1 and h_2 are the enthalpies of the steam at sections 1 and 2 respectively and $\pm\ {}_1Q_2$ is the heat added or abstracted and $\pm\ {}_1W_2$ the work done on or by the steam between the two sections, all in British thermal units per pound of steam. If heat is added in some device such as a boiler; where no work is done, the equation becomes

$${}_1Q_2 = h_2 - h_1 \tag{4}$$

This means that the difference in enthalpy is a measure of the heat added, and that therefore, to compute the amount of heat added one need only subtract the initial enthalpy from the final enthalpy. Similarly, if one considers the flow of steam through a steam turbine, where ${}_1Q_2 = 0$ and ${}_1W_2$ is positive, the equation becomes

$${}_1W_2 = h_1 - h_2 \tag{5}$$

The work done by the turbine is, therefore, obtained by merely subtracting the final enthalpy from the initial enthalpy. Also, when water is vaporized into steam at constant pressure the amount of heat required, called the latent heat of vaporization, is equal to the difference between the final and initial enthalpies $h_g - h_f$, designated as h_{fg}.

The above illustrations indicate the need for accurate information as to the value of enthalpy of steam as a function of pressure and temperature. These values have been determined and tabulated as will be observed in subsequent sections.

Entropy.—The entropy of steam is a property defined mathematically by the equation

$$s = s_0 + \int \left(\frac{\delta Q}{T}\right)_{\text{rev}} \tag{6}$$

A complete discussion of this property cannot be undertaken here because it would involve an extensive excursion into thermodynamics. However, it is one of the most useful properties of steam and is employed for evaluating that portion of heat energy which is unavailable for conversion into useful work.

Equation (6) states that the entropy s at a given state is equal to the entropy s_0 at an arbitrarily chosen initial state, plus the integral between the two states of the infinitesimal of heat added, divided by the absolute temperature at which the heat is added. The term δQ is used instead of the usual dQ to indicate that the quantity is not an exact differential but an infinitesimal which depends upon the thermodynamic path along which the heat is added. The subscript rev means that the process must be reversible, a thermodynamic term indicating the absence of any dissipative effects such as friction or turbulence plus complete thermal, chemical and mechanical equilibrium throughout the process.

In most steam tables the value of s_0 is arbitrarily taken as zero at 32° F. and the corresponding saturation pressure of 0.08854 p.s.i.a. As in the case of volume, internal energy and enthalpy, the entropy of saturated liquid and vapour are designated as s_f and s_g respectively. Similarly, in the two-phase region it is equal to

$$s = (1-x)s_f + xs_g = s_f + xs_{fg}.$$

Mollier Chart.—A useful tool for engineers is the Mollier chart in which the enthalpy of steam is plotted opposite its entropy. A sketch of such a chart showing the various regions of importance, is presented in fig. 12. This chart shows the general relationship between pressure, temperature, enthalpy and entropy for steam. On the left is the saturated liquid line along which values of enthalpy for saturated liquid h_f can be read. This region is not included in most Mollier charts since it can more accurately be covered with tables. Along the saturated vapour line, values of enthalpy of saturated vapour h_g can be read. Above this line the enthalpy of superheated vapour is obtainable. In the two-phase region where liquid and vapour coexist and where the temperature depends only on the pressure, the enthalpy can be obtained as a function of pressure and steam quality, x.

Steam Tables.—Our modern knowledge of the properties of steam is largely due to H. L. Callendar of England, and more recently to Joseph H. Keenan and numerous other investigators both in the United States and abroad. With the help of equations of state and other information, the basis of which is partly theoretical and partly empirical, Callendar and Keenan have rationalized the experimental data and compiled comprehensive tables intended for use of engineers and scientists.

Callendar's tables were first published in 1915 and later in enlarged form in 1924. Under the auspices of the American Society of Mechanical Engineers' special committee on thermal properties of steam, a thorough investigation of the properties of water was undertaken. From the data obtained a new set of tables, by Joseph H. Keenan, was published in 1930. Subsequent research and comparison of data obtained in the United States, England, Germany and Czechoslovakia led to a revision of previous data and publication of an entirely new and more comprehensive set of tables called *Thermodynamic Properties of Steam* by Joseph H. Keenan and Frederick G. Keyes (1936). The range of temperatures covered in these tables is from 32° F. to 1,600° F. and the range of pressures from saturation pressure at 32° F. to 5,500 p.s.i.a. The tables include values of v_f, v_{fg}, v_g, u_f, u_g, h_f, h_{fg}, h_g, s_f, s_{fg} and s_g for pressures and temperatures up to the critical. They also include values of v, h and s in the superheat region for a wide range of pressures and temperatures.

BIBLIOGRAPHY.—H. L. Callendar, *Properties of Steam and Thermodynamic Theory of Turbines* (New York, 1920); Joseph H. Keenan and Frederick G. Keyes, *Thermodynamic Properties of Steam* (New York, London, 1936); E. F. Leib, "Thermodynamic Properties of Vapors," *Trans. Amer. Soc. Mech. Engrs.* (Feb. 1941).

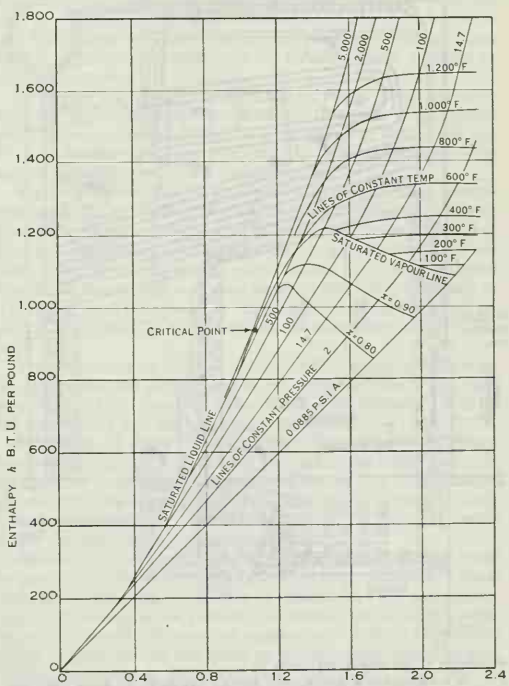
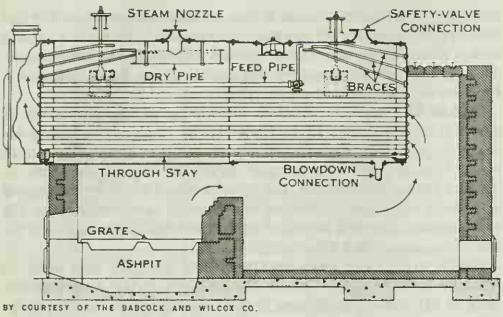


FIG. 12.—MOLLIER DIAGRAM FOR STEAM



BY COURTESY OF THE BABCOCK AND WILCOX CO.

FIG. 13.—EARLY HORIZONTAL-RETURN-TUBULAR BOILER

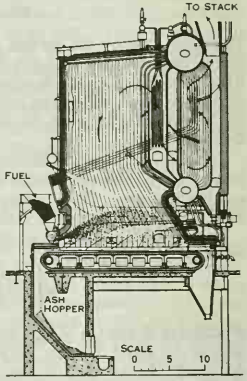
III. STEAM GENERATION

The generation of steam from water is one of the most important processes of our modern industrial age. Steam is used for the heating of nearly all large office buildings, hotels, apartment houses and factories. It is also widely used in industry for the manufacture of steel, aluminum, copper, nickel and other metals, for the refining of oil and for the production of chemicals and many other products. In addition, it drives the huge turbogenerators which furnish the major portion of our electric power.

Historical.—Although the use of steam was first discussed by Hero of Alexandria in his treatise *Pneumatica*, it was not until the invention of the steam engine that its generation and application became widespread. The first steam generators for this purpose were built in England in the early part of the 18th century. They were of rather simple design, usually consisting of a closed vessel heated by fire from underneath. Illustrations of a number of these early steam generators are contained in the preceding section *I. Steam Engine* to which the reader is referred for a discussion of early engines and their accompanying boilers.

With improvement of the steam engine there came a demand for higher pressure steam and for generators of increased capacity and efficiency. This stimulated the development of early versions of fire-tube and water-tube type boilers, forerunners of our present day units.

Fire-Tube Boiler.—This type boiler is characterized by the use of tubes through which hot gases flow on their way to the stack. The tubes pass through a vessel, usually cylindrical in shape, and are surrounded by the water to be converted into steam. The most popular boiler of this type is the horizontal-return-tubular boiler, an early hand-fired version of which is illustrated in fig. 13. With this unit fuel, usually coal, was shoveled onto the grate by hand. Air for combustion was drawn in from underneath the grate by natural draft created by the stack. The hot products of combustion first passed upward and over the bridgewall, thence to the rear of the boiler drum, and finally back through the fire tubes to the stack. The boiler drum was cylindrical in shape and arranged so that the steam generated could collect above the water level and be withdrawn through the dry pipe and steam nozzle shown at the top of the figure. The dry pipe was a tube with perforations near the top arranged so that leaving steam was forced to make a sharp turn in order to throw out entrained moisture. A safety valve, set to allow escape of steam at pressures somewhat above normal operating pressure, was also provided. This valve is an absolute necessity on all boilers since continued addition of heat to water in a closed vessel without means of escape will result in a rise in pressure and ultimate explosion of the boiler.



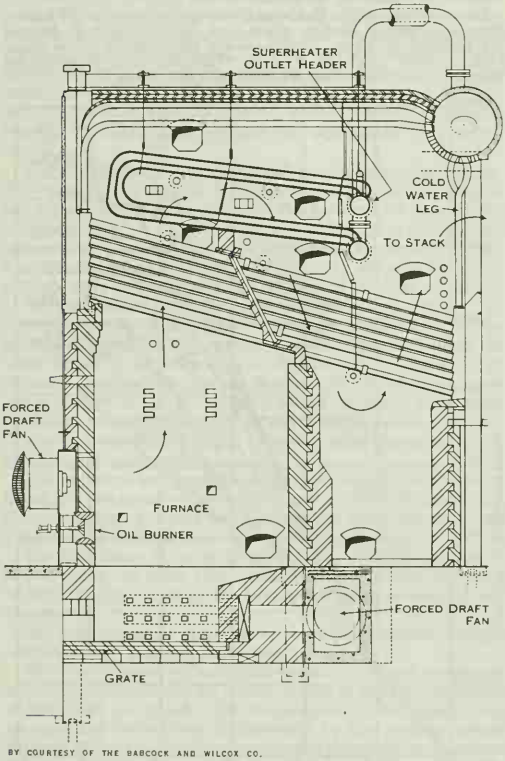
BY COURTESY OF COMBUSTION ENGINEERING, INC.

FIG. 15.—WATER-TUBE BOILER WITH SPREADER STOKER. CAPACITY 150,000 LB. PER HOUR AT 450 P.S.I.G. AND 750° F. ARROWS INDICATE THE MOVEMENT OF GASES OVER THE TUBES

Fire-tube boilers of many varieties were still being built in the mid-20th century. However, because of the need for a large drum and limited heating surface, fire-tube boilers are employed only for pressures below about 250 lb. per square inch gauge (p.s.i.g.) and for steam capacities not exceeding about 25,000 lb. per hour. They are widely used for supplying steam to large apartment buildings, hotels, small factories and other space heating tasks. Present day units can be made fully automatic and can be designed to operate on either oil, gas or coal.

Water-Tube Boiler.—The development of the steam turbogenerator early in the 20th century created a demand for large quantities of steam at pressures and temperatures far exceeding those possible with fire-tube boilers. This demand resulted in the evolution of modern water-tube boilers in which steam is generated inside water tubes from heat supplied on the outside by hot furnace gases. Boilers of this type are employed in a wide variety of sizes and designs and in capacities as high as 3,000,000 lb. of steam per hour and with pressures and temperatures as high as 5,000 lb. per square inch and 1,200° F. respectively. The largest of these units are found in central station power plants of public utilities.

Other units of substantial size are employed in steel mills, paper



BY COURTESY OF THE BABCOCK AND WILCOX CO.

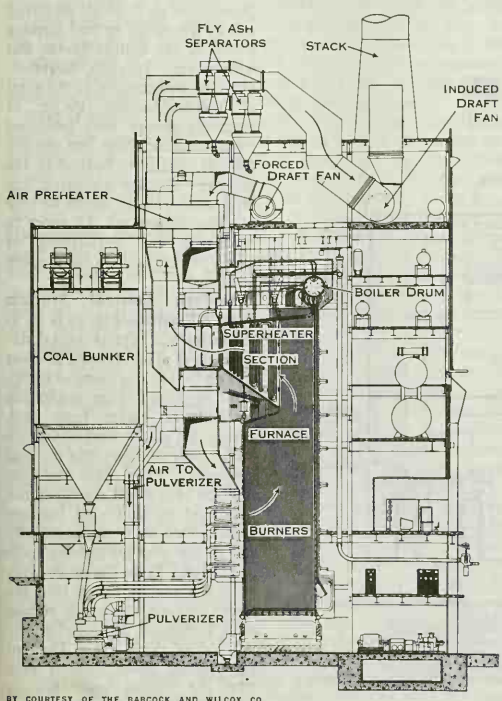
FIG. 14.—CROSS-DRUM WATER-TUBE BOILER DESIGNED TO BURN FUEL OIL AND COTTONSEED HULLS. CAPACITY 16,500 LB. PER HOUR OF 400 P.S.I.G., 790° F. STEAM

mills, oil refineries and large chemical and industrial manufacturing plants.

A modern cross-drum water-tube boiler of moderate size is illustrated in fig. 14. In this unit steam is generated in the inclined tubes and released into the drum above. Advantage is taken of the variation in density of the water in the cold leg on the right, and in the inclined tubes below, to create natural circulation. This improves heat transfer between the hot furnace gases and the water inside the tubes. Provisions are also made for superheating the steam above its boiling point in the hairpin bank of superheater tubes located in the gas path above the inclined tubes.

Another typical water-tube boiler equipped for burning coal of approximately pea size with the aid of a spreader stoker is illustrated in fig. 15. This unit contains two drums and a furnace completely surrounded by water tubes in which the steam is generated. Circulation is natural, and in a clockwise direction. Starting from the upper drum the water flows to the lower drum and then back again to the upper drum through the water tubes surrounding the furnace. The unit is capable of generating 150,000 lb. of steam per hour at 450 p.s.i.g. and 750° F. The spreader-type stoker throws the coal uniformly onto a moving grate which carries the ashes forward and deposits them in an ash hopper located below. These can be removed periodically by hand or by sluicing them out with water.

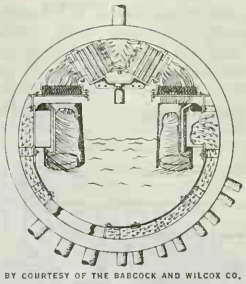
A typical central station type water-tube boiler employing pulverized coal fuel is illustrated in fig. 16. In this unit fuel stored in the bunker on the left is fed into pulverizers where it is reduced to dust size and blown into the burners with air supplied from a forced draft fan located above the furnace. Additional preheated air is supplied around the burners. Combustion takes place in the huge rectangular shaped furnace which is completely surrounded by water tubes in which steam is generated. The hot gases pass up the furnace and then to the left over the superheater tubes. From this point they move up through the air preheater in which a part of their heat is transferred to the air on its way to the burners. Leaving the air preheater the hot gases are drawn upward through fly ash separators into the induced draft fan and out the stack.



BY COURTESY OF THE BABCOCK AND WILCOX CO.

FIG. 16.—CROSS SECTION OF A TYPICAL CENTRAL STATION BOILER

Circulation of water is natural. Water from the drum near the top of the unit is carried down to the lower level of the furnace in huge pipes called downcomers. Here the water is fed into the water-wall tubes surrounding the furnace cavity. Steam is generated in these tubes and is finally returned to the drum where it is separated from the water by means of complex separators similar to those illustrated in fig. 17. The dry steam is then passed through superheater tubes before being piped to turbogenerators for the production of electric power.



BY COURTESY OF THE BABCOCK AND WILCOX CO.

FIG. 17.—STEAM SEPARATOR SHOWING CYCLONE SEPARATORS WITH SCRUBBER ELEMENTS AT TOP OF DRUM

In the 1950s pressures above the critical (see *II. Properties of Steam,* above) have been employed in some central station power plants. In these installations natural circulation is no longer feasible since at pressures above the critical, water does not boil. As a result there is insufficient difference in density between hot and cold water to effect natural circulation. In these units the water must be circulated through the unit by means of pumps.

Fuels.—Principal fuels employed for the generation of steam are coal, oil, natural gas, coke oven gas, blast furnace gas and wood. By far the major portion of coal used for this purpose is bituminous coal although some lignite and anthracite also are burned where plentiful. Most central station plants burn coal in pulverized form but many smaller industrial and heating plants burn it in lump or crushed form on traveling grate, spreader or other type stokers.

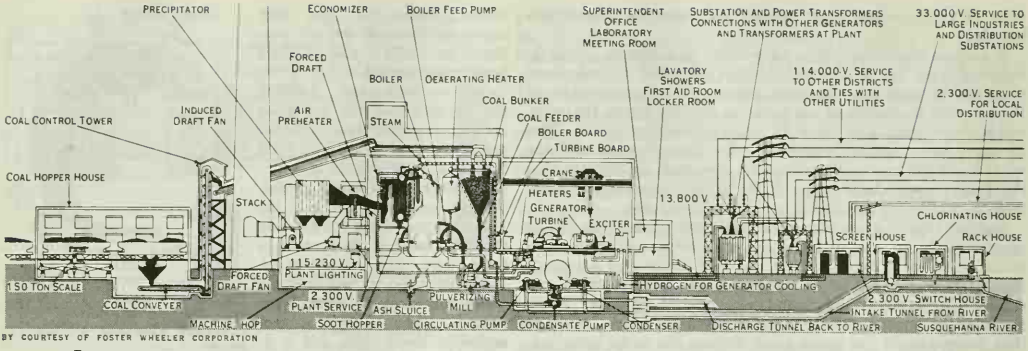
Oil and gas are burned in central station power plants whenever they can be purchased more economically than coal. Many plants are equipped to burn pulverized coal in the winter and natural gas in the summer. Some are equipped with burners capable of operating with either oil, pulverized coal or gas.

Water Treatment.—Water employed in modern steam generators must be completely free of all scale-forming or corrosive chemicals. It must also be free of dissolved oxygen and carbon dioxide. For this reason elaborate chemical treatment plants are necessary. In central station power plants the water generated into steam is used over and over again so that only about 1% or 2% new treated water need be supplied to make up for losses resulting from leakage, contamination, blowdown and other diversions.

The percentage of treated water added is called the makeup. In some industrial plants, such as oil refineries, nearly all steam generated is from makeup. This means that virtually 100% of the water going into the boiler must be treated.

Efficiency of Steam Generators.—The generation of steam involves the transfer of heat from the fuel to the boiler fluid. In this process some of the heat released by the fuel is lost and is therefore not absorbed by the steam. The ratio of the heat actually absorbed by the steam to the heating value of the fuel employed is called the efficiency of the steam generator.

Because of the high cost of fuel, a great deal of effort is made to reduce heat losses in steam generators to a minimum. These include the use of an economizer to preheat the feed water entering the boiler and an air preheater to preheat the air before being introduced into the furnace for combustion. Both of these items aid in cooling the hot flue gases to a lower temperature so that not so much heat is discharged from the stack. Other precautions involve the use of the exact amount of air for efficient combustion. This operation involves the supply of sufficient excess



BY COURTESY OF FOSTER WHEELER CORPORATION

FIG. 18.—TYPICAL CENTRAL STATION POWER PLANT, THE WESTOVER STATION OF THE NEW YORK STATE ELECTRIC AND GAS CORP.

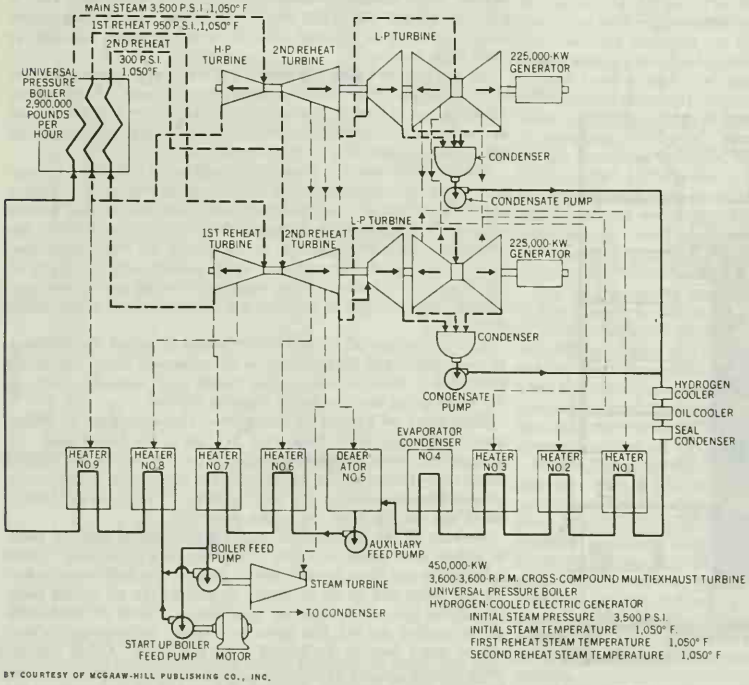
air to assure complete combustion of all carbon to carbon dioxide yet not so much that excess heat is thrown away by heating unnecessary air. This control of supply is accomplished by the use of intricate controls which automatically maintain the correct air-to-fuel ratio for maximum overall steam generator efficiency. In modern central station plants, steam generators are capable of achieving efficiencies as high as 90% based on the higher heating value of the fuel. Average efficiencies of about 87% are common. For small steam generators not employing elaborate control apparatus, economizers or air preheaters, efficiencies of the order of 70% to 80% are typical.

See also BOILER.

BIBLIOGRAPHY.—The Babcock and Wilcox Company, *Steam: Its Generation and Use,* (New York, 1955); Combustion Engineering, Inc., *Combustion Engineering* (New York, 1951); W. N. Barnard, F. O. Ellenwood and C. F. Hirshfeld, *Elements of Heat-Power Engineering,* vol. ii, 3rd ed. (New York, London, 1935); G. A. Gaffert, *Steam Power Stations,* 4th ed. (New York, London, 1952).

IV. STEAM POWER PLANTS

A steam power plant is a plant for the generation of electric power with steam. By far the major portion of electric power generated in the world comes from central station power plants. In the United States in 1956 total power generation by electric utilities amounted to about 600,000,000,000 kw.hr. Of this total, 80% was generated in fuel burning plants and 20% in hydroelectric plants. A breakdown of the types of fuel employed indicated that 70% of the power generated in fuel burning plants was generated from coal, 22% from gas and the remainder from oil. Since practically every plant burning coal or gas is a steam plant, over 92% of all electrical energy generated in fuel burning plants in the United States was with steam. In 1956, therefore, over 400,000,000,000 of the total of 600,000,000,000 kw.hr. of electrical energy were produced from steam and about 200,000,000 tons of coal were burned in the process. Similar statistics, but on a smaller scale, would be true of other countries. All serve to emphasize the importance of steam power plants in our modern economy.



BY COURTESY OF McGRAW-HILL PUBLISHING CO., INC.

FIG. 19.—FLOW DIAGRAM OF 450,000-KW. SUPERCRITICAL POWER PLANT

Modern Central Station Power Plants.—Fig. 18 is an illustration of a typical central station power plant showing in cross section the various items of equipment employed in the generation of electrical power from coal. On the left are the coal cars which dump their contents into the coal conveyer. From here the coal is lifted onto a belt conveyer and deposited in huge coal bunkers from which it is automatically fed to the coal pulverizing mills as needed. From the pulverizing mills the coal is blown into the boiler furnace where it is burned with preheated air supplied by the forced draft fan. The hot combustion gases formed pass over the various boiler heating surfaces, through the economizer, air preheater, precipitator, in-

duced draft fan and finally to the stack.

In the particular plant shown the saturated liquid water from the condenser enters the condensate pumps and is pumped through the first heater into the deaerating heater in which any entrained air is expelled. From the deaerating heater it flows to the boiler feed pumps which force it successively through three other feed-water heaters and then through the economizer into the boiler. Here it is generated into steam and after passing through the superheater flows into the turbine where the power is generated. From the turbine the steam exhausts to the condenser located immediately underneath and is divested of its latent heat by means of the circulating water pumped from a nearby river through the tubes of the condenser. The electric generator furnishes 60 cycle current at 13,000 volts. In the substation, transformers are employed to step the voltage up or down for transmission from the plant.

Modern steam power plants all employ regenerative cycles and many employ reheat. However, pressure and temperature conditions, number of heaters and details of arrangement vary greatly. During the decade 1940–50 most newly installed units employed one reheat and about five feed-water heaters. Pressures and temperatures varied considerably, but were of the order of 1,800 p.s.i.g. and 1,000° F. respectively. During the 1950s the trend was toward still higher pressures and temperatures with a few supercritical units of huge capacity being built. One unit under construction was designed to employ steam at 5,000 p.s.i.g. and 1,200° F. with two reheats, the first to 1,045° F. and the second to 1,050° F.; it was designed to generate 325,000 kw. Two other units were to operate at 3,500 p.s.i.g. and 1,050° F. with two reheats back to the same temperature. The capacity of these units was 450,000 kw. each. A flow diagram for the plant designed for one of these units is illustrated in fig. 19.

Fuel Consumption.—The efficiency of coal burning and other steam power plants has increased remarkably since the turn of the century. In 1900 the best plants were capable of achieving only about one-third of a kilowatt hour of electrical energy per pound of coal while average plants achieved only one-fifth or less. In the late 1950s, the best plants were capable of generating in the neighbourhood of one and one-third kilowatt hours per pound of coal, an increase of about four to one in 50 years. Average plants generated roughly one kilowatt hour per pound of coal.

Cost of Plants.—The cost of central station power plants varies widely with local conditions but in the mid-1950s an average figure was about $150 per kilowatt of generating capacity. This means that a typical modern central station plant capable of generating 600,000 kw. would involve an investment of about $90,000,000 for plant alone. To this must be added the cost of transmission, distribution and other miscellaneous capital equipment costs. During the years 1950 to 1955 the investment in steam plants alone was approximately $1,500,000,000 per year in the United States. During the same period the total investment in plant, transmission, distribution and miscellaneous equipment averaged about $3,500,000,000 per year.

Nuclear Power Plants.—Following World War II the use of nuclear energy for the generation of power became feasible. A number of plants were built in the United States, England, the U.S.S.R. and other countries. In these plants steam was still the motivating force and was generated with the aid of a reactor in which the controlled fission of uranium or plutonium furnished the necessary heat for its vaporization. Thus the reactor took the place of the steam generator of the conventional plant.

Problems associated with nuclear power generation were extremely complex and involved such considerations as shielding against radiation, disposal of fission products, hazard insurance and high cost of construction. Because of these factors, plants being built in the United States were not expected to be economically competitive with coal burning plants for some time. However, as more experience was gained this situation was expected to change.

Binary Cycle Plants.—To achieve maximum efficiency heat should be added to the cycle at the highest possible temperature. In conventional steam power plants a substantial portion of the heat is added at temperatures below the critical of 705° F. This is because the feedwater supplied to the boiler must first be heated to the boiling point and then evaporated at the saturation temperature corresponding to the pressure. In the binary cycle plant this limitation is overcome by simultaneously employing two separate cycles, one a high-temperature Rankine cycle using mercury vapour and the other a conventional steam cycle. The arrangement of the two cycles is such that most of the heat from the fuel is added at high temperatures to the mercury in a special boiler which need not withstand very high pressure. The mercury vapour is expanded through a mercury turbine and exhausted into a special condenser (called a condenser-boiler) that also serves as the boiler for the steam cycle. The condensing mercury thus supplies the latent heat of vaporization of the steam before it is pumped back into the mercury boiler. The steam generated is superheated in the furnace of the mercury boiler and then subjected to the conventional regenerative cycle, after which it is returned to the condenser-boiler for reuse. Although high efficiencies were obtained by this cycle, difficult design and heat transfer problems limited its development. During the 1930s several large plants were constructed but for economic reasons the binary cycle was abandoned in favour of high-pressure steam cycles employing multiple reheat, such as those already described.

Combination Steam and Gas Turbine Cycles.—Another means of increasing the efficiency and capacity of steam power plants is the use of a combination gas turbine and steam turbine cycle. There are a number of possible combinations but in general the objective is to direct the hot exhaust gases from a gas turbine, while they still contain approximately 15% oxygen, through the furnace of the steam generator, thus reducing the amount of preheated air required for combustion of the fuel. An increase of about 5% in thermal efficiency is realized by this means. Up to the latter part of the 1950s only a few units of this type had been built but more were expected to be installed as improvements were

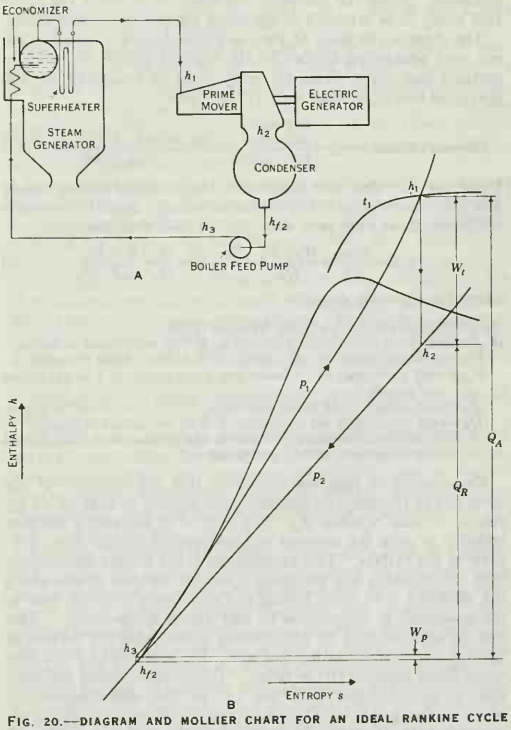


FIG. 20.—DIAGRAM AND MOLLIER CHART FOR AN IDEAL RANKINE CYCLE STEAM POWER PLANT

made.

Industrial Steam Power Plants.—In addition to central station plants there are a wide variety of industrial plants which generate their own electrical energy. Usually such plants require a great deal of steam and it is economical to generate this at higher pressure and temperature in order to permit passage through a noncondensing turbine before otherwise making use of it. In this manner by-product power may be obtained at low cost. Automatic extraction turbines are available for this purpose. With these, variable quantities of steam can be extracted at a fixed pressure regardless of electrical load carried. In industrial plants exhaust steam is also often used for space heating purposes.

Theory.—As may be inferred from fig. 20, the generation of power from steam is accomplished by means of a cycle of events involving the addition of heat in a boiler and the rejection of heat in a condenser. For the sake of simplicity in developing the theory one may start with the simple Rankine cycle named after the gifted English engineer and scientist, W. J. M. Rankine (1820–1872) of Glasgow university. In fig. 20 are illustrated the essential components of the Rankine cycle, together with an idealized diagram of the various processes involved as they would appear on a Mollier chart (see *II. Properties of Steam*, above).

Starting at the outlet of the condenser the water enters the boiler feed pump in the saturated liquid state and is pumped into the boiler at the enthalpy represented by h_3. In the boiler sufficient heat is added to vaporize the steam and to superheat it to the enthalpy represented by h_1 on the chart. From here the steam flows into the turbine through which it expands to the enthalpy h_2 turning the rotor and electric generator. After expanding through the turbine, the steam flows into the condenser where it is converted into saturated liquid with an enthalpy h_{f2}. This is accomplished by transferring the latent heat of the steam to large quantities of circulating water flowing through the tubes of the condenser. From the condenser the water again enters the boiler feed pump to be returned to the boiler for generation into steam.

The thermal efficiency of any power plant cycle is equal to the net work performed divided by the thermal energy expended to perform that work. Since all energy must be accounted for, and also since energy is added only in the boiler,

$$\text{Thermal efficiency} = \eta_T = \frac{\text{net work performed}}{\text{heat added}} = \frac{\text{heat added} - \text{heat rejected.}}{\text{heat added}} \quad (7)$$

Using the fact that heat added and heat rejected during steady flow processes is equal to the change in enthalpy (see *II. Properties of Steam,* above) one may write for the ideal Rankine cycle

$$\eta_{TR} = \frac{W_{\text{cycle}}}{Q_A} = \frac{W_t - W_p}{Q_A} = \frac{Q_A - Q_R}{Q_A} = \frac{(h_1 - h_2) - W_p}{(h_1 - h_{f2}) - W_p} \quad (8)$$

where for the cycle shown:

η_{TR} = thermal efficiency of ideal Rankine cycle,
W_{cycle} = net work performed by the cycle, B.T.U. per pound of steam,
W_t = work performed by the turbine, B.T.U. per pound of steam,
W_p = work performed on the boiler feed water pump, B.T.U. per pound of steam,
Q_A = heat added in the steam generator, B.T.U. per pound of steam,
Q_R = heat rejected in the condenser, B.T.U. per pound of steam,
h_n = enthalpy of the steam or water at the various sections indicated by subscripts, B.T.U. per pound.

Examination of equation (8) shows that the efficiency of the cycle can be increased by increasing the amount of heat added Q_A relative to that rejected Q_R. One method of increasing the heat added is to raise the pressure and temperature of the steam supplied to the turbine. This explains the trend toward increasingly high temperatures and pressures. Another method of increasing the efficiency is to avoid having to throw away so much heat to the condenser in comparison to that added in the boiler. This may be accomplished by withdrawing steam from the turbine at various stages of expansion and using this steam to heat the feed water being returned to the boiler. This process, called regeneration, is accomplished with the aid of feed-water heaters as indicated in fig. 21(A) which is drawn to represent a typical regenerative reheat cycle. Examination of the cycle reveals that all steam bled from the turbine for the purpose of heating the feed water goes through a Rankine cycle similar to that in fig. 21(A), but that the latent heat of this bled steam is not thrown away in the condenser. Instead it is absorbed by the condensate on its way back to the boiler. This means that more than one pound of steam is generated in the boiler per pound of steam condensed in the condenser, and that, therefore, more heat is added compared to that rejected.

In modern central station power plants about 30% of the steam entering the turbine is withdrawn for feed-water heating. The result is a 10% to 15% increase in thermal efficiency.

In fig. 21 the expansion through the turbine is shown much as it would actually occur. Because of losses in the turbine (*see* TURBINE: STEAM) the expansion takes place along the solid lines to the right of the ideal isentropic (constant entropy) lines shown dotted. Also one reheat is shown in which steam, after passing part way through the turbine, is returned to the boiler to be reheated before continuing its expansion through the remaining stages of the turbine.

From the diagram in fig. 21, and from the principle that work of expansion is equal to the weight of steam flowing times its change in enthalpy, the work performed by the steam in B.T.U. per hour can be computed as follows:

$$W_{\text{steam}} = w_1(h_1 - h'_{a1}) + (w_1 - w_a)(h_{a2} - h_b) + (w_1 - w_a - w_b)(h_b - h_c) + (w_1 - w_a - w_b - w_c)(h_c - h_d) + (w_1 - w_a - w_b - w_c - w_d)(h_d - h'_2) \quad (9)$$

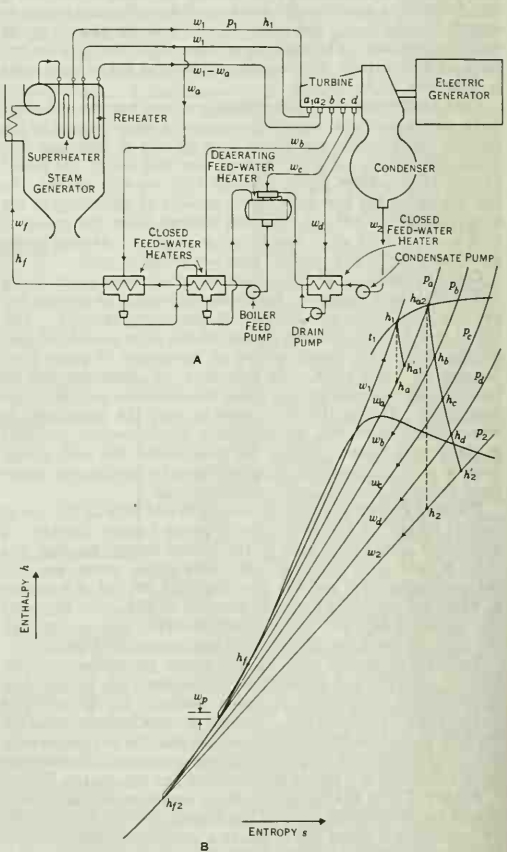


FIG. 21.—DIAGRAM AND MOLLIER CHART FOR A REHEAT-REGENERATIVE POWER PLANT CYCLE

where

w_1 = flow of steam to turbine, pounds per hour,
w_n = flow of steam from the various bleedpoints indicated by subscript a, b, c and d, pounds per hour,
h_1 = enthalpy of steam entering turbine, B.T.U. per pound,
h'_{a1} = enthalpy of steam returned for reheat, B.T.U. per pound,
h_{a2} = enthalpy of steam after reheat, B.T.U. per pound,
h_n = enthalpy of steam flowing from the various bleedpoints indicated by subscript a, b, c and d, B.T.U. per pound,
h'_2 = enthalpy of steam leaving turbine, B.T.U. per pound.

The net work of the power plant is obtained by subtracting the exhaust loss, mechanical loss, generator loss and auxiliary power requirements of the plant, including pump work, from W_{steam} (*see* TURBINE: STEAM).

The heat added by the fuel is obtained by multiplying the steam flow through the boiler by its increase in enthalpy in the boiler, superheater and reheater and dividing by the boiler efficiency. Thus,

$$Q_A = \frac{w_1(h_1 - h_f) + (w_1 - w_a)(h_{a2} - h'_{a1})}{\eta_{SG}} \qquad (10)$$

where:

h_f = enthalpy of boiler feed water entering the steam generator, B.T.U. per pound,
η_{SG} = efficiency of steam generator, B.T.U. per pound.

The thermal efficiency of the regenerative, reheat power plant is therefore

$$\eta_{TP} = \frac{W_{\text{plant}}}{Q_A} = \frac{(W_{\text{steam}} - W_{ex} - W_M - W_G - W_{\text{aux}})\eta_{SG}}{w_1(h_1 - h_f) + (w_1 - w_a)(h_{a2} - h'_{a1})} \qquad (11)$$

where

W_{ex} = exhaust loss, B.T.U. per hour,
W_M = mechanical loss, B.T.U. per hour,
W_G = generator loss, B.T.U. per hour,
W_{aux} = auxiliary power including pump work, B.T.U. per hour,
h_n = enthalpy at various points indicated by subscripts, B.T.U. per pound,
η_{SG} = steam generator efficiency.

Thermal efficiency of modern central station power plants varies widely depending on steam conditions, number of feed-water heaters and complexity. The range is from about 30% in the poorer stations to about 40% in the very best.

BIBLIOGRAPHY.—W. N. Barnard, F. O. Ellenwood and C. F. Hirshfeld, *Heat Power Engineering*, vol. i, ii and iii (1935); W. H. Severns, H. E. Degler and J. C. Miles, *Steam, Air, and Gas Power*, 5th ed. (1954); G. A. Gaffert, *Steam Power Stations* (1952); J. K. Salisbury, *Steam Turbines and Their Cycles* (1950); F. T. Morse, *Power Plant Engineering and Design*, 2nd ed. (1942); P. J. Potter, *Steam Power Plants* (1949); B. G. A. Skrotzki and W. A. Vopat, *Applied Energy Conversion* (1945); A. H. Zerban and E. P. Nye, *Steam Power Plants* (1952).
(RD. A. BR.)

STEAMBOAT: *see* MISSISSIPPI RIVER STEAMBOAT.

STEAM GAUGE, an instrument that utilizes the elasticity of a metal to indicate the pressure of steam in a boiler or other vessel. The action is explained in the article PRESSURE GAUGE.

STEAMSHIP: *see* SHIPBUILDING (MERCHANT AND NAVAL).

STEAMSHIP LINES: *see* SHIPPING INDUSTRY; *World Merchant Fleets.*

STEAM SHOVEL: *see* EXCAVATION.

STEAM TURBINE: *see* TURBINE: STEAM.

STEARIC ACID is one of the more important long-chain fatty acids (*see* CARBOXYLIC ACIDS) which occur in fats isolated from plants and animals. It is a predominant component of body fats; lard and tallow, for example, often contain up to 30% of stearic acid. Like other fatty acids, it does not occur in the free state, but is normally found as an ester of glycerol, *i.e.*, as a glyceride, or as an ester of a long-chain alcohol. Tristearin is a typical example of a triglyceride; it is a triester of glycerol and stearic acid.

Stearic acid, $CH_3(CH_2)_{16}.CO_2H$, is a colourless solid which exists in various crystalline forms of which that melting at 69.6° C. is the most stable. It is a monocarboxylic acid and shows all the reactions typical of this class of organic compounds.

Soaps are the sodium or potassium salts of long-chain carboxylic acids and are prepared by the alkaline hydrolysis of the glyceride components of natural fats. This process is often described as saponification. For the manufacture of soaps and edible fats, solid fats are preferable to liquids, and, since the compounds with lower melting points are unsaturated, hydrogenation of oils yields solid fats. This process is called hardening and is illustrated by the hydrogenation of the oil triolein (the triglyceride of oleic acid) to solid tristearin (melting point 71° C.). *See* PALMITIC ACID.

See K. S. Markley, *Fatty Acids, Their Chemistry and Physical Properties* (1947); A. E. Bailey, *Industrial Oil and Fat Products* (1945).
(W. D. Os.)

STEDMAN, EDMUND CLARENCE (1833–1908), U.S. poet and critic, who was one of the most highly respected and influential writers of his day, was born in Hartford, Conn., on Oct. 8, 1833. After attending Yale university for two years, he edited newspapers in Connecticut and later served on the staff of the *New York Tribune*, in which appeared, in 1859, the satirical "The Diamond Wedding," his first poem to attract wide attention. During the American Civil War he was a war correspondent for the *New York World*. Thereafter until his death he operated a brokerage firm in New York city. He served as president of such organizations as the American Copyright league, the National Institute of Arts and Letters and the New England society.

Although a capable craftsman, Stedman was an imitative poet, his longer poems all bearing a heavy Tennysonian cast. As critic and editor he reflected the conventional taste and methods of the period. His works include *Poems Lyrical and Idyllic* (1860); *Alice of Monmouth* (1864); *The Blameless Prince* (1869); *Victorian Poets* (1875); *Poets of America*, 2 vol. (1885); *A Library of American Literature* (edited with Ellen Hutchinson Cortissoz), 11 vol. (1889–90); *The Nature and Elements of Poetry* (1892); and *A Victorian Anthology, 1837–1895* (1895).

Stedman died in New York city on Jan. 18, 1908.

See Laura Stedman and G. M. Gould (eds.), *The Life and Letters of Edmund Clarence Stedman*, 2 vol. (1910).

STEED, HENRY WICKHAM (1871–1956), British journalist, was born on Oct. 10, 1871, at Long Melford, Suffolk. He joined the staff of the *Times* (London) in 1896 and was appointed foreign editor in Jan. 1914. During World War I, Steed was mainly responsible for the foreign policy of the *Times* and, in 1918, was chosen to frame the policy of Lord Northcliffe's department for propaganda in enemy countries.

He was sent on an official mission to the Italian front where he was authorized by the Allied governments to promise independence to the subject Habsburg peoples. In Feb. 1919 he was appointed editor of the *Times*, and held this post until Nov. 1922. From 1923 to 1930 he was editor of *The Review of Reviews*. Steed was lecturer on central European history at King's college, London, 1925–38, and broadcaster on world affairs in the overseas service, BBC, 1937–47. His publications include *The Habsburg Monarchy* (1913); *Through Thirty Years* (1924); *Hitler: Whence and Whither?* (1934); *That Bad Man* (1942); and *Words on the Air, 1938–1945* (1947).

Steed died at Wootton-by-Woodstock, Oxfordshire, on Jan. 13, 1956.

STEEL: *see* ALLOYS; ARCHITECTURAL ENGINEERING; CONVERTER STEEL; HIGH-SPEED STEEL; IRON AND STEEL INDUSTRY; MANGANESE STEEL; MOLYBDENUM; NICKEL-CHROMIUM STEELS; NICKEL STEEL; NITRIDING; STAINLESS STEEL; STEELS, ALLOY; TOOL STEEL; TUNGSTEN STEEL; VANADIUM STEEL; WIRE; and other specific headings.

STEEL CONSTRUCTION. Steel fabrication of buildings and bridges has been a natural consequence of progress in the production of steel itself and of improved steel alloys. Even before the development of the Bessemer and open-hearth steel-making processes in the latter half of the 19th century, which resulted in a rapidly expanding supply of cheap steel, cast and wrought iron had been used to a limited extent in bridge and building construction. As early as 1800 a technique had been developed for passing iron, heated to a plastic state, between rolls to form structural shapes. These shapes were small, however, and of a

profile that limited their structural usefulness. It was not until 1853 that rolled beams of wrought iron were used in floor construction, in the Cooper union in New York. Unlike the familiar I-beams of modern manufacture, these beams resembled in profile an inverted modern railroad rail. By the time structural steel became commercially available in the early 1880s, the basic structural shapes—I-beams, channels and angles—had proved their adaptability to the builder's needs. Combined with plates of various widths and thicknesses they could be fabricated into efficient members of any required size and strength, and these members could be joined together in the field, using hot-driven rivets.

Steel-Framed Buildings.—In 1889 the Eiffel tower, designed as a latticed girder with a complete system of vertical diagonal bracing, was erected in Paris to a height of 984 ft. (1,056 ft. high from the base of the structure, including a 55-ft. television mast added later). Since it was an open steel-framework construction, the total wind pressure against it was much less than it would have been against the sides of an enclosed building of the same height. This tower was to remain the highest man-made structure for more than 40 years. During that time other systems of bracing for tall enclosed buildings were evolved, depending in part upon the use of diagonal bracing, which in effect formed upended cantilever trusses, and in part upon the bending stiffness of all the beams and columns, joined to one another by means of very rigid connections.

Advent of the skeleton frame, with exterior curtain walls supported at each floor, released building construction from any height limitation due to the excessive thickness of masonry-bearing type walls. The most spectacular examples of this type of construction have been so-called skyscrapers, originally built to utilize effectively expensive urban real estate. (*See* ARCHITECTURE.) The Home Insurance building, erected in Chicago in 1885 and torn down in 1931, has been generally acclaimed the forerunner of the modern skyscraper. It was the first multistoried building having masonry walls supported on a complete structural frame. The columns of the Home Insurance building were cast iron. The first five floors were framed with wrought-iron beams, but the next five were framed in steel, which had only just become available and offered an attractive cost saving. In 1889 two more floors, framed in steel, were added. By mid-20th century hundreds of skeleton steel frame buildings higher than the Home Insurance building had been erected in the United States.

In 1931 the 102-story Empire State building was completed in New York, exceeding the height of the Eiffel tower by 266 ft. A 222-ft. television antenna mast, added in 1950, increased this height to 1,472 ft. This record height was exceeded in 1954, however, with the erection of a television antenna tower near Oklahoma City, Okla. Guyed with wire rope to provide lateral support, this structure rose 1,572 ft. above its ground-level foundations.

A number of interesting examples of steel dome construction can be cited. As early as 1894, for example, an exhibition building of spectacular clear span was constructed at Lyon, Fr. Its central portion consisted of a dome springing from a 361-ft.-diameter circle at ground level to an apex 180 ft. above ground. Encircling this large unobstructed area was a 131-ft. bay, having a roof that sloped down and away from the central dome. More recently, the Dome of Discovery at the Festival of Britain in London (1951) utilized a steel tension ring 342 ft. in diameter, resting 48 ft. above the ground on steel supports battered outward slightly to form a 365-ft.-diameter circle at ground level. Both these structures were subsequently razed. A 332-ft.-diameter dome formed the roof of the Civic centre at Charlotte, N.C., constructed in 1955. The principal framing consisted of 48 radial steel ribs, the outward thrust from which was taken by a circular steel tension ring. Approximately the same unencumbered ground area was provided within the 320 × 386-ft. elliptical Westfalenhalle erected in Dortmund, Ger., in 1952. In this case the roof was supported on steel ribs cantilevered inward from free-standing peripheral columns.

The aviation industry had been responsible for the construction of a number of very large buildings having wide areas free of supporting columns. Notable among these were the aircraft assembly building at Bristol, Eng., and the maintenance hangar at Kelly air force base, San Antonio, Tex. The former, completed in 1948, provided three adjoining clear aisles 331 ft. wide and a clear height of 65 ft., serviced by overhead cranes of 12 tons' lifting capacity. Latticed box frames provided the necessary support for the roof and cranes. The maintenance hangar at Kelly field covered a total area of approximately 23 ac., more than half of this in a single 300 × 2,000-ft. bay.

BY COURTESY OF (LEFT) AMERICAN INSTITUTE OF STEEL CONSTRUCTION, INC., (RIGHT) BENJAMIN SONNENBERG

THE EMPIRE STATE BUILDING, NEW YORK CITY, DURING THE EARLY STAGES OF ITS CONSTRUCTION (LEFT) AND AS IT LOOKED IN 1956 AFTER THE ADDITION OF AN ANTENNA MAST

Steel Bridges.—The first all-steel bridge, crossing the Missouri river at Glasgow, S.D., was erected in 1878, a century after the first iron bridge had been constructed at Coalbrookdale, Eng. Since that time steel has been used in thousands of bridges, employing various structural systems (*e.g.*, suspension cables, cantilevers and arches) to span great distances without intermediate support. The longest single span (4,200 ft.), that of the Golden Gate bridge in San Francisco, was opened to traffic in 1937. It is suspended from two 36½-in.-diameter steel cable catenaries, supported on steel towers 746 ft. above water level. At least four other suspension bridges have been built with spans longer than any supported by other structural systems.

The two record-making 1,700-ft. spans of the Firth of Forth bridge (Scot.), completed in 1889, each consist of a 350-ft.-long centre truss supported on 675-ft. cantilever arms. This cantilever-type construction was employed 28 years later to establish a new record with the 1,800-ft. main span of the Quebec bridge in Canada.

The Kill van Kull (Bayonne) bridge in New York city, spanning 1,652 ft. between masonry piers, established a record as the longest single steel arch. Completed in 1931, it exceeded in length by only a few feet a similar structure spanning Sydney harbour (Austr.), opened to traffic in 1932.

A bridge completed in Belgrade, Yugos., in 1956 contained the longest single-plate girder. Continuous for three spans, the load on its 856-ft. centre portion was largely counterbalanced by the weight of the 246-ft. end spans.

Technological Advances.—The use of welding made possible the introduction of lighter fabricated structural members. For example, open-web (latticed) steel joists were developed for use as floor beams and roof purlins in lieu of heavier rolled shapes. In Europe whole structures at mid-20th century utilized steel pipe and seamless tubing, welded together to form trusses, latticed columns and joists and supplementary bracing. Welding was also used in the fabrication of heavier structural members and frames, frequently from flat plate. Neater connections being possible,

exposed steelwork itself became a medium of architectural expression. High-strength bolts as a substitute for rivets in field connections meantime became popular in the United States.

Because of mass production techniques, steel floor and wall panels and roof decking came to play a major role in building construction. Cold-formed from light-gauge sheets, these products could span considerable distances with substantial saving in dead weight. Deeply corrugated for strength and stiffness, and backed up with a flat sheet, the cells of the panels thus formed were utilized as underfloor ducts for electrical wiring. Sandwichlike panels, consisting of one smooth and one ribbed or fluted sheet separated by a core of insulating material, were used extensively for the exterior walls of industrial buildings.

The façades of a number of modern office buildings consisted entirely of glass windows and stainless steel panels bolted to the steel building frame. These panels, press-formed from light-gauge sheets to various designs and backed up with thermal insulation about 4 in. thick, were lighter, more durable and less conductive of heat than older types of curtain wall construction. In one building alone the equivalent of 11 ac. of stainless-steel paneling was used in the exterior walls.

Approximately one-sixth of all the steel produced in the world is taken by the construction industry; only two-thirds of this amount, comprising those steel products required for structural purposes, comes within the category loosely referred to as steel construction. Of this amount, items fabricated from hot-rolled carbon steel structural shapes, bars and plates (as for example, bridges, building frames, supports for industrial equipment, towers and tanks) account for a larger tonnage of steel than all others put together.

BIBLIOGRAPHY.—A. C. Bossom, *Building to the Skies: the Romance of the Skyscraper* (New York, London, 1934); J. F. Baker, *Steel Skeleton*, 2 vol. (New York, London, 1954–56); J. E. Lothers, *Design in Structural Steel*, 2nd ed. (London, 1957); E. H. and C. N. Gaylord, *Design of Steel Structures* (New York, London, 1957); T. C. Shedd, *Structural Design in Steel* (New York, London, 1934); L. E. Grinter, *Design of Modern Steel Structures* (New York, 1941; London, 1942); H. Sutherland and H. L. Bowman, *Structural Design* (New York, London, 1938); C. D. Williams and E. C. Harris, *Structural Design in Metals* (New York, 1949). (T. R. H.)

STEELE, SIR RICHARD (1672–1729), Irish man of letters in the reign of Queen Anne, is inseparably associated in the history of literature with his personal friend Addison. The two were born in the same year. Steele, the senior by less than two months, was baptized on March 12, 1672 in Dublin. His father, also Richard Steele, was an attorney. He died before his son had reached his sixth year, but the boy found a protector in his maternal uncle, Henry Gascoigne, secretary and confidential agent to two successive dukes of Ormonde. Through his influence he was nominated to the Charterhouse in 1684, and there first met with Addison. Five years afterwards he proceeded to Christ Church, Oxford, and was a postmaster at Merton when Addison was a demy at Magdalen. Their schoolboy friendship was continued at the university, and probably helped to give a more serious turn to Steele's mind than his natural temperament would have taken under different companionship. Without waiting for a degree he volunteered into the army, and served for some time as a cadet "under the command of the unfortunate duke of Ormond" (*i.e.*, the first duke's grandson, who was attainted in 1715). This escapade was made without his uncle's consent, and cost him, according to his own account, "the succession to a very good estate in the county of Wexford in Ireland." He found a patron in Lord Cutts, to whom he dedicated a poem on the funeral of Queen Mary (1695). Cutts took the gentleman trooper into his household as secretary, made him an officer in his own regiment, and ultimately procured for him a captaincy in Lord Lucas's regiment of foot. A duel which he fought with Captain Kelly in Hyde Park in 1700, and in which he wounded his antagonist dangerously, inspired him with the dislike of duelling that he showed to the end of his life. Steele won William III.'s favour (too late to be of practical use) by a timely reference to him in *The Christian Hero*, his first prose treatise, published in April 1701. Steele complained that the reception of *The Christian Hero* by his comrades was not so respectful; they persisted in trying him by his own standard, and would not pass "the least levity in his words and actions" without protest. He therefore determined to clear his character of the charge of undue solemnity by writing a comedy, *The Funeral* (1701). With this play he began his work of reconciling wit, good humour and good breeding with virtuous conduct which he was afterwards to accomplish in the famous essays in the *Tatler* and the *Spectator*. In his next comedy, *The Lying Lover; or, the Ladies' Friendship* (1703), based on Corneille's *Menteur*, Steele's moral purpose was directly avowed, and the play, according to his own statement, was "damned for its piety." *The Tender Husband*, an imitation of Molière's *Sicilien* (in April 1705), was more successful. It was seventeen years before Steele again tried his fortune on the stage with *The Conscious Lovers*, the best and most successful of his comedies, produced in December 1722.

The Tatler and the Spectator.—In 1707 Steele was given the office of gazetteer. The *Gazette* gave little opening for the play of Steele's lively pen, his main duty, as he says, having been to "keep the paper very innocent and very insipid"; but the position gave him insight into journalism. The *Tatler* made its first appearance on April 12, 1709. It was partly a newspaper, a journal of politics and society, published three times a week. Steele's position as gazetteer furnished him with special advantages for political news, and as a popular frequenter of coffee-houses he was at no loss for social gossip. He gradually introduced into the *Tatler* as a special feature essays on general questions of manners and morality.

Steele was always in want of money. He had £300 a year from his gazetteership (paying a tax of £45), £100 as gentleman waiter to Prince George, £850 from the Barbadoes estates of his first wife, a widow named Margaret Stretch, and some fortune by his second wife—Mrs. Mary Scurlock, the "dear Prue" of his charming letters. But he lived in considerable state after his second marriage, and before he started the *Tatler* was borrowing money. The assumed name of the editor was Isaac Bickerstaff, but Addison discovered the real author in the sixth number, and began to contribute in the eighteenth. The success of the *Tatler* was established before Addison joined him, and Addison contributed to only forty-two of the two hundred and seventy-one numbers that had appeared when the paper was stopped, obscurely, in January 1711.

Only two months elapsed between the stoppage of the *Tatler* and the appearance of the *Spectator*, which was the organ of the two friends from March 1, 1711 to Dec. 6, 1712. Addison was the chief contributor to the new venture, and the history of it belongs more to his life. Nevertheless, it is to be remarked as characteristic of the two writers that in this as in the *Tatler* Addison generally follows Steele's lead in the choice of subjects. The first suggestion of Sir Roger de Coverley was Steele's although it was Addison that filled in the outline.

Steele projected various journals in the next years, most of them, such as the *Englishman* (1733), and the *Reader* (1714), *Town Talk*, *Tea Table* and *Chit Chat*, very short lived; though the *Guardian* (1713) had 176 numbers. Steele's most famous political paper, the *Plebeian* (1718) embroiled him in a controversy with Addison. A pamphlet, *An Apology for Himself and his Writings* (1714) is important biographically.

Later Life.—The fortunes of Steele as a zealous Whig varied with the fortunes of his party. Over the Dunkirk question he waxed so hot that he threw up a pension and a commissionership of stamps, and went into parliament as member for Stockbridge to attack the ministry with voice and vote as well as with pen. But he had not sat many weeks when he was expelled from the house for the language of his pamphlet on the *Crisis*, which was stigmatized as seditious. The *Apology* already mentioned was his vindication. With the accession of the House of Hanover his fortunes changed. Honours and substantial rewards were showered upon him. He was made a justice of the peace, deputy-lieutenant of Middlesex, surveyor of the royal stables, governor of the royal company of comedians—the last a lucrative post—and was also knighted (1715). After the suppression of the Jacobite rebellion he was appointed one of the commissioners of forfeited estates, and spent some two years in Scotland in that capacity. In 1718 he obtained a patent for a plan for bringing salmon alive from Ireland. Differing from his friends in power on the question of the Peerage Bill he was deprived of some of his offices, but when Walpole became chancellor of the exchequer in 1721 he was reinstated. With all his emoluments however the imprudent, im-

pulsive, ostentatious and generous Steele could never get clear of financial difficulties, and he was obliged to retire from London in 1724 and live in the country. He spent his last years on his wife's estate of Llangunnor in Wales, and, his health broken down by a paralytic seizure, died at Carmarthen, Sept. 1, 1729.

A selection from Steele's essays, with a prefatory memoir, has been edited by Austin Dobson (1885; revised 1896). *See* also *Selected Essays* (1907) ed., L. E. Steele; and *Letters of Richard Steele* (1928), ed., R. B. Johnson. Dobson contributed a biography to Andrew Lang's series of *English Worthies,* in 1886. In 1889 an exhaustive life was published by G. A. Aitken, who also edited Steele's plays (1898) and the *Tatler* (1898).

STEELS, ALLOY. It has been said that alloy steel is "a carefully made steel with sufficient alloying metals added to it to modify usefully the properties of a plain carbon steel." An alloying element or compound is one that is more or less soluble in the major metal and that participates in physical or chemical reactions at various temperatures and thus modifies the properties. Alloying metals may be various. The amount of alloying element added varies enormously, ranging from very low (0.0005% boron or 0.1% vanadium) to very high (20% chromium or more).

It is almost as hard to draw a clear line of demarcation between alloy steels and steels as it is between steel and iron. Diverse definitions have been set up legally for customs purposes, and commercially for statistical but more especially for pricing purposes.

Iron is produced from its ores by commercial processes that leave a number of impurities, either alloyed with it or included as insoluble nonmetallic particles. Carbon (in the chemical compound iron carbide, cementite, Fe_3C) is the principal alloying element in iron, proportionately increasing its strength and hardness, especially after correct heat treatment. Only a little iron carbide (carbon) is needed to convert iron into a hardenable steel, so a metallographer's definition of carbon steel might well be "a malleable alloy of iron and iron carbide" (malleable being inserted to distinguish plain carbon steel from cast iron, which is a nonmalleable alloy of the same substances). Fractional percentages of the alloying metals manganese and silicon also must be present in commercial plain carbon steels, and it is difficult to say, as these incidental elements increase, just where a carbon steel stops and a manganese alloy steel (or, more briefly, a manganese steel) and a silicon steel start. Likewise alloy steel practice consists of far more than adding alloying metals to the liquid steel. The raw material is selected for a low content of deleterious impurities, the furnace practice is extended to include a careful refining period, additions are made to control the grain size (which, in turn, controls the toughness), all surface defects are cut out of ingots or billets, inspection is much more stringent—to mention some of the major differences.

The earliest alloy steels used those alloys then available; after World War I useful steels were developed with alloying metals that were formerly unavailable or rare. (1) Chromium alloy steels were made commercially into hard grinding equipment by Julius Baur in 1869, and made their debut in structures in the arch members of the Eads bridge (1874) across the Mississippi river at St. Louis, Mo., and they now enter many steels for machine parts. The high chromium steels are corrosion and oxidation resistant (*see* STAINLESS STEEL). (2) Nickel, now another common constituent of alloy steels, was first incorporated in steels by Marabeau; the properties were described in 1889 by James Riley. Nickel steel was adopted by the U.S. navy for armour plate in 1891; it is said that its first application in the mass-production industries was some armour plate rolled into small bars for bicycle chains. (3) Tungsten is the essential element in the air-hardening tool steel of R. F. Mushet (1871). This was the precursor of modern high-speed tool steel. It is also used in magnets and steels for extra high-temperature service. (4) Molybdenum, its chemical cousin, was unavailable until large U.S. deposits began producing in Colorado about 1918. In addition to being a substitute for tungsten, molybdenum is also a most important element, in fractional percentages, in engineering alloy steels. (5) Manganese formed the first high alloy steel, 10% being added by Robert (later Sir Robert) Hadfield in 1882 to form a very tough steel of unique properties, largely used in items such as steam-shovel dipper teeth that must resist tremendous battering. Low manganese steels were the product of 1920–30 and were substituted in many uses for the more expensive chromium and nickel steels. (6) Silicon steels, especially useful for electrical transformers and other equipment, were devised in 1889 by Hadfield. (7) Vanadium and chromium-vanadium steels were investigated in England by Arnold and J. Kent-Smith at the turn of the century; the latter, at Canton, O., in 1907 made the first alloy steel by the open-hearth process for the Ford Motor Co., and this company pioneered the use of alloy steels in mass production. (8) Steels containing a fractional percentage of copper, for atmospheric corrosion resistance, also are a pre-World War I product.

The eight common elements referred to above, in various percentages and combinations, to say nothing of an equal number of less commonly used alloying metals, may be used to manufacture an infinite variety of steels. Since the beginning of the 20th century, literature about them has accumulated in a formidable mass. A group of metallurgists led by Edgar C. Bain systematized much of the existing information. Commercial requirements of both producers and consumers also demanded a limitation of types. The American Society of Automotive Engineers led the way in 1911 by adopting chemical specifications for 7 carbon and 11 alloy steels, a list extended jointly with the American Iron and Steel institute in 1937 to 107 steels alloyed with manganese, nickel, chromium, molybdenum and vanadium (singly or in combination), including 17 boron-treated steels using smaller amounts of alloys, plus 69 alloy steels specified by their hardenability as well as their chemical composition. There are also 71 standard varieties of stainless steels, wrought and cast. Many large consumers and the United States government have their own lists, and the total (exclusive of tool steels and high-alloy steels) doubtless exceeds 1,000.

Bain's scientific classification gave logic to the practical conclusion that many of the varieties are interchangeable. Users are slow to discard a steel that has served well in the past, fits into an existing heat treatment schedule and is approved by the staff, but economics slowly forces the adoption of the cheapest suitable raw material. This generally means steels of lower and lower alloy content. Such a change was forced on all users during World War II by the disruption of alloying metal production and the priority demands for arms and ordnance. In the United States it brought low-alloy steels whose hardenability was largely increased by adding a very small amount of boron (about ¼ oz. per ton). Aside from war stringencies, the engineering alloy steels made in the U.S. contained nearly 3% of total alloy in 1920 and only slightly over 1% in 1955.

While U.S. industry relies upon chemical composition (plus hardenability, as described below) to characterize its alloy steels, Europeans generally depend more on the strength of a steel after normal heat treatment, and give the steelmaker more latitude in the selection of an appropriate analysis.

General Influence of Alloys.—A statement frequently heard is that alloy steels are valuable chiefly because they may be hardened and strengthened by heat treatment without incurring the brittleness associated with hardened plain carbon steel. Further it is frequently said that to be economically justified, the alloy steels should be used in the quenched and tempered state. It will be shown below that the true function of alloys in the engineering or machine steels is to get high strength and toughness, or hardness and toughness, simultaneously, either more easily than they can be had in carbon steels, or in sizes and shapes in which they cannot be had in carbon steels. Certain alloys also impart unique properties to carbon steels (such as high chromium for corrosion resistance or high manganese for battering abrasion), but in defense of the plain carbon steel it may be said that high strength, hardness and toughness are not the sole property of alloys. Thus, the intensity of hardening primarily depends on carbon content, and no alloy steel has yet been discovered that can be made harder than properly quenched carbon steel containing 0.75% or more carbon (measured as C-67 by the Rockwell instrument, wherein hardness number is inversely proportional to

size of indentation by a loaded diamond point).

Since steel, as it has been defined, is essentially an alloy of iron and iron carbide, the alloys present in it can be located in one or more of three possible phases or microconstituents: (1) in solid solution in the iron—the alpha iron or ferrite; (2) combined with carbon, either replacing some iron in the iron carbide or forming a carbide of its own; (3) combined with oxygen, nitrogen or sulfur in nonmetallic, insoluble inclusions. Rarely does one alloy act in one way alone; partition among (1), (2) and (3) depends upon the amount of alloying element or elements, amount of carbon and temperature. However, some alloys, such as nickel and manganese, act notably to form solid solutions and strengthen the ferrite. Others act notably to enter and form carbides, as vanadium and molybdenum. Titanium is the outstanding example of the elements in category (3).

Strong Structural Steels.—Alloys falling in category (1) are particularly useful in improving the mild steels used for plates and structural shapes. These steels are usually fabricated by the user as they are delivered to him from the rolling mill, without further heat treatment. In fact, they must be rather insensitive to heat treatment, else flanging and especially welding will produce hard regions ruinous to the tools that punch or cut them, or regions lacking in required ductility. Extra strength and correspondingly lighter sections, such as are desirable for a long-span bridge or for railway car bodies, can, of course, be had by increasing the carbon content, but this aggravates the problem of hardness. The alternative is to add alloys that enter into solid solution with the iron and so strengthen the ferrite, which, by the way, is the major microconstituent of these low carbon steels. Bain lists these alloying elements in the order of their influence as ferrite strengtheners from strongest to weakest, as phosphorus, silicon, manganese, nickel and that portion of the chromium that is not combined with carbon. Each of them is used and in various combinations in commercial tonnage steels. Nickel steel was undoubtedly the first, used for long-span bridges, 3½% of that element greatly improving the tensile properties and doubling the impact strength. A cheaper alloy, known as structural silicon steel, is much used by structural engineers. An early type of such steel, used by constructors of the pre-World War I "Mauretania," is also shown in Table I, although the British admiralty now prefers D-steel. These strong steels, when weldability is important, must have low carbon content. Reductions in section thickness also demand a better degree of corrosion resistance than is afforded by the normal addition of 0.2% copper.

After 1940 there was a notable trend toward welded rather than riveted structures (ships, tanks, penstocks, bridges), and the problem of steel's weldability correspondingly arose. An early application of alloy steel, used for its strength in welded spheres containing hydrocarbons kept liquid at very low temperature, led to a disaster costing 128 lives in Cleveland, O., in 1944. Nor is plain carbon steel in rigid structures immune to catastrophic cracking, as witness the 25 or more major failures of welded "Liberty" ships during 1942–45 and 14 canal bridges in Belgium in 1947–50. Extensive investigation determined that for welded ships a fine-grained, low-alloy steel containing less than 0.2% carbon, 1% to 1.35% manganese and no more than 0.04% silicon has good toughness and the ability to stop the spread of an incipient crack at subzero temperatures. This crack-stopping ability is the characteristic necessary for safe performance. It is related to a low transition temperature, mentioned below.

Hardenability of Alloy Steels.—Alloys are, of course, of most importance to heat-treatable steels, for an alloy bar will harden under similar circumstances much more deeply than a comparable carbon steel bar. When steels are heated above their critical points—a range of about 725° to 825° C.—the crystal structure of alpha ferrite is transformed into a new one called gamma iron, which has far greater solubility for carbide. This high-temperature solid solution of the carbides in gamma iron is called austenite. On slow cooling the reverse action takes place at about the same temperature range, and the annealed, fairly soft and ductile steel is reconstituted, except that the process may have refined the grain. On more rapid cooling the precipitation of carbide from austenite is delayed—a little time is necessary for the action to incubate and then to proceed to completion—and occurs at a somewhat lower temperature. On very rapid cooling, as when quenching a small piece in cold water, austenite is trapped unchanged down to about 150° C. but then changes almost completely and almost instantaneously to a very hard structure called martensite, which seems to be alpha iron (ferrite) in which the iron carbide is dispersed in almost molecular particles. This is the essence of the hardening action. (For a fuller account of the metallographic change *see* IRON AND STEEL.)

Plain carbon steels have a low hardenability, in the sense that if a 1½-in. round bar is quenched in cold water only the outer layers of the steel are cooled rapidly enough to transform at low temperature into martensite; the inner core, cooling more slowly, is transformed at high temperatures near the critical into a much softer entity, a coarser mixture of laminated ferrite and carbide. The outer surface may be intensely hard, even the maximum possible for the carbon content of the steel, but the penetration of that hardness is low. To harden the bar more deeply, the cooling rate at the surface must be faster, so the deeper regions of the steel may cool rapidly enough to escape the high-temperature transformation; even the most drastic quenching, as in brines or water sprays, may be insufficient for complete hardening.

Fortunately, the presence of all the alloying metals except cobalt slows up both the solution of carbides in the austenite on heating, and their precipitation on cooling. This means that alloy steels do not need to be cooled nearly so rapidly in order for the austenite to be trapped and transformed at about 150° C. into hard martensite. Alloy steels therefore have a higher hardenability. Given an equal quench and an equal-sized bar, the alloy bar will harden much more deeply than the carbon steel bar of the same carbon content; or it may be satisfactorily hardened in a much milder quenching bath, as in oil or even in still air, with correspondingly less warpage and danger of cracking. As an example, the presence of 1¼% nickel and 0.6% chromium doubles the hardenability of a 0.45% carbon steel. Bain's list of the relative potency of alloys in this respect follows: very strong, vanadium; strong, molybdenum and tungsten; moderate, chromium, manganese, silicon; mild, copper, nickel, phosphorus. (At the top of this list would now be placed boron.)

Toughness of Alloy Steels.—Certain of the alloys have still another beneficial effect—that of refining the grain. According to the laws of crystallography, certain favoured austenite grains should grow at the expense of their neighbours, with the result that the hot steel should become coarser grained with longer and longer stay at higher and higher temperature. But first a considerable time is necessary to dissolve complex and sluggish carbides in the austenite; many seem to have a central particle that resists dissolution, and these refractory particles disseminated through the metal effectively prevent the natural grain growth of the solid solution. The advantage is that a fine-grained austenite, when quenched and tempered, has a very desirable degree of toughness. Bain's list of grain refining elements, arranged according to their effectiveness, is: titanium (very strong), vanadium, molyb-

TABLE I.—*Low-Alloy, High-Yield Steels for Riveted and Welded Structures*

Steel	Composition (%)				Tensile strength	Yield point	Elongation in 8 in.
	C	Si	Mn	Others			
Good structural	0.20	0.15	0.50	..	60,000–68,000	40,000–47,500	30%–25%
Nickel	0.33	0.06	0.60	3.25 Ni	105,000	63,000	20%
Structural silicon	0.30	0.25	0.80	..	85,000–90,000	50,000	>20%
"Mauretania"	0.25	1.10	0.75	..	92,000–105,000	65,000	30%–25%
D-steel	0.33	0.12	1.25	..	83,000–96,000	56,000	17%
"Cor-ten"	<0.10	0.75	<0.40	0.15 P, 1.0 Cr, 0.40 Cu	65,000–75,000	50,000–60,000	25% in 2 in.
"Double strength"	<0.12	....	0.75	1.0 Cu, 0.60 Ni, 0.20 Mo	75,000	65,000	30%
"Carilloy T-1"	0.15	....	1.00	0.90 Ni, 0.50 Cr, 0.50 Mo, 0.35 Cu + B, Al, V	130,000	90,000	..

denum, tungsten, chromium and manganese (mild).

Aluminum, added to the metal in proper small quantity in a properly deoxidized steel, apparently reacts with traces of oxygen and nitrogen dissolved in the steel and precipitates out as a cloud of insoluble particles in the solidifying steel. These inclusions are even more effective obstructions to austenite grain growth. Such fine-grained steels are popular in the United States. They are especially useful in low carbon, carburizing steels. In the carburizing operation, machined steel parts can acquire a high carbon skin or case, say 0.05 in. deep, in five hours' exposure to a highly carbonaceous gas at 1,700° F. Even after such a long and high heating the austenite remains so fine in grain size that the parts can be quenched directly from the carburizing heat to harden the high carbon case, the inner portions remaining fine grained and tough.

The toughness of alloy steels, or the ability to absorb rapid blows, is also improved by the circumstance that those containing the carbide-forming elements listed above may be tempered or reheated after quenching to considerably higher temperatures than carbon steels of the same carbon content and still retain the hardness of the plain carbon steel, quenched and tempered at a lower temperature. Evidently the sluggish diffusion of alloying elements, trapped in solution in the freshly quenched martensite, requires a higher temperature and a longer time before the complex carbides can re-form, in comparison with the formation of simple iron carbide (Fe_3C) by diffusion of trapped carbon. Since the toughening due to tempering is largely a matter of relieving the high internal stresses in freshly quenched steels, and since this stress relief is greater at higher temperature, it can be seen why alloy steels, after proper heat treatments, may give metal that is strong, hard and unusually tough.

Toughness, as the term is used in the preceding paragraph, is measured by a single-blow Charpy impact test, wherein a 10-mm. square bar, with carefully machined V-notch on one side, is struck by a swinging hammer, and the energy absorbed in breaking the small notched bar is measured. Various combinations of alloys respond differently to this test when quenched and tempered to the same hardness (and strength). For example, A.I.S.I. 5145 (0.45% C and 0.8% Cr) has 55 foot-pounds (ft.-lb.) at 70° F. and 30 ft.-lb. at −100° F. A.I.S.I. 4340 (0.4% C. 1.75% Ni, 0.8% Cr and 0.25% Mo) has 70 ft.-lb. at 70° F. and 60 ft.-lb. at −100° F. (Both these had equal hardness: 35 on the Rockwell C-scale.) Such circumstances indicate why steels containing considerable alloy are favoured for use at low temperatures and for shock-resisting pieces such as axles and cannon. Generally speaking, the carbon steels have high transition temperatures, around 70° F. (above which high impact strengths and fibrous fractures occur, and below which low impact strength and rapid extension of cleavage or brittle fractures occur). Low-alloy steels, like the ship plate mentioned above, have a transition temperature of −10° F. Higher engineering alloy steels such as A.I.S.I. 4340 have transition temperatures of −100° to −200° F. "Carilloy T-1," the weldable steel noted in Table I, has a transition temperature of −150° F. The austenitic stainless steels containing 18% Cr and 8% Ni have transition temperatures around that of liquid air.

The sluggish changes in the microstructure of the complex alloy carbides during heat treatment are utilized in steels for high-temperature, high-pressure equipment. Stable carbides enable the hot metal to resist creep at high temperature. The American Society of Mechanical Engineers' boiler code tentatively proposed a working stress of 5,750 lb. per square inch for plain carbon steels at 900° F. and 10,000 lb. per square inch (nearly double) for the same steels plus 0.5% molybdenum. To increase corrosion resistance, chromium is also added; 1.25% Cr and 0.5% Mo, or 2.25% Cr and 1% Mo are favoured compositions for pressure vessels serving at moderately high temperatures.

Tool Steels.—Since alloying elements, either in solid solution with iron or in the form of complex carbides, tend to strengthen and harden steel and also to bear up under heat without softening as readily as a plain carbon steel, cutting and forming tools are made of a family of complex alloy steels each suited for its particular duty. (*See* TOOL STEEL.) Excellent hardenability is desirable for forging or hot forming dies, for example, which must often be hardened in large blocklike masses; they must also be hard, to resist wear, and tough, to withstand blows. The tip of a lathe tool, cutting at a high rate of speed, is heated almost red hot by friction from the chip; it must retain useful hardness. Generally speaking, tungsten (up to 18%) and molybdenum (a useful substitute, at least in large part) are ordinarily depended upon for retention of hardness even when hot. Chromium (up to 12%) and carbon (up to 2%) give maximum hardness and wear resistance at ordinary temperature. Vanadium is the traditional element for refining the metallic grain size and thus promoting shock resistance. Manganese (up to 3%) is the main alloying constituent in dies that must be hardened without measureable change in size or shape, —the so-called nondeforming steels.

Interchangeability of Alloy Steels.—A definite combination of strength, hardness and toughness may be secured from any one of a large number of alloy steels, provided the size is not too large and the heat treatment properly adjusted. How true this is may be proved by transmission drive shafts used in six U.S. automobile plants in 1947. One firm used 0.6% carbon steel, oil quenched; another 1.75% nickel, 0.25% molybdenum, 0.2% carbon steel, carburized; two others used 1% nickel, 0.6% chromium, 0.75% manganese, 0.4% carbon steel, oil quenched; another used 0.25% carbon, 0.8% manganese, 0.2% molybdenum steel, carburized; and the last one used 1% chromium, 0.75% manganese, 0.35% carbon steel, oil quenched.

It will also be noted that the above alloys are all low alloys, and this represents an undoubted trend in practice for machine parts that are to operate at normal temperatures, in both the United States and Great Britain. The trend is especially strong in the smaller sizes of forgings or machined gears, where adequate penetration of hardness may be secured by not too drastic quenches. For devices that must operate at extremes of temperature and for massive forgings, steels with higher content of alloys and long-continued heat treatments to permit the sluggish reactions are necessary.

Large Forgings.—A rough forging for a 240-mm. howitzer weighs 30,000 lb., and the original ingot would probably weigh 75,000 lb. Rotor forgings for very large electrical generators may weigh over 100,000 lb. In such massive parts, good steelmaking practices are of prime importance; the basic process is used, since it can drive the impurities sulfur and phosphorus down to 0.015% each. To avoid tiny internal bursts or flakes, the gas content, especially hydrogen, must be reduced to as near zero as possible; in Germany, vacuum casting is resorted to. Extreme homogeneity and cleanliness will prevent small segregations of alloy within the metal that prevent uniform hardening reactions and introduce large internal stresses; nonmetallic inclusions ("dirt") are also dangerous stress raisers. U.S. specifications for gun tubes say nothing about chemical composition. The ordnance manufacturer can choose the alloy he thinks best for the required yield strength, reduction of area at fracture (ductility) and Charpy impact value at −40° F. (toughness). Test pieces are cut in a direction transverse to the axis of the gun so as to prove its ample strength against bursting. Principal stresses in generator rotors are centrifugal, from the high speeds of 3,600 r.p.m. A steel of 0.3% C, 2.5% Ni, 0.5% Mo and 0.5% Cr or 1% Mn has been widely used in the United States. German practice favours considerably less nickel and molybdenum.

Gear Steels.—As a specific application of the above general principles, the problems of gear manufacture may be considered, for the choice of steel and its treatment depend very intimately on the design of the part and its intended duty. The high-duty herringbone gear sets used on turbine-driven propeller shafts, for example, are marvels of precision; they are of cast alloy steels heat treated to relieve internal stresses and insure dimensional stability as well as induce as much hardness for wear resistance as is consistent with machinability. The steels contain 0.5% to 0.6% carbon, but alloys are necessary to give proper depth of hardening and fine grain size. Examples are 1.25% manganese plus one or more of the elements molybdenum (0.3%), nickel (1%),

chromium (0.5%) and vanadium (0.15%). In large sections a quenched and tempered chromium-nickel-molybdenum cast steel will have the following physical properties indicating medium strength and hardness with great toughness; yield point 105,000 lb. per square inch; tensile strength 135,000 lb. per square inch; elongation 20% in 2 in.; reduction of area 64%; Rockwell hardness C-32.

In the smaller gears utilized for automobile transmissions and rear axle drives the mass effect or deep hardening effect is not so necessary. Requirements are very diverse and exacting, and these, together with the desire to save weight, and therefore to have gears as small as possible, have preserved this field to alloy steels. The original conception of a gear tooth was that it should have a hard surface to resist wear and a tough centre or core to resist impact. Following this idea, gears were cut from forged and annealed blanks of low carbon alloy steel (less than 0.2% carbon); the machined gears were then carburized by heating in contact with carbonaceous solids or gases until their surfaces contain up to 0.9% or 1% carbon for a depth of 0.03 to 0.05 in. Subsequent heat treatment was at times rather intricate, consisting of a high-temperature quench to refine the low carbon core, a lower-temperature quench to harden properly the high carbon surface, and a tempering to improve the toughness. Steels widely used in this way included those shown in Table II. Representative properties of the core after an oil quench suitable to harden the carburized case are as shown. Evidently here is a variety of steels suitable for any designer's requirements, where he balances cost (alloy content, machinability and heat treatment schedule) against strength desired to carry the loads on the gear tooth, and toughness against impact loadings.

Carburizing all these analyses gives an outer surface containing 0.9% to 1% carbon, and since surface hardness of the quenched gear depends on carbon, rather than alloy, content, all these teeth can be made file hard—that is, too hard to be bitten by a testing file whose hardness is Rockwell C-65. Experience has also proved that maximum resistance to abrasion is secured from such high carbon cases on steel containing chromium or molybdenum carbides. One serious objection to the practice outlined above is the cost of heat treatment after carburization, and this has been attacked by the use of fine-grained steels that may be quenched directly from the carburizing heat and still possess a fine structure and considerable toughness. This practice was fairly common with the 1% chromium, 0.15% vanadium steels, which are inherently fine-grained, before the special fine-grained steels were regularly produced by the steel mills. One much used type is the nickel-molybdenum steel shown in Table II.

TABLE II.—*Properties of Low-Carbon Cores of Gear Teeth*
(Carburized and oil quenched from temperature shown)

Steel	Quenching temperature	Tensile streogth lb. per sq. in.	Rockwell hardness	Izod† impact ft.-lb.
0.2% carbon (no alloy)	780° C.*	80,000	C–0	70
3.5% nickel	800° C.	160,000	C–35	35
5% nickel	815° and 735° C.	110,000	C–20	75
1.25% nickel, 0.5% chromium	800° C.	155,000	C–34	30
1% chromium, 0.15% vacadium	890° C.	125,000	C–25	80
1.75%nickel,0.25%molybdenum	930° C.	140,000	C–32	35

*Water quench. †Energy of fracture from impact of swinging pendulum is measured.

High-speed transmission gears in automobiles must operate without much noise, and this requires a degree of accuracy of shape on the order of 0.0005 in. Warpage increases with the depth of carburization, so shallow cases were tried to give quieter gears. This was successful, but the hard case tended to flake off, due to fatigue failure of the weaker low carbon metal at junction of case and core. To strengthen this metal it was necessary to increase carbon content, and the result is that in mid-20th century American automobiles the great majority of transmission gears were of steels that used to be called oil quenching gear steels, high enough in carbon and alloy to quench to a file hard surface without carburization. An upper carbon limit is set by the demands of reasonable machinability. Whereas the old oil-quenched gears were heated in molten lead for quenching (in order to avoid any trace of decarburized soft skin), the more modern practice is to hold them in a cyanide bath at 815° C. for about 25 minutes, which is enough to induce a 0.002-in. case of carbides and nitrides, exceedingly hard and wear resistant, or to heat them in a controlled gas of carburizing and nitriding nature.

Austenitic Steels.—Hadfield's cast 10% manganese steel, discovered in 1882, was remarkable in that it did not harden on quenching. It is, in fact, an austenitic alloy, in the sense that the expected change from austenite solid solution to hard martensite does not occur at 200° C., but the quenched austenite is very stable at room temperature. Cobalt, but especially nickel, has similar effects. One result of this is that the most useful of the stainless steels, known as 18–8 and containing 18% chromium and 8% nickel, is unhardenable by quenching.

One characteristic of these commercial austenitic steels is that the austenite is not entirely stable. The small plastic movements in high manganese steel, associated with battering or attempted cutting, permit the martensitic reaction to occur in those restricted regions, causing intense hardness—hence their wear resistance in rough service. This is a favourable reaction, but many of the 18–8 steels are put into petroleum stills operating between about 450° and 850° C., and this temperature gives enough thermal agitation to the atoms so that chromium carbides form from the austenite solid solution, unfortunately not in a cloud throughout the mass to key and strengthen the metal, but localized at the grain boundaries, seriously depleting the surrounding matrix of its chromium necessary for corrosion resistance. This deterioration of 18–8 (even by the heat of welding) is prevented by introducing minor amounts of strong carbide-forming elements, like molybdenum, titanium or columbium. The latter two, especially, form carbides that are so insoluble in austenite that they act as inert nonmetallic inclusions.

Electrical Sheet.—The foregoing account relates primarily to strength of alloy steels. Other properties are often of more importance, such as corrosion resistance (*see* STAINLESS STEEL) and electrical properties. A prime example of the latter is silicon steel transformer sheet—really a very low carbon (0.07%) alloy of iron with 3.25% silicon, plus tiny additions of certain impurities such as nitrogen to control directional crystallization. The silicon goes into solution in the iron and this ferrite with cubic crystallinity has great magnetic superiority in one direction. The problem is to make sheet by the ton in which the tiny crystals—ordinarily oriented randomly—are lined up or regimented. The commercial method was discovered in 1933 by Norman P. Goss—a series of cold reductions each followed by annealing in hydrogen. The great utility of such silicon iron is illustrated by the fact that a 1.5-kva., 7,200-volt transformer made in 1935 weighed 275 lb.; in 1945 it weighed 120 lb. *See* IRON AND STEEL; ALLOYS; and the various types of steels under their own headings; *see* also Index references under "Steels, Alloy" in the Index volume.

BIBLIOGRAPHY.—Edgar C. Bain, *The Functions of the Alloying Elements in Steel* (1939); H. C. H. Carpenter and J. M. Robertson, *Metals*, 2 vol. (1939); E. C. Pigott, *Ferrous Analysis: Modern Practice and Theory*, 2nd ed. (1953); R. A. Grange *et al.*, *Boron, Calcium, Columbium, and Zirconium in Iron and Steel* (1957); W. F. Simmons and H. C. Cross, *Report on Elevated-Temperature Properties of Wrought Medium-Carbon Alloy Steels* (1957). *See* also comprehensive monographs issued by Alloys of Iron Research, New York. (E. E. T.)

STEEL WOOL, a mild abrasive commonly used to clean or polish metals and metallic objects. It is of great practical use in the japanning, plating and molding rooms of factories. In rubbing down metallic articles to attain a fine finish, steel wool achieves much the same results as pumice. Steel wool is also used to clean rusty and tarnished hollow ware.

STEELYARD, MERCHANTS OF THE, Hansa merchants who settled in London in 1250 at the steelyard on the riverside, near Cosin lane, now Ironbridge Wharf. Henry III in 1259, at the request of his brother Richard of Cornwall, king of the Romans, conferred on them important privileges, which were confirmed by Edward I. It was chiefly through their enterprise that the early trade of London was developed, and they continued to flourish till, on the complaint of the Merchant Adventurers in the reign of Edward VI, they were deprived of their privileges.

They succeeded in maintaining a footing in London till 1597.

See Stow, *Survey of London* (1598); Lappenburg, *Urkundliche Geschichte des hansischen Stahlhofes zu London* (1851); Pauli, *Pictures of Old London* (1851); Ehrenberg, *Hamburg und England im Zeitalter der Königin Elizabeth* (1896).

STEEN, JAN HAVICKSZ (1626–1679), after Rembrandt, the most varied and penetrating of the Dutch genre painters, was born at Leiden in 1626. His father was a brewer, and at various times Steen himself was a brewer and tavern keeper. He was a founder-member of the Leiden Guild of St. Luke in 1648. In 1649 he settled in The Hague and married Margareta van Goyen, daughter of the landscape painter. He moved to Delft in 1654 and to Haarlem in 1661. In 1670 he was back in Leiden; he married a second time in 1673. He was buried in Leiden, Feb. 3, 1679.

Steen's master was the Utrecht historical painter Nicolaus Knupfer, but there is little obvious influence of Knupfer even in his occasional history pieces, such as "Samson and Delilah" (Cologne) or the "Adoration of the Shepherds" (Rijksmuseum, Amsterdam). He is first and foremost a genre painter, heir to the "low life" tradition of Adriaen Brouwer and his school. In style and content he looks back on the one hand to the great early masters of genre, Hieronymus Bosch and Pieter Brueghel, and on the other adheres to the polished, colourful manner of his own generation, comparing favourably with Gerard Ter Borch and Gabriel Metsu in his delicacy of touch and his skill in rendering light and texture. However, his typical pictures are scenes of revelry, displaying a huge zest for living and a humour that ranges from subtlety to the crude or grotesque, or quieter scenes that yet pass a satirical comment on human character and behaviour, such as the numerous versions of "The Doctor's Visit." Often the coarse but engaging features of the painter himself grin familiarly at the spectator, inviting him to concur in this hilarious view of the world. The earthy jollity in Steen is balanced by an intellectual and literary element, with an extensive use of allegory, that is unique among the Dutch genre painters and gives a peculiar piquancy to his art.

See C. Hofstede de Groot, *Catalogue of Dutch Painters*, vol. i (1907); F. Schmidt Degener and H. E. van Gelder, *Jan Steen*, Eng. trans. (1927). (R. E. W. J.)

STEENKERKE (STEENKIRK), a village in Belgian province of Hainaut, on the Senne river, pop. (1961) 405; battle of Steenkerke (Steinkirk, Estinkerke) fought there July 23–Aug. 3, 1692 between allies (*see* GRAND ALLIANCE, WAR OF THE) under William III of England and the French commanded by the duke of Luxembourg. William set his army in motion before dawn on Aug. 3 to surprise the French right about Steenkerke. When the fight opened, Luxembourg was completely surprised, but the march of the allies' main body had been mismanaged. A few battalions only had come up to support the advanced guard when the real attack opened. The advanced guard had swept the first French line before it. The English and Danes stubbornly advanced, the second and third lines of the French infantry giving ground before them, but Luxembourg was rapidly massing his whole force to crush them. At the crisis Luxembourg had not hesitated to throw the whole of the French and Swiss guards into the fight, and as, during and after this supreme effort, more and more French troops came up, the allies were driven back, and, the coup having manifestly failed, William ordered the retreat. Over 8,000 men out of only about 15,000 engaged on the side of the allies were killed and wounded, and the losses of the French out of a much larger force were at least equal. Contemporary soldiers affirmed it was the hardest battle ever fought by infantry.

STEEPLE: *see* SPIRE.

STEEPLECHASING: *see* HORSE RACING AND BREEDING.

STEEVENS, GEORGE (1736–1800), English Shakespearean scholar, author of the first variorum edition, was born at Poplar on May 10, 1736, the son of a captain, later a director, of the East India company. He was educated at Eton and King's college, Cambridge, but left without taking a degree. He acquired a house on Hampstead Heath where he built up a great library, including over 30 early Shakespearean quartos and many Hogarth prints. Every morning he walked into London, visited his friend Isaac Reed, made a round of the booksellers and walked back to Hampstead. In 1766, when he was only 30, he published a reprint of 20 of the quarto editions of Shakespeare's plays, a notable achievement, especially at that time when the value of the quartos was not fully recognized. In 1773 he published a ten-volume edition of Johnson's *Shakespeare*, more fully annotated and including excerpts from other commentators. (Johnson's contributions to the edition were slight.) This therefore was an attempt at a variorum edition. In 1793, however, Steevens was persuaded by his jealousy of Edmund Malone to publish an edition of 15 volumes in which he used his wide knowledge of Elizabethan literature to offer many useful parallels from writers contemporary with Shakespeare. It is Isaac Reed's 1803 reissue of this edition, with additional notes left by Steevens, which is known as the first variorum edition.

Steevens was a learned man of an acrid and sometimes malignant humour. His practice of writing notes on obscene passages and expressions which he signed with the name of clergymen he disliked, earned him the title of the Puck of commentators. He was one of the exposers of the Chatterton-Rowley and Ireland forgeries. He died at Hampstead on Jan. 22, 1800.

See D. Nichol Smith, *Shakespeare in the 18th Century* (1928).

STEFAN, JOSEF (1835–1893), Austrian physicist, whose original contributions ranged over several important fields, including the kinetic theory of gases, hydrodynamics and, in particular, radiation, was born on March 24, 1835, at St. Peter near Klagenfurt and died on Jan. 7, 1893, in Vienna.

Stefan was educated in the University of Vienna, of which he became doctor of philosophy in 1858, then *Privatdozent* in mathematical physics, in 1863 professor ordinarius of physics and in 1866 director of the physical institute. He was a distinguished member of the Vienna Academy of Sciences, of which he was appointed secretary in 1875. Before Stefan's work, G. R. Kirchhoff had already described the perfect radiator as the "perfectly black body," namely, one that absorbed all the radiation that fell on it and reflected none, but emitted radiation of all wave lengths. Stefan showed empirically in 1879 that the radiation of such a body was proportional to the fourth power of its absolute temperature, a relationship since known as Stefan's law or as the Stefan-Boltzmann law after it had been deduced by L. Boltzmann in 1884 from thermodynamical considerations. Stefan's law was one of the first important steps leading to the understanding of black-body radiation from which the quantum idea of radiation sprang. (D. McK.)

STEFANSSON, VILHJALMUR (1879–1962), arctic explorer, was born on Nov. 3, 1879, at Arnes, Manitoba, of Icelandic parents. He graduated from the University of Iowa and later studied theology and anthropology for three years at Harvard university. After two archaeological voyages to Iceland in 1904 and 1905 he turned to arctic research and in 1906–07 he visited the Canadian arctic for the first time as ethnologist on an expedition to the Mackenzie river delta. The expedition ship, approaching from the west, was unable to penetrate the ice beyond Point Barrow, so Stefansson, who had traveled alone overland, never joined the expedition. Instead he spent a year living with the Eskimos, acquiring an intimate knowledge of their language and culture, and forming the belief, which he put into practice during his later travels, that Europeans could "live off the land" in the arctic by following the pattern of Eskimo life. From 1908 to 1912 he and R. M. Anderson carried out ethnographical and zoological studies among the Mackenzie Eskimo and the Copper Eskimo of Coronation gulf, including some groups hitherto unknown to Europeans.

He next led an expedition, sponsored by the Canadian government, spending five consecutive years (1913–18) north of the Arctic circle, exploring the western Canadian arctic and discovering Borden, Brock, Meighen and Longheed islands. During these years he acquired an extensive knowledge of the Canadian arctic and a profound belief in the potential economic importance of the area. In 1932 he became adviser on northern operations for Pan-American airways. In World War II he served the United States government in an advisory capacity, surveyed defense conditions

in Alaska and prepared reports and manuals for the armed forces, becoming arctic consultant to the Northern Studies program at Dartmouth college, Hanover, N.H., from 1947 until his death in 1962. His principal publications are: *My Life With the Eskimo* (1913); *The Friendly Arctic* (1921); *The Northward Course of Empire* (1922); *The Adventure of Wrangel Island* (1925); *The Standardization of Error* (1928); *Unsolved Mysteries of the Arctic* (1939); *Greenland* (1942); *Arctic Manual* (1945). (L. M. Fs.)

STEFFANI, AGOSTINO (1653–1728), Italian ecclesiastic, diplomatist and musical composer, was born at Castelfranco on July 25, 1653. At a very early age he was admitted as a chorister at St. Mark's, Venice. Count Tattenbach took him in 1667 to Munich, where his education was completed at the expense of Ferdinand Maria, elector of Bavaria, who gave him a court appointment. He was sent in 1673 to study in Rome, where Ercole Bernabei was his master, and among other works he composed six motets, the original manuscripts of which are now in the Fitzwilliam Museum at Cambridge. On his return to Munich in 1674 he published his first work, *Psalmodia vespertina*, a part of which was reprinted in Martini's *Saggio di contrappunto* in 1774. In 1675 he was appointed court organist. The date when he was ordained priest, with the title of Abbate of Lepsing, is not precisely known. Steffani's first opera, *Marco Aurelio*, which was written for the carnival and produced at Munich in 1681, was followed by other works now lost. In 1688 he became Kapellmeister at the court of Hanover, where he made many friends. He wrote for the opening of the new opera house in 1689, an opera called *Enrico il Leone*, which was produced with extraordinary splendour. For the same theatre he composed *La Lotta d'Ercole con Achilleo* in 1689, *La Superbia d'Alessandro* in 1690, *Orlando generoso* in 1691, *Le Rivali concordi* in 1692, *La Libertà contenta* in 1693, *I Trionfi del fato* and *I Baccanali* in 1695, and *Briseide* in 1696. Elevation of Ernest Augustus to the electorate in 1692 led to difficulties with the various German courts; Steffani was sent round as an envoy in 1696 and as a result received from Innocent XI the bishopric of Spiga for his services in securing privileges for Hanoverian Catholics. In 1698 he was sent as ambassador to Brussels, and after the death of Ernest Augustus in the same year he entered the service of the elector palatine, John William, at Düsseldorf, where he held the offices of privy councillor and protonotary of the Holy See.

In 1724 the Academy of Ancient Musick in London elected him its honorary president for life; and in return for the compliment he sent the association a magnificent *Stabat Mater*, for six voices and orchestra, and three fine madrigals. All these compositions are very much in advance of the age in which they were written; and in his operas Steffani shows a remarkable appreciation of the demands of the stage. Notable too are his beautiful chamber-duets, which, like those of his contemporary Carlo Maria Clari (1669–1745), are chiefly written in the form of cantatas for two voices, accompanied by a figured bass. The British museum possesses more than a hundred of these compositions, some of which were published at Munich in 1679. Steffani visited Italy for the last time in 1727. He died at Frankfurt on Feb. 12, 1728.

Steffani stands somewhat apart from contemporary Italian composers (*e.g.*, Alessandro Scarlatti) in his mastery of instrumental forms. His opera overtures show a combination of Italian suavity with a logical conciseness of construction which is due to French influence. His vocal music and his famous duets are inferior to those of Scarlatti, but his instrumental music is historically important as a factor in the artistic development of Handel.

STEFFEN, ALBERT (1884–1963), Swiss novelist and dramatist, one of the leading writers of the anthroposophical movement founded by Rudolf Steiner (*q.v.*), was born at Murgenthal on Dec. 10, 1884. His early works were compassionate messages of alarm at the disastrous effects of modern technological civilization and secularized thought in human relations. Moved by these problems, he joined the anthroposophical movement in 1907, settling at its centre in Dornach, near Basel. (Steffen was later president of the Anthroposophical Society and was editor of its review, *Das Goetheanum.*) From that time his numerous writings became visions of a world permeated by metaphysical powers of good and evil, as revealed in old and esoteric European and Asiatic traditions. His novels include *Die Erneuerung des Bundes* (1913) and *Aus Georg Archibalds Lebenslauf* (1950); his plays, *Hieram und Salomo* (1927), *Das Todeserlebnis des Manes* (1934), and *Barrabas* (1949; Eng. trans., *Christ or Barrabas?*, 1950); and his essays, *Der Künstler zwischen Westen und Osten* (1925; Eng. trans., *The Artist Between West and East*, 1946). *Buch der Rückschau* (1939) is autobiographical. Steffen died at Dornach on July 13, 1963.

BIBLIOGRAPHY.—A. von Sybel-Petersen, *A. Steffen* (1934); *Das A. Steffen-Buch*, ed. by P. Bühler (1944); F. Hiebel, *A. Steffen* (1960). (A. Bx.)

STEFFENS, HENRIK (1773–1845), German philosopher, of Norwegian extraction, was born on May 2, 1773, at Stavanger. He studied in Germany, and began to lecture at Copenhagen in 1802. His lectures were appreciated by Grundtvig and Oehlenschläger, but the authorities regarded them with disfavour, and Steffens removed to Germany, where he fought in the Napoleonic wars on the German side. He was professor of physics at Breslau (1811–32) and at Berlin (1832–45). He died in Berlin on Feb. 13, 1845. Steffens was the friend and adherent of Schelling and Schleiermacher. His profound knowledge of physics modified the speculations of Schelling. His chief works were *Beiträge zur innern Naturgeschichte der Erde* (1801); *Grundzüge der philos. Naturwissenschaft* (1806); *Anthropologie* (1824). He also wrote novels and poetry and an autobiography, *Was ich erlebte*, and after his death there was published *Nachgelassene Schriften* (1846).

STEFFENS, (JOSEPH) LINCOLN (1866–1936), U.S. journalist, lecturer, and political philosopher, was a leading figure among the writers whom Theodore Roosevelt called muckrakers, but his work was closer to documented sociological case study than mere sensational exposure. He was born in San Francisco, Calif., on April 6, 1866. After education at the University of California, graduate study with Wilhelm Wundt in Leipzig and with Jean Martin Charcot in Paris confirmed his basic positivist orientation. During nine years of New York newspaper work Steffens discovered abundant evidence of corruption of politicians by businessmen seeking special privileges. In 1901 he became managing editor of *McClure's Magazine*, and soon began to publish the influential articles later collected as *The Shame of the Cities*. Many nationwide lecture tours won him recognition as an American Socrates. He raised rather than answered questions, jolting his audience by comic irony rather than moral indignation into awareness of the ethical paradox of private interest in public affairs. He revealed the shortcomings of the popular dogmas that connected economic success with moral worth and national progress with individual self-interest by showing that the Horatio Alger hero often won his position by corrupting the community—and himself.

Developments in Mexico and Russia turned Steffens' attention from reform to revolution. After a trip to Petrograd in 1919 he wrote a friend, "I have seen the future, and it works." His unorthodoxy lost him his American audience during the 1920s; he continued to study revolutionary politics in Europe and became something of a legendary character for the younger expatriates. After the great success of his *Autobiography* he supported many Communist activities, but refused identification with any party or doctrine. His characteristic achievement, both as writer and as lecturer, was in persuading people to examine for themselves the character of contemporary civilization.

His *Autobiography* was published in 1931 and his *Letters* (2 vol.), edited by Ella Winter and G. Hicks, in 1938. (R. D. H.)

STEGOCEPHALIA, a term often applied to Paleozoic and Triassic amphibians with solidly roofed skulls, mainly members of the Labyrinthodontia. (*See* AMPHIBIA.)

STEGOSAURUS, a type of dinosaur found in the late Jurassic of North America. The name (plated lizard) refers to the armament of the reptile in the shape of a paired row of bony plates extending upward along the length of the back; the tail bore two pairs of bony spikes. *Stegosaurus* was an inoffensive, small-headed vegetarian with a maximum length of 30 ft. The hind legs are much longer than the fore, suggesting a descent from bipedal ancestors. *See* REPTILE; DINOSAUR. (A. S. RR.)

STEIN, SIR AUREL (1862–1943), British archaeologist, was born at Budapest on Nov. 26, 1862. Educated at Budapest and Dresden and at the universities of Vienna and Tübingen, he went to England for further study and then to India, where he became principal of the oriental college, Lahore, and registrar of the Punjab university in 1888. He was appointed in 1899 to the Indian education service, and for the next two years carried out archaeological explorations for the Indian government in Chinese Turkestan. In 1906–08 he made further explorations in central Asia and western China. From 1910 he was superintendent of the Indian archaeological survey, and in 1913–16 carried out explorations in Iran and central Asia, described by him in *The Geographical Journal* (1916). He was created K.C.I.E. in 1912. In 1926 he continued his explorations on the northwest frontier and identified the site of Aornus, thus correcting Abbott's view, and between 1926 and his death explored Iran, Iraq and Transjordan. He died in Kabul, Afghanistan, Oct. 28, 1943.

His writings include: *Chronicle of Kings of Kashmir* (1900); *Ancient Khotan* (1907); *Ruins of Desert Cathay* (1912); *Serindia*, 5 vols. (1921); *The Thousand Buddhas* (1921); *On Alexander's Track to the Indus* (1929); *A Catalogue of Paintings Recovered from Tun-Huang* (1931); *An Archaeological Tour in Gedrosia* (1931); *Archaeological Reconnaissances in S.E. Iran* (1937); *On Old Routes of Western Iran* (1940).

STEIN, CHARLOTTE VON (1742–1827), famous for her intimate friendship with Goethe (*q.v.*), was born at Weimar on Dec. 25, 1742, the eldest daughter of the *Hofmarschall* (master of ceremonies). When she was 15 she became lady in waiting to Duchess Anna Amalia, the mother of Karl August, the duke of Saxe-Weimar. In 1764 she married Freiherr Friedrich von Stein, master of the horse to the duke; they had seven children. Goethe, who moved to Weimar in 1775, was captivated by her and for more than ten years she was his constant companion and the inspiration of much of his poetry. On Goethe's return from Italy in 1788 this close relationship between them came to an end although their friendship was later renewed. Goethe's letters to Frau von Stein form one of the most interesting sections of his correspondence. She was also a friend of Schiller and his wife. Her husband died in 1793 and she continued to live at Weimar, where she died on Jan. 6, 1827.

Bibliography.—Goethe's *Briefe an Frau von Stein aus den Jahren 1776–1820* was ed. by A. Schöll, 3 vol. (1848–51), 3rd ed. by J. Wahle (1899–1900); and also by J. Petersen (1923) and H. H. Borcherdt, 2nd ed. (1934). *See* also W. Bode, *Charlotte von Stein* (1910); L. Voss, *Goethes unsterbliche Freundin Charlotte von Stein* (1921); A. Nobel, *Frau von Stein* (1939); M. Susman, *Deutung einer grossen Liebe, Goethe und Charlotte von Stein* (1951). (A. Gs.)

STEIN, GERTRUDE (1874–1946), expatriate U.S. author and one of the more colourful personalities of the post-World War I literary scene, was born in Allegheny, Pa., Feb. 3, 1874. She spent her infancy in Vienna and Paris and her girlhood in Oakland, Calif. At Radcliffe college, where she was a student from 1893 to 1897, she studied psychology under William James, and then attended Johns Hopkins medical school from 1897 to 1902. Of independent means, she went abroad to live in 1903. For more than a generation her home in Paris at 27 rue de Fleurus was a centre for artists and writers. She was, in the words of one of her protégés, Bravig Ihms, a "vital, headstrong and cordial personality," with "an insatiable appetite for people." They came and went in an unending procession, year after year, some famous, some unknown, some to become famous later: Henri Matisse, Pablo Picasso, Juan Gris, Clive Bell, Wyndham Lewis, Carl Van Vechten, Sherwood Anderson, Ezra Pound, Ford Madox Ford, Ernest Hemingway, André Gide, John Reed, Elliot Paul, Jo Davidson, Paul Robeson and scores of others.

Despite her German-Jewish origin, Miss Stein continued living in Paris during World War II under Nazi occupation. She died in Paris July 27, 1946.

A literary "cubist," Gertrude Stein carried to extremes her weird, repetitious and seemingly nonsensical manipulation of words,—one of her celebrated phrases being "Rose is a rose is a rose." "Widely ridiculed and seldom enjoyed," comments Edmund Wilson in a chapter about her in *Axel's Castle*, "she has yet played an important role in connection with other writers who have become popular." In *The Autobiography of Alice B. Toklas* (1933), Miss Stein, while ostensibly writing the life of her secretary and close companion since 1907, actually wrote her own. Alice Toklas, so the autobiography recounts, knew personally three geniuses in her lifetime: Picasso, Alfred North Whitehead and Gertrude Stein.

Among Miss Stein's works are: *Three Lives* (1908); *Tender Buttons* (1915); *The Making of Americans* (1925); *Useful Knowledge* (1928); *Ten Portraits* (1930); *Lucy Church Amiably* (1930); *Before the Flowers of Friendship Faded Friendship Faded* (1931); *How to Write* (1931); *Operas and Plays* (1932); *Matisse, Picasso and Gertrude Stein* (1932); *Four Saints in Three Acts*, an opera libretto (1934); *Lectures in America* (1935); *Narration* (1936); *Picasso* (1938); *Ida* (1941); *Brewsie and Willie* (1946). Carl Van Vechten edited her *Last Operas and Plays* (1949).

Bibliography.—D. Sutherland, *Gertrude Stein; a Biography of Her Work* (1951); H. R. Garvin, *Gertrude Stein: a Study of Her Theory and Practice* (1950); E. Sprigge, *Gertrude Stein; Her Life and Work* (1957).

STEIN, HEINRICH FRIEDRICH KARL, Baron vom und zum (1757–1831), German statesman, was born at the family estate near Nassau, on Oct. 26, 1757. He studied law at Göttingen, Wetzlar, Regensburg and Vienna, and settled in Berlin in 1780. There he entered the Prussian service.

After holding various diplomatic and other appointments he became (1787) Kammerdirektor, *i.e.*, director of the board of war and domains for the king's possessions west of the river Weser; and in 1796 he was appointed supreme president of all the Westphalian chambers dealing with the commerce and mines of those Prussian lands. Among the benefits of his administration was the canalization of the river Ruhr, which became an important outlet for the coal of that region. He also improved the navigation of the Weser, and kept up well the main roads committed to his care. On June 8, 1793 he married the countess Wilhelmine von Wallmoden-Gimborn, daughter of a natural son of King George II of Great Britain.

Stein's early training, together with the sternly practical bent of his own nature, made him completely impervious to the enthusiasm which the French Revolution had aroused in many minds in Germany. He disliked its methods as an interruption to the orderly development of peoples. Nevertheless he carefully noted the new sources of national strength which its reforms called forth in France. Under the untoward circumstances which followed the Peace of Basel Stein in 1804 took office at Berlin as minister of state for trade. He protested against the effects of the Gallophil policy of the chief minister, Haugwitz, and the evil influences which clogged the administration. Little, however, came of Stein's protests, though they were urged with his usual incisiveness and energy. Prussian policy continued to progress on the path which led to the disaster at Jena (Oct. 14, 1806).

The king then offered to Stein the portfolio for foreign affairs, which he declined, because he desired to see Hardenberg take that office and effect, with his own help, the necessary administrative changes. The king refused to accept Hardenberg, and, greatly irritated by Stein's unusually outspoken letters, dismissed him altogether, adding that he was "a refractory, insolent, obstinate and disobedient official." Stein now spent in retirement the months during which Napoleon completed the ruin of Prussia; but he saw Hardenberg called to office in April 1807 and important reforms effected in the cabinet system. During the negotiations at Tilsit, Napoleon refused to act with Hardenberg, who thereupon retired. Strange to say, the French emperor at that time suggested Stein as a possible successor, and on Oct. 4, 1807, Frederick William called Stein to office with very wide powers.

Stein's strong convictions led him to press on drastic reforms in a way which could not otherwise have been followed. First came the Edict of Emancipation, issued at Memel on Oct. 9, 1807, which abolished the institution of serfdom throughout Prussia from Oct. 8, 1810. All distinctions affecting the tenure of land (noble land, peasants' land, etc.) were also swept away, and the principle of free trade in land was established forthwith. The same famous edict also abrogated all class distinctions respecting occupations and callings of any and every kind, thus striking

another blow at the caste system which had been so rigorous in Prussia. Stein's next step was to strengthen the cabinet by wise changes and he also furthered the progress of the military reorganization which is connected more especially with the name of Scharnhorst (*q.v.*). Stein's efforts were directed more towards civil affairs; and in this sphere he was able to issue a measure of municipal reform (Nov. 19, 1808) which granted local self-government on enlightened and practical lines to all Prussian towns.

In August 1808 the French agents, who swarmed throughout the land, had seized one of his letters, in which he spoke of his hope that Germany would soon be ready for a national rising like that of Spain. On Sept. 10 Napoleon gave orders that Stein's property in the new kingdom of Westphalia should be confiscated, and he likewise put pressure on Frederick William to dismiss him. The king evaded compliance; but the French emperor, on entering Madrid in triumph, declared (December 16) *le nommé Stein* to be an enemy of France and the Confederation of the Rhine; and ordered the confiscation of all his property in the Confederation. Stein saw that his life was in danger and fled from Berlin (Jan. 5, 1809) to Bohemia.

For three years he lived in the Austrian empire, generally at Brünn; but in May 1812 he received an invitation from the emperor Alexander I to visit St. Petersburg, seeing that Austria was certain to range itself on the side of France in the forthcoming Franco-Russian War. At the crisis of that struggle Stein may have been one of the influences which kept the tsar determined never to treat with Napoleon. When the miserable remains of the Grand Army reeled back into Prussia at the close of the year, Stein urged the Russian emperor to liberate Europe.

Events now brought Stein rapidly to the front. On Dec. 30, 1812, the Prussian general Hans Yorck von Wartenburg signed at Tauroggen a convention with the Russian general Hans Diebitsch for neutralization of the Prussian corps at and near Tilsit, and for the free passage of the Russians through that part of the king's dominions. The tsar thereupon requested Stein to act as provisional administrator of the provinces of East and West Prussia. In that capacity he convened an assembly of representatives of the local estates, which Feb. 5, 1813, ordered the establishment of a militia (*Landwehr*), a militia reserve and a final levy (*Landsturm*). He took part in the drafting of a Russo-Prussian convention (March 19, 1813) respecting the administration of the districts which should be delivered from French occupation. The Allies conferred on Stein the important duties of superintending the administration of the liberated territories. After the battle of Leipzig (Oct. 16–19, 1813) Stein entered the city.

He now desired to see Germany reconstituted as a nation, in a union which should be at once strong for purposes of defense and founded on constitutional principles. His statesmanlike projects were foiled, partly by the short-sightedness of German rulers and statesmen, but also by the craft with which Metternich gained the alliance of the rulers of south and central Germany for the Austrian empire. During the Congress of Vienna Stein pressed in vain for an effective union of the German people. Austria and the secondary German states resisted all proposals in this direction; and Stein blamed the Prussian chancellor Hardenberg for an indefiniteness of purpose which probably resulted from the same defect in Frederick William of Prussia. Stein shared in the desire of all Prussian statesmen at that time to have Saxony wholly absorbed in their kingdom. In that, as in other matters, he was disappointed.

Stein passed into retirement after the Congress of Vienna, and saw with pain and disgust the postponement of the representative system of government which Frederick William had promised to Prussia in May 1815. His chief interest was in the study of history, and in 1818–1820 he worked hard to establish the society for the encouragement of historical research and the publication of the *Monumenta Germaniae historica*, of which his biographer, G. H. Pertz, became director. Stein died on June 29, 1831.

Possibly there has been a tendency to magnify the achievements of Stein. Research has shown that the credit for originating his reforms must be shared with Heinrich von Schön and many others.

BIBLIOGRAPHY.—For biographies of Stein, *see* G. H. Pertz, 6 vol. (1849–55); Sir J. Seeley, 3 vol. (1878); M. Lehmann (1902–03). *See* also H. F. K., Baron vom Stein, *Lebenserinnerungen* (1901); C. T. Perthes, *Politische Zustände und Personen in Deutschland zur Zeit der französischen Herrschaft*, 2 vol. (1862); *Denkwürdigkeiten des Staatskanzlers Fürsten von Hardenberg*, ed. by L. von Ranke, 5 vol. (1877); Varnhagen von Ense, *Denkwürdigkeiten*, 6 vol. (1837–42); A. Stern, *Abhandlungen und Aktenstücke aus der preussischen Reformzeit 1807–15* (1885); M. Philippson, *Geschichte des preussischen Staatswesens 1786–1813*, 2 vol. (1880); M. Lehmann, *Knesebeck und Schön* (1875); J. P. Hassel, *Geschichte der preussischen Politik, 1807–1815* (1881); Jean d'Ussel, *Études sur l'année 1813; la défection de la Prusse* (1907); A. D. White, *Seven Great Statesmen* (1910); G. Hasse, *Theodor von Schön und die Steinische Wirtschaftsreform* (1915); G. S. Ford, *Stein and the Era of Reform in Prussia, 1807–15* (1922); E. M. Arndt, *Meine Wanderungen und Wandelungen mit dem Reichsfreiherrn vom Stein* (1925); R. Huch, *Stein* (1925); H. M. Elster, *Freiherr vom Stein* (1927).

STEINBECK, JOHN ERNST (1902–), U.S. writer, who achieved fame in the 1930s for his powerful novels about agricultural workers in California and who received the Nobel prize for literature in 1962, was born in Salinas, Calif., on Feb. 27, 1902. He attended Stanford university intermittently between 1920 and 1926 but did not take a degree. His early novels, *Cup of Gold* (1929), *The Pastures of Heaven* (1932) and *To a God Unknown* (1933), were relatively unknown until *Tortilla Flat* (1935) brought him popularity. *In Dubious Battle* (1936) remains one of the best strike novels in the English language. *Of Mice and Men* (1937), a novelette about the strange, complex bond between two migrant labourers, was adapted for the stage and received the Drama Critics Circle award. His Pulitzer prize winning novel, *The Grapes of Wrath* (1939), about the migration of a dispossessed family from the Oklahoma dust bowl to California and their consequent exploitation by a ruthless system of agricultural economics, remains his best-known work. The furor occasioned by this novel's realistic picture of migrant workers and what the public assumed to be its political and economic theories became a colourful national event.

During World War II Steinbeck wrote some effective pieces of propaganda, among them *The Moon Is Down* (1942) and *Bombs Away* (1942), and he also served as war correspondent for the *New York Herald Tribune*. His immediate postwar work—*Cannery Row* (1945), *The Pearl* (1945, 1947), *The Wayward Bus* (1947)—reflected a bitterness against the greedy, rapacious elements of society that had made the horrors of war possible. In *Burning Bright* (1950) and *East of Eden* (1952) Steinbeck turned away from the naturalistic preoccupations and techniques of the 1930s and asserted that all men are brothers, keepers of one another, and that each man has the power to choose between good and evil. His later books, including *Sweet Thursday* (1954), *The Short Reign of Pippin IV* (1957), *The Winter of Our Discontent* (1961) and *Travels With Charlie* (1962), are comparatively slight works of entertainment and journalism, widely popular in their appeal.

Steinbeck's early reputation as a proletarian, naturalistic writer of sociological penetration, and his constant experimentation with materials and techniques, tended to obscure other facets of his writing, but under subsequent scrutiny much of his work has revealed itself rich in symbolic structures and mythopoeic qualities.

BIBLIOGRAPHY.—E. W. Tedlock and C. V. Wicker (eds.), *Steinbeck and His Critics* (1957); Peter Lisca, *The Wide World of John Steinbeck* (1958); Warren French, *John Steinbeck* (1961); F. W. Watt, *Steinbeck* (1962); Joseph Fontenrose, *John Steinbeck* (1963). (PE. L.)

STEINER, JAKOB (1796–1863), Swiss mathematician, one of the founders of modern synthetic geometry, was born at Utzenstorf, Bern on March 18, 1796. He grew up without schooling; only at 14 did he learn to write. Against the wish of his parents, at 18 he went to the Pestalozzi school at Yverdon where his extraordinary geometric intuition was discovered. Later he went to Heidelberg and Berlin to study, supporting himself precariously as a tutor. The founding of August Crelle's *Journal* (1826) gave him an opportunity to publish his geometric discoveries. In 1832 he received an honorary doctorate from the university in Königsberg, and in 1834 was appointed extraordinary professor at the university in Berlin, a post he held until his death on April 1, 1863, in Bern.

Steiner was one of the great developers of synthetic geometry and many of the basic concepts and results in this field are due to him. His writings were collected as *Gesammelte Werke*, 2 vol. (1881–82).

See J. H. Graf, *Der Mathematiker Jakob Steiner* (1897); L. Kollros, *Jakob Steiner* (1947). (O. OE.)

STEINER, RUDOLF (1861–1925), Austrian philosopher,

scientist and artist, who formulated the philosophic theory known as anthroposophy, was born on Feb. 27, 1861, at Kraljevic, Aus., the son of a railway stationmaster. After an education, supplemented by much self-education, completed at the Technische Hochschule in Vienna, he edited an edition of Goethe's scientific writings, and from 1889 to 1896 worked on the standard edition of Goethe's works at Weimar. During this period he wrote *Truth and Science,* which secured his Ph.D., and his best-known philosophical work, *Die Philosophie der Freiheit* (1894). He then moved to Berlin where he was successively editor of the *Magazin für Literatur* and lecturer in a workingmen's college. Meanwhile he was developing a faculty for spiritual perception independent of the senses. The result of these researches, which he called anthroposophy and described as "knowledge produced by the higher self in man," brought him in 1902 into connection with the Theosophical society (*see* THEOSOPHY), but he found himself out of sympathy with its subsequent policy, and founded the independent Anthroposophical society in 1912. Anthroposophy postulates the existence of a spiritual world comprehensible to pure thinking but fully accessible only to the higher faculties of knowledge latent in every man. Historically, Steiner regarded man as having originally participated in the spiritual processes of the world through a dreamlike consciousness out of which the more limited but awake consciousness of today has crystallized. He claimed that an enhanced consciousness can again achieve perception of spiritual worlds. In 1913 Steiner began to build the first *Goetheanum* at Dornach, near Basel, Switz., as a "school of spiritual science." This was burned in 1922, but Steiner immediately designed another, of molded concrete, which became the centre of the Anthroposophical society. He died there on March 30, 1925.

Activities derived from Steiner's work include: the Waldorf school movement, by the late 1950s comprising some 60 schools and 25,000 children in Europe and America; homes and schools for defective and maladjusted children; a therapeutic movement with a central clinic at Arlesheim, Switz.; the biodynamic method of farming and gardening; centres for scientific and mathematical research; eurythmy, an art of movement to speech and music; and schools of drama, speech, painting, sculpture and architecture. After World War I Steiner proposed a threefold organization of society as a solution of political, social and economic questions.

BIBLIOGRAPHY.—Of Steiner's numerous published works, which included lecture courses on man's relation to the earth and the universe, eurythmy, education, spiritual science, architecture, economics and studies in the Gospels and *Genesis,* the following are the most important. Dates and titles are of the latest English edition; the original date of publication is given in brackets. *The Philosophy of Spiritual Activity* (1949 [1894]); *Goethe's Conception of the World* (1928 [1897]); *Theosophy* (1955 [1904]); *Occult Science: An Outline* (1950 [1913]); *Knowledge of the Higher Worlds and Its Attainment* (1932 [1909]); *The Threefold Commonwealth* (1920); *Story of My Life* (1924).

See also A. P. Shepherd, *A Scientist of the Invisible* (1954); E. Lehrs, *Man or Matter* (1951); C. Waterman, *The Three Spheres of Society* (1946); H. Poppelbaum, *Man and Animal* (1951); G. Wachsmuth, *Rudolf Steiners Erdenleben und Wirken* (1951), Eng. trans. (1955).

(A. C. HD.)

STEINLE, EDUARD JAKOB VON (1810–1886), Austrian-born painter of the German Nazarene group, was born in Vienna, July 2, 1810. From 1828 to 1834 he worked in Rome, in the "Nazarene" (German Pre-Raphaelite) circle of Peter von Cornelius and Johann Friedrich Overbeck (*qq.v.*), which attempted to restore religious feeling to art. Steinle's fresco cycles in Strasbourg, Frankfurt and Cologne cathedrals are in the rather cold, linear Nazarene style. He also painted many Madonnas and other religious subjects. In smaller canvases, many of legendary subjects, and in water-colour and sepia illustrations for the fairy tales of Klemens Brentano and others, Steinle betrays his Viennese origin and romantic temperament in a fresher touch and more lyrical design. Steinle was professor of historical painting at the Städelschen Institut, Frankfurt, from 1850 until his death there on Sept. 18, 1886.

STEINMETZ, CHARLES PROTEUS (1865–1923), U.S. electrical engineer, one of the outstanding electrical geniuses of the United States, was born at Breslau, Ger., April 9, 1865. He was educated at Breslau, Zürich and Berlin, specializing in mathematics, electrical engineering and chemistry. His activities as a Socialist led him into difficulties with the authorities, and, after a short sojourn in other countries, he emigrated in 1889 to the United States and found work with the Osterheid and Eickemeyer factory at Yonkers, N.Y. In 1893 when that factory was absorbed by the General Electric Co., Schenectady, N.Y., he was given an appointment as consulting engineer. His knowledge won him speedy promotion. After 1902 he served also as professor of electrical engineering at Union college, Schenectady, N.Y. Steinmetz regarded his three greatest contributions to electrical science to be: (1) his investigations on magnetism, resulting in his discovery of the law of hysteresis, which enabled losses of electric power due to magnetism to be accurately forecast before starting the construction of motors, generators, transformers and other electrical apparatus employing iron; (2) the development of his symbolic method of calculating alternating-current phenomena, which simplified an extremely complicated field, understood by few, so that the average engineer could work with alternating current, an accomplishment which was largely responsible for the rapid progress made in the commercial introduction of alternating-current apparatus; (3) his investigation of lightning phenomena, which resulted in his theory of traveling waves and opened the way for his development of lightning arresters to protect high-power transmission lines. Though primarily a mathematical genius and a student of theory, Steinmetz had about 200 patents to his credit, including improvements on generators, motors, transformers, in electrochemical operations, and the invention of the induction regulator, the method of phase transformation and the metallic electrode arc lamp. He was the author of a number of standard textbooks, in addition to a large number of scientific papers.

Steinmetz died at Schenectady, N.Y., Oct. 26, 1923.

See J. W. Hammond, *Charles Proteus Steinmetz* (1924); J. N. Leonard, *Loki; the Life of Charles Proteus Steinmetz* (1929).

STELLENBOSCH, a town of the U. of S. Af., lying 31 mi. E. of Cape Town by rail and 364 ft. above sea level in the valley of the Eerste river and in the centre of a rich fruit and wine-producing area.

After Cape Town it is the oldest town in south Africa, having been founded in 1679 by the governor Simon van der Stel, after whom it was named.

Population (1951) 17,757, of whom 8,252 were Europeans, 7,487 coloured, 1,940 natives and 78 Asians.

The streets of Stellenbosch are lined with oak trees and open channels of water from the Eerste river. It is mainly a residential and educational town with a self-contained university, several research laboratories, a theological seminary, a conservatorium of music, a training college and several high schools. The medium used in the university is Afrikaans.

In 1926 the agricultural faculty of the university was amalgamated with the Elsenburg School of Agriculture to form the Stellenbosch-Elsenburg college of agriculture which is part of the University of Stellenbosch.

STEM, in popular language the stalk of a plant or trunk of a tree. In botany a stem may be defined as an axis bearing leaves. The stem with its leaves is known as the shoot. Structurally it differs from a root in having no cap (rootcap) over the growing point, in terminating in a bud and in having nodes and internodes. Even when the stem has a superficial resemblance to a root, as in some subterranean forms, it can be distinguished from one on the basis of its internal structure. The chief difference is the spatial relationship of the primary vascular tissues (xylem and phloem), which is usually collateral or bicollateral in stems of gymnosperms and angiosperms and radial in roots. There are also differences in the manner in which the primary xylem is differentiated in the two organs; *e.g.,*

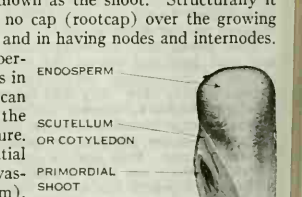
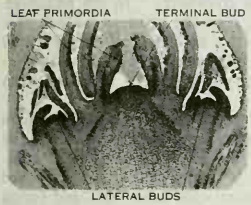


FIG. 1.—LONGITUDINAL SECTION OF GRAIN OF CORN (ZEA), WITH EMBRYO, SHOWN AT LEFT, ATTACHED TO ONE SIDE OF NUTRITIVE TISSUE. ENDOSPERM

mesarch or endarch in many stems, exarch in roots.

Types of Stems.—In general, the habit of the stem is erect or ascending, but it may lie prostrate on the ground, thus becoming procumbent, as in the sweet potato and strawberry. The scandent stem may climb on rocks or plants by means of rootlets, as in ivy; other vines have twining stems which twist around a supporting plant in a spiral manner. Twining plants turn either anticlockwise, as in morning-glory, dodder and gourd; or clockwise, as in honeysuckle, twining polygonum and hop. In other cases, climbing plants are supported by tendrils which may be specialized stems, as in the grape and passionflower; tendrillike leafstalks, as in clematis; or specialized leaflets, as in the pea. In warm climates twining plants often form thick woody stems and are called lianas, while in temperate regions they are generally herbaceous vines.

In some plants the stem does not elongate during its early development but forms a short conical structure from which a crown of leaves arises. These may form a bulb (onion, lily), a head (cabbage, lettuce) or a rosette, as in dandelion and plantain. Later this type of stem may grow rapidly into an elongated axis. Subterranean stems include rhizomes, corms and tubers. These are described in a later section.



FROM J. E. SASS, "BOTANICAL MICROTECHNIQUE," © IOWA STATE COLLEGE PRESS, 1958

FIG. 2.—SHOOT TIP OF SUGAR MAPLE (ACER SACCHARUM). MAGNIFIED ABOUT 20 TIMES

The character of the stem may be related to the longevity of the plant. Thus many herbaceous plants or herbs are annuals and complete their life cycle in one growing season, after which the entire plant dies. In biennial herbs the whole plant dies after two years; the lower part of the stem or crown persists after the first growing season and bears buds from which an erect stem arises the second season. In perennial herbs the short crown stem may produce new shoots for many years. Plants producing permanent woody stems are called trees and shrubs. The latter produce branches from or near the ground, while the former have conspicuous trunks. Shrubby plants of small stature are called undershrubs or bushes. The limits between these different kinds of stem are not always well defined; and there are some plants occupying an intermediate position between shrubs and trees, sometimes called arborescent shrubs.

The stem is not always conspicuous. Plants with a distinct stem are caulescent; those in which it is inconspicuous are acaulescent, as the primrose, cowslip and dandelion. A similar term is given in ordinary language to plants whose stems are buried in the soil, such as cyclamen and primula. Some plants are truly stemless. These may be thallophytes or bryophytes (*see* ALGAE; FUNGI; LIVERWORTS).

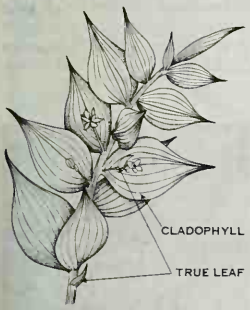
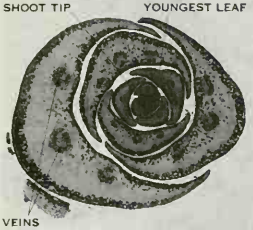


FIG. 3.—TWIG OF BUTCHER'S-BROOM (RUSCUS ACULEATUS)

Development of Stem.—The first rudiment of the young shoot of the embryo appears from the seed after the radicle (young root) has protruded. It is termed the plumule and differs from the radicle in the absence of a rootcap and in its tendency to ascend (*see* SEED). The apical growing portion constitutes the terminal bud of the plant, and by continued development the stem increases in height; projections appear at intervals, which are the rudimentary leaves. Lateral buds are produced which develop into lateral shoots more or less resembling the parent stem, and by these the branching of the plant is determined. A bud may be found in the axil of a previously formed leaf or, in other words, in the angle formed between the stem and leaf. Such buds are called axillary and are produced like the leaves from the tissues of the stem. During their development vascular bundles are formed which are continuous with those of the stem. Ultimately branches are produced that in every respect resemble the axis from which the bud was differentiated. The place of origin of the leaf is called a node; the intervals between nodes are called internodes.

Functions of the Stem.—The primary functions of the stem are support of the leaves, conduction of water and nutrient substances to the leaves and movement of elaborated foods from them. Other functions which may be performed by stems are storage of reserve food, photosynthesis and vegetative reproduction. The regular arrangement of leaves on the stem distributes them in such a manner that they are equally exposed to the light, as is necessary for photosynthesis (*see* LEAF: *Arrangement or Phyllotaxy*). The stem also supports the flowers.



BY COURTESY OF KATHERINE ESAU

FIG. 4.—TRANSVERSE SECTION OF PEACH LEAF BUD, SHOWING SHOOT TIP IN CENTRE SURROUNDED BY LEAF PRIMORDIA. MAGNIFIED ABOUT 20 TIMES

Conduction of water and nutrient salts from regions of absorption in the roots to the leaves takes place through certain vascular elements of the xylem (vessels and tracheids), while the movement of synthesized foods from the leaves to other organs of the plant occurs chiefly through the phloem. Radial movement of food within the stem may take place through ray and other parenchymatous cells. Storage of food and water frequently occurs in stems. Examples of food-storing stems are specialized forms, such as tubers, corms and rhizomes, and the woody stems of trees and shrubs. Water storage in stems is developed to a high degree in the Cactaceae.

All green stems carry on some of the photosynthetic work of the plant, but in most cacti this function is performed chiefly by the stem, the leaves being present often only for brief periods or as spines. Other plants in which food making is carried on in large measure by the stem are *Asparagus* and various species of *Euphorbia*. Where the stem is specialized so that it is flattened and leaflike it is termed a cladophyll, as in butcher's-broom or asparagus. Vegetative reproduction is accomplished by many underground stems and is especially common among the grasses and many members of the lily family.

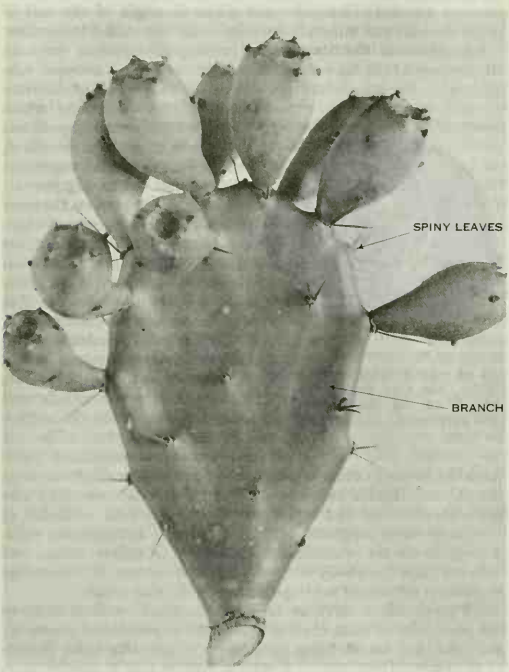
Growth and Bud Formation.—Growth in length of the stem results from the activity of the apical meristem and elongation of the internodes; the zone of most rapid growth is at some distance below the apex; below this the rate of growth gradually diminishes until the part is reached where growth in length no longer takes place. In some stems, as in those of grasses, growth in length persists for a longer time in a small region at the base of the internodes; this is known as intercalary growth. In shoots of limited growth, such as those of the larch, the internodes elongate only slightly and the leaves remain close together. Lateral buds give rise to branches, from which others, called branchlets or twigs, arise. The terminal bud sometimes dies at the end of one season, and the whole plant, as in annuals, perishes; or part of the axis is persistent and remains for two or more years, each of the leaves before its decay producing a bud in its axil, which continues the growth in spring.

REPRINTED WITH PERMISSION FROM W. W. ROBBINS, T. E. WEIER AND C. R. STOCKING, "BOTANY," COPYRIGHT 1950, JOHN WILEY & SONS, INC.

FIG. 5.—FIRE THORN (PYRACANTHA)

In ordinary trees, in which numerous lateral buds are formed, any injury done to a few branches is easily repaired; but in palms, which form mainly terminal buds and have no other provision for a lateral formation of them, destruction of the terminal bud results in no further growth of the

tree. In the trees of temperate and cold climates the buds which are developed during one season lie dormant during the winter, ready to open in the spring. They are generally protected by specialized leaves in the form of scales, which frequently exhibit a firmer and coarser texture than the foliage leaves. They serve a temporary purpose and usually fall off sooner or later after the leaves are expanded. The bud is often protected by a coating of resinous matter, as in the horse chestnut and balsam poplar, or by a thick downy covering, as in the willow.



J. HORACE MCFARLAND CO.

FIG. 6.—CACTUS (OPUNTIA MONACANTHIA)

In plants of warm climates the buds are often devoid of any protecting appendages; such buds are called naked.

When the terminal bud is injured or arrested in its growth, the elongation of the main axis stops, and the lateral branches often acquire increased activity. By continually cutting off the terminal buds or branches, a woody plant is made to assume a bushy appearance, and thus pollard trees are produced. The peculiar bird's nest appearance often presented by the branches of the common birch depends on an arrestment in the terminal buds, a shortening of the internodes, and a consequent clustering of the twigs (*see* GALL, PLANT).

Branching.—The mode in which branches come off from the stems gives rise to various forms of trees, as pyramidal, spreading or weeping—the angles being more or less acute or obtuse. In the Italian poplar and cypress the branches are erect, forming acute angles with the upper part of the stem; in the oak and cedar they are spreading, forming nearly a right angle; in the weeping ash and elm they develop at an obtuse angle; while in the weeping willow and birch they are pendulous, owing to their flexibility. The comparative length of the upper and under branches also gives rise to differences in the contour of trees, as seen in the conical form of spruce, and the umbrellalike form of the Italian stone pine (*Pinus pinea*). If no branches are formed, as in palms, the form of growth is referred to as columnar (fig. 7, centre). Repeated branching, as in oak, results in deliquescent growth (fig. 7, left).

Trees with a single main shaft and relatively small lateral branches, as exemplified by many conifers, are said to have an excurrent growth pattern (fig. 7, right).

Branches are sometimes long and slender and run along the ground, producing buds and adventitious roots at their extremity. This is seen in the runner of the strawberry. In the houseleek (*Sempervivum tectorum*) there is a similar prostrate branch of a shorter and thicker nature, known as an offset, producing a bud at its extremity capable of independent existence. In many instances the branch decays, and the young plant is separated from the parent plant.

Gardeners propagate plants by the process of layering, which consists in bending a twig, fixing the central part of it into the ground and, after the production of roots, cutting off its connection with the parent. A stolon is a branch which curves toward the ground and, on reaching a moist spot, takes root and forms an upright stem and ultimately a separate plant. This is a sort of natural layering, and the plant producing such branches is called stoloniferous. In the rose and mint a subterranean branch arises from the stem, which runs horizontally to a certain extent and ultimately sends up an aerial stem, which may become an independent plant.

Such branches are denominated rhizomes or suckers, and the gardener cuts the connection between the sucker and the parent stem to propagate these plants.

In asparagus and other plants which have a perennial stem below ground, subterranean buds are annually produced which appear above ground as shoots or branches covered with scales at first, and ultimately with cladophylls resembling true leaves. These branches are herbaceous and perish annually, while the main stem remains below ground to send up fresh shoots the following season.

In bananas the aerial stem is a shoot sent up by an underground stem and perishes after fruiting. Branches are sometimes arrested in their development and, in place of forming leaves, develop as spines or thorns, as in the hawthorn. Plants which have spines in a wild state, as the apple and pear, often are less spiny when cultivated; in some cases, as in the fire thorn (*Pyracantha coccinea*), a branch bears leaves at its lower portion and terminates in a spine.

(LEFT, RIGHT) RUTHERFORD PLATT, (CENTRE) E. AUBERT DE LA RUE

FIG. 7.—GROWTH HABIT OF TREES

(Left) Deliquescent or spreading growth of white oak (*Quercus alba*); (centre) columnar growth of assai palm (*Euterpe edulis*); (right) excurrent or pyramidal growth of larch (*Larix laricina*)

Rhizomes, Corms, Tubers and Bulbs.—Some stems are compressed or flattened laterally, while others are angular. Various terms are applied to the forms of stems, as cylindrical or terete, quadrangular or square, jointed or articulated, etc. The following are some of the more important modifications of stems: A rhizome or rootstock is a horizontal stem usually sending out numerous adventitious roots and leaf buds from its upper surface. It occurs in ferns, the iris, *Acorus* or sweet flag, ginger, the water lily, many species of *Carex*, rushes, grasses, anemone, etc. The leaves develop as scales, and by their presence, the absence of a rootcap and the presence of a bud at the apex, a rhizome can be distinguished from a root.

A rhizome such as occurs in Solomon's-seal is not the product of a single bud but is composed of portions of successive axes, the aerial parts of which have died off, leaving their scars. A rhizome sometimes assumes an erect form, as in *Scabiosa succisa*, in which the so-called praemorse root is in reality a rhizome, with the lower end decaying. The erect rhizome of water hemlock (*Cicuta virosa*) has hollow internodes, separated by partitions.

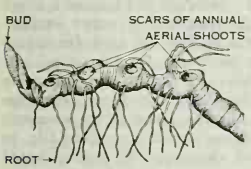


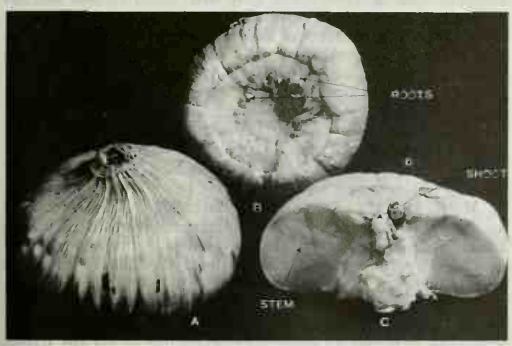
FIG. 8.—RHIZOME OF SOLOMON'S-SEAL (POLYGONATUM MULTIFLORUM)

In the coralroot orchid (*Corallorhiza*), which grows in soil rich in humus, no roots at all are developed, the corallike branching rhizome acting as the absorbing organ. A tuber is a short thickened stem or branch in which there is little internodal elongation, as in the potato. The eyes of the potato are leaf buds. Tubers are sometimes aerial, occupying the place of branches. The ordinary herbaceous stem of the potato, when cut into slips and planted, sends off branches from its base, which assume the form of tubers.

Tubers frequently store up a quantity of starch, as in *Maranta arundinacea*, whence West Indian arrowroot is derived.

Another form of thickened underground stem is the corm, as seen in the autumn crocus (*Colchicum*), gladiolus, etc. Structurally it is composed of a solid more or less rounded axis covered by thin membranous scales.

A corm is often only of one year's duration, giving off buds annually in the form of young corms. In the axil of an upper leaf of one (or more) of the bud corms the flowering stem develops and



J. HORACE MCFARLAND CO.

FIG. 9.—CORM OF GLADIOLUS (GLADIOLUS COMMUNIS), SHOWING: (A) WHOLE CORM; (B) UNDERSIDE; (C) CUTAWAY VIEW, WITH FIBROUS LEAF BASES REMOVED

bears the flowers. The flowering stem dies down and the bud corm, from which it arose, enlarges greatly at the expense of its parent corm, which thus becomes shriveled. In spring the corm that flowered the spring before sends out large leaves which were mere rudiments the previous autumn. At the end of spring these leaves die down; only the bases of the lower ones remain as thin brown scales around the corm.

This corm, in its turn produces in the axils of its leaves new bud corms, which form rudimentary leaves in the autumn. The following spring one (or more) of these bud corms produces the flowering stem and, eventually, leaves and new bud corms; and thus the cycle goes on.

The bulb is a bud which is often subterranean. The axis in this case is much shortened, and the internodes are unelongated. The bases of the leaves rising from the stem are close together and become succulent and enclose the axis. In the lily the thick and narrow scales are arranged separately in rows, and the bulb is called scaly; while in the leek, onion and tulip the scales are broad and enclose one another in a concentric manner, the outer ones being thin and membranous, and the bulb is tunicated. In the axils of these fleshy scales new lateral shoots arise, forming new bulbs.

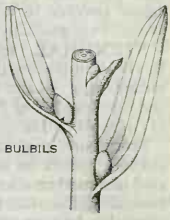
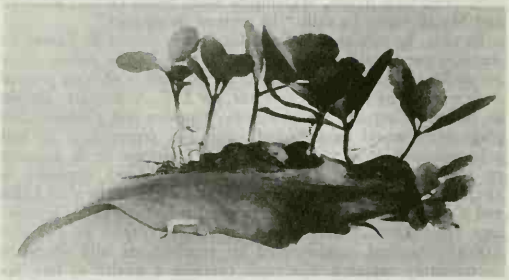


FIG. 10.—STEM OF BULBIFEROUS LILY (LILIUM BULBIFERUM)

The lateral buds may remain attached to the axis and produce flowering stems, so that apparently the same bulb continues to flower for many years, as in the hyacinth and tulip; in other species the young bulbs form separate plants.

In the axil of the leaves of *Lilium bulbiferum*, *Dentaria bulbifera* and some other plants, small conical or rounded bodies are produced, called bulbils or bulblets. They resemble bulbs and consist of a small number of thickened scales enclosing a growing point. These scales are frequently united closely together, so as to form a solid mass. The scales in bulbs vary in number. In *Gagea* there is only one scale; in the tulip and *Fritillaria imperialis* they vary from two to five; while in lilies and hyacinths there are a great number of them.

Adventitious Shoots.—Adventitious shoots are those which arise elsewhere than in the normal place, as from old stems or roots. Such shoots are frequent on the roots of elm, poplar, plum and other fruit trees. Occasionally, adventitious buds are produced



J. HORACE MCFARLAND CO.

FIG. 11.—PLANTLETS BEING PRODUCED IN LEAF NOTCHES OF LIFE PLANT (BRYOPHYLLUM CALYCINUM)

on the edges of leaves, as in *Bryophyllum calycinum*, *Malaxis paludosa* and various species of *Asplenium*, and on the surface of leaves, as in *Ornithogalum thyrsoideum*. These are capable of forming independent plants.

Similar buds may be induced to develop on the leaves of begonia, gesneria, gloxinia and achimenes by wounding various parts of them and placing them in moist soil, a method pursued by gardeners in their propagation.

See also PLANTS AND PLANT SCIENCE: *Anatomy of Plants.*

BIBLIOGRAPHY.—A. F. Blakeslee and C. D. Jarvis, *Trees in Winter* (1926); K. Esau, *Plant Anatomy* (1953); A. S. Foster and E. M. Gifford, Jr., *Comparative Morphology of Vascular Plants* (1959); H. E. Hayward, *The Structure of Economic Plants* (1948); E. Maino and F. Howard, *Ornamental Trees* (1955); W. Rauh, *Morphologie der Nutzpflanzen* (1950); R. W. Schery, *Plants for Man* (1952); W. Troll, *Praktische Einführung in die Pflanzenmorphologie* (1954).

(H. E. HD.; K. E.)

STENDAL, a town in the district of Magdeburg, Ger., on the Uchte, 70 mi. W. of Berlin on the main line of railway to Hanover and at the junction of lines to Bremen, Magdeburg and Wittenberge. Pop. (1950) 40,618. Stendal was founded in 1151 by Albert the Bear, on the site of a Wendish settlement, and soon afterwards acquired a municipal charter. Becoming capital of the Altmark, it rose to a considerable degree of prosperity, in part recently restored to it by its railway connections. The original Wends were gradually fused with the later Saxons.

STENDHAL (1783–1842) the pseudonym of MARIE HENRI BEYLE, who borrowed the name from a small German town, the birthplace of Johann Winckelmann, was one of the most original and distinguished of French essayists and novelists. He was born at Grenoble on Jan. 23, 1783.

From Nov. 1796 to 1799, he attended the Ecole Centrale at Grenoble, where he was brilliantly successful. Here he was initiated into mathematics and accurate thinking. He arrived in Paris soon after the 18th Brumaire to study for the Ecole Polytechnique, but he soon gave up this scheme. Pierre Daru, the future *grand administrateur* of the empire, was a relation and his patron, and obtained a post for him in the War Office; later, in 1799, he sent him to Italy which became the chosen country of this cosmopolitan. He lived at Novara and Milan, in the autumn of 1800 became a sub-lieutenant, but resigned in 1802 and lived in Paris, where he frequented the *salons* and the theatres. After attempting business in Marseilles, in Aug. 1805, out of infatuation for an actress, he became a deputy to the *commissaire des guerres* (1806). He then lived for two years beyond the Rhine in Brunswick. He asked vainly for a post in Spain, was appointed auditor to the *Conseil d'Etat* in 1810, took part in the Russian campaign and became *intendant* at Sagan in 1813. He had received his baptism of fire in the Alps; he had heard the cannon at Marengo, and seen close at hand the horrors of war at the burning of Ebersberg, Smolensk and Moscow. During the tragic hours at the Beresina, he presented himself, freshly shaved, to Daru, who by this recognized in him an *homme de coeur*. After the empire fell, having lost all hope of being appointed a prefect, he went to live in Italy (1814–21); he spent six years in Milan, where he became acquainted with Byron, Madame de Staël, Silvio Pellico, Manzoni and Monti, whom he calls "the greatest living poet."

In 1814, he published in Paris *les Lettres écrites de Vienne en Autriche sur le célèbre compositeur, Joseph Haydn, suivies d'une vie de Mozart et de considérations sur Métastase et l'état présent de la musique en France et en Italie, par L. C. A. Bombet.* Herein he pillaged Carpani, Schlichtegroll, Winckler and Cramer. His plagiarisms would have brought him less reproach if he had acknowledged them more generously. *L'histoire de la peinture en Italie, par M.B.A.A.* (Beyle, *ancien auditeur*), appeared in 1817. Stendhal explains and animates the works of the abbé Lanzi and Vasari; he adds trivial or profound reflections to his borrowings from Lavater and Bossi. In *Rome, Naples et Florence* (1817), Stendhal supplemented his recollections with abundant documentary research; the work contains a wealth of fine, ingenious and vivid impressions on life, music, Italian patriotism, and on the charm of the delicious life he was enjoying. In this respect, this work seems to contain the matter of and to be a commentary on *La Chartreuse de Parme.*

In 1821 Stendhal was suspected of *carbonarisme* and espionage, and had to return to Paris. Exiled from his terrestrial paradise in Italy, he visited the houses of La Pasta, de Tracy and Delécluze, and made friends of Mérimée and Jacquemont. A brilliant and paradoxical talker, he both pleased and scandalized. He attempted to reconcile his dandyism with his poverty. He travelled in England in 1821 and 1826 and sent regular accounts and articles to the various English reviews, and to French journals.

He continued to publish literary hack-work: the amusing and anecdotal *Vie de Rossini* (1823) in which his dilettante spirit demanded from music a sentimental pleasure and poetic reverie. *Les Promenades dans Rome* (1829) was a partial success. The "cicerone of intelligence and taste" was not content with anecdotes; he described manners, customs and the art of the *"chasse au bonheur"* in Italy; and the guide may again be seen in the account of the landscapes and monuments which enchanted the author. In 1838 appeared *Les Mémoires d'un Touriste.* Stendhal had travelled extensively through France, either alone or in company with Mérimée, who instructed him in the Gothic. In addition to *choses vues,* the description of his native province, the Dauphiné, and the fair of Beaucaire, the work contains many stories clipped from journals or taken from the impressions of others. But it is saved by its vivid and feeling style, and by its "precious nothings." Stendhal also wrote some more original works. *De l'amour* (1822) constitutes under the guise of a psychological and documented analysis, a study and a preface for his novels; physiology is mingled with psychology. In it is also found a curious and original theory of "crystallization," a symbolic picture of the birth and growth of love.

Stendhal's first novel was an interesting study of this "crystallization," *Armance ou quelques scènes d'un salon de Paris en 1827* (1827). This was followed by a novel dealing with love and ambition—*Le Rouge et le Noir* (1830). In reading the *Gazette des Tribunaux,* which he cynically called "the golden book of French energy in the 19th century," Stendhal had noticed a tragic incident. The son of a farrier of the Dauphiné, a former seminarist, had been guillotined at Grenoble on Feb. 23, 1828, for having shot and wounded a lady whom he loved. On this scandalous material the novelist constructed his story and even wove in some personal confessions, for his Julien Sorel sometimes resembles the author. The interest of this novel is found in the energetic spirit of his characters, full of intense and unscrupulous life, in the dramatic adventures, traits of character and passion noted with precision and clearness, and in its psychological analysis.

Stendhal had already shown himself to be a pamphleteer. It was by a real pamphlet that he took his position in the Romantic battle, when he published in 1823 *Racine et Shakespeare,* which was re-published in 1825 with more vigour and more mordancy than ever. This work was impregnated with the Italian Romanticism of Manzoni, Visconti and Berchet. Stendhal attempted to free the theatre from the "unities" of place and time. He condemned Alexandrine verse and recommended a national tragedy which would be modern, even topical, and by preference liberal. These ideas were not all new, but the tone and the accent were completely personal. Stendhal had protested against the industrialism of Saint Simon in his witty work *D'un nouveau complot contre les industriels* (1825).

Shortly after the revolution of July 1830 Stendhal was appointed consul at Trieste, but since Metternich refused him his *exequatur,* he was sent in April 1831 to Civita-Vecchia. There he longed for Paris, the city of conversation. He wandered about Naples, made excavations in the papal territory, continually visited Rome and took holidays in Paris between 1836 and 1839. He had not yet had any true success when he was hailed as a great novelist by Balzac in an article published on Sept. 25, 1840, in the *Revue Parisienne,* on the *Chartreuse de Parme* (Eng. trans. *The Charterhouse of Parma*). He had just completed an advantageous contract with the *Revue des deux Mondes* for "nouvelles" when he was overcome on the boulevard by an attack of apoplexy on March 22, 1842. Three friends, including Mérimée, followed his funeral to the cemetery of Montmartre. He counted on obtaining a little glory about 1880 or even later: "I take a ticket in the lottery, the grand prize of which may be summed up as—to be read in 1935."

He was read long before this, and *La Chartreuse de Parme* is one of the great romances of the 19th century. The hypothesis which saw in it only an ingenious reproduction of ancient Italian chronicles which Stendhal had dug out of the libraries has been exploded. The value of the novel lies in the analysis of the *grandes passions* and in his pictures of contemporary manners. It differs widely from the rude and simple stories which he liked to take as his inspiration. Stendhal has put in this book his experiences and intimate feelings, his impressions of Italy revived by nostalgia. This book was seven years in contemplation and it was finally written in Paris in seven weeks, appearing in April 1839. The first part of the book consists of memories rearranged by the author—of Milan on the morrow of Lodi, of the adventures of Fabrice,

present at a great battle without seeing much of it, of France after Waterloo, of the intrigues of a little Italian court in 1820. The book is full of Stendhal's reading on Italy of former days which gives rise to certain discrepancies. The actors are original characters, even if one sees the prototype of Count Mosca in Metternich or in Count Sauron, the governor of Lombardy. The Duchess of Sanseverina differs greatly from la Vannoza, the mistress of the future Pope Alexander VI, and Fabrice del Dongo, the Napoleonic marquis who became a priest after taking part in a conspiracy, has more resemblance to Julien Sorel than to Alexander Farnese. Stendhal excels in the delineation of souls and of exceptional and complex characters, scheming and energetic. Portraits, landscapes and Italian life play a large part in this crowded and disconcerting romance. The style is always natural, clear and limpid. The "little precise phrases, worthy of a code or of algebra" raised the enthusiasm of Taine who regarded Stendhal as a "superior spirit." From his "Italian chronicles" he repeatedly drew dark and bloody stories and horrible and romantic novels. He was amused by these stories which he published between 1829 and 1838 in the *Revue de Paris: Vanina Vanini, Le coffre et le revenant, Le Philtre. Vittoria Accoramboni, Les Cenci, La duchesse de Palliano, L'Abbesse de Castro, San Francesco a Ripa,* appeared in the *Revue des deux Mondes* (1837–53).

Stendhal's admirers have published his unpublished works and even some illegible fragments preserved in the library at Grenoble. With audacious disquieting sincerity, sometimes bordering on cynicism, Stendhal wrote autobiographies entitled *La Vie d'Henri Brulard* (1890), *Souvenirs d'Egotisme* (1892), *Lucien Leuwen* (1894), *Journal d'Italie* (1911), in order to enjoy his past and to rescue from oblivion his "so different" personality. *Lamiel* (1889) is a romance in the style of Balzac and Sue of a foundling child of a seduced peasant woman. In his correspondence, vivid, primitive, amusing, passing from irony to tenderness, he paints himself with somewhat crude freedom. A place apart must be given to the *Lettres à Pauline,* his younger sister, whom Stendhal made his confidante. *La Vie de Napoléon* (1876) is less original than might have been expected. History is mingled with legend and gossip.

Almost unknown during his lifetime, Stendhal the psychologist, an expert in analyzing human sentiments, careful to "write philosophically" for a fastidious *élite,* for the "happy few," was one of the introducers of romanticism and of the psychological novel.

(J. Gir.)

Bibliography.—Paul Hazard, *La vie de Stendhal* (1927), trans. by Eleanor Hard (1929); Lytton Strachey in *Books and Characters* (1922); Matthew Josephson, *Stendhal; or the Pursuit of Happiness* (1946); Howard Clewes, *Stendhal* (1951); Robert Sage (ed. and tr.), *Private Diaries of Stendhal* (1954).

STENO, NICOLAUS (Nils Steensen) (1638–1686), Danish physician and theologian who also wrote on geology. He was born at Copenhagen on Jan. 10, 1638. He studied medicine at the University of Copenhagen and in 1660 he went to Amsterdam and then to Leyden to continue his studies of anatomy. While at Amsterdam he discovered the parotid duct known as *Ductus Stenonianus.* In 1664 he went to Paris, where he delivered a discourse on the anatomy of the brain that was printed in 1669. This paper was notable for its intelligibility and freedom from traditional interpretation. In 1665 he went to Florence, where he was appointed as physician to the Grand Duke Ferdinand II.

Steno had time for travel and study and in 1669 published his geological observations in *De Solido Intra Solidum Naturaliter Contento Dissertationis Prodromus.* In this short work he recognized fossils as remains of organisms and concluded that many rocks had formed in muddy marine or fresh waters. He rejected the idea that mountains grew, recognized the nature of volcanoes, and declared that some mountains had been shaped by running water. He attributed other mountains to "exhalations" and thought strata had been tilted and deformed as the roofs of great caverns collapsed.

Steno became a Catholic in 1667 and soon abandoned science for religion. In 1676 he was made a bishop and was appointed apostolic vicar of northern Germany and Scandinavia. After a stormy career in Germany he died at Schwerin in November 1686.

See *The Prodromus of Nicolaus Steno's Dissertation Concerning a Solid Body Enclosed by Process of Nature Within a Solid,* Eng. trans. with introd. by J. G. Winter (1916). (M. A. F.; C. L. Fe.)

STENOTYPY is the substitution of a printed character for the usual shorthand phonogram. This is accomplished by means of a machine called the stenotype. The machine is equipped with 23 keys and a numeral bar and prints 16 letters of the alphabet, plus 11 combinations of these letters to form the complete alphabet, also with the numerals from zero to nine and an asterisk. The mechanism is such that any one or any combination of the 16 letters or numbers can be printed at one time on one line of a strip of paper which is fed automatically through the back of the machine. This permits the printing of several short words, a complete word or the contraction of a word at a single stroke. The keys are so arranged that the initial consonants are under the fingers of the left hand, the vowels under the thumbs, and the final consonants under the fingers of the right hand. Duplication of consonants accounts for the five additional keys.

STENSIÖ, ERIK HELGE OSWALD ANDERSSON (1891–), Swedish paleontologist, was born on Oct. 2, 1891, at Döderhult, Kalmar. Educated at Uppsala university, in 1923 he became professor and director of the department of paleozoology, Swedish Museum of Natural History, Stockholm. He led a series of seven expeditions to Spitsbergen and described the fossil fish collected in the Devonian and Trias rocks there. His doctorate thesis on fossil fishes from Spitsbergen set an entirely new standard for the interpretation of such materials, and a later paper described and interpreted the structure of the head of Cephalaspids with extraordinary completeness, showing that they were very closely related to the still-existing lampreys. Stensiö's methods of detailed study of such materials were applied to fossil fish from many geological horizons in scattered places over the world and revolutionized knowledge and understanding of the evolution of fish and their precursors. He was awarded many distinctions and became a foreign honorary member of the Royal societies of London and Edinburgh. (D. M. S. W.)

STENTOR, one of the Greeks before Troy (*Iliad* v. 783), whose "stentorian" voice was as loud as that of 50 men.

STEPHANITE, which was an important ore of silver in the famous Comstock lode in Nevada, is a mineral consisting of silver, antimony and sulfur, Ag_5SbS_4, containing 68.5% of silver. It occurs with other ores of silver in metalliferous veins. Under the name *Schwarzerz* (black ore) it was mentioned by G. Agricola in 1546. Stephanite has a hardness of 2.5 and is very brittle; the specific gravity is 6.3.

STEPHANUS BYZANTINUS (fl. probably early and middle 6th century), was the author of *Ethnica,* a geographical encyclopaedia. This gave the names of the countries and places of the known world, listed alphabetically, with information on their topography, history, mythology and etymology. The *Ethnica* became a standard work of reference for the Byzantines, but of the whole work, only a later bare epitome has come down to us. The scope of the original can be seen in a few remaining fragments. One of these (some under the letter Δ) has survived in its own right; the rest are preserved in other authors, especially Constantine Porphyrogenitus, *De thematibus* (a description of Sicily), and *De administrando imperio* (an account of Spain, and a part of Persia). (T. A. R.)

STEPHEN, in some senses the greatest figure in primitive Christianity prior to Paul's conversion, was one of "the Seven" (Acts xxi. 8, nowhere called "deacons") set over the "daily ministration" toward the needy members of the Jerusalem community. Like Philip and perhaps others of his colleagues (vi. 3), he had higher gifts than his office would suggest. He was "full of faith and of holy Spirit"; and as his spiritual power showed itself in mighty deeds as well as words (vi. 5, 8), he became a marked man in Jerusalem. Himself a Jew of Greek culture, he naturally tried to win over his fellow Hellenists (vi. 9).

It is here that Stephen's advance upon prior Apostolic teaching becomes apparent. His special "wisdom" lay in greater insight into the merely relative nature and value of the externals of Israel's religion, particularly those connected with the Temple. His fellow Hellenists, as a body, were eager to disprove the feel-

ing of the native "Hebrews" that they were only half Jews. Hence teaching which minimized the value of the sacred "customs which Moses had delivered" (vi. 14), by making "salvation" depend simply on faith in Jesus as Messiah, would cause deep resentment in such circles, in spite of their more liberal attitude to things non-Jewish. For in Jerusalem the Temple overshadowed men's thoughts touching the Divine presence. To this he would reply in the spirit of the prophets, that the heart is the true seat of the Shekinah; and that if they refused God manifest in His Messiah, no holy "customs"—no, not the Temple itself—could save from the displeasure of the living God. Nay, he argued, quoting words of Jesus (Matt. xxvi. 61, Mk. xiv. 58, Jo. ii. 10) which were easy to misquote (Acts vi. 14), that the Temple might even be destroyed, as it had been in the past, without loss to true religion. But they could not rise to this conception and treated his words as "blasphemous," and roused "the people and the elders and the scribes" against him.

He was seized and brought before the Sanhedrin on the charge of speaking "against the Temple and the Law" (vi. 11–14). His defence took the form of a survey of Israel's religious past, with a view to show: (1) that "the God of Glory" had covenant relations with their forefathers before they had either Holy Place (Land or Temple) or Law (vii. 1–17); (2) that the first visible meeting place between God and His people was far other than that for which absolute sanctity was now claimed. Indeed, the form of "the tabernacle of testimony in the wilderness" (no Holy Land) had more divine sanction than any later Temple (44–47); (3) that, after all, the presence of "the Most High" was not bound up with any structure of human hands, as Isaiah witnessed (48–50). The moral of all this was plain: Israel's *forms* of fellowship with the Most High had all along been relative and subject to change. Hence there was no "blasphemy" in suggesting that in the Messianic age yet another change might come about, and that observance of Temple services could prove little as to acceptance with God. But there is another and more urgent line of pleading. This is found in the elaborate section dealing with the person and work of Moses, the great lawgiver (17–38)—a section full of extra-biblical touches—followed by one on Israel's hardness of heart towards him, together with its result, the Exile (39–43). Pure and original Mosaism is represented as something which in its full spiritual intention had been frustrated by Israel's stiff-neckedness (39, 42 *seq.*). The figure of Moses is made to stand forth in ideal outlines, the thinly-veiled Christian application shining through. "This is that Moses who said unto the children of Israel, 'A prophet shall God raise up unto you . . . like unto me:' who received living oracles to give unto us: to whom our fathers would not be obedient, but thrust him from them, and turned back in their hearts. . . ." (38 *seq.*). Here we have the very situation as between Stephen and his hearers; and it is made unmistakable by the speaker's closing words (51–53). Had they kept the Law dutifully, they would have believed on Him in whom true Mosaism was fulfilled and transcended. The author of Acts probably owed his report of Stephen's speech (as of his whole story) to Philip the "evangelist," who had been one of the Seven (xxi. 8). Possibly also Paul had spoken in Luke's hearing of Stephen's death and his own part in it (vii. 58, 60, *cf.* vi. 9).

Stephen's martyrdom is described as tumultuary in character, though the legal forms of stoning for blasphemy were observed (vii. 58); nor is it inconceivable that an act exceeding Jewish rights under the Romans should have taken place at the sudden impulse of religious fanaticism.

See articles in the *Ency. Bib.* vol. iv., and Hauck's *Realencykl. f. protestant. Theol. u. Kirche*, vol. xix. (J. V. B.)

STEPHEN, the name of ten popes, sometimes numbered as nine, for reasons explained under *Stephen II.*

St. Stephen I (d. 257), pope from 254 to 257, a Roman by birth, succeeded Lucius I. He is known principally from three controversies reported in the letters of St. Cyprian of Carthage. Stephen restored two Spanish bishops, accused of having "lapsed" under persecution (*libellatici*), and apparently he refused to depose Bishop Marcianus of Arles, accused of rigourism (Novatianism). Stephen also objected to the practice of rebaptizing heretics and threatened to excommunicate bishops in Africa (including Cyprian) and in Asia Minor unless they discontinued that practice. His death on Aug. 2, 257, almost coincided with a renewal of persecution under Valerian, but it is doubtful that he was martyred. His feast day is Aug. 2. (F. X. G.)

Stephen II (d. 752), a Roman, was elected March 24, 752 and died two days later, before having been consecrated. He is listed neither in the *Liber pontificalis* nor in the catalogue of the popes. His successor, who also took the name Stephen is called both Stephen II and Stephen III.

Stephen III (II) (d. 757), pope from 752 to 757, was a deacon when chosen in March 752 to succeed Zacharias in place of Stephen II. The central act of his pontificate was to free the papacy from Byzantium and to ally it with the Franks. Twice he appealed to the Frankish king Pippin III for assistance against the Lombard king Aistulf, who was trying to conquer all Italy. In 754 he anointed Pippin at St. Denis and obtained from him the Donation of Quiercy. In 756 Pippin subdued the Lombards and conferred on Stephen territory in the exarchate of Ravenna, the duchy of Rome and the districts of Venetia and Istria. Thus the papal states were founded, independent of any temporal power, and Stephen became the first temporal sovereign. He died in April 757.

Stephen IV (III) (d. 772), pope from 768 to 772, was a Sicilian, a Benedictine, and succeeded Paul I, on Aug. 7, 768. His pontificate was mainly occupied with reaction to the antipapal policies of the Lombard king Desiderius, who attempted an alliance with the Franks against Stephen. He died in Feb. 772.

Stephen V (IV) (d. 817), pope from 816 to 817, succeeded Leo III in June 816. Immediately after his consecration he ordered the Romans to swear fidelity to Emperor Louis the Pious whom he personally visited in the following August. His crowning of Louis at Reims solidified the alliance between the papacy and the Franks. He died in Jan. 817.

Stephen VI (V) (d. 891), pope from 885 to 891, succeeded Adrian III. In his dealings with both east (Photius and the Slavonic Church) and west (Ravenna and Bordeaux) he acted with the utmost independence and authority.

Stephen VII (VI) (d. 897), pope from 896 to 897, succeeded Boniface VI. His conduct toward the remains of Pope Formosus which he exhumed, placed on trial, degraded and desecrated, ended in his own imprisonment and strangulation in Aug. 897. His treatment of Formosus may have been influenced by Emperor Lambert.

Stephen VIII (VII) (d. 931), pope from 928 to 931, was born of a Roman family and as cardinal priest of St. Anastasia was active in the administration of the Roman Church before his election to the papacy in Dec. 928. The history of his pontificate is practically unknown. He died in Feb. 931.

Stephen IX (VIII) (d. 942), pope from 939 to 942, was a Roman elected on July 14, 939, to succeed Leo VII. His pontificate was dominated by the Roman prince Alberic so that Stephen had little opportunity for independent action. All his political efforts were directed toward the support of the last Carolingians. He died in Oct. 942.

Stephen X (IX) (d. 1058), pope from 1057 to 1058, succeeded Victor II in Aug. 1057, five days after the latter's death. His baptismal name was Frederick, and he was young brother of Godfrey, duke of Lorraine and marquis of Tuscany. Frederick, who had been raised to the cardinalate by Leo IX, acted for some time as papal legate at Constantinople and was with Leo in his unlucky expedition against the Normans. He showed great zeal in enforcing the law of clerical celibacy. During his pontificate the reform program became clearer, stronger and more centralized. He died on March 29, 1058, at Florence. (R. E. McN.)

STEPHEN (1097?–1154), king of England, was the third son of Stephen Henry, count of Blois and Chartres, and, through his mother Adela, a grandson of William the Conqueror. Born some time before 1101, he was still a boy when he was taken into favour by his uncle, Henry I of England, and received the honour of knighthood and the county of Mortain. In 1118 he severed his connection with Blois and Chartres, renouncing his hereditary claims in favour of his elder brother Theobald. But he acquired the county of Boulogne by marrying Matilda (*c.*

1103–1152), the heiress of Count Eustace III and a niece of Henry's first wife. The old king arranged this match after the untimely loss of his son, William Atheling, in the tragedy of the White Ship; until 1125 Stephen was regarded as the probable heir to the English throne. But on the return of the widowed empress Matilda (*q.v.*) to her father's court, Henry compelled Stephen and the rest of his barons to acknowledge her as their future ruler (Jan. 1, 1127). Seven years later these oaths were renewed; and in addition the ultimate claims of Matilda's infant son, Henry of Anjou, were recognized (1133). But the death of Henry I found the empress absent from England. Stephen hurried across the Channel and began to canvass for supporters, arguing that his oaths to Matilda were taken under coercion, and that she, as the daughter of a professed nun, was illegitimate.

He was raised to the throne by the Londoners, the official baronage and the clergy; his most influential supporters were the old justiciar, Robert, bishop of Salisbury, and his own brother Henry, bishop of Winchester. Innocent II was induced by Bishop Henry to ratify the election, and Stephen thus cleared himself from the stain of perjury. Two charters of liberties, issued in rapid succession, confirmed the King's alliance with the Church and earned the good will of the nation. But his supporters traded upon his notorious facility and the unstable nature of his power. Extortionate concessions were demanded by the great barons, and particularly by Earl Robert of Gloucester, the half-brother of the empress. The clergy insisted that neither their goods nor their persons should be subject to secular jurisdiction. Stephen's attempt to create a mercenary army and a royalist party led at once to a rupture between himself and Earl Robert (1138), which was the signal for sporadic rebellions. Soon afterward the king attacked the bishops of Salisbury, Ely and Lincoln—a powerful family clique who stood at the head of the official baronage—and, not content with seizing their castles, subjected them to personal outrage and detention. The result was that the clergy, headed by his brother, the bishop of Winchester, declared against him (1139).

Stephen was thoroughly discredited when the empress at length appeared in England (Sept. 30, 1139). Through a misplaced sense of chivalry he declined to take an opportunity of seizing her person. She was therefore able to join her half-brother at Gloucester, to obtain recognition in the western and southwestern shires, and to contest the royal title for eight years. Stephen's initial errors were aggravated by bad generalship. He showed remarkable energy in hurrying from one centre of rebellion to another, but he never ventured to attack the headquarters of the empress. In 1141 he was surprised and captured while besieging Lincoln castle. The empress in consequence reigned for six months as "Lady (*Domina*) of the English"; save for her faults of temper, the cause of Stephen would never have been retrieved. But, later in the year, his supporters were able to procure his release in exchange for the earl of Gloucester. After an obstinate siege he expelled Matilda from Oxford (Dec. 1142) and compelled her to fall back upon the west.

The next five years witnessed anarchy such as England had never before experienced. England north of the Ribble and the Tyne had passed into the hands of David of Scotland and his son, Prince Henry; Ranulf earl of Chester was constructing an independent principality; on the west the raids of the Angevin party, in the east and midlands the excesses of such rebels as Geoffrey de Mandeville, earl of Essex, turned considerable districts into wildernesses. Meanwhile Geoffrey of Anjou, the husband of the empress, completed the conquest of Normandy (1144). In 1147 the situation improved for Stephen; Robert of Gloucester, the ablest of the Angevin partisans, died, and the empress left England in despair. But her son soon appeared in England to renew the struggle (1149) and conciliate new supporters. Soon after his return to Normandy Henry was invested by his father with the duchy (1150). He succeeded to Anjou in 1151; next year he acquired Aquitaine by marriage.

Stephen struggled hard to secure the succession for Eustace, his elder son. But he had quarrelled with Rome respecting a vacancy in the see of York; the pope forbade the English bishops to consecrate Eustace (1151); and there was a general unwillingness to prolong the civil war. Worn out by incessant conflicts, the king bowed to the inevitable when Henry next appeared in England (1153). Negotiations were opened; and Stephen's last hesitations disappeared when Eustace was carried off by a sudden illness. Late in 1153 the king acknowledged Henry as his heir, only stipulating that the earldom of Surrey and his private estates should be guaranteed to his surviving son, William. The king and the duke agreed to co-operate for the repression of anarchy; but Stephen died soon thereafter (Oct. 1154). (H. W. C. D.)

STEPHEN I (ST. STEPHEN) (977–1038), king of Hungary, was the son of Geza, duke of Hungary, and of Sarolta, one of the few Magyar Christian ladies. His tutors included the German priest Bruno, the Czech priest Radla, and an Italian knight, Theodate of San Severino, who taught him arms and letters. In 996 Stephen married Gisela, the daughter of Duke Henry II of Bavaria. In the following year his father died and the young prince was confronted by a formidable pagan reaction between the Drave and Lake Balaton. Stephen hastened against the rebels, bearing before him the banner of St. Martin of Tours, whom he chose to be his patron saint, and routed the rebels at Veszprem (998), assuming the royal title immediately afterward.

The rest of Stephen's life was spent largely in a struggle against the pagan nobles, which engrossed his energies, and compelled him to adopt a pacific policy toward the emperors of the east and west; but when attacked by the emperor Conrad in 1030 Stephen repelled the invasion successfully. He died at Esztergom in 1038 and was canonized in 1083. For an account of his reforms *see* HUNGARY: *History*.

STEPHEN, ALEXANDER (1795–1875), Scottish shipbuilder who reconstructed the firm of Alexander Stephen & Sons, now of Linthouse, Glasgow, which, from World War I onward, supplied the British admiralty with more than 50 ships, was born in 1795 at Aberdeen. In 1828 he took over the family firm which his grandfather Alexander Stephen, had started at Burghead *c.* 1750. As it prospered the yard was moved successively to Arbroath, Dundee, the river Clyde at Kelvinhaugh, Glasgow, in 1850 and then to Linthouse, two miles away, in 1869. Stephen's first iron ship was completed in 1852 and his first steamship, the "William McCormick," of 840 tons, in 1854. In 1863, the "Sea King," a full-rigged auxiliary screw steamer of 1,152 tons, was built for the China tea trade; this later became the "Shenandoah," a Federal raider in the American civil war. Among other famous ships built by the firm were the "Terra Nova" (1884), the "Nimrod" (1866) and the "Bear" (1874), used respectively by Capt. Robert Scott, Sir Ernest Shackleton and Adm. Richard Byrd in their expeditions to the antarctic. During the 1950s the chairman of the firm was Sir Alexander Murray Stephen (1892–), great grandson of Alexander, who moved the firm to Glasgow and who died at Dundee on April 24, 1875. (J. F. BR.)

STEPHEN, SIR JAMES FITZJAMES (1829–1894), English judge, law reformer and historian whose work in the codifying of criminal law was of wide and lasting influence, was born in London on March 3, 1829. He was the son of James Stephen, for many years undersecretary of state for the colonies, and an elder brother of Leslie Stephen, English philosopher and critic. Stephen was educated at Eton, London university and Trinity college, Cambridge. He was called to the bar in 1854, was made recorder of Newark in 1859, and became queen's counsel in 1868. In his early years at the bar, Stephen became one of the leading controversial publicists of his day. His miscellaneous writings, which encompass a wide range of topics, appeared frequently in such journals as the *Pall Mall Gazette* and the *Saturday Review*. Stephen's intellectual position may, in some respects, reflect the utilitarian tradition; but his thought was of a highly individual cast, and he was in no sense the disciple of any school. In 1863 he published his *General View of the Criminal Law of England*, a work of great importance, representing the first attempt since Blackstone to give systematic statement to the principles of English criminal jurisprudence. Ten years later, in reply to J. S. Mill's *On Liberty*, Stephen published *Liberty, Equality, Fraternity*, a work antidemocratic in tendency and displaying de-

fects of expression and argument, which, nevertheless, remains one of the important statements of the conservative political philosophy.

In 1869 Stephen accepted the post of legal members of council in India, and, like Macaulay and Sir Henry Maine before him, became engrossed in the reform and recodification of Indian law. Throughout these labours Stephen displayed remarkable energy and efficiency. He returned to England in 1872 and almost immediately became identified with the movement to codify the criminal law of England. His *Digest* of the criminal law (1877) was enthusiastically received. He was the principal draftsman of the Indictable Offences bill, which was, in effect, a criminal code. Because of powerful opposition, the bill was never introduced in parliament. Although failing of their immediate objectives, Stephen's efforts influenced the drafting of legislation in the nations of the British commonwealth, and, even today, are consulted in attempts to reformulate the criminal law of English-speaking countries. In 1883 Stephen published his great work, *History of the Criminal Law of England*. Although dogmatic in tone, often deficient in the critical handling of sources, and in many respects incomplete, the work represents a remarkable achievement which, on the whole, has never been surpassed.

In 1879 Stephen became a judge in the queen's bench division of the high court. He resigned in 1891, after a breakdown, and received a baronetcy in recognition of his services. He died in Ipswich on March 11, 1894. A selection of his letters, edited by his daughter, Caroline Emelia Stephen, was published in 1907.

Bibliography.—Sir Leslie Stephen, *Life of Sir James Fitzjames Stephen* (1895) with bibliographical appendix, is one of the outstanding biographies in the English language; a shorter account by the same author will be found in the *Dictionary of National Biography*. *See* also Sir C. P. Ilbert, "Sir James Stephen as a Legislator," *Law Quart. Rev.*, x. 222; L. Radzinowicz, *Sir James Fitzjames Stephen*, Selden Society Lecture (1957). (F. A. A.)

STEPHEN, SIR LESLIE (1832–1904), English critic, man of letters and first editor of the *Dictionary of National Biography*, was born in London on Nov. 28, 1832, into a distinguished intellectual family. His grandfather was a member of the Clapham sect, his father a fine colonial servant, his brother a great criminal lawyer and Anglo-Indian administrator; the Venns and Diceys were his first cousins. He was educated at Eton, at King's college, London, and at Trinity hall, Cambridge, where he was elected to a fellowship in 1854 and became junior tutor in 1856. He won fame as a rowing coach and inaugurated the Oxford and Cambridge athletics contest; he also tried unsuccessfully to reform the outdated university curriculum. He had been ordained in 1855, but his philosophical studies, combined probably with the controversy that followed the publication of Charles Darwin's *Origin of Species* (1859), caused him to lose his faith; in 1862 he resigned his tutorship and two years later left Cambridge to live in London.

Through his brother, James Fitzjames, a contributor to the *Saturday Review*, Stephen gained an entry to the literary world, contributing to many periodicals. In politics he was a radical and supported the North during the American Civil War, and his articles in defense of agnosticism were republished in *Essays on Freethinking and Plainspeaking* (1873) and *An Agnostic's Apology* (1893). In 1871 George Smith offered him the editorship of the *Cornhill*, for which he wrote literary criticism later republished in the three series of *Hours in a Library* (1874–79). Stephen was the first serious critic of the novel and his work still deserves consideration by historians of literary criticism. Thomas Hardy, R. L. Stevenson, Edmund Gosse and Henry James were among those whom Stephen, as an editor, encouraged. After 11 years he resigned but continued to write for periodicals. *Studies of a Biographer* (4 vol., 1898–1902) and *Social Rights and Duties* (1896) contain the best of this work. He also wrote a short book on *Hobbes* (1904) and his Ford lectures, *English Literature and Society in the Eighteenth Century* (1904), were pioneer work in the sociological study of literature.

His greatest learned work was his excellently written *History of English Thought in the Eighteenth Century* (1876). *The English Utilitarians* (1900) was less compact and successful though it is still a useful source. His philosophical contribution to rationalist tradition, *Science of Ethics* (1882), attempted to wed evolutionary theory to ethics. His most enduring bequest to posterity is the *Dictionary of National Biography*, of which he edited the first 26 volumes and to which he contributed many biographies. In recognition of this service to letters he was created knight commander of the Bath in 1902 and received other honours. He was a leader among Victorian mountaineers and was president of the Alpine club (1865–68) and editor of the *Alpine Journal* (1868–71). Some of his more famous climbs are recorded in *The Playground of Europe* (1871). He was also the leader of "The Tramps," a club of intellectuals who enjoyed 30-mi. Sunday walks.

Stephen was highly respected in intellectual society and remembered for many acts of kindness to his friends. He was formidable in appearance, dispassionate in judgment and the enemy of whatever he considered to be cant or wilful obscurity which was designed to conceal the truth. He was shy and given to silence, the more so after the death in 1875 of his first wife Minny, the second daughter of Thackeray. In 1878 he married Julia Jackson, the widow of Herbert Duckworth, and among their four children were the painter Vanessa Bell and the novelist Virginia Woolf. He died in London on Feb. 22, 1904.

Bibliography.—F. W. Maitland, *Life and Letters of Leslie Stephen* (1906); N. G. Annan, *Leslie Stephen: His Thought and Character in Relation to His Time* (1951); and—for a list of Stephen's writings—L. Stephen, *Men, Books and Mountains*, ed. by S. O. A. Ullmann (1956). (N. G. A.)

STEPHEN (ISTVÁN) BÁTHORY (1533–1586), king of Poland and prince of Transylvania. Báthory spent his early years at the court of the emperor Ferdinand, subsequently attached himself to Janos Zapolya, and won equal renown as a soldier and diplomatist. Zapolya rewarded him with the voivodeship of Transylvania, where his defense of the rights of his patron's son, John Sigismund, incurred the animosity of the emperor Maximilian, who kept him in prison for two years. On May 25, 1571, on the death of John Sigismund, Báthory was elected prince of Transylvania by the Hungarian estates, in spite of the opposition of Vienna. He expelled the rival candidate, Gaspar Békesy, from Transylvania (1572). In 1575 the Polish nobility, at the instigation of Zamoyski (*q.v.*), elected Báthory king of Poland, in opposition to the emperor Maximilian, the candidate of the senate. Báthory persuaded the Transylvanian estates to elect his brother Christopher in his stead, hurried to Cracow, espoused the princess Anne, the sister of the last Jagiello, and was crowned on April 29, 1576. The leading events of Báthory's reign can here only be briefly indicated. All armed opposition collapsed with the surrender of Danzig, after a six months' siege, on Dec. 16, 1577. Stephen was able to devote himself to foreign affairs. The difficulties with the sultan were temporarily adjusted by a truce signed on Nov. 5, 1577; and the Diet of Warsaw having voted supplies, Stephen embarked on war with Muscovy. Penetrating to the heart of the country, he besieged Pskov (Aug.–Dec. 1581), and forced Ivan the Terrible to cede him Polotsk and Livonia (peace of Zapoli, Jan. 15, 1582). The chief domestic event of Stephen's reign was the establishment in Poland of the Jesuits, in furtherance of his designs of uniting Poland, Muscovy and Transylvania into one great state. The project was dissipated by his sudden death, of apoplexy, Dec. 12, 1586.

Bibliography.—1. Polkowski, *The Martial Exploits of Stephen Báthory* (1887); Paul Pierling, *Un Arbitrage pontifical au xvime siècle* (1890); Lajos Szadeczky, *Stephen Báthory's Election to the Crown of Poland* (1887).

STEPHENS, ALEXANDER HAMILTON (1812–1883), American statesman, vice president of the Confederate States during the Civil War, was born in Wilkes (now Taliaferro) county, Georgia, on Feb. 11, 1812. In 1828 he was sent by the Georgia Educational society to Franklin college (University of Georgia), where he graduated in 1832.

In 1836 he was elected to the Georgia house of representatives after a campaign in which he was vigorously opposed because he had attacked the doctrine of nullification, and because he had opposed all extra-legal steps against the abolitionists. He was annually re-elected until 1841; in 1842 he was elected to the State senate, and in the following year, on the Whig ticket, to the national House of Representatives. In this last body he urged

the annexation of Texas, chiefly as a means of achieving more power for the South in Congress. He vigorously supported the Compromise of 1850 and continued to act with the Whigs of the North until they, in 1852, nominated Gen. Winfield Scott for the presidency without Scott's endorsement of the compromise. In 1854 Stephens helped to secure the passage of the Kansas-Nebraska bill. He strongly opposed secession, but when Georgia seceded he "followed his state" and was elected vice president of the Confederacy. Throughout the war, he was so concerned about states' rights and civil liberty that he opposed the exercise of extra-constitutional war powers by President Jefferson Davis lest the freedom for which the South was fighting should be destroyed. His policy was to preserve constitutional government in the South and strengthen the anti-war party in the North by convincing it that the Lincoln administration had abandoned such government; to the same end he urged, in 1864, the unconditional discharge of Federal prisoners in the South. Stephens headed the Confederate commission to the peace conference at Hampton Roads in Feb. 1865. In the following May, after the fall of the Confederacy, he was arrested at his home and taken to Fort Warren, in Boston harbour, where he was confined until Oct. 12. In 1866 he was elected to the U.S. Senate, but was not permitted to take his seat. He was a U.S. representative (1873–82) and governor of Georgia in 1882–83, dying in office, at Atlanta, on March 4, 1883.

From 1871 to 1873 Stephens edited the *Atlanta Daily Sun*, and he published *A Constitutional View of the Late War Between the States* (1868–70), perhaps the best statement of the Southern position with reference to state sovereignty and secession; *The Reviewers Reviewed* (1872), a supplement to the preceding work; and *A Compendium of the History of the United States* (1875; new ed., 1883).

STEPHENS, JAMES (1882–1950), Irish poet and storyteller whose pantheistic philosophy is revealed both in his fairy tales set in the Dublin slums and in his compassionate poems about animals, was born in Dublin on Feb. 2, 1882. He was a solicitor's clerk when *The Crock of Gold* (1912), with its fabulous Celtic theme, established his fame. But he also knew urban poverty (his theme in *The Charwoman's Daughter*, 1912), and his astringent irony suggests affinities with James Joyce. He wrote *The Demi-Gods* (1914) in a similar vein, and (having learned Gaelic) *Deirdre* (1923) in more formal, rhythmic prose. Short stories and lyric poems constitute the remainder of his work. Stephens was actively connected with the Irish nationalist movement. By 1940 he was living in London, and having discovered himself as a broadcaster of genius, was frequently on the air from 1939 until his death in London on Dec. 26, 1950. (P. At.)

STEPHENS, JOHN LLOYD (1805–1852), U.S. traveler who studied Indian life and antiquities in Central America, was born at Shrewsbury, N.J., on Nov. 28, 1805, and was trained as a lawyer. Appointment to a mission in Central America gave him the opportunity to start explorations in company with his friend Frederick Catherwood, who made drawings of many of the monuments that they examined. Stephens wrote *Incidents of Travel in Central America, Chiapas and Yucatan* (2 vol., 1841), and *Incidents of Travel in Yucatan* (2 vol., 1843). These works reveal Stephens' interest in the ways of the Indians, and the impression made on him by the mixture of paganism with Christianity in their religion. Stephens died in New York city on Oct. 12, 1852.

STEPHENSON, the family name of three English engineers prominent in the development of railways.

George Stephenson (1781–1848), inventor and founder of railways, was born at Wylam, Northumberland, on June 9, 1781. His early life was spent in extreme poverty, but in 1804 he moved to Killingworth, Northumberland, supplementing his meagre income by repairing watches and clocks, and in 1812, after seriously considering emigrating, was appointed engine-wright of High Pit colliery at a salary of £100 per year. It was there that, in 1815, he designed and produced his version of the miners' safety lamp, while Sir Humphry Davy was still carrying out experiments. When Davy's lamp did appear there was violent controversy as to who should receive the credit for the invention.

About this time also Stephenson was given financial backing to produce for the colliery a "travelling engine" from pit-head to the port. It was a success and when, in 1822, the Stockton and Darlington railway was under construction, he persuaded the directors to use steam instead of animal traction. The railway opened on Sept. 27, 1825, and the first public passenger train in the world was drawn by Stephenson's "Active," later renamed "Locomotion." When the Liverpool and Manchester railway was nearing completion, Stephenson again persuaded the directors to try locomotives: a prize of £500 was offered for a suitable locomotive and was won by his "Rocket." Eight engines in all were employed when this line opened on Sept. 15, 1830, all having been built at Newcastle, in the works he founded with his cousin, Thomas Richardson, and a colleague, Edward Pease. Stephenson was engineer of many other railways in the midlands, but strongly disapproved of the "railway mania" which reached its climax in 1844. He was also consultant to many foreign companies, notably in Belgium and Spain. His last years were spent in retirement at Tapton house, Chesterfield, where he died on Aug. 12, 1848. (*See* also Railway: *Locomotives;* Steam.)

Robert Stephenson (1803–1859), who specialized in the construction of railway bridges, was born at Willington Quay, Northumberland, Oct. 16, 1803, the only son of George Stephenson. Educated at Bruce's academy, Newcastle-on-Tyne, and Edinburgh university, in 1819 he was apprenticed to Nicholas Wood at Killingworth. In 1821 he assisted his father in a survey for the Stockton and Darlington and Liverpool and Manchester railways, but failing health in 1824 led to his accepting a mining appointment in South America. Recalled to England in 1827 to take up work on locomotives at Newcastle, he was able to overcome the problem of securing boiler tubes, a problem that had been giving much trouble to the designers, and he was responsible for most of the improvements to locomotives at this time. In 1833 he was appointed engineer of the London and Birmingham railway; while holding this appointment he carried out a number of engineering feats, such as the Blisworth cutting and the Kilsby tunnel. This railway proved to be a great success and was the first to run into London from the industrial north.

Robert Stephenson, however, was particularly concerned with the construction of tubular girder railway bridges and his work extended to Sweden, Denmark, Belgium, Switzerland and Egypt. The more notable examples were the Britannia tubular bridge over the Menai straits, the tubular bridge over the river Conway and the Victoria tubular bridge over the St. Lawrence river at Montreal. In 1847 he entered parliament as conservative member for Whitby, retaining the seat until his death in London on Oct. 12, 1859. (*See* also Railway.) (T. M. S.)

George Robert Stephenson (1819–1905), the nephew of George Stephenson (1781–1848), was born at Newcastle-on-Tyne on Oct. 20, 1819. After working in colliery workshops near Manchester, he was sent to King William's school, Isle of Man. After his father's death in 1837 he worked for his uncle in the drawing office of the Manchester and Leeds railway. In 1860, when he was consulting engineer to the provincial government of Canterbury, New Zealand, the Lyttelton-Christchurch line was built under his direction. He returned to England and with Sir John Hawkshaw in 1864 built the East London railway. He was partly responsible for the construction of the Victoria tubular bridge across the St. Lawrence and designed many bridges in England and abroad. In 1859 he succeeded his cousin Robert Stephenson at the locomotive works at Newcastle-on-Tyne, and at the Snibston and Tapton collieries. He died at Cheltenham on Oct. 26, 1905.

STEPNEY, a metropolitan borough of London, Eng., which under the London Government Act 1963 was to be amalgamated on April 1, 1965, with Bethnal Green and Poplar to form the new London borough of Tower Hamlets (*see* London). The metropolitan borough of Stepney is bounded by Bethnal Green, Poplar, the river Thames, the City of London and Shoreditch. Pop. (1961) 91,940. Area 2.8 sq.mi. The thoroughfares of Mile End road, Whitechapel road and Commercial road traverse the borough from the east and converge near the City boundary, where stood the ancient Aldgate. Stepney marks the beginning of what is known as the East End of London. The area of Spitalfields

(*q.v.*) in the north has an old association with the silk-weaving industry which was established there from 1685. Also in the north are Whitechapel and Mile End. At Mile End Wat Tyler and Richard II met in 1381 and it was there that William Booth began the work of the Salvation Army in July 1868. In the south are Wapping, Shadwell, Ratcliff and Limehouse (*q.v.*). Just off Wapping stood Execution dock, once the hanging place for pirates. At Wapping Wall still stands the famous "Prospect of Whitby" inn, the original of which is said to date back to 1543. Wharves and factories line the river bank. In the extreme west the borough includes the Tower of London (*q.v.*), the Royal mint and the northern half of Tower bridge. In this area is Middlesex street with its Sunday street market, called Petticoat Lane, the former name of Middlesex street. Thomas Barnardo founded his first children's home in Stepney causeway in 1870. The Bernhard Baron St. George's Jewish settlement was opened in 1930 and in the Mile End road is Queen Mary college, of the University of London, part of which was formerly the People's palace. Stepney is a parliamentary constituency and a bishopric suffragan in the diocese of London. (HE. WA.)

STEPNIAK, SERGIUS (1852–1895), Russian revolutionist, whose real name was Sergius Michaelovich Kravchinski, was born in South Russia, of noble parents. He became an officer in the artillery; but his sympathy with the peasants, among whom he had lived during his boyhood in the country, developed in him at first democratic and, later, revolutionary opinions. He began secretly to sow the sentiments of democracy among the peasants, but in 1874 he was arrested. He succeeded in making his escape and immediately began a more vigorous campaign against autocracy. His sympathetic nature was revolted by the brutal methods adopted toward prisoners, especially political prisoners, and by the stern measures of repression of revolutionary sentiment. For a time he supported a terrorist policy, and in 1880 he was obliged to leave the country. After a few years he went to London. He was already known in England by his book *Underground Russia*, which had been published in London in 1882. He followed it up with other works on the condition of the Russian peasantry, on Nihilism, and on the conditions of life in Russia. Stepniak was killed by a railway engine at a level crossing at Bedford Park, Chiswick, on Dec. 23, 1895.

STEPPE, a term originally applied to the level, treeless, grassy plains in southern Russia. The term has been generalized to designate any dry grassy plain outside the tropics with a semi-arid climate as, for example, the drier sections of the North American Great Plains, large parts of southwest Africa, and the desert margins of Australia. It is also used to describe the climate of such areas—a steppe climate. *See* PLAIN; PRAIRIE. (W. C. C.)

STERCORARIIDAE: *see* SKUA.

STERCULIACEAE, the cacao or chocolate family of plants, consisting of about 50 genera with 1,200 species of tropical or subtropical trees, shrubs and herbs. Five genera (*Ayenia, Fremontodendron, Hermannia, Melochia* and *Waltheria*) have representatives in southern United States. The leaves of Sterculiaceae are alternate and simple or palmately compound; flowers, often unisexual, are solitary or clustered and have three to five sepals and five or no petals. The usually numerous stamens are separate or joined into a column and the superior ovary has two to five cavities or separate parts. The fruit is fleshy, or dry and generally capsular. Important plants of the family include cacao (*Theobroma cacao*), source of cocoa (*q.v.*) and chocolate; cola (*see* KOLA NUT) (*Cola nitida* and *C. acuminata*), which yields a stimulant used in cola drinks; and a few *Sterculia* species (especially, *Sterculia urens*), which yield karaya gum, an important substitute for tragacanth (*see* GUMS, PLANT). Species of a few genera are cultivated as ornamentals, notably bottle tree (*Brachychiton*), dombeya (*Dombeya*), Chinese parasol tree (*Firmiana*), flannel bush (*Fremontodendron*) and sterculia (*Sterculia*). In geologic times the family was distributed far beyond the limits of the present tropics and subtropics; *e.g.*, into Alaska. (J. W. TT.)

E. AUBERY DE LA RUE

BOTTLE TREE (BRACHYCHITON RUPESTRIS), COMMON TO AUSTRALIA

STEREOCHEMISTRY. Following are the main sections and divisions of this article:

I. Introduction
II. Early History of Stereochemistry
  1. Optical Activity
  2. Specific Rotation
  3. The Work of Pasteur
III. General Discussion of Stereoisomerism
  1. Symmetry Elements of Molecules
  2. Enantiomorphs
  3. Diastereomers
  4. Meso Forms
  5. Racemic Modifications
  6. A Word About Nomenclature
  7. Racemization
  8. Resolution
  9. Asymmetric Syntheses and Decompositions
  10. The Origin of Optically Active Substances in Nature
IV. The Stereochemistry of Carbon
  1. The Tetrahedral Carbon Atom
  2. Asymetric Carbon Atoms
  3. The Stereochemistry of Olefinic Compounds
  4. The Stereochemistry of Acetylenic Compounds
  5. Free Rotation
  6. Plane Projection Diagrams
  7. Use of Plane Projection Diagrams for the Prediction of the Number of Stereoisomers
  8. Optically Active Allenes and Their Analogues
  9. Optically Active Biphenyls
  10. Optical and Geometrical Isomerism
  11. Walden Inversion
  12. Optically Active Series
  13. Strain Theory
  14. Conformational Analysis
  15. Steric Hindrance
V. The Stereochemistry of Nitrogen
VI. The Stereochemistry of the Remaining Elements
  1. Elements of Co-ordination Number 3
  2. Elements of Co-ordination Number 4
  3. Elements of Co-ordination Number 6

I. INTRODUCTION

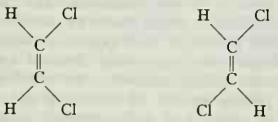
The term stereochemistry (from Gr. "solid," + chemistry) is applied in a wide sense to all that part of physical science which deals with those properties of chemical compounds which are believed to be due to the relative positions of the different atoms within a single molecule, or to the relative positions of different molecules in a larger aggregate of material. In this article, however, the term is used in a narrower sense. The following discussion is almost entirely limited to such aspects of the subject as are included in the field of stereoisomerism (*see* ISOMERISM), and it deals only with the stereochemical properties of individual molecules or ions. (For other aspects of the subject, which are largely ignored here, *see* CRYSTALLOGRAPHY: *Crystal-Structure Analysis;* ELECTRON DIFFRACTION; SPECTROSCOPY.)

One important goal of the structural theory of organic chemistry (knowledge of which is here assumed, *see* CHEMISTRY: *Organic Chemistry;* ISOMERISM) is to associate with the molecule (or with each of the ions, *see* MOLECULE) of each distinct compound a diagram, known as its structural formula or structure. The purpose of such a diagram is not to represent the true Euclidean geometry of the molecules of the substance; it is rather to specify the valences of the various atoms (*i.e.*, the numbers of lines which terminate at the corresponding atomic symbols) and the numbers of bonds between the various pairs of atoms (*i.e.*, the numbers of lines drawn between the corresponding pairs of atomic symbols).

About the middle of the 19th century, it was discovered that in certain instances the methods of structural chemistry led to the assignment of the same structural formula to more than one compound. It was, therefore, evident that the rules of structural chemistry were, in these instances, insufficient for the complete

identification of the compounds in question. The theories of stereochemistry (in the narrower sense) were devised as the corrections necessary to remedy this defect. It proved possible to account for the number of distinct compounds to be associated with each structural diagram if two molecules of the same substance were assumed always to have not only the same structure, but also the same arrangement of their atoms in space. This geometrical interpretation has received independent support from the data derived from studies of X-ray and electron diffraction and of spectra. Consequently, although the possibility of an alternative, nongeometrical explanation of the observed numbers of isomers has not been seriously explored, the correctness of the geometrical interpretation is generally accepted without question. In the present article, it is assumed explicitly.

For the sake of definiteness, but at the expense of anticipating and duplicating a small part of the discussion in subsequent sections, a specific example of the way in which stereoisomers are considered to differ geometrically may be given here. The two different 1,2-dichloroethylenes, which possess the same structure CHCl=CHCl, are thought to have the different geometrical forms represented by the following planar diagrams.



II. EARLY HISTORY OF STEREOCHEMISTRY

1. Optical Activity.—The original approach to the subject of stereochemistry arose from a study of certain optical phenomena. In 1811 D. F. Arago discovered the fact that crystals of quartz are optically active (*i.e.*, that they rotate the plane of polarization of plane polarized light). That this optical activity (or rotatory power) is not restricted to crystalline solids was demonstrated by J. B. Biot, who found in 1815 that oil of turpentine and aqueous solutions of tartaric acid or of sugar are optically active, and in 1817 that even gaseous turpentine is also active.

2. Specific Rotation.—For the characterization of an optically active substance, the so-called specific rotation (introduced by Biot in 1836) has proved useful. This quantity is defined by the equation

$$[\alpha]_{\lambda}^{t} = \frac{\alpha}{ld}$$

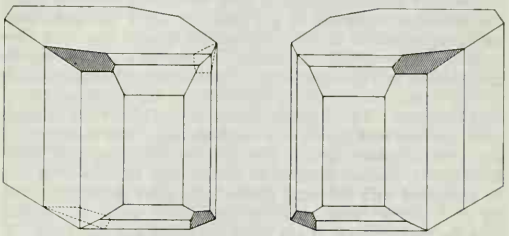
where $[\alpha]_{\lambda}^{t}$ is the specific rotation for light of wave length λ at the temperature $t°$; and α is the degrees of rotation produced by passage of the light through a tube l decimetres in length, which is filled with the optically active substance (either pure or in solution) at a concentration of d grams per cubic centimetre. The rotation is taken with the positive sign if it is to the right (*i.e.*, in the sense which is clockwise when regarded in the direction contrary to that in which the light is traveling), and with the negative sign if it is to the left (*i.e.*, in the opposite, counterclockwise sense).

Ideally, the specific rotation, as so defined, should be a characteristic property of the substance to which it applies, and its value should be independent of the experimental conditions. In practice, however, it varies somewhat in magnitude with the wave length of the light, with the temperature, with the solvent (if any) and with the concentration. In a few instances, even the sign of rotation is found to change with conditions.

With colourless compounds, the variation in specific rotation with the wave length of the light used is ordinarily small, but nevertheless great enough to make necessary the precise statement of the wave length. The remaining variations are usually counteracted by standardization of the experimental procedure to as great an extent as possible. Thus, unless statements are made to the contrary, specific rotations are considered to apply to room temperature and to dilute solutions in specified solvents. Within these limitations, specific rotations may be employed for the characterization, identification and even quantitative analysis of optically active substances.

3. The Work of Pasteur.—After the discoveries of polarized light and of optical activity, the next important advance toward the modern stereochemical theories was made by Louis Pasteur in a series of investigations on tartaric and racemic acids, carried out chiefly during the years 1849–53. These two acids not only have identical chemical compositions but also possess the same structure $HO_2C—CHOH—CHOH—CO_2H$. Yet they show certain marked differences in properties. For example, racemic acid is much less soluble in water than tartaric acid and crystallizes with water of crystallization, whereas tartaric acid separates in anhydrous crystals; the most remarkable difference between the two substances, however, is in their effect on plane polarized light. Tartaric acid and its salts (the tartrates) are optically active; more specifically, they are dextrorotatory (*i.e.*, their aqueous solutions rotate the plane of polarization to the right). On the other hand, racemic acid and its salts (the racemates) are optically inactive.

Pasteur's discoveries arose out of a crystallographic investigation of tartaric acid. He observed that there was a certain lack of symmetry in the crystalline form of this substance of such a kind that the crystal was nonsuperposable on its mirror image.



FROM STEWART, "STEREO-CHEMISTRY"; REPRODUCED BY PERMISSION FROM LONGMANS GREEN & CO.

FIG. 1.—(A AND B) ENANTIOMORPHOUS CRYSTALS OF SODIUM AMMONIUM TARTRATE

The crystal and its image thus differ in the same kind of way as the right and left hands. He prepared and examined 19 different salts of tartaric acid and found that each of these showed a similar lack of symmetry in its crystalline form. Thus fig. 1(A) represents the crystalline form of one of these salts, sodium ammonium tartrate. If the crystal possessed the full symmetry of the system to which it belongs, two faces, corresponding with the two shaded areas, should be found in the positions indicated by broken lines. These faces, instead of occurring eight times in the crystal, occur only four times (there are of course two corresponding faces on the back of the crystal not shown in the figure); such faces, occurring only half as frequently as the full symmetry of the crystal demands, are termed hemihedral faces. The sodium ammonium tartrate crystal, because of the presence of these hemihedral faces, is evidently nonsuperposable on its mirror image represented in fig. 1(B). In contrast to tartaric acid and the tartrates, racemic acid and such of its salts as Pasteur examined (with the exception of sodium ammonium racemate, *see* below) crystallize in symmetrical forms superposable on their mirror images. The optically active tartrates thus crystallize in dissymmetric forms, whereas the optically inactive racemates yield symmetrical crystals. (According to accepted terminology, an object is said to be dissymmetric if it is not superposable on its mirror image, and to be asymmetric if it possesses no element of symmetry whatever, *see* below. In discussions of stereochemistry, the word asymmetric is often used loosely in places where dissymmetric is intended; in the present article, however, the distinction is observed.)

Pasteur's chief discovery was made through investigating the exceptional racemate, sodium ammonium racemate, referred to above. Crystallized from water at a temperature below 28° C., this salt yields crystals which, unlike those of the other race-

mates, exhibit hemihedral faces; the crystals, however, are of two kinds. Some are identical in form with those of dextrorotatory sodium ammonium tartrate (fig. 1[A]) and are in fact crystals of that salt. The other crystals differ from these only in that their hemihedrism is in the opposite sense; they are thus the realization of the mirror image of sodium ammonium tartrate and have the form represented in fig. 1(B). When picked out from the mixture and dissolved in water, they give a levorotatory solution (*i.e.*, a solution which rotates the plane of polarization to the left). The acid obtained from these crystals has a crystalline form which is the mirror image of that of tartaric acid, and its solutions are levorotatory. It is to be regarded as a second variety of tartaric acid.

This new form of tartaric acid and the original one have exactly the same melting point and specific gravity; moreover, they exhibit exactly the same chemical reactivities toward all optically inactive reagents. Two substances related in this manner are said to be enantiomorphs of each other. (The further expressions antimers and optical antipodes are also frequently used.) A more detailed discussion of the ways in which such pairs of substances resemble each other, and of those in which they differ from each other, is given below (see *Enantiomorphs*, below).

On the basis of his discoveries, Pasteur postulated that, in consequence of the arrangement of their constituent atoms in space, the individual molecules of any optically active substance must be nonsuperposable on their own mirror images (*i.e.*, that they must be dissymmetric). Moreover, he postulated further that the molecules of any optically active substance must be superposable on the mirror images of the molecules of the enantiomorph of that substance. The situation visualized by Pasteur can be made more concrete with the aid of a rough analogy. The molecules of, say, dextrorotatory tartaric acid may be compared with right-hand gloves, whereas those of its levorotatory enantiomorph may be compared with left-hand gloves. No glove is identical with its own mirror image, but any right-hand glove is identical with the mirror image of any left-hand glove, and vice versa. If the analogy is carried a step further, a crystal of racemic acid can be compared with a large package of pairs of gloves; since each pair consists of one right-hand and one left-hand glove, the package as a whole may be considered superposable on its mirror image. These postulates of Pasteur form the basis of the following discussion; their justification lies in the fact that, as is shown below, they permit uniformly correct predictions to be made regarding the properties of, and relationships among, optically active substances.

III. GENERAL DISCUSSION OF STEREOISOMERISM

1. Symmetry Elements of Molecules.—It is desirable at this point to examine the characteristics of a molecule which make it dissymmetric, or prevent its being so, as the case may be.

The symmetry of any rigid body can be described with the aid of its so-called symmetry elements. These elements are of several different kinds, of which only the following need be mentioned here. (1) An object is said to possess a plane of symmetry if a plane mirror can be imagined passing through it so that one-half of the object coincides with the reflection in the mirror of the other half. The plane of the mirror is then the plane of symmetry. (2) An object is said to possess an n-fold axis (or an axis of order n) if it possesses an axis, about which rotation by $360°/n$ brings the object into a position indistinguishable from its original one. The order n can take on any positive integral value, or it may be infinite if the object has, for example, the symmetry of a right circular cylinder. In general, n rotations by $360°/n$ bring the object into a position identical with, and not merely indistinguishable from, the original one. Any straight line passing through any object is clearly a onefold axis; such onefold axes may usually be ignored, however, in discussions of symmetry since an infinite number of them are always present. (3) An object is said to possess a centre of inversion (or of symmetry, or of reflection) if it contains a point within it, such that any straight line through that point encounters exactly the same environment in each of its two directions. Further discussion, with examples, of each of these types of symmetry element is given in the article Crystallography. A number of additional examples are also mentioned below.

The above-mentioned plane and centre of symmetry are examples of elements with what may be termed reflection symmetry. The most general type of element possessing this kind of symmetry is the so-called alternating (or mirror) axis, of which the plane and centre of symmetry are only special cases. Although no detailed description of alternating axes need be given here, mention should be made of the fact that an object may possess such an axis without having either plane or centre of symmetry. (*See* below.)

Mathematical analysis has shown that an object is superposable on its mirror image if, and only if, it possesses reflection symmetry of some sort. In stereochemistry, this rule finds its only important applications in those instances in which the molecules have either planes or centres of symmetry, or both. Not until 1956 was there reported (G. E. McCasland and S. Proskow) the preparation of an optically inactive substance with neither of these two symmetry elements, but with a fourfold alternating axis. Accordingly, the statement that a compound must be optically active unless it is a mixture of enantiomorphs (like racemic acid) or else its molecules possess either planes or centres of symmetry is adequate for most purposes. It is especially to be noted that the presence of simple (not alternating) axes of symmetry is not sufficient to ensure optical inactivity; or, in other words, that a molecule may possess an axis of symmetry and still be dissymmetric. (For an example of a dissymmetric molecule with both twofold and threefold simple axes, *see* the cobalt compound XXXII in the section below, *Elements of Co-ordination Number 6*.)

2. Enantiomorphs.—Since enantiomorphs have not only identical structures but also very similar configurations (*i.e.*, atomic arrangements in space), such pairs of substances should possess extremely similar properties. Indeed, in consequence of the mirror-image relationship assumed to exist between enantiomorphs, each property of either enantiomorph must be (in the sense discussed below) the mirror image of the corresponding property of the other enantiomorph; in other words, two enantiomorphs must be identical with respect to all those properties which are themselves (in this same sense) identical with their own mirror images, and can differ with respect to only those other properties which are different from their mirror images. For example, the melting point of any substance may be considered identical with its own mirror image, since the temperature at which solid and pure liquid are in equilibrium must be the same whether the experiment is viewed directly or by reflection in a mirror. Consequently, two enantiomorphs must have identical melting points. In a similar way, it can be shown that two enantiomorphs must also have identical vapour tensions at any specified temperature, identical boiling points at any specified pressure, identical densities under the same conditions, identical solubilities in any specified optically inactive solvent, identical rates of reaction with any specified optically inactive reagent (or reagents), identical equilibriums in their reactions with any specified optically inactive reagent (or reagents), and so on. On the other hand, the mirror image of a rotation of the plane of polarization by a specified number of degrees to the right (or left) may be considered to be a rotation by exactly the same number of degrees to the left (or right), respectively. Consequently, two enantiomorphs, under identical conditions, must rotate the plane of polarization by equal amounts but in opposite directions. Similarly, the mirror image of right (or left) circularly polarized light is identical not with itself but instead with left (or right) circularly polarized light. Consequently, two enantiomorphs may be expected to possess different optical properties (*e.g.*, absorption spectra, indexes of refraction, etc.) when examined with circularly polarized light, but not when examined with ordinary light or even (except for the optical rotation) when examined with plane polarized light; conversely, the optical properties of a single optically active substance may be expected to be different with right and left circularly polarized light. The expected differences in spectrum have been

observed experimentally (circular dichroism, Cotton effect); and the differences in index of refraction, although so small that their direct observation is difficult, are known always to exist since the observed angle of rotation by an optically active substance can be shown to be proportional to the difference in the indexes of refraction for right and left circularly polarized light. Also, measurable differences have been found in the rates and extents of adsorption (*q.v.*) of enantiomorphs upon optically active solid adsorbents, and in the rates and equilibriums of their reactions with optically active reagents (see *Asymmetric Syntheses and Decompositions*, below). Finally, the crystals of two enantiomorphs may be nonsuperposable mirror images of each other (*cf.* the enantiomorphic tartaric acids and tartrates, above); in many instances, however, the forms of the crystals of enantiomorphs are symmetrical and identical.

3. Diastereomers.—Examples of stereoisomers which are not enantiomorphs of each other are well known. In 1853 Pasteur discovered that the cinchonine (*see* CINCHONA BARK, ALKALOIDS OF) salt of either tartaric or racemic acid, when heated for several hours at 170° C., is transformed into the salt of still a different stereoisomeric acid. This acid, known as mesotartaric acid, resembles racemic acid in being optically inactive, but it differs from racemic acid in having a different melting point, different solubilities, etc., and also in not being separable into two enantiomorphic forms. The individual molecules of mesotartaric acid, therefore, must be symmetrical (*i.e.*, superposable on their mirror images). The substance, although a stereoisomer, is not an enantiomorph, of either of the active tartaric acids.

In subsequent years, a large number of nonenantiomorphic stereoisomers have been discovered. An extreme example of this is provided by the aldohexoses (*see* CARBOHYDRATES), of which the pyranose structure permits no less than 32 stereoisomeric forms, divisible into 16 pairs of enantiomorphs; all are known.

No general term has come to be universally adopted for the description of the relationship between two stereoisomers which are not enantiomorphs of each other. The word "diastereomer" (or diastereoisomer, or diamer) is often applied to certain, but usually not to all, such pairs of substances. For purposes of brevity and convenience throughout the remainder of this article, any nonenantiomorphic stereoisomer will be termed a diastereomer. It is to be noted, however, that, in this respect, the present article departs somewhat from the most customary terminology.

Unlike enantiomorphs, diastereomers (in the sense just defined) need not have closely similar physical and chemical properties. Indeed, they may differ as greatly as do structural isomers. For example, either of the optically active tartaric acids melts at 187° C., whereas mesotartaric acid melts at 143° C.

4. Meso Forms.—As with the tartaric acids, it frequently happens that a set of stereoisomers contains both optically active and optically inactive members. The inactive members of such sets are frequently distinguished from their active isomers by being called meso forms. (This is, of course, the significance of the name mesotartaric acid.)

5. Racemic Modifications.—A mixture of equal quantities of two enantiomorphs, which is optically inactive, is known as a racemic modification. The name is derived from racemic acid, the first example of such a mixture to be carefully studied. (This double use of the word racemic is found in practice to lead to little if any confusion.) Any process by which an optically active substance is transformed into the corresponding racemic modification is known as a racemization; conversely, any process by which a racemic modification is separated into the two enantiomorphs is known as a resolution. (See *Racemization* and *Resolution*, below.)

In the liquid and gaseous states, a racemic modification seems always to be merely a mechanical mixture of its two components. In the solid state, however, three different situations have been encountered. (1) Sometimes, as with sodium ammonium racemate crystallized from water at a temperature below 28° C. (*see* above), the solid is a conglomerate (or racemic mixture) consisting of separate crystals of the two enantiomorphic forms. (2) Sometimes, as with the dimethyl ester (*see* ESTERS) of racemic acid, the two enantiomorphs form a 1:1 compound (or racemate) with each other. Such compounds, however, decompose into their constituents at temperatures above their melting points. (3) Sometimes, as with camphor oxime (*see* OXIMES), the enantiomorphs form a solid solution (*q.v.*) (or pseudoracemic mixed crystal).

6. A Word About Nomenclature.—Dextrorotatory and levorotatory enantiomorphs were originally designated by the prefixes *d* and *l*, respectively, so that the dextrorotatory and levorotatory tartaric acids, for example, were called *d*-tartaric acid and *l*-tartaric acid, respectively. Somewhat later, however, these prefixes were made to refer also to the so-called optically active series to which the substances belonged, as well as to the directions of rotation. (For the significance of these series, see *Optically Active Series*, below.) As a result of this ambiguity in nomenclature, there were found to be many substances which had to be assigned the prefix *d* because they were dextrorotatory, and, at the same time, the prefix *l* because they belonged to the *l*-series (or the opposite). The confusion which was thus caused has been alleviated to some extent by the adoption of the convention that the prefixes D and L are used for the designation of the optically active series, and that the further prefixes (+) and (−) are employed to represent dextrorotation and levorotation, respectively.

Unfortunately, the convention just discussed is not always followed consistently by all chemists. If a substance does not belong to any of the accepted optically active series, the prefixes *d* and *l* are often used, as originally, to indicate the direction of rotation. Thus, the naturally occurring levorotatory quinine (*q.v.*) is often referred to as *l*-quinine rather than as (−)-quinine. No confusion can arise in such instances since no *d*- and *l*-series, or D- and L-series, have been defined for alkaloids like quinine. However, an expression like "*d*-lactic acid" may refer either to the dextrorotatory form, which belongs to the L-series (*see* above), or to its enantiomorph in the D-series. Consequently, such an expression is configurationally meaningless, unless accompanied by a statement as to which convention is being employed. In this article, the prefixes D and L are used to represent the optically active series, and the prefixes (+) and (−) are used to represent the direction of rotation (even for substances like quinine, for which no series have been defined).

Optically inactive meso forms are frequently designated by the prefix *i* (for inactive) or *m* (for meso). Thus mesotartaric acid may be called either *i*-tartaric acid or *m*-tartaric acid. Racemic modifications are similarly designated by the prefix *r* (for racemic) or (±) or DL (to indicate that the compound is a mixture of the two enantiomorphs). Thus, racemic acid may be called *r*- , (±)- , or DL-tartaric acid.

7. Racemization.—The ease with which an optically active substance can be transformed into the racemic modification varies within wide limits. On the one hand, the racemization of certain compounds (for example, an optically active paraffin hydrocarbon) is extremely difficult; on the other hand, the racemization of certain other compounds (for example, a tertiary amine, see *Stereochemistry of Nitrogen*, below) is so easily accomplished that the optically active forms cannot be obtained. With innumerable further compounds, however, the ease of racemization lies between these two extremes. Such compounds can be separated (see *Resolution*, below) into the active enantiomorphs, which can then be racemized more or less readily.

The experimental conditions under which the racemization of racemizable substances can be effected also vary widely. In all instances, however, the racemization is presumed to occur as a result of a reversible transformation into an unstable substance which is symmetric and so incapable of optical activity; the reverse transformation back from this hypothetical inactive intermediate then leads to the inactive (*i.e.*, racemic) modification of the original substance.

The transformations which thus lead to racemization are considered most commonly to involve chemical reactions of a familiar type, but occasionally to involve merely geometrical deformations.

8. Resolution.—Whenever a substance capable of optical ac-

tivity is prepared under the usual experimental conditions from exclusively inactive reagents, the product obtained is always the inactive racemic modification. This is because, in view of the complete symmetry of the experimental situation, the probability that a molecule of the dextrorotatory enantiomorph will be formed is exactly the same as the probability that one of the levorotatory enantiomorph will be formed. Consequently, equal numbers of molecules of the enantiomorphs must be formed, so that the product must be racemic. If an optically active product is desired instead a resolution must therefore be performed.

The separation of a racemic modification into its two optically active components is made rather difficult by the close similarity of enantiomorphs in nearly all their properties. Two important methods of resolution were developed by Pasteur. The first of these, known as the method of spontaneous separation, has already been described in connection with Pasteur's work on sodium ammonium racemate. It can be employed if the solid racemic modification is a conglomerate composed of observably different crystals. Only a few instances in which this condition is satisfied have been reported in the organic chemical literature. Consequently, this method, although of great historical and theoretical interest, is seldom applicable.

Pasteur's second method of resolution is of much greater generality than his first, and it is the only one commonly used. It is based upon the transformation of the mixture of enantiomorphs into a mixture of diastereomers, which, since they differ in physical properties, can be separated relatively easily. This transformation requires the use of some previously obtained optically active substance. For example, Pasteur showed in 1853 that when racemic acid is treated with a naturally occurring optically active base, such as cinchonine, strychnine, brucine or quinine (*see* ALKALOIDS), the resulting salt is a mixture of diastereomers and no longer one of enantiomorphs. This is because (if, for definiteness, the base is assumed to be dextrorotatory) the salt formed from the (+)-acid and (+)-base is not the mirror image of the one formed from the (−)-acid and the same (+)-base. (The mirror image would instead be the salt formed from the (−)-acid and (−)-base.) The two salts present in the mixture may, therefore, be expected to have different solubilities and so to be separable by crystallization. After the separation has been carried out, the salt from the (+)-acid and (+)-base gives the pure (+)-acid when treated with a strong mineral acid, and the other salt gives the pure (−)-acid when treated similarly.

In the form just outlined, Pasteur's second method can be applied to the resolution of only racemic acids. Various modifications have been developed, however, and these permit it to be greatly extended. Thus, a racemic base can be resolved by means of salt formation with an optically active acid, a racemic aldehyde or ketone can be resolved by means of hydrazone formation with an optically active hydrazine, and so on. Moreover, a racemic alcohol can be resolved if it is first transformed into the acid ester of a dibasic acid (such as phthalic acid, *q.v.*), which is neutralized as before with an optically active base; the resulting ester salt can then be separated into the diastereomeric forms and finally hydrolyzed back to the alcohol.

Similar schemes of greater or less complexity can be devised for most remaining types of substance.

9. Asymmetric Syntheses and Decompositions.—If a compound capable of optical activity is prepared in a dissymmetric environment, the product obtained need not be completely racemic since factors are present which may favour the formation of one enantiomorph over that of the other. Such reactions are known as asymmetric syntheses (although dissymmetric syntheses might seem a more appropriate expression). For example, although the reduction of inactive phenylglyoxylic acid C_6H_5—CO—CO_2H leads to completely racemic mandelic acid C_6H_5—CHOH—CO_2H, the reduction of the (−)-menthyl ester of the acid, followed by hydrolysis of the resulting menthyl mandelate gives a mandelic acid which is levorotatory although it contains a considerable amount of the (+)-enantiomorph. This slightly active acid can be separated by crystallization into the optically pure (−)-mandelic acid and *r*-mandelic acid. A different type of asymmetric synthesis is illustrated by the base-catalyzed reaction between benzaldehyde C_6H_5—CHO and hydrogen cyanide HCN to give mandelonitrile C_6H_5—CHOH—CN. Since the reagents are optically inactive, the product is necessarily racemic if the basic catalyst is also inactive; however, if an optically active catalyst, such as the alkaloid quinine, is used instead, the resulting mandelonitrile is somewhat optically active.

The asymmetric syntheses mentioned above are relative, since they require the previous existence of at least one optically active substance. A few absolute asymmetric syntheses, which are not subject to this requirement, have been reported. These syntheses are based upon photochemical reactions (*see* PHOTOCHEMISTRY) initiated by circularly polarized light; since circularly polarized light is not identical with its own mirror image (*see* above), it provides the necessary dissymmetric environment.

Asymmetric decompositions, either relative or absolute, can be performed under conditions analogous to those employed with the above asymmetric syntheses.

Asymmetric syntheses and decompositions are of especially great importance in biological reactions, since the reactions which occur in living systems are so frequently brought about by optically active reagents or are catalyzed by optically active catalysts (enzymes). For example, the fact that green plants are able to transform inactive water and carbon dioxide into active carbohydrates (*see* PHOTOSYNTHESIS) is due to the occurrence of asymmetric syntheses. Similarly, the fact that the natural D(+)-glucose is easily metabolized by animals and is fermented by yeast, whereas its enantiomorph undergoes neither reaction, must be attributed to the occurrence of asymmetric decompositions. An important application of such biological reactions was discovered by Pasteur, who found (1856–60) that the mold *Penicillium glaucum*, when allowed to grow in the presence of racemic (tartaric) acid, destroys the (+)-enantiomorph preferentially, so that the acid which remains becomes levorotatory. This method of obtaining optically active substances is of considerable generality since other microorganisms have been found to have similar effects, and since other racemic modifications have been made active in similar ways; it has, however, serious limitations since many substances are not attacked by any known organisms, and since often the organism destroys just the enantiomorph which is desired. It should be noted that the method cannot be called a resolution since one of the two enantiomorphs is always destroyed. (*See* also CARBOHYDRATES.)

10. The Origin of Optically Active Substances in Nature.—The frequent occurrence of optically active substances in animal and vegetable sources can be explained only as the result of asymmetric syntheses or decompositions. These reactions are made possible by the presence of pre-existing optically active substances, which must themselves have been made by asymmetric syntheses or decompositions; and so on. If life started on this earth many millions of years ago by the coming together of simple inorganic and optically inactive materials, there must have been a time at which the optical activity began. The question therefore arises how it came into being in the absence of any pre-existing optically active substance. (If life is assumed not to have started on this planet but to have come from some other body, the difficulty is merely pushed back in time and not avoided.) Naturally, the answer to this question is unknown, and possibly unknowable. Scientists have nevertheless speculated about it for many years.

Two types of theory have been advanced to explain the origin of optical activity. The first of these is based upon the fact of spontaneous separation. It is not inconceivable that an optically active substance produced in small amount in this way may have been involved somehow in the origin of life. An objection raised to this explanation is that it depends too much upon the operation of chance. The supposition that the universal and uniform occurrence of optical activity could have arisen in such a manner seems unlikely a priori.

The second kind of theory is based upon the existence of absolute asymmetric syntheses and decompositions. For example, since light can be circularly polarized by reflection, life may conceivably have started at such a time and place that the first

organisms were more or less constantly subjected to the action of either right or left circularly polarized light. This theory avoids the difficulty of assigning too great a role to chance. It encounters a different one, however, in that the activities producible photochemically in the laboratory have always been extremely small, and also in that the natural circular polarization of light could at most be slight. A further theory, which assumes that the earth's magnetic field produced the necessary dissymmetric environment enjoys a similar advantage, but it suffers from similar disadvantages.

IV. THE STEREOCHEMISTRY OF CARBON

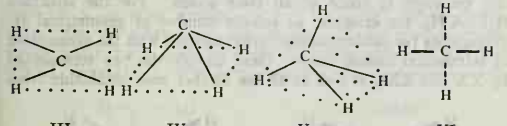
1. The Tetrahedral Carbon Atom.—The above extremely general considerations require no assumptions in regard to the actual geometrical forms of the molecules under discussion. Such assumptions become necessary, however, if the attempt is made to account for the numbers of stereoisomers corresponding to each structural formula. The first successful theory based upon explicit geometrical assumptions was brought forth simultaneously and independently in 1874 by J. H. van't Hoff and J. A. Le Bel. The ideas of van't Hoff were somewhat more definite than those of Le Bel, but they differed in no respect important to the present article. The following discussion makes no effort to follow either van't Hoff's or Le Bel's development very closely; it attempts instead to present the fundamental principles in logical rather than in historical order.

For the molecular formula CH_4, only the single structure I can be drawn (if, as usual, carbon and hydrogen are considered to be quadrivalent and univalent, respectively), and only the

I II

single substance, methane, is known. Similarly, for the formula CH_3Cl, only the single structure II can be drawn, and only the single substance, methyl chloride, is known. Now, if one hydrogen atom in methane were different from any one of the others in any respect, the substance produced by the replacement of the first hydrogen atom by a chlorine atom should be different from the one produced by the replacement of the second hydrogen atom. Consequently, at least two distinct substances, CH_3Cl, should exist. The fact that only one such substance is known strongly suggests, therefore, that all four hydrogen atoms of methane are completely equivalent to each other, not only structurally but also geometrically.

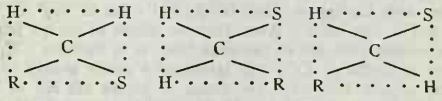
The requirement that all four hydrogen atoms of methane be geometrically equivalent restricts the atomic arrangements to the following: (1) The hydrogen atoms might lie at the corners of a rectangle, with the carbon atom at the centre, as in III. The dotted lines outlining the rectangle are given only for the sake of clarity; they do not represent valence bonds of any kind. This model is known as the planar model. (2) The hydrogen atoms might lie at the corners of the base of a square pyramid, with the carbon atom at the apex, as in IV. (It should be noted that,



III IV V VI

if the base of the pyramid were rectangular rather than square, two enantiomorphic forms of each monosubstituted derivative CH_3R would be possible. Such a configuration can, therefore, be excluded.) The dotted lines are again given only for the sake of clarity. This model is known as the pyramidal model. (3) The hydrogen atoms might lie at the corners of a tetrahedron, with the carbon atom in the centre, as in V. The dotted lines which outline the imaginary tetrahedron are given, as before, only for the sake of clarity. A more convenient, but equivalent, representation of this so-called tetrahedral model, which is employed throughout the remainder of the article, is shown in the diagram VI. The two broken lines in this latter figure represent valence bonds to hydrogen atoms lying behind the plane of the paper; the two heavy lines represent bonds to hydrogen atoms lying in front of the plane of the paper.

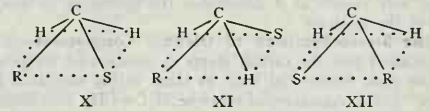
The decision among the planar, pyramidal and tetrahedral models is made possible by a consideration of the numbers of isomeric substances CH_2RS, where R and S represent any two specified, but not necessarily different, atoms or groups. Only one known substance, methylene chloride, exists with the formula CH_2Cl_2; only one, chloroacetic acid, exists with the formula $CH_2Cl—CO_2H$, etc. For no formula of this type is more than a single substance known. Now, the planar model would lead to the three distinct spatial arrangements or configurations VII, VIII and IX, of which VII and VIII are identical if the rectangle



VII VIII IX

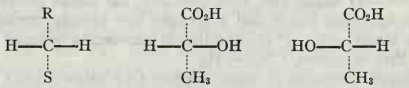
is a square. Since IX cannot be identical with VII or VIII under any circumstances, however, this model requires the existence of at least two stereoisomeric forms and so can be excluded from further consideration.

The pyramidal model might be expected, like the square planar one, to permit two stereoisomeric substances. Thus, the two configurations X and XI are possible. If R and S are not identical, however, the configuration X is dissymmetric, and so its nonsuperposable mirror image (enantiomorph) XII represents still a third stereoisomer. Since stereoisomerism has never been observed with structures of the present type and, moreover, since optical activity has never been encountered, the pyramidal model must also be incorrect.



X XI XII

Only the tetrahedral model remains. If the tetrahedron is regular, this model permits only a single substance CH_2RS, whether R and S are identical or different. In particular, the configuration, as shown in XIII, is symmetrical, since it has a plane of symmetry passing through R, S and C, and so no optical activity is possible and no enantiomorphs can exist. The regular tetrahedral model, therefore, reproduces the experimentally observed numbers of isomers.



XIII XIV XV

The regular tetrahedral model is the only one in agreement with the observed numbers of isomers; consequently, it must be assumed correct. Confirmation of this model is provided by the fact that it correctly predicts the occurrence of optical activity, and hence also of enantiomorphs, with substances with the formula CPRST, where no two of the atoms or groups P, R, S and T are identical. Lactic acid, $CH_3—CHOH—CO_2H$, for example, occurs

in enantiomorphic dextrorotatory and levorotatory forms. For the molecules of these substances, the nonsuperposable, mirror-image configurations XIV and XV are possible. One of these must therefore represent the molecule of the dextrorotatory lactic acid, and the other must represent that of the levorotatory form. Although it was long uncertain which of these two configurations belongs to each isomer, the problem was finally solved in 1951 by J. H. Bijvoet and his coworkers. By means of a complex and unusual modification of the methods of X-ray crystal-structure analysis, these authors showed that configurations XIV and XV describe, respectively, (−)- and (+)-lactic acid.

Physical evidence has confirmed the conclusions reached above. Studies of spectra and of X-ray and electron diffraction have shown that a saturated carbon atom is indeed tetrahedral, in the sense that the four single bonds which it forms are directed toward the corners of a tetrahedron, of which it occupies the centre. (The tetrahedron itself, of course, is only an intellectual construction without physical reality.) Moreover, the tetrahedron has been found to be regular, or very nearly regular, so that the angle between any two of the single bonds formed by the central carbon atom has the so-called tetrahedral value of 109° 28′, or very nearly that value. When the four atoms or groups joined to the carbon atom are of the same kind, as in methane CH_4 or carbon tetrachloride CCl_4, the tetrahedron seems always to be strictly regular, but, when the atoms or groups are not of the same kind, as in methylene chloride, CH_2Cl_2, some small distortions may occur.

2. Asymmetric Carbon Atoms.—A carbon atom is said to be asymmetric if it is joined by four different atoms or groups. Thus, of the three carbon atoms in lactic acid, the central one is seen to be asymmetric. (The terminology here is correct since an asymmetric carbon atom possesses no element of symmetry and so is actually asymmetric, and not merely dissymmetric.) The majority of all known optically active substances consist of molecules which contain one or more asymmetric carbon atoms. However, the presence of such atoms is neither necessary nor sufficient for optical activity. As will be discussed below, many substances containing no asymmetric atoms are active. Conversely, many containing asymmetric atoms are inactive. Optical activity is caused by the dissymmetry of each individual molecule as a whole; the presence or absence of asymmetric atoms is important only insofar as it determines the symmetry or dissymmetry of the molecule.

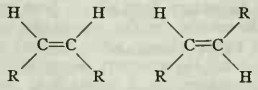
3. The Stereochemistry of Olefinic Compounds.—Van't Hoff assumed that the carbon atoms in unsaturated compounds have the same tetrahedral form as those in the saturated substances. The configuration of ethylene, $H_2C{=}CH_2$, thus becomes XVI. If this configuration is correct, the six atoms in the molecule

XVI

lie in the same plane (the plane of the paper in XVI), the H—C—H bond angles have the tetrahedral value of 109° 28′, the C—C—H angles (or, more precisely, the angles between the carbon–hydrogen bonds and a straight line passing through the carbon atoms) are 125° 16′, and the carbon–carbon distance is 0.58 times that in ethane, CH_3—CH_3.

Although erroneous in some respects (*see* below), van't Hoff's configuration of ethylene appears to be correct in all essential stereochemical details. In particular, it leads to correct predictions of the numbers of stereoisomeric forms of suitably substituted ethylenes. Indeed, its significant features can be deduced directly, and without reference to the tetrahedral carbon atom, from considerations of isomer numbers. For example, with a molecule with the structure HRC=CHR, this planar model permits the two configurations XVII and XVIII. Now, XVII and XVIII must represent optically inactive diastereomers since, with each, the plane of the molecule is a plane of symmetry.

In the particular case in which R is the carboxyl group, CO_2H, two substances, fumaric and maleic acids (*q.v.*), are known; each of these acids is optically inactive, and the two differ markedly



XVII XVIII

in nearly all their properties (*see* above). Consequently, they must be diastereomers and cannot be enantiomorphs. Since numerous further examples of similar type are known, and since optical activity has never been encountered in substances with the structure HRC=CHR (unless R represents a dissymmetric group), the necessary conclusion is that the planar model is correct, in agreement with van't Hoff. The configuration XVII is commonly said to be cis (Lat. *cis*, "on this side"), whereas XVIII is said to be trans (Lat. *trans*, "across").

From considerations of isomer number alone, no information can be obtained regarding the values of the bond angles or of the interatomic distances in ethylene and its derivatives; consequently, the angles and distances derived from van't Hoff's configuration, with the tetrahedral carbon atoms, cannot be verified by purely chemical procedures.

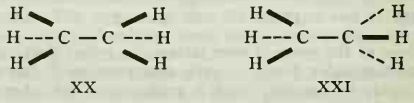
Physical investigations (*see* SPECTROSCOPY; ELECTRON DIFFRACTION) have shown, however, that it is precisely in such respects that van't Hoff's model is incorrect. Thus, in ethylene, the H—C—H and C—C—H angles are more nearly equal to 120° than to 109° 28′ and 125° 16′, respectively; and the carbon–carbon distance is more nearly 0.87 than 0.58 times that in ethane. Moreover, the modern quantum-mechanical picture of the double bond is rather different from that provided by the tetrahedral model XVI. Nevertheless, these defects in no way limit the usefulness of the simple model for the prediction and interpretation of the observed numbers of stereoisomers.

4. The Stereochemistry of Acetylenic Compounds.—Just as with ethylene, van't Hoff assumed that the carbon atoms in acetylene, H—C≡C—H, are tetrahedral. He accordingly wrote the configuration as XIX. This linear model is in agreement with

XIX

the observed numbers of isomeric substitution products and also with recent physical evidence. Consequently, although van't Hoff's model of acetylene is inaccurate in several respects (analogous to those in which his model of ethylene is inaccurate, *see* above), it leads always to correct predictions of the numbers of isomers.

5. Free Rotation.—Certain limitations must be imposed upon the general statement that two molecules correspond to different substances unless they are exactly alike with respect to the relative positions in space of all their atoms. For the structure H_3C—CH_3, for example, an infinite number of geometrical arrangements (or conformations) are consistent with the assumption of tetrahedral carbon atoms; these include the two represented by XX and XXI, as well as all the further ones obtainable from



XX XXI

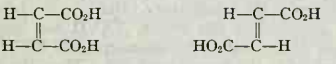
either XX or XXI by rotation of one CH_3 group with respect to the other about the carbon–carbon bond. If each of these

conformations represented a distinct substance, then an infinite number of stereoisomers with the structure in question should exist. Only a single substance, ethane, is known however. Similar considerations apply to nearly all further structures in which multivalent atoms are joined by single bonds. It is evident, therefore, that, if the geometrical interpretation of stereoisomerism is to be maintained, some further assumption, which restricts the predicted numbers of isomers, is required.

The assumption generally adopted for this purpose is contained in the rule that any two conformations (such as XX and XXI, or any other two of the infinite number of possible ones) are to be considered equivalent if they differ only by the rotation of one or more parts of the molecule about one or more single bonds. Thus, XX, XXI, etc. are equivalent; all these conformations correspond to the same substance, ethane, and all are said to represent the same configuration. Although this rule has a number of exceptions (see *Optically Active Biphenyls,* below), it is sufficiently general to find application throughout the entire field of stereochemistry. It has indeed already been assumed implicitly at several points of the preceding discussion, and it is assumed hereafter throughout the following discussion without further comment (except in those instances in which it is invalid).

A common alternative statement of the above rule is that rotation about single bonds is free. The use of the term "free rotation" suggests that, in ethane for example, no forces operate to make any one conformation more stable than any other, so that each molecule of the substance spends equal fractions of its time in each of its possible geometrical arrangements. This interpretation is obviously consistent with the fact that stereoisomeric forms of ethane, differing by rotation about the carbon–carbon bond, do not occur; nevertheless, it is probably incorrect. From a combination of spectroscopic and thermochemical data, the conclusion has been drawn that the "eclipsed" conformation XX is less stable (*i.e.,* has a higher energy) than the "staggered" one XXI by about 3,000 cal. per mole. If this conclusion is correct, most of the molecules at any one time must have arrangements rather close to the staggered one XXI. Stereoisomerism is still impossible, however, since each molecule must spend part of its time in each of the conformations, and since the transition from any one conformation to any other is rapid and easy.

In nearly all other substances for which data are available, the situation seems to be similar to that in ethane. Certain conformations always seem to be appreciably more stable than others differing from them by rotation about one or more single bonds, but stereoisomers are not encountered because the transition from one conformation to another is easy. In general, stereoisomers can be expected to occur only if there are two or more essentially different geometrical arrangements of the atoms, and if transitions between these arrangements are not easy. A transition may be considered easy at ordinary temperatures if it can be undergone by a molecule possessing an energy much less than about 20,000 cal. per mole above the average. This is because, as a result of the statistical variations in energy (*see* Kinetic Theory of Matter), the number of molecules possessing the required excess energy is then sufficiently great to permit an appreciable rate of interconversion; consequently, the separation of the isomers would be difficult or impossible. Stereoisomerism is, therefore, impossible in ethane, since a molecule of this substance requires an excess energy of only approximately 3,000 cal. per mole for a complete rotation of 360° about the carbon–carbon bond. On the other hand, the existence of the stereoisomeric maleic and fumaric acids, to which the configurations XXII and XXIII, respectively, have been assigned, shows that only such



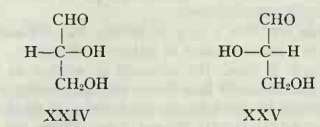
XXII XXIII

molecules as possess at least 20,000 cal. per mole of excess energy can undergo the transition. Just as the nonexistence of stereoisomers in ethane (and the like) is attributed to the freedom of rotation about the carbon–carbon single bond, so also the existence of stereoisomerism between maleic and fumaric acids (and the like) is attributed to the lack of free rotation about the double bond.

It is interesting that van't Hoff's models of ethane and ethylene (XX and XVI, respectively) are consistent with the presence of free rotation about the single bond and the lack of it about the double bond. Since the two carbon atoms in ethane are joined by just one linkage, no forces exist in the model XX to favour any one conformation over any other. Consequently, rotation might be expected to be easy in the molecule also. On the other hand, since the two carbon atoms in ethylene are joined by two bonds, the model XVI is held rigidly in the planar form. Consequently, rotation might be expected to be impossible in the molecule. The modern quantum-mechanical pictures of the single and double bonds are rather different from those provided by the models of van't Hoff, but they are equally in agreement with the observed freedom and restriction, respectively, of rotation.

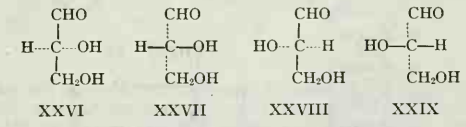
6. Plane Projection Diagrams.—The use of three-dimensional models of the sort employed by van't Hoff becomes extremely cumbersome if the molecule to be represented is even moderately complicated. Moreover, such models cannot be depicted very conveniently on a plane surface, such as a printed page; even the relatively simple diagrams XX and XXI, for example, are not too easily intelligible, and the situation rapidly becomes worse as the complexity of the molecule increases. Consequently, some simpler method of representation is highly desirable.

It has become conventional to represent the three-dimensional models by what are called plane projection diagrams or, more simply, plane configurations. Thus, by definition, the diagrams XXIV and XXV represent the dextrorotatory and levorotatory forms,



XXIV XXV

respectively, of glyceraldehyde. It is to be noted that these diagrams do not in themselves specify the absolute configurations (*i.e.,* the true three-dimensional forms of the actual molecules). The diagram XXIV, for example could be considered the projection of either XXVI or XXVII; similarly, XXV can be considered the projection either of XXVIII (identical with XXVII) or of its enantiomorph XXIX (identical with XXVI). Nevertheless, the



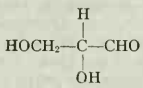
XXVI XXVII XXVIII XXIX

work of Bijvoet *et al.* (*see* above) has shown that the plane configurations XXIV and XXV actually correspond to the absolute configurations XXVII and XXIX, respectively.

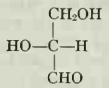
The usefulness of plane configurations like XXIV and XXV lies in the fact that they make possible the unique representation of relative configurations. In other words, the plane projection diagrams define the configurations of the substances to which they refer by comparing them with those of the glyceraldehydes. Since the absolute configurations of the glyceraldehydes are now known, the relative configurations of other substances define the absolute ones.

Considerable care must be exercised in the use of the plane configurations, since errors and inconsistencies can easily be introduced. It might appear, for example, that the diagrams XXIV and XXV are identical since a rotation of 180° about a vertical axis lying in the plane of the paper would transform either into

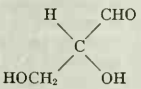
the other. Consequently, if XXIV and XXV are to represent enantiomorphs, a rotation which takes the figure out of the plane in which it is drawn must be considered to invert the configuration (*i.e.*, to change it into that of the enantiomorph). Although a rotation of 180° out of the plane of the paper inverts the configuration, as stated above, any rotation within the plane of the paper leaves the configuration unchanged. Thus, XXX, XXXI and XXXII are equivalent to XXIV and, like it, represent dextrorotatory glyceraldehyde.



XXX



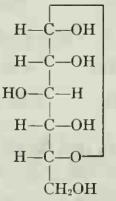
XXXI



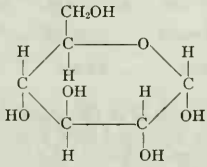
XXXII

It is to be noted that freedom of rotation does not in general exist about the single bonds in plane projection diagrams of the present type (see *Free Rotation,* above). This is because a rotation about any bond in such a diagram requires that that part of the diagram which is rotated leave the plane of the paper. If this part is dissymmetric, then, according to the convention adopted, it is inverted in the process, so that the resulting configuration is not equivalent to the original one. For example, a rotation by 180° of the lower part of XXIV about the upper carbon–carbon bond changes this configuration into the enantiomorphic one XXV. (The CH_2OH group is not affected by the rotation since it is symmetric and so identical with its own mirror image.) These remarks do not imply any corresponding lack of freedom of rotation about the single bonds in the actual molecules or in the three-dimensional models; they have reference instead only to the conventions which must be adopted if the plane configurations are not to lead to self-contradictions.

If a molecule contains a ring of atoms, its configuration may be expressed on a plane surface in either of two ways. When one of these methods is used, the diagram is written as if no ring were present, and then the fact that a ring is present is indicated, more or less as an afterthought, by the drawing in of the necessary bonds as long, and frequently curved, lines. Thus, the configuration of α-D-glucose (*see* CARBOHYDRATES) is often written as XXXIII. When the second method of representation is employed, the ring is drawn as if it were a plane polygon approximately at right angles to the plane of the paper, and the substituents attached to the ring atoms are represented by the appropriate symbols above and below the plane of the ring. Thus, the configuration of α-D-glucose can be written as XXXIV; this diagram is to be considered equivalent to XXXIII. Plane projection dia-



XXXIII



XXXIV

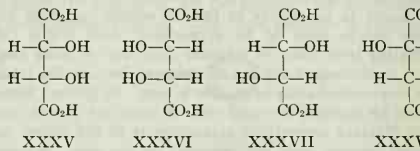
grams like XXXIV do not imply knowledge of the corresponding absolute configurations. A molecule of glucose in which the upper horizontal line (of XXXIV) is in front of the paper, and the lower line is behind it, would be the enantiomorph of a molecule in which the upper line is behind the plane of the paper, and the lower line is in front of it. Nevertheless, since determination of the absolute configurations of the glyceraldehydes (*see* above) has shown that the hydrogen atoms and hydroxyl groups which are joined to the asymmetric carbon atoms lie in front of the plane of the paper in diagram XXXIII, consistency requires that the lower horizontal line of XXXIV be in front of, and that the upper horizontal line be behind, the paper. These diagrams do not imply that the rings are completely planar (see *Strain Theory,* below); they imply rather that the average positions of the atoms of the ring lie within a single plane.

The representation of an ethylenic compound is relatively easy since the two carbon atoms joined by the double bond, and the four other atoms joined by single bonds to these two, lie in the same plane. Consequently, diagrams like XXII and XXIII for maleic and fumaric acids, respectively, represent the absolute configurations. Such diagrams need no further discussion.

If a molecule contains several rings, several double bonds, or both rings and double bonds, the representation of its configuration may be rather difficult. In such instances, the conventional scheme just outlined often must be supplemented by the use of heavy and broken lines to represent bonds to atoms in front of and behind the paper, respectively, or by other devices analogous to those employed in preceding sections of this article. Examples of such molecules are given below.

7. Use of Plane Projection Diagrams for the Prediction of the Number of Stereoisomers.—The foregoing rather general considerations can be made more definite by a discussion of some specific examples of stereoisomerism. The dextrorotatory and levorotatory glyceraldehydes are represented by the diagrams XXIV and XXV, as has already been mentioned. That these two diagrams correspond to enantiomorphs is shown (according to the conventions adopted) by the fact that each can be considered identical with the reflection of the other in a mirror which is halfway between them and perpendicular to the plane of the paper. Since there is only one asymmetric atom in the molecule, and no further source of molecular dissymmetry, these two enantiomorphs are the only possible stereoisomeric forms.

A molecule of tartaric acid, $HO_2C\overset{*}{C}HOH\overset{*}{C}HOH—CO_2H$, has two asymmetric carbon atoms (the two designated by asterisks). Since there are two possible configurations about each such atom, one might anticipate altogether 2×2, or 4, distinct stereoisomers, which could be represented by the diagrams XXXV–XXXVIII. Of these, however, the first two are equivalent inasmuch as either is transformed into the other by a rotation

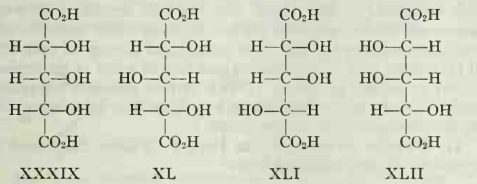


XXXV XXXVI XXXVII XXXVIII

of 180° in the plane. Since these two equivalent diagrams are mirror images of each other, they must represent a substance which is identical with its own mirror image, and so optically inactive; they are, therefore, equivalent forms of the plane configuration of mesotartaric acid (*see* above). The optical inactivity of this substance follows also from the presence of a plane of symmetry in either of the two diagrams (*i.e.*, from the fact that the upper half of either diagram is identical with the reflection of the lower half in a mirror perpendicular to both the plane of the paper and the central carbon–carbon bond). On the other hand, the last two configurations, XXXVII and XXXVIII, are not equivalent since neither can be transformed into the other by any rotation in the plane. Since they are mirror images of each other, they must be enantiomorphs, and so one must represent the dextrorotatory, and one must represent the levorotatory, tartaric acid. In a way which cannot be described in this article, XXVII has been shown to belong to the dextrorotatory enantiomorph, and XXVIII has been shown to belong to the levorotatory one. Thus, the existence of the three known tartaric acids can be accounted for.

The trihydroxyglutaric acids, $HO_2C\overset{*}{C}HOH—\overset{*}{C}HOH—\overset{*}{C}HOH—CO_2H$, have the three asymmetric carbon atoms marked by

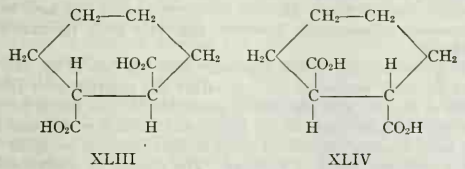
asterisks. (The central carbon atom may appear not to be asymmetric since two of the groups attached to it are structurally identical. It is customary, however, to regard such atoms as asymmetric since the attached groups may differ in configuration. Atoms of this type are sometimes said to be pseudoasymmetric.) One might expect 2^3, or eight stereoisomeric forms, but, as with the tartaric acids, the number is reduced somewhat by the occurrence of meso forms. Only the four nonequivalent configurations XXXIX-XLII can be drawn, and, in agreement with prediction, only four stereoisomeric substances are known. It is easily seen that XXXIX and XL represent two different meso forms since each has a plane of symmetry (passing through the central CHOH group and perpendicular to the paper), and that XLI and XLII represent a pair of enantiomorphs.



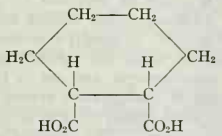
XXXIX XL XLI XLII

The aldohexoses (*see* CARBOHYDRATES) have five asymmetric atoms in the pyranose structure (*cf.* the configurations XXXIII and XXXIV of α-D-glucose). Since no meso forms are possible, 2^5, or 32, different optically active stereoisomers should exist. As has already been mentioned, all these forms are known.

The cyclohexane-1,2-dicarboxylic acids, like the tartaric acids, contain two equivalent asymmetric carbon atoms; consequently, they also exist in two enantiomorphic forms, XLIII and XLIV, and one meso form XLV. It is readily verified that the configurations XLIII and XLIV are mirror images of each other, and that neither possesses a plane or centre of symmetry. On the other hand, XLV has a plane of symmetry perpendicular both to the ring and to the bond joining the two asymmetric atoms; it is, therefore, identical with its mirror image.



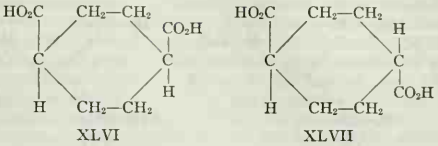
XLIII XLIV



XLV

The isomerism of the cyclohexane-1,2-dicarboxylic acids (as well as that of other analogous cyclic compounds) is frequently described by the statement that the optically active forms XLIII and XLIV are trans, whereas the meso form XLV is cis. This nomenclature is convenient since it expresses clearly and concisely the relative positions of the substituent groups with respect to the plane of the ring. However, it has given rise to much confusion since it suggests a false analogy with the cis-trans isomerism in ethylenic compounds like maleic and fumaric acids (*see* above).

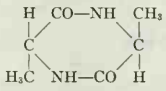
In cyclohexane-1,4-dicarboxylic acid, cis and trans forms again exist. Here, however, both the cis form XLVI and the trans form XLVII have planes of symmetry and so are optically inactive. The isomerism of these two substances (like that of the 1,2-dicarboxylic acids) is not due to the rigidity of the ring; no conceivable rotation about any bond or bonds in the ring could



XLVI XLVII

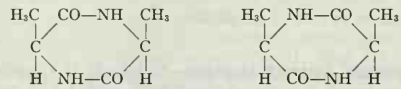
transform either isomer into the other, even if the ring were perfectly flexible.

Alanine anhydride (dimethyldiketopiperazine), with a six-membered ring, exists in cis and trans forms. Here, the trans form XLVIII has a centre of symmetry (at the centre of the ring) and



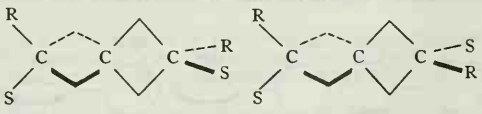
XLVIII

so it is optically inactive. (The presence of the centre of symmetry is apparent if the ring in XLVIII is visualized as lying in a plane perpendicular to that of the paper. The cis form has neither plane nor centre of symmetry, and so it exists in the two enantiomorphic forms XLIX and L.



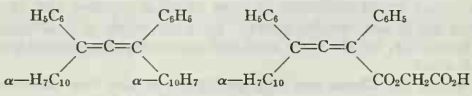
XLIX L

8. Optically Active Allenes and Their Analogues.—Van't Hoff pointed out that, if his tetrahedral model of the double bond (*see* above) is correct, the two CH_2 groups in allene, $H_2C{=}C{=}CH_2$, should lie in planes which are at right angles to each other; and that a substance with the structure RSC=C=CRS, where R and S are any two different atoms or groups, should exist in the enantiomorphic forms LI and LII. It was not until 1935, however,



LI LII

that this prediction was verified by P. Maitland and W. H. Mills, who obtained the optically active diphenyldi-α-naphthylallene LIII. In the same year, E. P. Kohler, J. T. Walker and M. Tishler reported the preparation of the similar optically active allene LIV.

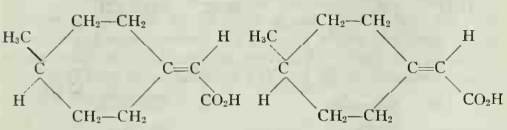


LIII LIV

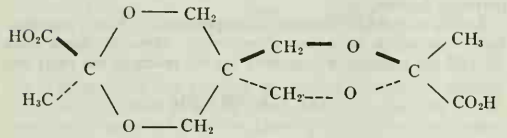
It is of interest that neither of these substances contains an asymmetric atom.

Although the existence of optically active allenes was not confirmed until comparatively recently, several analogous substances have been known for a longer time. These substances may be considered related to the allenes proper by the replacement of one or both of the two-membered rings (*i.e.*, double bonds) by larger rings. Two typical pairs of enantiomorphs are represented

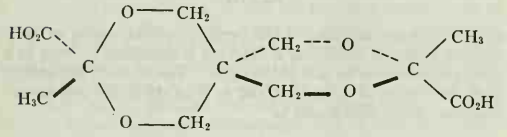
by the configurations LV and LVI, and LVII and LVIII. As with the allenes, none of these substances contains an asymmetric atom.



LV LVI

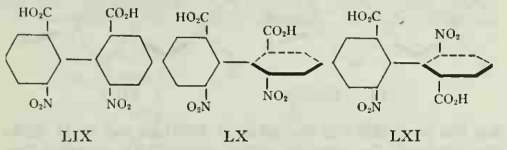


LVII



LVIII

9. Optically Active Biphenyls.—In 1922 G. H. Christie and J. Kenner found that the dinitrodiphenic acid LXIX can be separated into a pair of enantiomorphs. The present interpretation of this fact is that the molecules of the two substances have the configurations LX and LXI. The two rings here are considered to lie in approximately perpendicular planes. Since a rotation of 180° about the central carbon–carbon bond would transform either enantiomorph into the other, such rotation must be assumed to be impossible. The factor responsible for this restriction of rotation is presumed to be the size of the four substituents attached to the rings. In order for a molecule to pass from one enantiomorphic form to the other, it must, at some time during



LIX LX LXI

the process, assume a configuration in which the two rings are coplanar; if this should happen, however, the substituent groups mentioned would be forced to come so close together that they would get in one another's way. In other words, it is difficult, or impossible, for the large nitro and carboxyl groups to slip past each other.

As with the allenes and their analogues, neither this substance nor any of the numerous other ones similar to it possesses an asymmetric atom.

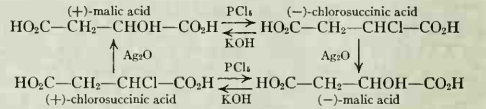
10. Optical and Geometrical Isomerism.—Stereoisomerism is often subdivided into optical and geometrical isomerism. The distinction between these two types of isomerism does not appear to be so significant as when it was first introduced; nevertheless, it has been retained by most chemists for historical reasons. A simple and precise definition of optical and geometrical isomerism appears to be impossible. On the one hand, many pairs of substances must be considered both optical and geometrical isomers of each other at the same time; and, on the other hand, there is no complete unanimity as to exactly where the dividing line between these two types of isomerism is to be drawn.

A typical example that is usually classified as illustrating geometrical isomerism is provided by maleic and fumaric acids, to which, as was stated above, the configurations XXII and XXIII, respectively, have been assigned. The characteristic feature of

XXII XXIII

these two substances which is of interest here is that, as a result of the rigidity of the double bond, the two carboxyl groups, the two remaining carbon atoms and the two remaining hydrogen atoms are held in the same plane. In many other instances also, two stereoisomers are said to be geometrical isomers of each other if they differ only in the relative positions in space of six coplanar atoms or groups of atoms. (With certain nitrogen compounds, either one or two of these atoms or groups may be missing. See *Stereochemistry of Nitrogen*, below.)

11. Walden Inversion.—In 1893, P. Walden discovered the cycle of reactions indicated below:



Thus, by treating either of the optically active malic acids first with phosphorus pentachloride (in chloroform solution) and then with potassium hydroxide (in aqueous solution), he was able to transform it into its own enantiomorph; similarly, by treating either of the active chlorosuccinic acids first with silver oxide and then with phosphorus pentachloride, he was able to transform it also into its enantiomorph. (In each sequence of reactions, considerable racemization occurs, so that the activities of the final products are not great. In other similar sequences that have been discovered subsequently, however, relatively little racemization occurs.)

It is evident that, in one (and only one) reaction of the two in each of these sequences (*i.e.*, in either the reaction with silver oxide or the one with phosphorus pentachloride, but not in both), the entering substituent at the asymmetric atom does not take up the same relative position in space as that formerly occupied by the substituent which it replaces. The change in configuration which thus results is called an inversion or, more precisely, a Walden inversion. (The expression Walden inversion was introduced by E. Fischer in 1906. Originally it referred to the complete sequence of reactions by which the optically active substance is converted into its enantiomorph, and not to the single reaction in which the inversion of configuration occurs. Later, however, the expression seems more usually to refer to the single reaction; it is so used in this article.) The problem of identifying the particular reaction in which the inversion occurs is a difficult one. From evidence which cannot be described here, however, it has been shown that inversions occur in the reactions of the above scheme in which phosphorus pentachloride and potassium hydroxide participate, but that there is retention of configuration in the ones with silver oxide. It is to be especially noted that the fact that the sign of rotation changes in the first two types of reaction but not in the last is purely accidental; the sign of rotation often changes in reactions which are considered to proceed with retention of configuration, and, conversely, it often remains the same in reactions which are considered to proceed with inversion. No necessary general relation exists between the configurations and signs of rotation of nonenantiomorphic substances.

With relatively few exceptions every reaction in which one substituent atom or group is replaced by another is believed to be

accompanied by an inversion of configuration about the atom at which the replacement occurs. Reactions, like that between chlorosuccinic acid and silver oxide, in which no net inversion is considered to occur, are interpreted as proceeding in two steps, with an inversion in each; two inversions, of course, are equivalent to a retention. No inversion is considered to occur at atoms at which no substitution takes place.

A probable mechanism of the Walden inversion is shown schematically in fig. 2. In fig. 2(A), representing an early stage of the reaction, the reagent R approaches the carbon atom from the side opposite the substituent S; in fig. 2(B), representing the midpoint of the reaction, the atoms or groups R and S are equidistant from the carbon atom, which is no longer tetrahedral; and in fig. 2(C), representing the conclusion of the reaction, the displaced substituent S is departing, and the entering one R has become attached to the carbon atom. In the process, the molecule has been "turned inside-out" like an umbrella in a high wind.

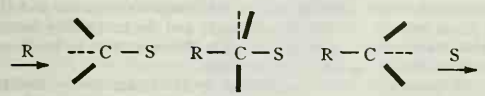
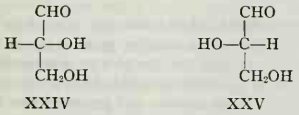


FIG. 2.—A PROPOSED MECHANISM OF THE WALDEN INVERSION
(A) An early stage; (B) mid-point; (C) conclusion

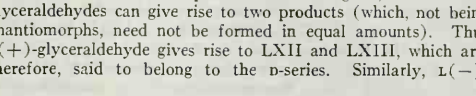
12. Optically Active Series.—The sign of the rotation produced by an optically active substance can occasionally be reversed by a change in the experimental conditions (see *Specific Rotation*, above). A further factor which introduces confusion into stereochemical classification is that a parent substance often gives derivatives of identical configuration but with opposite sign of rotation. For example, the salts and esters of the levorotatory lactic acid, CH_3—CHOH—CO_2H, are dextrorotatory; those of the dextrorotatory acid are levorotatory. For these reasons, it would seem desirable to include in the complete name of an optically active substance a symbol which refers directly to the configuration of the substance and is independent of the sign of rotation. This symbol would then remain unchanged when the substance, under different conditions, exhibits different signs of rotation, or when it forms derivatives of opposite rotation.

It may be doubted whether any completely general and completely unambiguous classification of configurations of the type desired is possible; certainly, no such classification has as yet been devised. Nevertheless, a scheme has been developed which permits the classification of configurations (but not without some ambiguity) within certain narrowly defined groups of compounds. This scheme consists in dividing configurations into those of the so-called D-series and those of the L-series on the basis of the genetic relationships among them. A substance is then said to belong to the D-series (or to the L-series) if it is a derivative of a substance belonging to the D-series (or L-series), respectively, and so on.

In order that an optically active series may be set up, some particular substance must be defined as belonging to the D- or L-series, so that the remaining substances may be related to it. Originally, Emil Fischer chose the naturally occurring sugars and their derivatives as the prototypes of the series. Later, the glyceraldehydes, XXIV and XXV, were chosen instead (M. A. Rosanoff, 1906); the two conventions lead to the same assignment of series, so that no ambiguity results.



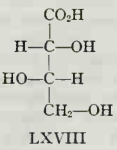
The way in which the optically active series are built up can be illustrated by a few examples. As has been mentioned before, (+)- and (−)-glyceraldehyde are by definition assigned the configurations XXIV and XXV, respectively. Also by definition, they are said to belong to the D- and L-series, so that their complete names are D(+)- and L(−)-glyceraldehyde, respectively. By well-known methods (*see* CARBOHYDRATES), the aldehyde group (CHO) of each of these substances can be changed into a CHOH—CHO group. Since this larger group contains an asymmetric carbon atom not present in the original, each of the glyceraldehydes can give rise to two products (which, not being enantiomorphs, need not be formed in equal amounts). Thus D(+)-glyceraldehyde gives rise to LXII and LXIII, which are, therefore, said to belong to the D-series. Similarly, L(−)-glyceraldehyde gives rise to LXIV and LXV, which are said to belong to the L-series.



The procedure can be continued further. Each of the tetroses LXII-LXV can be transformed into two pentoses, each of which can in turn be transformed into two hexoses, and so on. In each reaction, the two products belong to the same optically active series as the substance from which they were made. In general, and without explicit reference to the sequence of reactions starting with the glyceraldehydes, the optically active series to which any monosaccharide belongs may be said to be determined by the configuration about the asymmetric carbon atom farthest removed from the carbonyl group (C=O); when the chain of carbon atoms is written vertically and with the carbonyl group at the top, the hydroxyl group upon the atom in question is to the right of the chain in the D-series, and to the left in the L-series.

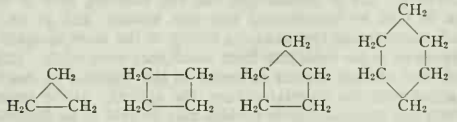
If the glyceraldehydes are oxidized, they give the glyceric acids LXVI and LXVII. The first of these, even though it is levorotatory, is said to belong to the D-series because of its genetic relationship to D-glyceraldehyde, it is accordingly called D(−)-glyceric acid. Similarly, the substance LXVII is called L(+)-glyceric acid.

With the α-amino acids, and frequently also with the α-hydroxy acids, the optically active series to which a given substance is assigned is determined by the configuration of the carbon atom adjacent to the carbonyl group, and not by that of the one farthest from this group. The two conventions are distinguished by the subscripts *s* and *g* which refer, respectively, to the simple amino acid serine, HO—CH_2—CN(N^+H_3)—CO_2^-, and to the simple sugar, glyceraldehyde XXIV (or XXV). For example, the substance with configuration LXVIII may be described either as D_s- or as L_g-trihydroxybutyric acid.



13. Strain Theory.—In 1885 A. Baeyer called attention to the fact that nearly all the then known carbocyclic compounds (*i.e.*,

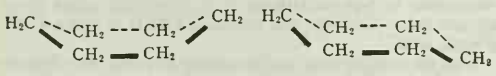
cyclic compounds with only carbon atoms in the ring, or rings) contained either five- or six-membered rings. To explain this fact, he brought forth a group of ideas now called the strain theory. The angle between any two single bonds formed by a saturated carbon atom had been postulated by van't Hoff to have the tetrahedral value of 109° 28′ (*see* above). Consequently, Baeyer considered that an angle of this magnitude is most stable, so that, if one or more angles in a molecule were forced in some way to have a widely different value, the molecule would be "strained," and, therefore, unstable. Now, in cyclopropane LXIX, the C–C–C angles of the three-membered ring must be equal to 60°; since this value is much smaller than the preferred 109° 28′, the molecule is strained. Thus, Baeyer considered that the difficulty of preparing such a substance is accounted for. Similarly in cyclobutane LXX, the angles would have to be equal to 90° if the four-membered ring is planar; again considerable strain is present, and again the difficulty of preparing the substance appears to be explained. The minimum strain is reached with cyclopentane LXXI, in which the planar ring requires bond-angles of 108°, almost identical with the tetrahedral value. With still larger rings, the strain increases regularly with the size if, as before, the rings are considered to be planar. The six-membered ring of cyclohexane LXXII requires an angle of 120°, the seven-membered ring of cycloheptane requires one of 128° 34′, and so on. Thus,



LXIX LXX LXXI LXXII

the least strain occurs in just those rings which are easiest to prepare and best known.

Baeyer's original ideas, although still considered essentially correct, have undergone extensive modification (H. Sachse, 1890). The strain postulated by Baeyer apparently exists only in the three- and four-membered rings, but no strain exists in the rings of six or more atoms. With the small rings, the least strain results if the rings are planar as assumed by Baeyer, but, with the larger rings, the strain can be relieved completely if the ring is permitted to be puckered. With cyclohexane, for example, two different strain-free configurations are possible; these are shown in LXXIII (the so-called boat-form or C-form) and LXXIV (the so-called chair form or Z-form). Transitions among the various



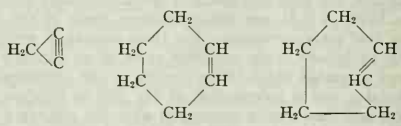
LXXIII LXXIV

forms of these puckered rings are so easy that isolation of stereoisomeric substances, differing only in the form of the ring, is impossible.

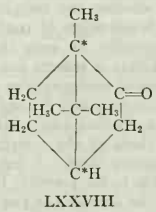
The assumption that the rings with six or more atoms are puckered and strain-free permits the conclusion that such large rings should be as stable as those with five atoms, and, in fact, as stable as the noncyclic chains. This conclusion is verified experimentally. There is no significant difference, referable to strain, between, say, cyclotriacontane, with 30 atoms in the ring, and cyclopentane, with only 5; nor is there any significant difference between either of these substances and any of the typical paraffin hydrocarbons (*see* Paraffin Hydrocarbons). In particular, neither ring can be broken easily by ordinary chemical reagents, and the heats of combustion of the two cyclic substances (*see* Thermochemistry), per CH_2 group, are practically identical. (In contrast, the three-membered ring in the strained cyclopropane LXIX is easily broken by reagents like bromine, hydrogen in the presence of a platinum catalyst, and hydrogen chloride, which have no similar effects upon cyclopentane, cyclotriacontane, or a paraffin hydrocarbon; and the heat of combustion of cyclopropane, per CH_2 group, is appreciably higher than that of either cyclopentane or cyclotriacontane.)

The reason why Baeyer was misled into believing the large rings to be strained was that, experimentally, such rings are difficult to prepare. More recently this fact has been interpreted differently. The difficulty of preparing cyclotriacontane, for example, is explained as the result, not of any assumed instability of the product, attributable to strain, but rather to the improbability that the two ends of a chain of 30 carbon atoms will happen to come close enough together to permit being joined to each other by a bond. The relative ease of forming the five-membered ring in cyclopentane, on the other hand, is attributed to the relatively great probability that the ends of a much shorter chain of only five carbon atoms will come close enough. In other words, the fact that the five- and six-membered rings are the most easily formed of all is explained as the result of two opposing effects: the strain, which operates against the formation of three- and (to a lesser extent) four-membered rings; and the probability factor, which causes the ease of formation to decrease more or less regularly as the size of the ring increases.

An important service performed by the strain theory consists in providing a reasonable explanation of the nonexistence of certain substances. For example, cyclopropyne LXXV is unknown although, as far as the ordinary rules of valence are concerned, it would appear to be a perfectly reasonable substance. However, if it existed, it would be extremely strained since the normal value of C—C≡C bond angles is 180°. Similarly, cyclohexene might be expected to be obtainable in cis and trans forms, analogous to maleic and fumaric acids, respectively. Only the cis form LXXVI is known; the trans form LXXVII would be too highly strained to be capable of existence. Moreover, camphor LXXVIII



LXXV LXXVI LXXVII

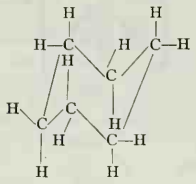


LXXVIII

(*see* Camphor) has the two asymmetric carbon atoms marked with the asterisks. Since these atoms are nonequivalent, meso forms are impossible, and so two pairs of enantiomorphs might be expected. Only one such pair is known. This is the one in which the attachment of the —$C(CH_3)_2$— "bridge" to the outer six-membered ring is cis; the second pair of enantiomorphs, in which the attachment is trans, would be too strained to be capable of existence.

14. Conformational Analysis.—If the chair configuration LXXIV of cyclohexane is written more fully as LXXIX, the hydrogen atoms are seen to be of two geometrically different types. Six of these atoms are joined to the carbon atoms by bonds which are parallel to one another and perpendicular to the average plane of the ring; these atoms and bonds are commonly described as axial. On the other hand, the remaining six hydrogen atoms are joined to the carbon atoms by bonds which lie approximately in the average plane of the ring; these atoms and bonds are commonly described as equatorial.

If one of the hydrogen atoms of cyclohexane is replaced by a substituent, which may be explicitly a methyl group, a halogen atom, or the like, two conformations consistent with the Z-form



LXXIX

of the ring (*see* above) are possible inasmuch as this substituent may occupy either an axial or an equatorial position. Usually, however, the equatorial position is favoured because it permits the widest possible separation of the atoms that are not bonded to each other.

It will be possible here only to mention, without discussion, two of the many important applications of conformational analysis, which have been made since about 1950. (1) With a disubstituted cyclohexane, cis and trans isomers (*see* above) are possible. Of these, the trans form is usually the more stable when the substituents are either 1,2 or 1,4 with respect to each other, but the cis form is usually the more stable when the substituents are 1,3 to each other. The reason for these generalizations is that the favoured configurations are the ones permitting the maximum numbers of equatorial substituents. (2) When two or more saturated six-membered rings are fused together, as in decahydronaphthalene, the steroids and numerous other compounds, the molecules are often more or less rigid, so that substituents in given positions are required to be either axial or equatorial. The conformations of the substituents, determined in this way, have been related to the relative stabilities of stereoisomeric forms and to the observed differences in reactivity. The details are, however, too complex to be given here.

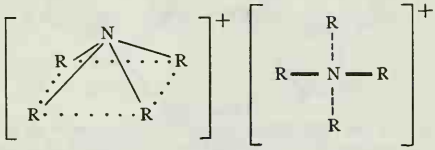
15. Steric Hindrance.—The rate of a chemical reaction is often decreased by the presence of nearby bulky substituents. For example, the esterification (*see* ESTERS) of benzoic acid (*q.v.*) becomes very slow if methyl groups, bromine atoms, or the like are introduced into the positions adjacent to the carboxyl group. This regularity is not limited to esterifications but is encountered throughout organic chemistry; an explanation for it was proposed by F. Kehrmann (1888) and by V. Meyer (1894). According to these authors, the substituents fill up the available space about the reactive centres and so make more difficult the approach of the reagent molecules.

This picture, although it contains an important element of truth, is doubtless oversimplified. Specific electrostatic interactions, resonance effects (*see* RESONANCE, THEORY OF), etc., would have also to be taken into account in any complete treatment of the relation between structure and reactivity, but further discussion cannot be given here.

V. THE STEREOCHEMISTRY OF NITROGEN

The general principles essential to the understanding of the stereochemistry of the elements other than carbon differ in no significant respect from those essential to the understanding of the stereochemistry of carbon itself. Since these principles are described in considerable detail in the preceding sections of this article, they need not be discussed further here. Moreover, since the factual material relating to the stereochemistry of the remaining elements is relatively restricted (although rather extensive in an absolute sense), a brief survey of the field is sufficient to bring out the important features.

In all optically active compounds known up to 1899, the molecular dissymmetry was dependent upon the spatial arrangement of the valencies of carbon. In that year, however, W. J. Pope and S. J. Peachey discovered that quaternary ammonium salts of the type $[NQRTZ]^+X^-$ (where Q, R, T and Z represent four dissimilar hydrocarbon radicals) could be resolved into enantiomorphs, and hence that the valency configurations of other elements besides carbon could have sufficient permanence to make their stereochemical investigation possible. (J. A. Le Bel had stated earlier that he had obtained an optically active substance of this type, but his work was shown to be incorrect.) Thus a wide field of research was opened. Further work on quaternary ammonium salts showed: (1) that, since molecular dissymmetry could no longer be demonstrated when two of the hydrocarbon radicals were identical, the ion $[NQRRT]^+$ probably has a plane of symmetry (no centre of symmetry is possible here, whatever the configuration about the nitrogen atom may be); and (2) that, since different methods of formation of a given salt, for instance $NQRT + ZX \rightarrow [NQRTZ]^+X^- \leftarrow NRTZ + QX$, never yielded isomeric modifications, the valencies linking the four hydrocarbon radicals to the nitrogen atom are most probably equivalent. Only two configurations of the ammonium ion satisfy these conditions, namely, the pyramidal configuration I and the tetrahedral configuration II; of these, the first has been definitely disproved by W. H. Mills and E. H. Warren (1925).

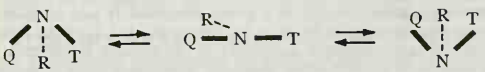


I II

The amine oxides are another class of compound in which a nitrogen atom forms bonds to four distinct other atoms or groups. In these also, the radicals appear to be tetrahedrally disposed about the nitrogen atom, for amine oxides of the type III have been obtained in optically active forms by J. Meisenheimer (1908).

III

The configuration of the trivalent nitrogen atom is of special interest because of the number and importance of the compounds in which the element is present in this state. The numerous attempts made to obtain optically active compounds of the type NQRT have all given negative results, but it is not safe to conclude from this fact that these compounds have a planar configuration; it is possible that instead they have a dissymmetric pyramidal configuration but are racemized easily by passage through an intermediate planar configuration, thus



That this second explanation of the optical inactivity of amines of the form NQRT is the correct one has been shown by studies of spectra, of dipole moments, of crystal structure, and of electron diffraction.

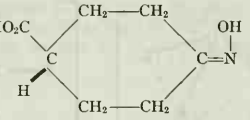
In compounds of trivalent nitrogen of the type —X=N—Y, where the nitrogen atom is directly linked to only two other atoms, and, therefore, to one of them by a double bond, the stereochemistry of the nitrogen atom was established at an early date by the classical stereochemical methods (*i.e.*, by consideration of isomer numbers and not with the use of the physical methods employed with the amines). Thus, oximes can generally be prepared in two isomeric modifications if the two substituents attached to the carbon atom doubly bonded to the nitrogen are dissimilar. The two isomeric forms are represented by the con-

figurations IV and V.

IV V

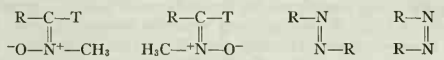
In these configurations the carbon, nitrogen and oxygen atoms, and the atoms of the groups R and T joined directly to the central carbon atom are considered to lie in the same plane. The situation is thus closely analogous to that in the ethylenic compounds considered previously, like maleic and fumaric acids; the only significant difference is that the nitrogen atom here forms three bonds whereas the corresponding carbon atom in the ethylenic compounds forms four. (Sometimes, the unshared pair of electrons on the nitrogen atom represented by the pair of dots, is said to act as a fourth valence bond.)

The above interpretation of the isomerism of the oximes, due to A. Hantzsch and A. Werner (1890), is confirmed by the optical activity of the cyclic oxime VI, demonstrated by W. H. Mills and A. M. Bain (1910).

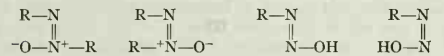


VI

Further examples of stereoisomeric nitrogen compounds, which owe their isomerism to restricted rotation about a double bond to the nitrogen atom, include the N-methyl oximes VII and VIII, the azo compounds IX and X and the azoxy compounds XI and XII. The so-called stable and labile diazohydrates (*see* Diazo Compounds) were also considered by Hantzsch to be stereoisomers



VII VIII IX X



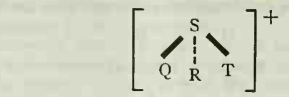
XI XII XIII XIV

of each other with the configurations XIII and XIV, respectively; this view, although probably correct, has been contested by other authors.

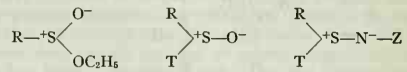
VI. THE STEREOCHEMISTRY OF THE REMAINING ELEMENTS

For reasons of convenience, the various remaining elements are classified below on the basis of their co-ordination numbers. In general, a given element in a given compound is said to have a co-ordination number N, or to be N-co-ordinated, if in the molecules of that compound each atom of the element in question is joined to N distinct other atoms by bonds that are not broken ionically in solution. For example, in sulfuric acid, H_2SO_4, the co-ordination number of sulfur is four, since each sulfur atom is joined to four oxygen atoms by bonds of the type stated; similarly, the co-ordination number of oxygen in the same substance is one, since the oxygen–hydrogen bonds are broken ionically in solution and so do not contribute to the co-ordination number. In those instances in which the nonionizable union between each atom and each of its neighbours consists of a covalent single bond, the co-ordination number of each atom is equal to its valence.

1. Elements of Co-ordination Number 3.—Sulfonium salts $[SQRT]^+X^-$, in which three dissimilar hydrocarbon radicals are linked to the sulfur atom, were shown to be dissymmetric by W. J. Pope and S. J. Peachey, and S. Smiles (1900); the sulfonium ion, therefore, has a nonplanar configuration as in XV. Other analogous compounds of sulfur are the sulfinic esters XVI, the sulfoxides



XV



XVI XVII XVIII

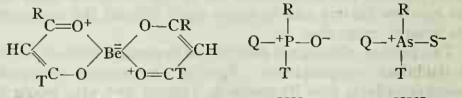
XVII and the sulfilimines XVIII; each of these has been shown to have a similar nonplanar configuration, since optically active representatives of the three classes have been obtained, principally by H. Phillips and J. Kenyon (1925–27).

Optically active selenonium salts (analogous to the above sulfonium salts) were obtained by W. J. Pope and A. Neville (1902), and optically active telluronium salts were obtained by T. M. Lowry and F. L. Gilbert (1929). Consequently, the 3-co-ordinated selenium and tellurium atoms must be nonplanar.

That trivalent arsenic is able to maintain a pyramidal configuration has been shown by the resolution of an arsine, AsRQT, with three dissimilar substituents. In this respect, arsenic resembles sulfur, selenium and tellurium rather than nitrogen. Optically inactive stereoisomers of presumably the cis-trans type have been obtained when two trivalent arsenic atoms are present in a ring.

2. Elements of Co-ordination Number 4.—In the first short period (*see* Periodic Law, The), besides carbon and nitrogen, the elements beryllium and boron can have co-ordination number four, and there is evidence that the atoms of each of these elements then have the four radicals tetrahedrally disposed about them. This was demonstrated for beryllium by the production of an optically active compound of the type XIX (W. H. Mills and R. A. Gotts, 1926), and for boron in a similar way by J. Boëseken and J. Meulenhoff (1924).

The molecular dissymmetry of silicon compounds of the type SiQRTZ has been established by F. S. Kipping, and the investigation of the configuration of 4-co-ordinated zinc by the method used for 4-co-ordinated beryllium has shown that its valencies are also tetrahedrally distributed. Phosphorus and arsenic, like nitrogen, have also been shown to have the tetrahedral configuration when 4-co-ordinated—phosphorus through the optical activity



XIX XX XXI

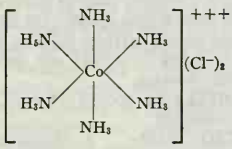
of a phosphine oxide XX (J. Meisenheimer and L. Lichtenstadt, 1911), and arsenic through that of an arsine sulfide XXI. Optically active arsine oxides $RQTAs^+—O^-$ and arsonium salts $[RQTZAs]^+X^-$ are also known. The existence of diastereomers and of meso forms has been demonstrated with compounds containing two such asymmetric phosphorus or arsenic atoms.

It is not universally true, however, that when an atom is linked directly to four other atoms these must be arranged around it tetrahedrally. Thus, many compounds of 4-co-ordinated platinum containing a complex of the type PtA_2B_2 occur in two inactive, but evidently stereoisomeric, modifications; typical examples include the doubly charged cation $[Pt(NH_3)_2(C_5H_5N)_2]^{++}$ and

XXII XXIII

the neutral molecule $Pt(NH_3)_2Cl_2$. This isomerism is clearly incompatible with a tetrahedral arrangement of the radicals about the platinum atom and is to be explained most simply by the supposition that 4-co-ordinated platinum has a plane configuration. The isomers are of the types XXII (cis) and XXIII (trans).

3. Elements of Co-ordination Number 6.—Werner's proof that substances like the compounds of metallic salts with ammonia (formerly regarded as "molecular compounds") are co-ordination compounds in which a certain number of atoms or groups, most frequently six, are directly attached to the metallic atom ($CoCl_3.6NH_3$, for example, having the structure XXIV)



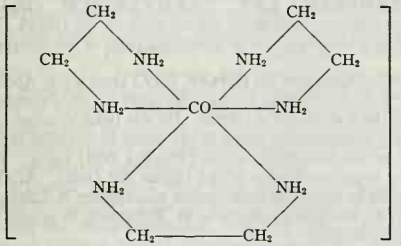
XXIV

brought with it the problem of determining how these six groups are disposed about the central atom. If it is assumed that the valencies by which they are attached are equivalent (and this is supported by the absence of isomerism which should otherwise occur), three arrangements are possible: the plane hexagonal XXV, the prismatic XXVI and the octahedral XXVII. (For

XXV XXVI XXVII

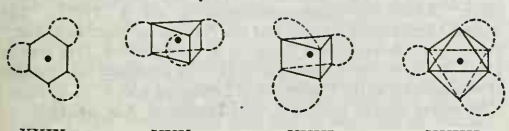
ease of visualization, the respective geometrical figures are outlined, the central atom is represented by a large dot and the valence bonds are not shown. A substituent is supposed to occupy each of the positions marked by a corner of the figure.)

A decision among these three possible configurations was obtained by investigating the compounds formed by linking the co-ordination positions together in pairs through four-membered chains, as can be done, for example, by the threefold replacement of two molecules of ammonia. NH_3, by one of ethylenediamine, $H_2N—CH_2—CH_2—NH_2$ (or of two nitrito groups, NO_2^-, by an oxalato group $—O_2C—CO_2^-$, or the like) so as to produce a complex in which the metallic atom forms a common member of three symmetrical five-membered rings, as in XXVIII. Since it seems



XXVIII

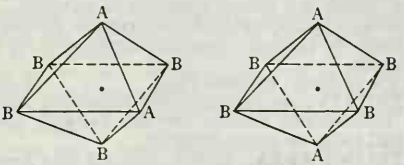
most probable that only adjacent co-ordination positions can be bridged in this way, the possible configurations for the tricyclic ion are those indicated by the dotted lines in the four different



XXIX XXX XXXI XXXII

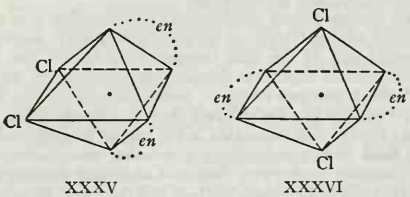
diagrams XXIX–XXXII. Of these, the first three are evidently symmetrical, but the last is dissymmetrical, for it possesses only twofold and threefold axes of rotation. Experimental investigation of the salts of these tricyclic ions has shown that they can be produced in optically active forms. The ions are thus dissymmetric and must consequently have the last configuration XXXII. It is, therefore, to be concluded that in such compounds the six valencies of the central metallic atom are octahedrally disposed.

This conclusion is confirmed by the occurrence of a different type of stereoisomerism shown by complexes of the type MA_2B_4, such as are contained in the diacidotetrammine salts of cobalt and chromium. These compounds should exist in three isomeric forms if their configuration is hexagonal and in four if it is prismatic, but in only two if it is octahedral; the alternative arrangements are then those in which the two groups A lie at the ends



XXXIII XXXIV

(1) of an edge as in XXXIII (cis compound), and (2) of a diagonal as in XXXIV (trans compound) of the octahedron. The number of isomers actually found is two. Moreover, in dicyclic derivatives of this type like $[en_2CoCl_2]^+ Cl^-$, where *en* represents ethylenediamine, the isomer with the cis configuration can be obtained optically active, whereas the other, the trans isomer, cannot be resolved. These facts are in accordance with the con-



XXXV XXXVI

figuration XXXV for the cation of the cis isomer, and XXXVI for the cation of the trans isomer.

BIBLIOGRAPHY.—J. H. van't Hoff, *La Chimie dans l'Espace* (1875), *Arrangement of Atoms in Space* (1877); A. Werner, *Lehrbuch der Stereochemie* (1904); F. M. Jaeger, *Lectures on the Principle of Symmetry* (1917); K. Freudenberg, *Stereochemie* (1933); R. L. Shriner, R. Adams and C. S. Marvel, "Stereoisomerism," ch. 4 of H. Gilman's *Organic Chemistry*, 2nd ed. (1943); G. W. Wheland, *Advanced Organic Chemistry*, 2nd ed. (1949); W. Klyne (ed.), *Progress in Stereochemistry* (1954); M. S. Newman (ed.), *Steric Effects in Organic Chemistry* (1956). (G. W. WD.; W. H. MI.)

STEREOCOMPARATOR, an instrument for comparing two astronomical photographs of the same star field, commonly combined with a Blink microscope (*q.v.*).

STEREOSCOPE. An optical instrument used to combine two pictures of the same object taken from slightly different points of view, thereby giving the impression of a single picture having depth.

See BINOCULAR INSTRUMENT.

STEREOTYPING, in printing, is the process of making a duplicate metal or plastic printing plate cast from a mold or matrix of composed type.

In 1725, William Ged (*q.v.*) of Edinburgh developed a metal plate with which he could print on a standard letterpress. By impressing a page of set type into moist gypsum, Ged produced a matrix from which he could cast his plate of type metal. His efforts to publish books by stereotype failed and after his death his invention was almost forgotten. It was reintroduced in 1804

by Charles Mahon, 3rd earl of Stanhope, who used plaster of paris for his molds.

Stereotyping was not markedly successful until after the introduction of the papier-mâché matrix by Claude Genoux, a French printer, in 1829. In the Genoux process, layers of wet blotting and tissue paper, called flong, were placed on the composed type and the wet paper was forced down around the type by beating with a broad, stiff brush.

New and improved equipment was developed as a result of the adoption of the paper matrix. Paper molds of set type were prepared by revolving rollers or by direct-pressure machines. The curved casting box was developed in 1854 and, with the production of the curved plate, printing by rotary presses became practicable. The dry-mat process, which substituted dry mats for wet ones and is the process employed in most modern stereotyping, appeared in the last decade of the 19th century. Modern matmaking, casting and printing equipment is automatic. Plastic materials have been utilized in the preparation of printing plates; these plates offer lightness, strength, a smooth surface and faithful reproduction.

See also PRINTING; PRINTING PRESS. (E. L. Y.)

STERILITY: *see* FERTILITY AND FECUNDITY; REPRODUCTION; REPRODUCTIVE SYSTEM: *Sterility; Spermatic Cord.*

STERLING: *see* POUND STERLING.

STERLING AREA. In Sept. 1939 and more formally in July 1940 a group of British Commonwealth and associated countries took steps to ensure the continuance of co-operation in the management of their exchange reserves and external monetary policies. Following the devaluation of sterling in Sept. 1931 the British and other countries that continued to maintain parity with sterling and to hold their exchange assets in London had become known as the sterling bloc. When World War II brought exchange control and prospective exchange shortage, particularly of dollars, those countries with the closest ties to London adopted parallel policies of monetary collaboration.

There was no formal constitution of the sterling area and no centralized administration or elaborate machinery. Members in the second half of the 20th century were, besides the United Kingdom and the dependent colonies and protectorates, Australia, New Zealand, Southern Rhodesia, India, Pakistan, Ceylon, Burma, Libya, Iraq, Jordan, Ireland and Iceland. South Africa was a member insofar as it contracted to sell part of its gold production to the Bank of England. Egypt and Sudan were members until 1947, Palestine until 1948, Belgian Congo and Ruanda-Urundi 1941–44, Syria and Lebanon 1941–44 and the Faroe Islands 1941–48. Some of the colonies that were dependent in 1939—Gold Coast, Nigeria, Malaya, Northern Rhodesia, the West Indies—subsequently became self-governing; all remained members but, since they were large dollar-earners, it was anticipated that their independence would test the viability of the sterling area.

Within the sterling area there was relative freedom of transferability, though at times restrictions were imposed, particularly in respect to blocked sterling balances. Members kept their currencies aligned with sterling, but not necessarily at full prewar parity. The overseas members kept their exchange assets in London. Generally the independent members drew from the dollar pool as much as they put into it.

The dependent members, including until the late 1950s Ghana, Nigeria and Federation of Malaya, earned dollars and drew (or accumulated) sterling. Their continued membership rested upon access to the British capital market as well as upon traditional currency and trading ties.

Except for the dependent colonies, the administration of exchange control and the use of dollars rested in the hands of the central banks who followed parallel policies and maintained close consultative co-operation with the Bank of England and with each other. In 1947 the Sterling Area Statistical committee was formed to facilitate the exchange of information.

Periodic meetings of the Commonwealth finance ministers maintained liaison among governments, but actual administration was decentralized and dependent upon co-operative understanding, particularly among the central banks.

BIBLIOGRAPHY.—Philip W. Bell, *The Sterling Area in the Postwar World* (1956); Economic Co-operation Administration, *The Sterling Area: an American Analysis* (1951); A. R. Conan, *The Sterling Area* (1952). (J. B. CE.)

STERLITAMAK, a town of the Bashkir Autonomous Soviet Socialist Republic in the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the left bank of the Belaya river, at its confluence with the Sterlya, about 75 mi. S. of Ufa. Pop. (1959) 111,575. Founded in the 18th century, it only acquired importance with the industrialization of the Urals after the Revolution and especially with the development of the Second Baku oil field after World War II. It is a crossing point of railways southward from Ufa and westward from Magnitogorsk. Oil refining is important and other industries include machine building, production of synthetic rubber, soda (using local salt) and cement, flour milling, distilling and leather working. (R. A. F.)

STERN, DANIEL: *see* AGOULT, MARIE CATHERINE SOPHIE DE FLAVIGNY, COMTESSE D'.

STERN, OTTO (1888–), theoretical physicist and Nobel prize winner, was born in Sorau, Upper Silesia, on Feb. 17, 1888. He received the degree of doctor of philosophy in physical chemistry from the University of Breslau in 1912. Interested in the research work of Albert Einstein, he joined Einstein at the University of Prague and followed him to the University of Zürich, where he received the *venia legendi* for physical chemistry. Early in 1914 Stern joined the faculty of the University of Frankfurt, where he remained until 1921 except for a period of military service during World War I.

After a short term as associate professor of theoretical physics at the University of Rostock, he was appointed to the chair of physical chemistry at the University of Hamburg in 1923. Resigning this position in 1933 he accepted shortly thereafter an invitation to join the Carnegie Institute of Technology in Pittsburgh, Pa., as research professor of physics. In 1945 he retired as emeritus professor to Berkeley, Calif.

Stern's earlier scientific contributions belong to the field of theoretical physics, the most important ones dealing with statistical thermodynamics. After 1919, however, his research was more experimental and was mostly devoted to the development and application of the molecular beam method, a very powerful tool for the investigation of the properties of atoms, molecules and atomic nuclei. A series of brilliant experiments designed to answer crucial questions of physical theory won him international recognition.

Stern was awarded the Nobel prize for physics in 1943 for the first measurement of the magnetic moment of the proton.

(I. EN.)

STERNBERG, LEV YAKOVLEVICH (1861–1927), Russian ethnographer, noted for his work on the Gilyak (Nivkhi) of Sakhalin Island and the neighbouring parts of the Siberian mainland.

Born in Zhitomir in the Ukraine, May 3 (new style; April 21, old style), 1861, he studied at the University of St. Petersburg and later in the law faculty in Odessa. He was twice arrested for revolutionary activities and exiled to the island of Sakhalin for participation in the Narodnaya Volya (People's Will) party. While in exile he began his studies of the Gilyak and Oroki. After return from exile he remained under police surveillance in Zhitomir until in 1899 he was able to return to St. Petersburg to pursue his ethnographic studies.

From 1902 through 1917 he continued his work, at the same time being active in Russian and Jewish social movements. He traveled to museums in Berlin, Leipzig and elsewhere, and he also studied the Ainu and the Tungus peoples ("The Ainu Problem," *Anthropos*, vol. 24, pp. 755–799 [1929]; "Die Religionen der Giljaken," *Archiv für Religionswissenschaft*, vol. 8 [1905]; "The Turano-Ganowanian System and the Nations of Northeast Asia," *Proceedings*, International Congress of Americanists, 18th session [1912]). After the Revolution of 1917 he organized the teaching of ethnography at the University of Leningrad with V. G. Bogoraz.

Sternberg died at Dudergof, near Leningrad, Aug. 14, 1927.

(L. K.)

STERNE, LAURENCE (1713–1768), English novelist, the author of *Tristram Shandy,* was born at Clonmel, Tipperary, on Nov. 24, 1713. His father Roger Sterne was a poor and improvident soldier who belonged to a gifted and hardheaded family which had produced an archbishop of York, a great-grandfather of the novelist. Sterne disliked his mother, who was of partly French descent; but he idolized his father, and his early memories of him, and of the scenes of military life that he saw as he moved about with his mother after his father's regiment, were to serve him as inspiration later in life. In 1723 or 1724 he was sent to a school near Halifax, Yorkshire, and in 1733 he entered Jesus college, Cambridge, as a sizar. At Cambridge, where he is said to have "read a little, laugh'd a great deal" and amused himself by "puzzling his tutors," he became friendly with John Hall (later Hall-Stevenson), a wealthy young wit of dissipated habits whose home at Skelton castle (renamed "Crazy castle") he was often to visit. He delighted in exploring the extensive library there, which was rich in what booksellers term "curiosa." From an early age Sterne had been an inquisitive reader: from his own writings it is clear that he knew such writers as Rabelais, Cervantes, Burton, Montaigne, Erasmus, Horace and Swift. He was also a great admirer of "the sagacious Locke," whose theory of the "association of ideas" lies behind *Tristram Shandy.*

Quite early in life Sterne realized that he was tubercular, and this discovery increased his propensity to snatch at life and its pleasures while there was still time. It was no sense of vocation that led him to make his career in the church, but rather the influence of his uncle Jacques, a pluralist who was both archdeacon of Cleveland and precentor of York. Having taken his degree in 1737 Sterne entered the church the following year. Although his vicarage was at Sutton-on-the-Forest near York he preferred to live in York, then a centre of fashionable life and church politics. In 1741 he married a wealthy young woman, Elisabeth Lumley. She was a difficult woman, irritable and self-righteous, while Sterne was a capricious and unfaithful husband, so that the marriage was not a success. In the year of his marriage Sterne became a prebendary of York. He frequently preached in York cathedral, with considerable success: although he had a weak voice he studied the art of preaching, and it is partly to this that the interest in posture and gesture that is so evident in *Tristram Shandy* must be attributed.

But for an ecclesiastical controversy his famous book might never have been written. Sterne began as a controversialist in support of his uncle. Then the two men fell out, and Jacques Sterne managed to do his nephew permanent harm by spreading a story that the novelist's mother and sister were lodging in a debtor's prison because he was too mean to support them. For a while Sterne withdrew from ecclesiastical politics, but in 1758 he returned and wrote *A Political Romance,* published early in 1759 and later renamed *The History of a Good Warm Watch-Coat.* This satire on a Dr. Topham who had drawn into his own hands most of the legal offices connected with the diocese of York is a precursor of *Tristram Shandy,* and it must have been while he was writing it that he realized his genius for comedy and satire. At the beginning of 1759 he wrote the opening volumes of *Tristram Shandy,* which were published on Jan. 1, 1760.

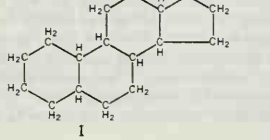
Like *Hudibras* and *Don Juan* (two satires with which it has something in common), *Tristram Shandy* came out in parts (nine in all) and was never completed. The full title—*The Life and Opinions of Tristram Shandy, Gentleman*—gives one a hint of the nature of the book: this is not (like so many novels of the time) *The Life and Adventures* of somebody: it deals with opinions, not with things. The reason for this is given in the motto of the first two volumes: "It is not things, but opinions about things, that throw mankind into confusion." Talk rather than action constitutes the "plot" of the book. Thus *Tristram Shandy* could not possibly have been the first novel: it is (among other things) a sort of satire on its own precursors: it cocks a snook at the rigorous morality of Samuel Richardson, the careful plot-construction of Henry Fielding and the violent adventures of Tobias Smollett. Its own morality may best be described as "relaxed": it has (in the usual sense) no plot whatever: while its characters, so far from being involved in exciting adventures, seldom even venture out-of-doors. The book opens with an account of the begetting (not the birth) of its hero, so providing Sterne with a splendid occasion for a discussion of heredity, education and life in general. It takes the hero a long time to get born, but the time is most agreeably filled in by an account of the conversations of his father and Uncle Toby *de omnibus rebus et quibusdam aliis.* The most important structural device is that of the association of ideas: instead of limiting himself to one subject, as most writers do, Sterne allows his mind to wander freely from one topic to another, as happens in a state of reverie. The result is that his book, while it lacks all logic of the usual sort, has yet a subliminal logic of its own which gives it a remarkable hold on the minds of its readers.

On the publication of the first part of *Tristram Shandy* Sterne found himself famous, though many readers were shocked when they discovered that it was the work of a cleric. He moved to a new house near Stillington and went on with his book. He delighted in his celebrity, dining out constantly on his visits to London; but his health declined, and in 1762 he set out on an extended trip in France and Italy in search of the sun and a renewal of health. In France he found himself as famous as in his native country.

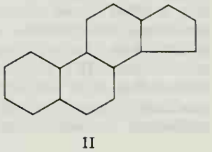
The main events during the remainder of his life were the publication of further installments of *Tristram Shandy* and his infatuation with Elizabeth Draper, a deeper and more sentimental love affair than most in his life. *The Journal to Eliza,* which he wrote for her, reveals the pathological state of the emotions which inspired *A Sentimental Journey Through France and Italy* (1768), a travel book of a new kind in which Sterne concentrates on the effect of his travels on himself, dealing particularly with such "sentimental" experiences as the flirtations in which he became involved as he traveled about the continent. As a sort of riposte to Smollett's *Travels Through France and Italy,* Sterne concentrates on whatever had called forth his affections. The result is a work much slighter than *Tristram Shandy* which yet had a great influence on European literature and helped to give the word "sentimental" its international currency. He died in London on March 18, 1768.

BIBLIOGRAPHY.—The first good edition of his *Works* (including his *Sermons* and *Letters*) appeared in 10 volumes in 1780. There is a 12-vol. edition by Wilbur L. Cross (1904). The only authentic text of *Tristram Shandy,* however, is that edited by J. A. Work (1940). The *Letters* were edited by L. P. Curtis (1934). J. Ferriar's *Illustrations of Sterne* (1798; rev. ed. 1812) is an interesting early study. Much the best biographical and critical studies are *The Life and Times of Laurence Sterne,* by W. L. Cross (1909; rev. ed., 1929) and *Laurence Sterne, de l'homme à l'oeuvre,* by H. Fluchère (1961). (I. J.)

STEROIDS in chemistry and biology comprise a large group of naturally occurring substances and their synthetic derivatives. All possess as a fundamental structure the four-membered hydrocarbon ring system known as perhydrocyclopenta-(a)-phenanthrene (I), for convenience represented by the diagram (II).



I



II

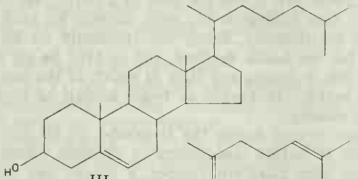
The steroids are colourless crystalline solids that differ among themselves in the nature of substituent groups attached to this basic nucleus. Steroid compounds are intimately concerned with a variety of life processes and may be divided into six principal classes: sterols, bile acids, sex hormones, adrenal cortical hormones, cardiotonic agents and sapogenins as discussed below.

Because of the large variety of known steroid compounds, the problem of naming them is formidable. Two systems of steroid nomenclature are employed. So-called trivial nomenclature is based on the arbitrary names given to important natural steroids

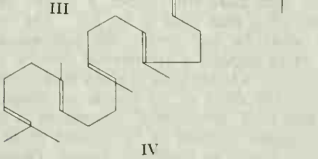
upon their discovery; compounds closely related in structure to these substances are named as derivatives thereof. In systematic nomenclature compounds are designated as derivatives of certain reference hydrocarbons, and the systematic name denotes unambiguously the complete structure of the compound. Both types of names are employed in the following section; an explanation of steroid nomenclature is given in the section *Steroid Nomenclature* below.

CLASSES OF STEROIDS

Sterols.—The sterols are constituents of the lipid material of the cells of practically all living organisms. In all these compounds the steroid nucleus is substituted with a β-hydroxyl group at position 3 and with a branched side chain containing eight, nine or ten carbon atoms. The predominant sterol of the higher animals is cholesterol (III), whereas lower animals and plants



III



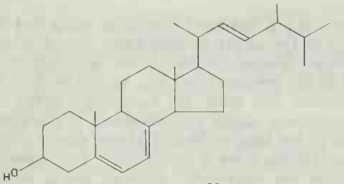
IV

possess a variety of other sterols, most of them with the longer side chains. Some of the more than 30 different sterols that have been isolated from various species are listed in Table I.

TABLE I.—*Representative Sterols*

Trivial name	Systematic name	Source
Cholesterol	cholest-5-en-3β-ol	animal cells
Cholestanol (dihydrocholesterol)	5α-cholestan-3β-ol	animal cells
Coprostanol	5β-cholestan-3β-ol	feces
Lathosterol	5α-cholest-7-en-3β-ol	with cholesterol
7-Dehydrocholesterol (provitamin D_3)	cholesta-5,7-dien-3β-ol	with cholesterol
Ergosterol (provitamin D_2)	ergosta-5,7,22-trien-3β-ol	yeast, ergot
Stigmasterol	stigmasta-5,22-dien-3β-ol	soybean
Lanosterol	4,4,14-trimethyl-5α-cholesta-8, 24-dien-3β-ol	wool fat, yeast, liver

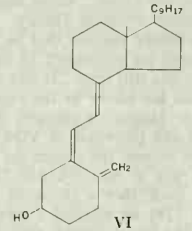
Cholesterol (*q.v.*) is particularly abundant in hair, nervous tissue, adrenal glands and egg yolk, occurring both in the free form and esterified with various fatty acids. Cholesterol is obtained commercially by extraction from the brains and spinal cords of cattle. In animals cholesterol is synthesized in the liver, and probably in other tissues as well, from acetic acid which, by way of acetoacetic and mevalonic acids, is converted to five-carbon isoprenoid units. Six of these units combine to form the linear molecule squalene (IV), which by appropriate folding and ring closure leads to lanosterol and finally to cholesterol. Cholesterol is metabolized to cholestanol and coprostanol, which are excreted. One established biochemical function of cholesterol is to serve as a precursor for the biosynthesis of steroid hormones in the sex and adrenal glands; other physiological functions of cholesterol remain obscure. In contrast to the cholesterol of most tissues, which is in continual metabolic turnover, the cholesterol of the brain and nervous tissue is metabolically inert.



V

In the pathological condition of atherosclerosis, the deposition of cholesterol occurs within the lining of the aorta and arteries, a situation that often leads to coronary thrombosis and death. The amount and nature of fat in the diet markedly influences this phenomenon, as do the steroid sex hormones, since this disease is much less common in premenopausal women than it is in men and in postmenopausal women.

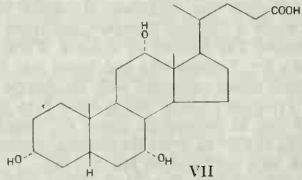
Certain sterols that possess double bonds in both the 5 and 7 positions have an important relationship to vitamin D, a factor that is necessary for the proper deposition of calcium in the growing bone; absence of vitamin D leads to the pathological condition of rickets. On irradiation with ultraviolet light, ergosterol (V) is converted to calciferol or vitamin D_2 (VI); 7-dehydrocholesterol similarly yields vitamin D_3, which differs from calciferol only in possessing the cholesterol (C_8H_{17}) rather than the ergosterol (C_9H_{17}) side chain. Both substances are of about



VI

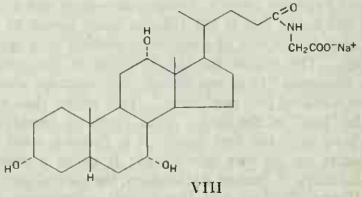
equal antirachitic potency in humans, but vitamin D_3 is especially potent in chicks. The preparation of ergosterol from yeast and the synthesis of 7-dehydrocholesterol from cholesterol for transformation to the D-vitamins are important commercial processes. The antirachitic effects obtained by exposure of an individual to sunlight result from the photochemical conversion to vitamin D_3 of the small amounts of 7-dehydrocholesterol that accompany cholesterol in the skin. Similarly, the irradiation of milk endows it with vitamin D_3 formed from the 7-dehydrocholesterol present.

Bile Acids.—The bile acids constitute a group of acidic steroids that are present in animal bile in the form of bile salts in which the bile acid is linked through a peptide bond to the amino acids glycine or taurine. The predominant bile acid in human or cattle



VII

bile is cholic acid (VII), which occurs chiefly as the glycocholate bile salt (VIII). The acids of human (and cattle) bile are listed



VIII

in Table II; related bile acids have been isolated from other species.

TABLE II.—*Human Bile Acids*

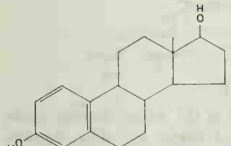
Trivial name	Systematic name
Cholic acid	3α,7α,12α-trihydroxy-5β-cholanic acid
Deoxycholic acid	3α,12α-dihydroxy-5β-cholanic acid
Chenodeoxycholic acid	3α,7α-dihydroxy-5β-cholanic acid
Lithocholic acid	3α-hydroxy-5β-cholanic acid

The bile acids and salts are synthesized in the liver from cholesterol and are discharged with the bile into the duodenum. The bile salts are surface-active agents that promote the absorption of fats through the intestinal wall. In conditions such as obstructive jaundice when sufficient amounts of bile do not reach the intestine, absorption of lipids is impaired.

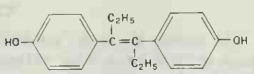
The bile acids are obtained commercially from cattle bile. Deoxycholic acid is an important raw material for one industrial process for the production of cortisone and other related anti-inflammatory steroids. Cholic acid, the most abundant bile acid, can be transformed into deoxycholic acid by chemical removal of the 7α-hydroxyl group.

Sex Hormones.—The steroid sex hormones are produced primarily in the gonads (ovary in the female and testis in the male) under the stimulation of protein substances called gonadotropins secreted by the pituitary. The sex hormones are necessary for the growth and development of the reproductive tract and accessory sex organs and for their continued maintenance and function.

Two types of ovarian hormones are produced at different stages of the ovarian cycle. The follicle containing the ripening ovum produces the estrogenic hormone, estradiol (estra-1,3,5(10)-triene-3,17β-diol; British: oestradiol) (IX), which is primarily respon-

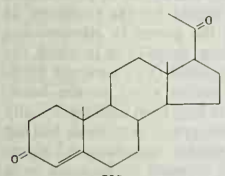


IX

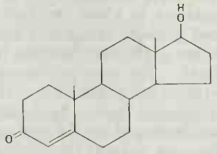


X

sible for the development of the female organs and the general feminine characteristics. After the mature ovum has been expelled into the Fallopian tube, the wall of the ruptured ovarian follicle forms a yellowish body (corpus luteum) that secretes the progestational hormone, progesterone (pregn-4-ene-3,20-dione) (XI).



XI



XII

This substance exerts a marked stimulatory effect on the growth and secretion of the uterine lining. It is necessary for the proper implantation of the fertilized ovum, for the maintenance of pregnancy and for the preparation of the mammary gland for lactation.

The testis in the male produces chiefly the androgenic hormone, testosterone (17β-hydroxyandrost-4-en-3-one) (XII). Testosterone causes the development and function of the male sex organs and such masculine characteristics as facial hair, deepening of the voice, etc. In addition to promoting marked growth of specific sexual tissues, testosterone exhibits a general anabolic or tissue-building effect.

The activity of an estrogenic hormone is evaluated by its ability to promote growth of the uterus or cellular changes in the vaginal epithelium of a castrated or immature rat or mouse. Progestational activity is determined by the extent of proliferation evoked in the uterine endometrium of a rabbit that has been previously stimulated with estrogen. Androgenic activity is measured by the stimulation of growth either of the comb in chicks and capons or of the prostate gland and seminal vesicles of the immature or castrated rat.

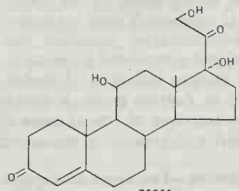
Since insufficiency of progesterone production in an individual during pregnancy may result in abortion, the chief clinical use of progestational hormones is for the maintenance of pregnancy and prevention of miscarriage. Progesterone also is used to alleviate pain and headache accompanying menstruation. Estrogenic hormones are employed for the relief of physical and psychic symptoms accompanying the menopause, for the suppression of lactation and for the treatment of prostatic and some types of breast cancer. There are indications that estrogenic hormones may be of benefit in the prevention or amelioration of atherosclerosis. A combination of estrogenic and progestational hormones are used in the treatment of ovarian hypofunction (amenorrhea).

Testosterone and similar androgenic hormones are used clinically for the treatment of testicular insufficiency (eunuchoidism), the suppression of lactation, the therapy of certain types of breast cancer and the treatment of frigidity in women. Because of their general anabolic effect, androgenic hormones are sometimes administered to promote tissue building in debilitated conditions.

Since progesterone and testosterone are not very active when given by mouth, certain related steroids are employed clinically for oral administration. 17α-methyltestosterone and 9α-fluoro-11β-hydroxy-17α-methyltestosterone are used as oral androgens, whereas 6α-methyl-17α-acetoxyprogesterone, 17α-ethynyltestosterone (ethisterone), 17α-ethynyl-19-nortestosterone (norethisterone) and 17α-ethynyl-17β-hydroxyestr-5(10)-en-3-one (norethynodrel) are used as oral progestins. These latter substances have the ability to inhibit ovulation and show promise as oral contraceptive agents. The estrogens are unique among the steroid hormones in that their physiological actions can be duplicated by certain simple organic compounds; one such synthetic estrogen that finds extensive clinical use is diethylstilbestrol (X). The steroid sex hormones are manufactured commercially from more abundant steroids, chiefly sapogenins; diethylstilbestrol is synthesized inexpensively from simple organic chemicals.

The biochemical precursor of the sex hormones is cholesterol. Estradiol is converted metabolically to estrone (3-hydroxyestra-1,3,5(10)-trien-17-one), also an active estrogen, estriol (estra-1,3,5(10)-triene-3,16α,17β-triol) and 16-epiestriol (estra-1,3,5(10)-triene-3,16β,17β-triol), all of which are excreted in the urine. The principal urinary metabolite of progesterone is pregnanediol (5β-pregnane-3α,20α-diol), which furnishes a measure of progesterone production in an individual. During pregnancy, urinary pregnanediol levels usually are checked periodically to determine whether additional progesterone need be administered to prevent abortion. Testosterone is metabolically converted to androsterone (3α-hydroxy-5α-androstan-17-one) and etiocholanolone (3α-hydroxy-5β-androstan-17-one), which are excreted in the urine; the latter substance is pyrogenic when administered to humans, and its production in abnormal amounts has been implicated in certain hitherto unexplained conditions of fever.

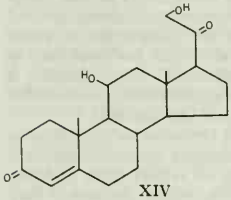
Adrenal Cortical Hormones.—The cortex or outer capsule of the adrenal gland secretes a variety of steroid hormones when stimulated by a protein hormone from the pituitary called adrenocorticotropin (ACTH). The major functions of the adrenal secretions are the regulation of salt and water balance of the body by promoting the retention of sodium and excretion of potassium



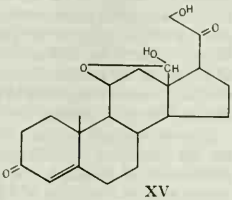
XIII

ions and, by their effect on carbohydrate metabolism, the enabling of the individual to cope with situations of stress. Unlike the sex hormones the adrenal secretions are necessary for life; an animal or individual cannot survive adrenalectomy for long unless adrenal hormones are administered.

Although more than 30 different steroids have been isolated from adrenal tissue, the principal hormones secreted by this gland are cortisol (hydrocortisone, compound F) (XIII), corticosterone (XIV) and aldosterone (XV). The first two of these substances



XIV



XV

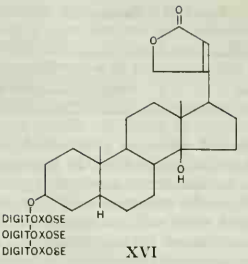
are known to be produced from cholesterol by way of progesterone. Cortisol exerts the greatest effect on carbohydrate metabolism and stress resistance (glucocorticoid activity), aldosterone has the most powerful effect on salt balance (mineralocorticoid activity) and corticosterone shows both properties to a lesser extent. Among other active steroids isolated from adrenal tissue are the 11-keto analogue of cortisol, cortisone (compound E, 17,21-dihydroxy pregn-4-ene-3,11,20-trione), which shows nearly as much glucocorticoid activity as cortisol, and 11-deoxycorticosterone (DOC, 21-hydroxypregn-4-ene-3,20-dione, 21-hydroxyprogesterone), which possesses about $\frac{1}{25}$ the salt-retaining activity of aldosterone.

The glucocorticoid activity of an adrenal hormone is evaluated in the adrenalectomized rat by its ability to cause deposition of glycogen in the liver, to cause involution of lymphatic tissue, to maintain the ability of the leg muscle to do work or to enable the animal to withstand exposure to cold. Mineralocorticoid activity is evaluated by prolongation of life in adrenalectomized animals, or by sodium retention following salt administration.

The earliest clinical use of adrenal hormones was for the treatment of the usually fatal condition of adrenal insufficiency known as Addison's disease. At first crude extracts of beef adrenal glands were used for this purpose, but pure compounds later became available by chemical synthesis. With the discovery in 1949 that administration of cortisone or cortisol could effect remarkable amelioration of rheumatoid arthritis and other inflammatory conditions, a new era was opened in both medicine and steroid chemistry. A demand arose for adrenal steroids in amounts far exceeding those in which steroid hormones ever had been produced. An intensive search ensued for practical methods for the large-scale synthesis of cortisone and cortisol, in particular for the difficult task of introducing the 11-oxygenated substituent characteristic of the glucocorticoids. The earliest procedure starting with deoxycholic acid involves chemical transfer of the hydroxyl group from C_{12} to C_{11}, and this process is still utilized. Since the discovery that certain strains of mold and fungus, when incubated with progesterone, can introduce a hydroxyl group directly in position 11, most current processes for adrenal hormone manufacture utilize this fermentation step starting with progesterone derivatives obtained from sapogenins.

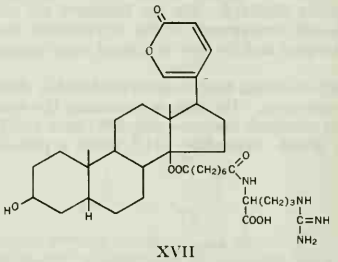
Certain synthetic modifications of the cortisol structure have been found to yield compounds with markedly increased glucocorticoid and anti-inflammatory properties that make them of considerable clinical and commercial importance. Molecular features that increase potency include a 1,2 double bond (prednisolone) a 6α methyl or fluorine group, a 9α-fluorine substituent and, especially, combinations of two or more of these features, in some instances with a 16-hydroxyl or methyl substituent present as well.

Cardiotonic Agents.—Two types of steroids found in nature exert a unique effect on the performance of the heart muscle in



XVI

higher animals. These are the cardiac glycosides, illustrated by digitoxin (XVI), which are found in plants and consist of a steroid genin or aglycone combined in glycosidic linkage through its 3-hydroxyl group with one or more molecules of a sugar, and the toad poisons, illustrated by bufotoxin (XVII), which occur in the skin and parotid gland secretions of toads and consist of a steroid genin conjugated through its 14-hydroxyl group with suberylarginine, an amino acid derivative.



XVII

The cardiac glycosides are found in the leaves, flowers, seeds, roots or bark of a large variety of plants including digitalis, *Strophanthus,* squill, oleander and ouabaio. Extracts of such plants have long been recognized as possessing toxic properties and have been employed as arrow poisons (inée), rodenticides (red squill), drugs for medieval ordeal trials and folk remedies (foxglove). When administered in sublethal doses, the cardiac glycosides exert a powerful effect on the heart muscle, causing a slowing of the rate but an increase in the force and extent of contraction. Thus, these substances are of great benefit for the treatment of conditions in which the heart lacks the power to circulate the blood properly (congestive heart failure, dropsy) or in which the rhythm of the heart beat has been disturbed (*e.g.,* auricular fibrillation). The toad poisons exert a similar effect on the heart, and dried and powdered toad skins have long been used medicinally in China and Japan. By the early 1960s, practical methods had not been developed for the synthesis of cardiac aglycones and glycosides, and material for clinical use is obtained by isolation from plant sources. Crystalline glycosides, as well as crude powers and extracts of digitalis leaves, find therapeutic use.

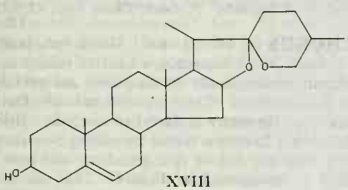
Hydrolysis of the cardiac glycosides by acid or by enzyme preparations removes the sugar groups and yields the steroid agly-

TABLE III.—*Typical Cardiac Aglycones*

Aglycone	Systematic name	From glycoside
Digitoxigenin	3β,14-dihydroxy-5β-card-20(22)enolide	digitoxin, digilanide-A
Periplogenin	3β,5,14-trihydroxy-5β-card-20(22)enolide	periplocin, perplocymarin
Sarmentogenin	3β,11α,14-trihydroxy-5β-card-20(22)enolide	sarmentocymarin
Digoxigenin	3β,12β,14-trihydroxy-5β-card-20(22)enolide	digoxin, digilanide-C
Gitoxigenin	3β,14,16β-trihydroxy-5β-card-20(22)enolide	gitoxin, digilanide-B
Strophanthadin	3β,5,14-trihydroxy-19-oxo-5β-card-20(22)enolide	cymarin, k-strophanthoside
Ouabagenin	1β,3β,5,11α,14,19-hexahydroxy-5β-card-20(22)enolide	ouabain
Scillarenin	3β,14-dihydroxybufa-4,20,22-trienolide	scillabiose
Scilliglaucosidin	3β,14-dihydroxy-19-oxobufa-4,20,22-trienolide	scilliglaucoside

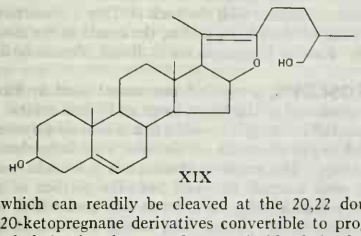
cones. The simple aglycones are much less active physiologically than the complete glycosides. Some of the commonest aglycones are listed in Table III. Characteristic features of this group of steroids are the unsaturated cyclic lactone side chain containing either four or five carbon atoms, 3β- and 14β-hydroxyl groups and the 14β ring junction that is unique among natural steroids.

Sapogenins.—The sapogenins are the steroid portions of a type of glycosidic compound found in the leaves, seeds and roots of a number of plants (*Dioscorea*, yucca, agave, sarsaparilla, digitalis) and called saponins because they form a soapy lather when mixed with water. Treatment of saponins with acid yields the sapogenins, which are characterized by a 3β-hydroxyl group and an eight-carbon side chain with two atoms of oxygen arranged in a bicyclic "spiroketal" structure (XVIII).



XVIII

On treatment with hot acetic anhydride, sapogenins are converted to the acetate esters of the so-called pseudo (ψ) sapogenins



XIX

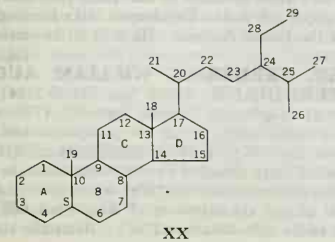
(XIX), which can readily be cleaved at the 20,22 double bond to form 20-ketopregnane derivatives convertible to progesterone. Because of their abundance and ease of side-chain degradation, the sapogenins are the chief raw materials for the manufacture of steroid hormones. The most important substance in this regard is diosgenin (XVIII), which is obtained from the roots of a species of *Dioscorea* that abounds in southern Mexico. The structures of typical sapogenins are given in Table IV.

TABLE IV.—*Typical Steroid Sapogenins*

Trivial name	Systematic name
Diosgenin	25a-spirost-5-en-3β-ol
Yamogenin	25b-spirost-5-en-3β-ol
Smilagenin	5β,25a-spirostan-3β-ol
Sarsasapogenin	5β,25b-spirostan-3β-ol
Hecogenin	3β-hydroxy-5α,25a-spirostan-12-one
Botogenin	3β-hydroxy-25a-spirost-5-en-12-one
Gitogenin	5α,25a-spirostane-2α,3β-diol
ψ-diosgenin	furosta-5,20(22)-diene-3β,27-diol
ψ-sarsasapogenin	5β-furost-20(22)ene-3β,27-diol

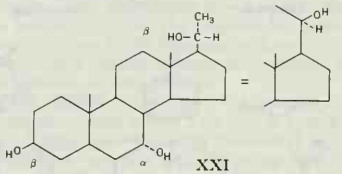
STEROID NOMENCLATURE

Numbering.—The four rings of the steroid nucleus are designated by letters and the individual carbon atoms by numbers (XX). Naturally occurring steroids possess an "angular" methyl group (18) attached at position 13, and most of them a similar methyl (19) at position 10; these are indicated in the formula by the short lines. Many steroids also possess a side chain containing two, five, eight, nine or ten carbon atoms attached to the nucleus at position 17 and numbered as shown.



XX

α and β Configuration.—Since the four valence bonds of a carbon atom possess a tetrahedral relationship, the single substituent (I) at each of the tertiary positions (ring junctions) of the steroid nucleus will project toward one side or the other of the approximate plane formed by the steroid ring system. At all other (secondary) positions, one of the two substituents will project toward one side of this plane and the other substituent toward the opposite side. When the steroid nucleus is represented in the conventional manner with the ring system lying in the plane of the paper, the configuration at C_{13} in all the known natural steroids is such that the 18-methyl group projects upward. Substituents on the nucleus projecting in the same direction as the C_{18} methyl group are said to possess a "β" orientation and are denoted in the formula by attaching that group with a solid line (XXI). Conversely, substituents projecting toward the opposite side of the molecule (downward) have an "α" orientation denoted by a broken or dotted line. Orientation at position 20 is defined as β if the substituent projects toward the left of C_{20} when the longest chain attached to this position is swung to the rear and as α if it projects to the right (XXI). In the sapogenins C_{25} is asymmetric, and two series of compounds exist (25a or "iso" and 25b or "neo") that differ in the configuration at that position.

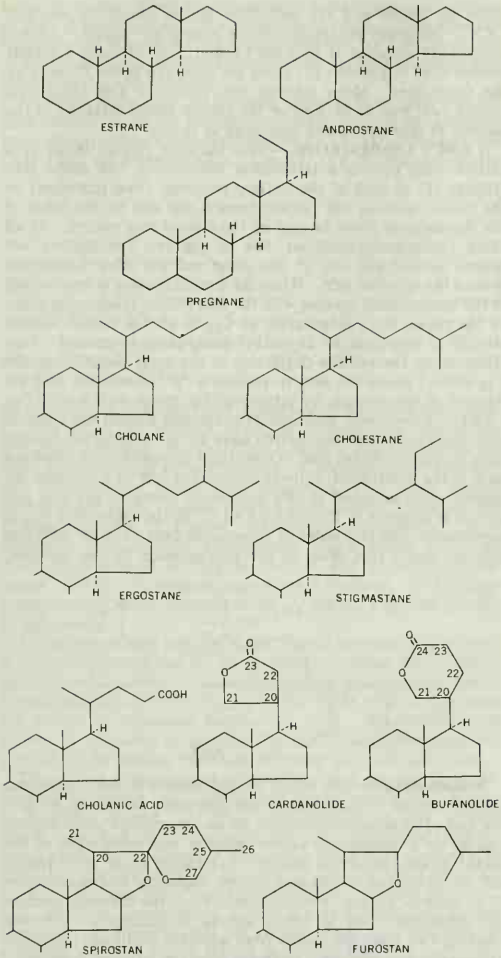


XXI

Nomenclature.—In trivial steroid nomenclature, compounds closely related in structure to the commonly occurring steroids are named as derivatives thereof by an appropriate prefix accompanied by a designation of the position and orientation of the added, missing or altered feature. In addition to the usual chemical notations for substituent groups replacing hydrogen atoms (*e.g.*, methyl-, chloro-, hydroxy-, oxo-, etc.), the following prefixes are commonly used in trivial steroid nomenclature: dehydro- (lacking two hydrogen atoms from adjacent positions); dihydro- (possessing two additional hydrogen atoms in adjacent positions); anhydro- (lacking the elements of water from adjacent positions); deoxy- (hydroxyl group replaced by a hydrogen atom); epi- (differing in configuration at a secondary position); iso- (differing in configuration at a tertiary position); nor- (lacking one carbon atom); and homo- (possessing one additional carbon atom).

Systematic nomenclature, though often cumbersome, affords an unambiguous name for any steroid structure. Here compounds are designated as derivatives of reference hydrocarbons whose names imply the structure and stereochemistry of the complete carbon skeleton. Since certain classes of steroids possess characteristic oxygen-containing groupings in the side chain, it is convenient also to use certain reference structures that include such groups.

Except for cardanolide and bufanolide, which imply the 14β configuration characteristic of the cardiotonic agents, these parent compounds possess the "backbone" configuration common to most natural steroids, namely, 8β, 9α, 10β, 13β, 14α and 17β. Thus, configuration at any of these positions is not specified in the name nor indicated in the formula unless it differs from that of the foregoing. Since compounds with both 5α or 5β configurations are found in nature, the configuration at position 5 must be



REFERENCE COMPOUNDS FOR SYSTEMATIC NOMENCLATURE (*see* TEXT)

indicated, giving rise to two series of steroids for each parent hydrocarbon; 5α-androstane, 5β-androstane, etc. Until 1956 systematic nomenclature employed the names androstane and cholestane only for 5α compounds and used pregnane and cholane for 5β compounds. 5β-Androstane was known as etiocholane or testane, 5β-cholestane as coprostane, 5α-pregnane as allopregnane and 5α-cholane as allocholane. Names based on this former system are still encountered.

The presence of unsaturation is indicated by changing the syllable "ane" to "ene" for one double bond, "adiene" for two double bonds, etc., and inserting before this syllable the number of the lowest numbered carbon atom attached to the double bond. If the other carbon atom is not the next higher numbered, its number is also given enclosed in parentheses. If there is no other numerical prefix to the parent name, it is common in the American but not the British literature to place the numbers locating the unsaturation before the name (*e.g.*, 4-androstene for androst-4-ene). Formerly, the symbol Δ was placed before the number to indicate unsaturation, but this practice, though still encountered, is unnecessary and outmoded. Substituent groups with their position and orientation are indicated by the appropriate prefix. Alternatively, hydroxyl, carbonyl and carboxyl substituents may be indicated by the suffixes -ol, -one and -oic acid, respectively. The final "e" of the parent name is dropped when such a suffix begins with a vowel. Although often disregarded in the older steroid literature, it is not proper to employ more than one suffix in a single name (*e.g.*, 3β-hydroxy-5α-androstan-17-one, not 5α-androstan-3β-ol-17-one).

BIBLIOGRAPHY.—L. F. Fieser and M. Fieser, *Steroids* (1959); G. Pincus and K. V. Thimann (eds.), *The Hormones*, vol. i (1948), vol. ii (1950), vol. iii (1955); W. Klyne, *The Chemistry of the Steroids* (1957); C. W. Shoppee, *Chemistry of the Steroids* (1958); F. Radt (ed.), *Elsevier's Encyclopaedia of Organic Chemistry*, series 3, vol. 14, supp. 3, *Steroids*, pt. 1, p. 1,347 ff. (1954); pt. 2, p. 1,869 ff. (1956), pt. 3, p. 2,215 ff. (1959); R. H. Williams (ed.), *Textbook of Endocrinology*, 2nd ed. (1955); D. Kritchevsky, *Cholesterol* (1958); E. Heftmann and E. Mosettig, *Biochemistry of Steroids* (1960); "Definitive Rules for Nomenclature of Steroids," *J. Amer. Chem. Soc.*, 82:5,577 (1960).
(E. V. JE.)

STESICHORUS (*c.* 640–555 B.C.), Greek lyric poet, a native of Himera in Sicily, or of Mataurus, a Locrian colony in the south of Italy. About 30 fragments of his poems are extant, none of them longer than six lines. They are written in the Doric dialect, with epic licences; the metre is dactylico-trochaic. Brief as they are, they show what Longinus meant by calling Stesichorus "most like Homer"; they are full of epic grandeur, and have a stately sublimity that is reminiscent of Pindar. This use of lyric for heroic subjects is what Quintilian means by saying that he "sustained the burden of epic poetry with the lyre." Several of his poems sang of the adventures of Heracles; one dealt with the siege of Thebes, another with the sack of Troy. Stesichorus completed the form of the ode by adding the epode to the strophe and antistrophe. For the fragments *see* T. Bergk, *Poetae lyrici graeci*, iii.

STETHOSCOPE, a medical instrument used in listening to the sounds produced in the heart, lungs and other organs. This is known as auscultation (*q.v.*), which is a means of physical examination used in the diagnosis of diseases, especially those of the heart and lungs. The single stethoscope is a straight wooden or metal tube with a small flattened bell, the surface of which is usually covered with ivory or bone at one end, which is placed against the body of the patient, and a large flat disk at the other end, for application to the ear of the observer. In the "binaural" stethoscope, which has the advantage of flexibility, the tube is divided above the bell into two flexible tubes which lead to both ears. (F. L. A.)

STETTIN: *see* SZCZECIN.

STETTINIUS, EDWARD REILLEY, JR. (1900–1949), U.S. industrialist and statesman generally credited as one of the leading figures in the formation of the United Nations, was born on Oct. 22, 1900, at Chicago, Ill. Stettinius attended the University of Virginia, Charlottesville, but did not take a degree. In 1924 he was employed by a subsidiary of General Motors corporation and by 1931 he had risen to an executive position in the corporation. He joined the United States Steel corporation and in 1938, at the age of 37, became chairman of the board of directors of that firm. He was named chairman of Pres. Franklin D. Roosevelt's War Resources board (1939), Lend-Lease administrator (1941–43) and undersecretary of state (1943). Stettinius served as U.S. secretary of state in 1944 but he resigned the following year to become the first U.S. delegate to the general assembly of the United Nations. Earlier, he had been one of President Roosevelt's chief advisers at the Yalta conference, and he played a principal role in the Dumbarton Oaks meetings prior to formation of the United Nations. He died at Greenwich, Conn., Oct. 31, 1949. (H. J. Sc.)

STEUBEN, FREDERICK WILLIAM AUGUSTUS HENRY FERDINAND, BARON VON (1730–1794), German soldier, born at Magdeburg, Prussia, Sept. 17, 1730, was a descendant of a noble family, which for generations had produced soldiers. From his 14th year, Steuben led a soldier's life. Reared in the rigorous military school of Frederick the Great, he took part in many battles of the Seven Years' War, where he so distinguished himself as to attract the attention of King Frederick, who appointed him as his aide-de-camp (1762). After the close of the

Seven Years' War, he resigned from the army and became court chamberlain for the Prince of Hohenzollern-Hechingen, and, after ten years applied unsuccessfully for a similar position at the court of the margrave of Baden. In 1777, his old friend the count of St. Germain, then the French minister of war, persuaded him to go to the assistance of the American colonists.

Steuben arrived at Portsmouth, N.H., on Dec. 1, 1777, and offered his services to congress as a volunteer. In March 1778 he began drilling the inexperienced soldiers at Valley Forge; and by May he was made inspector-general or drill master. He trained the soldiers admirably, adapting Prussian military ideas to the needs of his pupils. Results of his work were shown in the next campaign, particularly at Monmouth, where he rallied the disordered, retreating troops of Gen. Charles Lee. His *Regulations for the Order and Discipline of the Troops of the United States* (1779) was the official manual until 1812. He served on the court-martial which tried Maj. John André in 1780, and after Gen. Horatio Gates's defeat at Camden was placed in command of the district of Virginia, with instructions "to collect, organize, discipline and expedite the recruits for the Southern Army."

In April 1781 he was superseded in command of Virginia by La Fayette, and later took part in the siege of Yorktown. Retiring from the service after the war, he passed the last years of his life at Steubenville, N.Y., where he died on Nov. 28, 1794.

(W. M. We.; X.)

STEUBENVILLE, a city of eastern Ohio, U.S., on the Ohio river is located 26 mi. N. of Wheeling, W.Va., the seat of Jefferson county. It is one of the state's oldest cities. The site was settled temporarily as early as 1765 by Jacob Walker and was occupied again with the building in 1786 of Ft. Steuben, named for Baron von Steuben, soldier and patriot of the American Revolution. The fort was abandoned with the outbreak of the Indian wars in 1790. In 1797 a permanent settlement was laid out by two Pennsylvania land speculators, Bezaleel Wells and James Ross, who named it Steubenville. It was incorporated in 1805.

Adjacent to one of the nation's principal waterways in an area rich in coal and clay, the city early became an important manufacturing and commercial centre. Though the textile industry was its most important in the period after the War of 1812, by mid-20th century Steubenville's economy centred upon the production of steel, chemicals, dinner ware, brick and tile, heaters, stoves, electrical equipment, paperboard and ferroalloy metals. In the 1960 census Steubenville was named a central city of the Steubenville-Weirton standard metropolitan statistical area, comprising Jefferson county in Ohio and Brooke and Hancock counties in West Virginia, with a population of 167,756. For comparative population figures for the two cities *see* tables in Ohio: *Population* and West Virginia: *Population. See* also Weirton.

(P. R. S.)

STEVENAGE, a market town and urban district of Hertfordshire, Eng., on the Great North road 31 mi. N. of London. Pop. (1961) 42,968. Area 7.2 sq.mi. Silk stockings are manufactured. Beside the old town, with its coaching inns, Norman-Early English church and 16th-century grammar school, is Stevenage New Town, planned (1946) to cover nearly 9½ sq.mi.

STEVENS, ALFRED (1828–1906), Belgian painter, was born on May 11, 1828, in Brussels. After 1844 Stevens mainly worked, and from 1849 exhibited, in Paris, which also provided material for his pictures, usually conversation pieces and small, informal, full-length portraits of fashionable women set in second empire *hôtels*. In the elegant genre tradition of Gabriel Metsu, J. E. Liotard and L. L. Boilly, they reflect neither Ingres' academism, Eugène Delacroix's romanticism nor Gustave Courbet's new realism. Typical are "L'Invitation" (late 1860s), "Les Derniers Jours de veuvage" (1870s; both in Brussels private collections) and "Le Sphynx parisien" (*c.* 1880). Stevens died in Paris on Aug. 24, 1906.

(D. C. T. T.; X.)

STEVENS, ALFRED GEORGE (1817–1875), English designer, painter and sculptor, notable for the Michelangelesque vigour of his work, was born on Dec. 30, 1817, at Blandford, Dorset. He traveled in Italy, 1833–42, spending a year under Bertel Thorvaldsen (*q.v.*) in Rome. In 1845–47 he taught at the new School of Design, Somerset house, London, and in 1850–52 designed cast-metal hearth furniture, stoves, etc., for Hoole and company, Sheffield, some of which attracted attention at the 1851 Great exhibition. In the 1857 competition for the Wellington monument in St. Paul's cathedral, London, he was placed sixth, though his design was the only one adapted to the nave arch under which the monument was to stand.

In 1862, however, Stevens was given the commission and, despite vicissitudes, the complex marble structure with its allegories and bronze equestrian statue was nearly complete at his death (in Hampstead, London, May 1, 1875). For Dorchester house, London, Stevens carried out decorations in 1858–62 (caryatid mantlepiece and many drawings for Dorchester house and other projects now in the Tate gallery, London), and these best display the freshness he brought to his forms. Stevens also painted portraits (*e.g.*, "Mrs. Leonard Collman," 1854, National gallery, London), but is probably best remembered for his red chalk figure drawings.

Bibliography.—Sir Walter Armstrong, *Alfred Stevens* (1881); H. Stannus, *Alfred Stevens* (1891); K. R. Towndrow, *Alfred Stevens, Architectural Sculptor, Painter and Designer* (1939) and (comp.) *The Works of Alfred Stevens at the Tate Gallery* (1952). (D. C. T. T.)

STEVENS, HENRY (1819–1886), U.S. bibliographer and bookseller, was born at Barnet, Vt., on Aug. 24, 1819, and was graduated from Yale University in 1843. Peter Force, the annalist of the American Revolution, employed him to collect books, pamphlets, and manuscripts for Force's *American Archives*. Stevens' commissions from Force and other U.S. collectors subsequently took him to London, where he established an important rare-book firm, supplying Americana to the Smithsonian Institution, the Library of Congress, and such private buyers as John Carter Brown and James Lenox (*q.v.*); he also supplied Sir Anthony Panizzi (*q.v.*), librarian of the British Museum, with many U.S. books.

Stevens published various bibliographic studies. He died at Hampstead, London, on Feb. 28, 1886.

STEVENS, JOHN (1749–1838), U.S. inventor, pioneered in U.S. transportation. Born in New York City in 1749, he graduated from King's College (now Columbia University) in 1768, subsequently studied law, and was admitted to the New York bar in 1771. From 1777 to 1782 he served as treasurer of New Jersey.

The problem of safe and speedy transportation across the Hudson River drew his interest early to the possibilities of steam navigation. In 1788 he built the first multitubular boiler on record for his marine engines. In a desire to protect this invention, he petitioned Congress for a patent law, which he outlined, and which was passed as the Patent Law of 1790, the foundation of the present system. He was the first to apply the principle of screw propulsion to navigation, building in 1802 a steamboat with two underwater propellers of screw type which crossed the Hudson several times and failed only because the boilers could not generate enough power. He turned then toward the development of high-pressure steam engines and boilers. A few days after Fulton's successful trial of the "Clermont," Stevens launched the "Phoenix," the engine of which was built by Stevens himself whereas that of the "Clermont" was imported from England. Fulton having a monopoly grant of navigation rights on the Hudson, Stevens sent the "Phoenix" by sea to Philadelphia, Pa., the first voyage of a steamship on ocean waters.

Stevens was also interested in railways. On Feb. 6, 1815, he obtained the first railway charter granted in America for a railway between the Delaware and Raritan rivers, but it was not until 1830 that he formed the Camden and Amboy Railway Company and not until 1832 that he completed the first railway across New Jersey. He died at Hoboken, N.J., March 6, 1838.

STEVENS, ROBERT LIVINGSTON (1787–1856), U.S. mechanical engineer, was born at New York City on Oct. 18, 1787, the son of John Stevens (*q.v.*). Whereas the father spent most of his thought on the improvement of motive power and machinery of steamboats, the son made many advances in construction, strengthening hulls, reducing weight, distributing strain, and improving the lines to gain speed. He designed the false bow, which, when introduced on the "New Philadelphia," enabled that steamer

to break records and be the first to go from Albany to New York during daylight. In 1822 he built the ferryboat "Hoboken" on modern lines and introduced the piled slip which directed the boat's entrance to the landing place.

In 1830 he was made first president of the Camden and Amboy Railway. On a voyage to England he designed the "T" rail with a broad base, long known as the Stevens or American rail and now universally used. He contracted in England for 500 tons of such rails, the first made, and also placed an order with the Stephensons for a locomotive made according to his specifications. This was the famous "John Bull," which he brought back to the U.S. and which is now in the Smithsonian Institution.

He made many improvements on railway locomotives, introducing the pilot truck in 1832, the bogie truck shortly afterward, and, finally, the use of eight instead of four or six wheels in order to promote adhesion to the rails and drawing power. When stone blocks could not be furnished fast enough for his roadbed he resorted to logs laid crosswise with broken stone and gravel between, a bed more serviceable and comfortable than any known previously. He devised spikes "six inches long with heads" to hold the rails in place. In 1844 he designed the "Maria," for 20 years the swiftest yacht afloat. Later he designed the "America" for his brother John Cox Stevens (*see* YACHTING). He died at Hoboken, N.J., April 20, 1856.

STEVENS, THADDEUS (1792–1868), U.S. political leader, was born in Danville, Vt., April 4, 1792. He graduated from Dartmouth college in 1814, moved to York, Pa., was admitted to the bar (in Maryland) and for 15 years practised at Gettysburg, Pa. He frequently appeared in behalf of fugitive slaves before the Pennsylvania courts. He was a Whig representative in congress in 1849–53 and was leader of the radical Whigs and Free-soilers, strongly opposing the Compromise measures of 1850 and being especially bitter in his denunciations of the Fugitive Slave law. In 1855 he took a prominent part in organizing the Republican party in Pennsylvania. He became chairman of the ways and means committee on July 4, 1861, and until his death was, as James G. Blaine said, "the natural leader who assumed his place by common consent," During the Civil War he was instrumental in having necessary revenue measures passed. He was not in harmony with Lincoln, who was far more conservative. He introduced from the joint committee what became the 14th amendment, and also the Reconstruction act of Feb. 6, 1867. He advocated the Freedmen's bureau bills and the Tenure of Office act, and went beyond congress in favouring the confiscation of property in Confederate states. He led congress in the struggle with Pres. Andrew Johnson, reported the impeachment resolution and was chairman of the committee appointed to draft the articles of impeachment. He died at Washington, D.C., Aug. 11, 1868.

STEVENS, WALLACE (1879–1955), U.S. poet whose work is characterized by subtle intellectuality and originality beneath a dandified and elegant manner, was born on Oct. 2, 1879, in Reading, Pa. He attended Harvard for three years, worked briefly for the *New York Herald Tribune,* then won a degree (1904) at the New York Law school. After practising law in New York city, he joined an insurance firm in Hartford, Conn., in 1916. In 1934 he became a vice-president.

In 1914 Harriet Monroe published several of his poems in *Poetry.* Thereafter he was a frequent contributor to the "little magazines." *Harmonium* (1923), his first book, sold poorly but received some favourable critical notices; it was reissued in 1931 and in 1947. Stevens thereafter published several volumes, including his *Collected Poems* (1954). A volume of critical essays, *The Necessary Angel* (1951), was devoted to the chief concern of his poetry—the relationship between "imagination" and "reality."

Late in life Stevens won major literary awards—the Bollingen award (1949), the National Book award (1950), and the Pulitzer prize (1955). He died in Hartford, Conn., on Aug. 2, 1955.

(W. V. O'C.)

STEVENSON, ADLAI EWING (1835–1914), vice-president of the United States, 1893–97, was born in Christian county, Ky., on Oct. 23, 1835. He studied at Centre college, read law and commenced practice at Metamora, Ill. He was a member of the 44th and 46th congresses and served as first assistant postmaster general in Grover Cleveland's first administration. He was given the Democratic nomination for vice-president in 1892 so that the west would be represented on the ticket which Cleveland, an easterner, headed. In 1897 he was chairman of a commission sent to Europe to work for international bimetallism. He was again nominated for vice-president on the ticket with W. J. Bryan in 1900, and in 1908 he was unsuccessful as the Democratic candidate for governor of Illinois. He wrote a pleasant book of reminiscences, *Something of Men I Have Known.*

STEVENSON, ADLAI EWING (1900–1965), U.S. lawyer, Democratic presidential candidate in 1952 and 1956, and U.S. representative at the United Nations (1961–65), was born in Los Angeles, Calif., on Feb. 5, 1900. He was the son of Lewis Green Stevenson, then an executive of Hearst newspapers in California, and the grandson of Adlai Ewing Stevenson (*q.v.*), vice-president of the U.S. during Grover Cleveland's second term (1893–97).

After the family returned to its home in Bloomington, Ill., in 1906, Stevenson attended local public schools, and for college preparatory study was sent to the Choate School in Wallingford, Conn. During World War I he served for about a year as an apprentice seaman in the U.S. Naval Reserve. He then entered Princeton, where he graduated in 1922. After studying at Harvard Law School for the next two years, he returned to Illinois to become assistant managing editor of the *Bloomington Daily Pantagraph,* a newspaper that had been owned by his family for several generations. Meanwhile, he commuted to Chicago to continue his study of law at Northwestern University Law School, graduating in 1926.

In the same year Stevenson was admitted to the Illinois bar and began the practice of law in Chicago, but his legal work was frequently interrupted by periods of public service. During the years before World War II he served two terms as president of the Chicago Council on Foreign Relations, was chairman of the Civil Rights Committee of the Chicago Bar Association, and chairman of the Chicago chapter of the Committee to Defend America by Aiding the Allies. From July 1941 to April 1944 he was special assistant to the secretary of the navy and traveled extensively throughout the world on official business. Meanwhile, in 1943, Stevenson headed a Foreign Economic Administration mission to Italy to develop a U.S. program for relief and reconstruction after the invasion of Italy by the Allied armies. He became an assistant to the secretary of state in 1945 and was adviser to the U.S. delegation at the United Nations Conference on International Organization at San Francisco, Calif. He was named senior adviser to the U.S. delegation at the first meeting of the UN General Assembly in London in 1946, and was a U.S. delegate at the General Assembly meetings in New York, 1946–47.

In 1948 Stevenson was elected governor of Illinois by the largest majority any candidate had ever received in the state's history. His regime as governor was distinguished by far-reaching reforms, both legal and administrative; it included establishment of a merit system for state police, improved care and treatment of patients in state mental hospitals, greater state aid for public schools, and a revitalized Civil Service.

At the Democratic national convention held in Chicago in 1952, Stevenson, in spite of his refusal to seek the nomination, was chosen as the party's candidate for president. He waged a vigorous campaign and his speeches were widely applauded for their thoughtful eloquence and wit, but in the November election he was defeated by the Republican nominee, Dwight D. Eisenhower. In 1956 he was again his party's nominee for president and was again defeated in the election.

After this second defeat Stevenson returned to the practice of law and also traveled extensively throughout Africa and Europe, partly on business. In 1960, in company with William Benton with whom he was associated as a member of the Board of Directors of Encyclopaedia Britannica, Inc., he traveled throughout Central and South America.

There was some support in the Democratic Party for nominating Stevenson for president in 1960 but the national convention held that year nominated Sen. John F. Kennedy of Massachusetts, and Stevenson supported him during the campaign. Shortly after his election, President-elect Kennedy announced that he would appoint Stevenson to be U.S. representative to the United Nations, with cabinet rank and the title of ambassador. He served in this posi-

tion until his sudden death in London on July 14, 1965.

In 1928 Stevenson married Ellen Borden, who divorced him in 1949. The couple had three sons, Adlai Ewing, Borden, and John Fell.

Among Stevenson's published works were *Major Campaign Speeches of Adlai E. Stevenson, 1952* (1953), *Call to Greatness* (1954), *What I Think* (1956), *The New America* (1957), *Friends and Enemies* (1959); *Putting First Things First* (1960), and *Looking Outward: Years of Crisis at the United Nations* (1963).

(W. McC. B.)

STEVENSON, ROBERT LOUIS BALFOUR (1850–1894), Scottish novelist, essayist and poet, was born at Edinburgh on Nov. 13, 1850, only son of Thomas Stevenson, prosperous civil engineer, and his wife Margaret Isabella Balfour. His poor health made regular schooling difficult, but he did attend Edinburgh academy and other schools before entering Edinburgh university at the age of 16. There he was expected to prepare himself for the family profession of lighthouse engineering. But Stevenson had no desire to be an engineer, and he eventually agreed with his father, as a compromise, to prepare instead for the Scottish bar.

He had early shown a desire to write, and once in his teens he set himself deliberately to learn the writer's craft by imitating a great variety of models in prose and verse. His youthful enthusiasm for the Covenanters led to his writing *The Pentland Rising*, his first printed work. During his years at the university he rebelled violently against his parents' religion and set himself up as a liberal Bohemian who abhorred the cruelties and hypocrisies of bourgeois respectability.

In 1873, in the midst of painful differences with his father, he visited a married cousin at Cockfield, Suffolk, where he met Sidney Colvin, who became a lifelong friend, and Mrs. Fanny Sitwell (the future Lady Colvin). Mrs. Sitwell, an older woman of charm and talent, drew the young man out and won his confidence. Soon Stevenson was deeply in love, and on his return to Edinburgh wrote her a series of letters in which he played the part first of lover, then of worshiper, then of son. One of the several names by which Stevenson addressed her in these letters was "Claire," a fact which many years after his death was to give rise to the erroneous notion that Stevenson had had an affair with a humbly born Edinburgh girl of that name. The passion modulated eventually into a lasting friendship.

Later in 1873 Stevenson suffered severe respiratory illness and was sent to the French Riviera, where Colvin later joined him. He returned home the following spring. In July 1875 he was called to the Scottish bar, but he never practised. Stevenson was frequently abroad, most often in France. Two of these journeys produced *An Inland Voyage* (1878) and *Travels With a Donkey in the Cevennes* (1879). His career as a writer developed slowly. His essay "Roads" appeared in the *Portfolio* in 1873, and in 1874 "Ordered South" appeared in *Macmillan's*, a review of Lord Lytton's *Fables in Song* appeared in the *Fortnightly*, and his first contribution (on Victor Hugo) appeared in the *Cornhill*, then edited by Leslie Stephen. It was these early essays, carefully wrought, quizzically meditative in tone, unusual in sensibility, that first drew attention to Stevenson as a writer.

Stephen brought Stevenson into contact with Edmund Gosse, who became a good friend, and later, when in Edinburgh, Stephen introduced him to W. E. Henley. Henley and Stevenson became warm friends, and remained so until 1888, when a letter from Henley to Stevenson containing a deliberately implied accusation of dishonesty against the latter's wife precipitated a quarrel which Henley, jealous and embittered, perpetuated after his friend's death by a venomous review of Graham Balfour's official biography.

While staying at Fontainebleau in 1876 Stevenson met Fanny Vandegrift Osbourne, an American lady separated from her husband, and the two fell in love. Stevenson's parents' horror at their son's involvement with a married woman subsided somewhat when Mrs. Osbourne returned to California in 1878, but it revived with greater force when Stevenson decided to join her in Aug. 1879. Stevenson arrived in California ill and penniless (the record of his arduous journey appeared later in *The Amateur Emigrant*, 1895, and *Across the Plains*, 1892). His adventures, which included coming very near death and eking out a precarious living in Monterey and San Francisco, culminated in marriage to Fanny Vandegrift (divorced from her first husband) early in 1880. About the same time a telegram from his relenting father offered much-needed financial support, and after a honeymoon by an abandoned silver mine (recorded in *The Silverado Squatters*, 1883) the couple sailed for Scotland to achieve reconciliation with the Thomas Stevensons.

Soon after his return Stevenson went, on medical advice (he had tuberculosis), to Davos in Switzerland, accompanied by his wife and his stepson, Lloyd Osbourne. They left there in April 1881 and spent the summer in Pitlochry and then in Braemar. There, in spite of bouts of illness, Stevenson embarked on *Treasure Island* (begun as a game with Lloyd), which started as a serial in *Young Folks*, under the title *The Sea-Cook*, in Oct. 1881. Stevenson finished the story in Davos, to which he returned in the autumn, and then started on *Prince Otto* (1885), a more complex but a less successful work. *Treasure Island* is an adventure presented with consummate skill, with atmosphere, character and action superbly geared to each other. The book is both a "rattling good yarn" for boys and a wry comment on the ambiguity of human motives.

In 1881 Stevenson published *Virginibus Puerisque*, his first collection of essays, most of which had appeared in the *Cornhill*. The winter of 1881 he spent at a chalet in Davos. In April 1882 he left Davos, but a stay in the Scottish Highlands, while it resulted in two of his finest short stories, "Thrawn Janet" and "The Merry Men," produced lung hemorrhages and in September he went to the south of France. There the Stevensons finally settled at a house in Hyères where, in spite of intermittent illness, Stevenson was happy and worked well. He revised *Prince Otto*, worked on *A Child's Garden of Verses* (first called *Penny Whistles*), and began *The Black Arrow*, a piece of "tushery" for *Young Folks*. In May 1885 he was again critically ill. On his recovery he went to England and eventually settled in Bournemouth. The Stevensons lived at Bournemouth from Jan. 1886 until July 1887, during which period his frequent bouts of dangerous illness proved conclusively that the British climate, even in the south of England, was not for him. The Bournemouth years were fruitful, however. There he got to know and love Henry James. There he revised *A Child's Garden* (first published in 1885) and wrote "Markheim." *Kidnapped* (1886) and *The Strange Case of Dr. Jekyll and Mr. Hyde* (1886). The poems in *A Child's Garden* represent with extraordinary fidelity an adult's recapturing of the emotions and sensations of childhood: there is nothing quite like them in English literature. In *Kidnapped*, the fruit of his researches into 18th-century Scottish history and of his feeling for Scottish landscape, history, character and local atmosphere mutually illuminate one another. But it was *Dr. Jekyll*—both moral allegory and thriller —that established his reputation with the ordinary reader.

In Aug. 1887 Stevenson set out for America with his wife, mother and stepson, still in search of health. On arrival at New York, he found himself famous, with editors and publishers offering lucrative contracts. He stayed for a while in the Adirondacks, where he wrote essays for *Scribner's* and began *The Master of Ballantrae* (1889). This novel, another exploration of moral ambiguities, contains some of Stevenson's most impressive writing, but it falls away in the contrived conclusion.

In June 1888 Stevenson, with his wife, his mother and his stepson, sailed from San Francisco in the schooner yacht "Casco," which he had chartered, on what was intended to be an excursion for health and pleasure; but in fact he was to spend the rest of his life in the South Seas. They went first to the Marquesas, then to Fakarava atoll, then to Tahiti, then to Honolulu where they paid off the "Casco" and stayed nearly six months, leaving in June 1889 on the trading schooner "Equator" for the Gilberts and then to Samoa, where he spent six weeks.

During his months of wandering around the South Sea islands Stevenson made intensive efforts to understand the local scene and the inhabitants. As a result, his writings on the South Seas (*In the South Seas*, 1896; *A Footnote to History*, 1892) are admirably pungent and perceptive. He was writing first-rate jour-

nalism, deepened by the awareness of landscape and atmosphere such as that so notably rendered in his description of the first landfall at Nuku Hiva, in the Marquesas. Once, in his violent *Father Damien* (1890), passionately defending the memory of the priest who had given his life in the service of the lepers and as passionately attacking the Rev. Dr. Hyde, his detractor, Stevenson wrote in a wholly personal tone about South Seas affairs.

In Oct. 1890 he returned to Samoa from a voyage to Sydney, and established himself at Vailima, his house in Samoa, with patriarchal status, with his wife, mother, stepson and stepdaughter. The climate suited him; he led an industrious and active life; and when he died suddenly on Dec. 3, 1894, it was of a cerebral hemorrhage, not of the long-feared tuberculosis. His work during these years was moving toward a new maturity. While *Catriona* (U.S. title, *David Balfour;* 1893) marks no advance in technique or imaginative scope on *Kidnapped,* to which it is a sequel, *The Ebb-Tide* (1894), a grim and powerful tale written in a style of studied neutrality (it was a complete reworking of a first draft by Lloyd Osbourne) shows that Stevenson had reached an important transition in his literary career. The next phase was demonstrated triumphantly in *Weir of Hermiston* (1896), the unfinished masterpiece on which he was working on the day of his death. *The Beach of Falesa* (first published 1892; included in *Island Night's Entertainments,* 1893), a story with a finely wrought tragic texture, as well as the first part of *The Master of Ballantrae,* pointed in this direction, but neither approaches *Weir.* Drawing on the scenery of the Lowland hills and of Edinburgh for setting, using the character of Lord Braxfield as a central figure to set off the sensitive hero, his son; bringing in those plangent overtones of tragic destiny to be found in the Scottish border ballads—Stevenson achieved in this work a remarkable richness of tragic texture in a style stripped of all superfluous fat. The dialogue contains some of the best Scots prose in modern Scottish literature. Fragment though it is, *Weir of Hermiston* stands as a great work and Stevenson's masterpiece.

Stevenson was an indefatigable letter writer and his letters (edited by Sidney Colvin in 1899) provide a lively and enchanting picture of the man and his life. But Colvin omitted many of the most interesting letters, and compressed and dovetailed others, with the result that many important facts about Stevenson's emotional life remained unknown until the true text of all the letters was available. Colvin presented Stevenson's letters to Mrs. Sitwell to the Advocates' library (now the National Library of Scotland) with the proviso that they were not to be used until 1949; the revealing and often fascinating letters to Charles Baxter, given by Baxter to the Savile club, were much later acquired by Edwin Beinecke who deposited them (along with much other Stevenson material) in the Yale university library. Stevenson's biography suffered from his early canonizing; later writers built up a counter-picture, which eventually emerged as that of an immoral swaggerer restrained into reluctant respectability by a jealous wife. Access to the crucial letters yielded a picture of a Stevenson who is neither the "seraph in chocolate" against whom Henley protested, nor a low-living rake, nor an optimistic escapist, nor the happy invalid, but a sensitive and intelligent writer who had no illusions about life and wryly made the best of a world to which he did not profess to have the key.

His literary reputation has also fluctuated. The reaction against him set in soon after his death: he was considered a mannered and imitative essayist, or only a writer of children's books. But eventually the pendulum began to swing the other way, and by the 1950s his reputation was established among the more discerning as a writer of originality and power; whose essays at their best are cogent and perceptive renderings of aspects of the human situation; whose novels are either brilliant adventure stories with overtones of moral subtlety or original and impressive presentations of human action in terms of history and topography as well as psychology; whose short stories produce some new and effective permutations in the relation between romance and irony or manage to combine horror and suspense with moral diagnosis; whose poems, though not showing the highest poetic genius, are often skillful, occasionally (in his use of Scots, for example) interesting and original, sometimes (in *A Child's Garden*) valuable for their exhibition of a special kind of memory or sensibility.

Stevenson's other works include: *New Arabian Nights* (1882); *Familiar Studies of Man and Books* (1882); *Memories and Portraits* (1887); *Underwoods* (poems) (1887); *St. Ives,* completed by Sir A. Quiller Couch (1897). Stevenson's *Collected Poems* were edited by Janet Adam Smith (1950). The letters were edited by Sidney Colvin in 1899; *see* also *Vailima Letters,* to Sidney Colvin (1895) and, of great interest and importance, *R. L. S.: Stevenson's Letters to Charles Baxter,* ed. by De Lancy Fergusson and Marshall Waingrow (1956).

BIBLIOGRAPHY.—Bibliographies of the works of Stevenson were published by J. S. Slater (1894) and W. F. Prideaux (rev. by J. F. Livingston, 1917). *See* also *The Stevenson Collection of E. J. Beinecke, a Catalogue,* comp. by George L. Mackay (1950–52). Among the biographies and critical studies are Graham Balfour, *The Life of Robert Louis Stevenson* (1901); Rosaline Masson, *I Can Remember Robert Louis Stevenson* (1922); Lloyd Osbourne, *An Intimate Portrait of R. L. S.* (1924); G. S. Hellman, *The True Stevenson* (1925); G. K. Chesterton, *Robert Louis Stevenson* (1928); J. A. Smith, *R. L. Stevenson* (1927), *Henry James and Robert Louis Stevenson* (1948); A. B. Fisher, *No More a Stranger* (1946); David Daiches, *Robert Louis Stevenson* (1947); A. R. Issler, *Happier for His Presence* (1949); J. C. Furnas, *Voyage to Windward* (1951); Fanny Osbourne Stevenson, *Our Samoan Adventure,* with a diary by Mrs. Stevenson (1955).
(D. Ds.)

STEVENSON, WILLIAM (d. 1575), possibly the author of *Gammer Gurton's Needle,* memorable as one of the earliest English vernacular comedies extant, was born at Hunwick, Durham; entered Christ's college, Cambridge, 1546; was ordained deacon 1552; and took his M.A. in 1553. From 1551 to 1554 he held a fellowship and was concerned in the presentation of college plays. He received his B.D. in 1560 and was probably installed prebendary of Durham on Jan. 28, 1560/61.

Gammer Gurton's Needle is described in the earliest printed text (1575) as having been played at Christ's college "not long ago," and "made by Mr. S. Mr. of Art." Evidence, however, points to a date before 1563 for composition and first performance. In 1553–54, William Stevenson could have been called "Mr. S., Master of Arts," and might have been the author of one of the plays he presented. Nevertheless, contemporary report attributed *Gammer Gurton's Needle* to John Bridges, a Pembroke man (M.A. 1560); and, on the assumption that—being dean of Salisbury when his opponents in the Marprelate controversy of 1588 twitted him with authorship of this vigorous but coarse play—he would have denied it if he could, his silence has been taken for acknowledgement. The last payment to William Stevenson as fellow was recorded 1554; the Stevenson who appears on the college books 1559–60 was apparently another man. These dates seem to forbid the reconciliation of the claims of "Mr. S." and the reputed author by supposing collaboration.

Stevenson died in 1575.

BIBLIOGRAPHY.—On the authorship of *Gammer Gurton's Needle,* see Henry Bradley in his edition of the play for C. M. Gayley's *Representative English Comedies* (1903); John Peile, *Biographical Register of Christ's College,* 2 vol. (1910–13); F. S. Boas, *University Drama in the Tudor Age* (1914); H. F. B. Brett-Smith in his edition of the play, Percy reprints (1920). (M. M. Ls.)

STEVENSTON, a small burgh (1952) of Ayrshire, Scot., on the Firth of Clyde, 27 mi. S.W. of Glasgow by road. Pop. (1951) 9,392. In 1678 coal mining began to succeed agriculture but the mines have closed and the industrial life now centres on the explosives and chemical factory of Imperial Chemical Industries at Ardeer. Holiday makers bathe on the sandy shore.

STEVIN (STEVINUS), **SIMON** (1548–1620), Dutch mathematician, who helped standardize the use of decimal fractions, was born in 1548 at Bruges and died in 1620, either at The Hague or in Leiden. He was director of the department for roads and waterways, and afterward quartermaster general. Stevin was known to his contemporaries by his military methods and inventions; he invented a defense by a system of sluices to release water through the dikes to flood land before the enemy which was of great importance to Holland. He also invented a carriage with sails which was used on the seashore and carried 26 passengers.

In his *Statics and Hydrostatics* (1586), he enunciated the important theorem of the triangle of forces. This gave a new impetus

to the study of statics, which had previously been founded on the theory of the lever. He discovered the hydrostatic paradox that the downward pressure of a liquid is independent of the shape of the vessel, and depends only on its height and base.

In 1586 he published a pamphlet of a few pages, *Thiende*, in which he treated decimal fractions. Decimal fractions and fractions had been employed for the extraction of square roots five centuries before his time, but no one before Stevin had established their daily use. He declared that the universal introduction of decimal coinage, measures and weights would be only a question of time. His notation is rather unwieldy. He printed little circles round the exponents of the different powers of one-tenth. For instance, $237\frac{5}{10}\frac{7}{00}\frac{8}{0}$ was printed 237 ⓪ 5 ① 7 ② 8 ③.

There are early editions of Stevin's works (1608, 1634); the first volume of a Dutch-English edition, with a biography, appeared in 1955. (O. Oe.)

STEWART, Stuart or Steuart, the surname of a family which inherited the Scottish and ultimately the English crown. Their descent is traced to a Breton immigrant, Alan the son of Flaald, which Flaald was a brother of Alan, steward (or seneschal) of Dol in Brittany. This elder Alan, whose name occurs in Breton documents before 1080, went on crusade in 1097, and was apparently succeeded by his brother Flaald, whose son, the younger Alan, enjoyed the favour of Henry I, who bestowed on him Mileham and its barony in Norfolk, where he founded Sporle priory. By the daughter of Ernulf de Hesdin (in Picardy), a Domesday baron, he was father of at least three sons: Jordan, who succeeded to the family office of steward of Dol; William, who inherited Mileham and other estates in England, and who founded the great baronial house of Fitz Alan (afterward earls of Arundel); and Walter, who was made by David I steward (*dapifer*) or seneschal of Scotland. The Scottish king conferred on Walter various lands in Renfrewshire, including Paisley, where he founded the abbey in 1163. Walter, his grandson, third steward, was appointed by Alexander II justiciary of Scotland and, dying in 1246, left four sons and three daughters. The third son, Walter, obtained by marriage the earldom of Menteith, which ultimately came by marriage to Robert, duke of Albany, son of Robert II. Alexander, fourth steward, the eldest son of Walter, third steward, inherited by his marriage with Jean, granddaughter of Somerled, the islands of Bute and Arran, and on Oct. 2, 1263, led the Scots against Haakon IV, king of Norway, at Largs. He had two sons, James and John. The latter, who commanded the men of Bute at the battle of Falkirk in 1298, had seven sons: (1) Sir Alexander, whose grandson George became in 1389 earl of Angus, the title afterward passing in the female line to the Douglases, and in 1761 to the duke of Hamilton; (2) Sir Alan of Dreghorn, ancestor of the earls and dukes of Lennox, from whom Lord Darnley, husband of Queen Mary, and also Lady Arabella Stuart, were descended; (3) Sir Walter, who obtained the barony of Garlies, Wigtownshire, from his uncle John Randolph, earl of Moray, and was the ancestor of the earls of Galloway, younger branches of the family being the Stewarts of Tonderghie, Wigtownshire, and also those of Physgill and Glenturk in the same county; (4) Sir James, who fell at Dupplin in 1332, ancestor of the lords of Lorn, on whose descendants were conferred at different periods the earldoms of Athole, Buchan and Traquair, and who were also the progenitors of the Stewarts of Appin, Argyllshire, and of Grandtully, Perthshire; (5) Sir John, killed at Halidon Hill in 1333; (6) Sir Hugh, who fought under Edward Bruce in Ireland; and (7) Sir Robert of Daldowie, ancestor of the Stewarts of Allanton and of Coltness. James Stewart, the elder son of Alexander, fourth steward, succeeded his father in 1283 and, after distinguishing himself in the wars of Wallace and of Bruce, died in 1309. His son Walter, sixth steward, who had joint command with Sir James Douglas of the left wing at the battle of Bannockburn, married Marjory, daughter of Robert the Bruce, and during the latter's absence in Ireland was entrusted with the government of the kingdom. He died in 1326, leaving an only son who as Robert II ascended the throne of Scotland in 1371. Sir Alexander Stewart, earl of Buchan, fourth son of Robert II, who earned by his ferocity the title of the "wolf of Badenoch," inherited by his wife the earldom of Ross, but died without legitimate issue, although from his illegitimate offspring were descended the Stewarts of Belladrum, of Athole, of Garth, of Urrard and of St. Fort. On the death of the "wolf of Badenoch" the earldom of Buchan passed to his brother Robert, duke of Albany, also earl of Fife and earl of Menteith, but these earldoms were forfeited on the execution of his son Murdoch in 1425, the earldom of Buchan again, however, coming to the house of Stewart in the person of James, second son of Sir James Stewart, the black knight of Lorn, by Joan or Joanna, widow of King James I. From Murdoch, duke of Albany, were descended the Stewarts of Ardvoirlich and other families of the name in Perthshire, and also the Stuarts of Inchbreck and Laithers, Aberdeenshire. From a natural son of Robert II were descended the Steuarts of Dalguise, Perthshire, and from a natural son of Robert III the Shaw Stewarts of Blackhall and Greenock. The direct male line of the royal family terminated with the death of James V in 1542, whose daughter Mary was the first to adopt the spelling "Stuart." Mary was succeeded in her lifetime in 1567 by her only son James VI, who through his father Lord Darnley was also head of the second branch, there being no surviving male issue of the family from progenitors later than Robert II. In James V, son of James IV by Margaret, daughter of Henry VII, the claims of Margaret's descendants became merged in the Scottish line, and on the death of Queen Elizabeth I of England, the last surviving descendant of Henry VIII, James VI of Scotland, lineally the nearest heir, was proclaimed king of England, in accordance with the arrangements made by Lord Burghley and Elizabeth's other advisers. The accession of James was, however, contrary to the will of Henry VIII, which favoured the heirs of his younger sister Mary, wife of Charles Brandon, duke of Suffolk.

By the usurpation of Cromwell the Stuarts were excluded from the throne from the defeat of Charles I at Naseby in 1645 until the restoration of his son Charles II on May 8, 1660. On the death of Charles II without legitimate issue in 1685, his brother James, duke of York, ascended the throne as James II, but he so alienated the sympathies of the nation by his unconstitutional efforts to further the Roman Catholic religion that an invitation was sent to the prince of Orange to come "to the rescue of the laws and religion of England." Next to the son of James II, still an infant under his father's control, Mary, princess of Orange, elder daughter of James II, had the strongest claim to the crown; but the claims of the prince of Orange also, even apart from his marriage, were not very remote, since he was the son of Mary, eldest daughter of Charles I. The marriage had strengthened the claims of both, and they were proclaimed joint sovereigns of England on Feb. 12, 1689, Scotland following the example of England on April 11. They left no issue, and the Act of Settlement, passed in 1701, excluding Roman Catholics from the throne, secured the succession to Anne, second daughter of James II, and on her death without issue (1714) to the Protestant house of Hanover, descended from the princess Elizabeth, daughter of James I, wife of Frederick V, count palatine of the Rhine. George, elector of Hanover, eldest son of Sophia (youngest child of the princess Elizabeth), and Ernest, elector of Brunswick-Lüneburg, or Hanover, became sovereign of Great Britain and Ireland. The female issue of James II ended with the death of Anne. James, called James III by the Jacobites and the Old Pretender by the Hanoverians, had two sons—Charles Edward, the Young Pretender, who died without legitimate issue in 1788, and Henry Stuart, titular duke of York, commonly called Cardinal York, on whose death in 1807 the male line of James II came to an end. Henry was also the last descendant in the lineal male line of any of the crowned heads of the race, so far as either England or Scotland was concerned. In the female line, however, there are among the descendants of James I representatives of the royal Stuarts who are senior to the house of Hanover, for Philip, duke of Orléans (brother of Louis XIV), married, as his first wife, Henrietta, daughter of Charles I, and, as his second, Charlotte, granddaughter and heiress of the princess Elizabeth (daughter of James I). By the former, through their daughter, the queen of Sardinia, he was ancestor, among others, of the princess Maria Theresa of Bavaria, who in 1910 was "heir of line" of the house of Stuart, her eldest son, Prince

Rupert, being at that time heir to the throne of Bavaria.

BIBLIOGRAPHY.—Sir George Mackenzie, *Defence of the Royal Line of Scotland* (1685), *Antiquity of the Royal Line of Scotland* (1686); Crawfurd, *Genealogical History of the Royal and Illustrious Family of the Stuarts* (1710); Duncan Stewart, *Genealogical Account of the Surname of Stewart* (1739); Andrew Stuart, *Genealogical History of the Stuarts* (1798); Stodart, *House of Stuart* (privately printed, 1855); *An Abstract of the Evidence to Prove That Sir William Stewart of Jedworth, the Paternal Ancestor of the Present Earl of Galloway, Was the Second Son of Sir Alexander Stewart of Darnley* (1801); Riddell, *Stewartiana* (1843); W. Townend, *Descendants of the Stuarts* (1858); R. W. Eyton, *History of Shropshire*, vol. vii (1858); Bailey, *The Succession to the English Crown* (1879); Skelton, *The Royal House of Stuart* (1890); J. H. Round, *Studies in Peerage and Family History* (1901); S. Cowan, *The Royal House of Stuart* (1908); and T. F. Henderson, *The Royal Stewarts* (1914). The best chart pedigree of the house was prepared for the Stuart Exhibition by W. A. Lindsay.

STEWART, CHARLES (1778–1869), U.S. naval officer, first senior flag officer of the United States, was born in Philadelphia on July 28, 1778. He went to sea at the age of 13 and rose to be master of a merchantman. He entered the navy in 1798 as a lieutenant in the "United States" under Commodore John Barry, his first naval command being the "Experiment," July 16, 1800. He became a captain in 1806. During the War of 1812 Stewart won a gold medal from congress for his capture of two British sloops of war, the "Cyane" and "Levant," while he was in command of the "Constitution." Afterward he commanded squadrons in the Mediterranean and the Pacific. An act of congress of 1859 provided for a senior flag officer, and this office Stewart held until July 16, 1862, when he became a rear admiral on the retired list. The Irish nationalist statesman Charles Stewart Parnell (*q.v.*) was his grandson. Stewart died Nov. 6, 1869, in Bordentown, N.J. (J. B. HN.)

STEWART, SIR DONALD MARTIN (1824–1900), British field marshal, son of Robert Stewart of Forres, Elginshire, was born at Mount Pleasant, near Forres, on March 1, 1824. He entered the Bengal army in 1840, and served in 1854 and 1855 in the frontier expeditions against the Mohmands, and Afridis Aka and Bari Khel (medal and clasp). In the Indian Mutiny in 1857 Stewart, after a famous ride from Agra to Delhi with despatches, served on the staff at the siege and capture of Delhi and Lucknow, and afterwards through the campaign in Rohilkhand. For nine years he was assistant and deputy-adjutant-general of the Bengal army, commanded the Bengal brigade in the Abyssinian expedition in 1867, and became a major-general in 1868. He reorganized the penal settlement of the Andaman Islands, where he was commandant when Lord Mayo was assassinated, and, after holding the Lahore command, was promoted lieutenant-general in 1877, and commanded the Kandahar field force in the Afghan War in 1878. On hearing of the Maiwand disaster he despatched Roberts with a division on his celebrated march from Kabul to Kandahar, and himself led the rest of the army back to India by the Khyber Pass. Promoted general in 1881, he was for five years commander-in-chief in India, and afterwards member of the council of the secretary of state for India until his death. He was promoted to be field marshal in 1894, and appointed governor of Chelsea Hospital in 1895. He died at Algiers on March 26, 1900.

See G. R. Elsmie, *Sir Donald Stewart* (1903).

STEWART, DUGALD (1753–1828), Scottish philosopher, was born in Edinburgh on Nov. 22, 1753. His father, Matthew Stewart (1715–1785), was professor of mathematics in the university of Edinburgh (1747–1772). Dugald Stewart was educated at Edinburgh and Glasgow. At 19 he acted as substitute for his father in the chair of mathematics at Edinburgh. From 1785 to 1820 he was professor of moral philosophy in the university, though he retired from the active duties of the chair after the death (1809) of his second son, which was a great blow to him. He died at Edinburgh on June 11, 1828.

Stewart's principal work is *Elements of the Philosophy of the Human Mind* (3 vols. 1792, 1814, 1827). He also wrote *Philosophical Essays* (1810), and *The Philosophy of the Active and Moral Powers* (1828). His philosophical views are mainly the reproduction of those of his master Reid.

STEWART, SIR WILLIAM (*c.* 1540–*c.* 1605), Scottish politician, began the life of a soldier in the Netherlands, where he became a colonel. In 1582 he was in Scotland, where James VI made him captain of his guard. Having visited the English court in the king's interest in 1583, Stewart helped to free James from William Ruthven, earl of Gowrie, and to restore James Stewart, earl of Arran, to power; he was made a privy councilor and for a time assisted Arran to govern Scotland. In 1584 he captured Gowrie at Dundee. In 1594 he was knighted and was given lands at Houston. He died before 1606.

STEWART, WILLIAM DOWNIE (1878–1949), New Zealand lawyer and politician, was born in Dunedin, N.Z., on July 29, 1878. He was educated at Otago university and later entered the legal profession. Though taking part in local politics he maintained an early interest in the study of political economy and was joint author with J. E. Le Rossignol of *State Socialism in New Zealand* (1910). He entered parliament in 1914 as member for Dunedin West, but shortly afterward joined the New Zealand expeditionary forces. He was invalided from France in Dec. 1916, crippled with rheumatism, from which he never completely recovered. This was a handicap, particularly in New Zealand, where politicians are expected to travel widely. In spite of this, however, his intellectual merit was so outstanding that he was invited to join the cabinet in 1921 as minister of customs. He later held the portfolios of internal affairs (1921–24) and industries and commerce (1923–26). From 1926, when he acted as prime minister, to 1928, he was minister of finance in J. G. Coates's ministry. In 1933 he resigned from G. W. Forbes's ministry, in which he was again minister of finance, because of differences over policy. Stewart was a delegate to the Ottawa conference in 1932 and to the Commonwealth Relations conferences at Toronto in 1933 and Sydney in 1938. His speeches in the debate preceding and following the conferences were marked by clearness of thought and a close grasp of constitutional principles. He died at Dunedin on Sept. 29, 1949. His books include *Life and Times of Sir Francis Bell* (1937); and *Life of William Rolleston* (1940).

(A. T. CL.)

STEYN, MARTINUS THEUNIS (1857–1916), last president of the Orange Free State, was born at Winburg in that state on Oct. 2, 1857. He was a student in Holland and later in England at the Inner Temple, and was called to the English bar in Nov. 1882. After his return to South Africa he practised as a barrister at Bloemfontein, and in 1889 was appointed state attorney of the Free State. A few months afterward he became second puisne judge, and in 1893 first puisne judge of the high court. His decisions won him a reputation for ability and sound judgment. In 1895, upon the resignation of Pres. F. W. Reitz, Steyn was the candidate of the pan-Dutch party for the vacant post. The election resulted (Feb. 1896) in a decisive victory for Steyn. As president he linked the fortunes of his state with those of the Transvaal, a policy which led to the extinction of the republic. He took part in the peace negotiations at Klerksdorp in April 1902, but was prevented by illness from signing the instrument of surrender at Pretoria on May 31. In 1908–09 he was vice-president of the Closer Union convention, where he showed a conciliatory attitude.

STEYNING, a small town and country parish in West Sussex, England, 11 mi. W.N.W. of Brighton by road. Pop. of parish (1951) 2,500. Area 5.5 sq.mi. A borough by prescription, Steyning returned two members to parliament from Edward I to 1832; it is now in the Arundel and Shoreham division. The Anglo-Saxon church mentioned in Domesday Book is attributed to St. Cuthman whose shrine became a resort for pilgrims. The restored church of St. Andrew contains Norman and Early English work, and the vicarage stands on the site of a Benedictine priory suppressed by Henry IV. The Caen stone for this church was brought by boat, for Steyning was a river port and market town until Henry VIII granted the commissioners of the river Adur a charter to embank the river, after which the port ceased to exist.

STEYR, a town in Oberösterreich *Bundesland* at the junction of the Steyr with the Enns. It is an old town, dominated by a 10th century castle. Two important new suburbs, Steyrdorf and Ennsdorf, lie across the streams. Steyr is a centre of the iron and

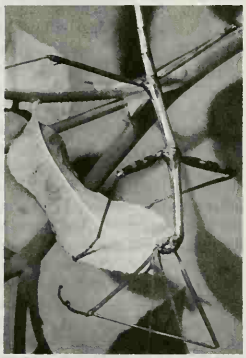
steel industry. Pop. (1951) 36,818.

STIBNITE, a mineral consisting of antimony sulfide, is an important ore of antimony. The powdered mineral was used by the ancients for darkening the eyebrows to increase the apparent size of the eyes. In addition to its use as an ore of antimony, it is employed in the manufacture of matches, fireworks and percussion caps. The formula of stibnite is Sb_2S_3. The crystals are prismatic in habit, longitudinally striated and usually terminated by acute pyramidal planes. There is a perfect cleavage parallel to the length of the crystals, and the basal plane is a plane of gliding; it is the latter property that facilitates the bending of the crystals, and stibnite crystals are often found in a deformed condition. The colour is lead gray and the lustre is metallic and brilliant; crystals become dull on prolonged exposure to light. The mineral is quite soft; hardness is 2 and specific gravity is 4.6. Thin flakes of the mineral transmit a small amount of red light, but are more transparent in the infrared. Stibnite is a photoconductive material which shows maximum sensitivity at 770 mμ.

Stibnite is found associated with quartz in low-temperature hydrothermal veins in Hunan province, China; Shikoku Island, Jap.; and Lubilhac, Haute-Loire, France.

See also ANTIMONY. (A. J. F.)

STICK INSECT (WALKING STICK), an insect of the family Phasmatidae (*see* ORTHOPTERA), so named because of its resemblance to twigs of trees and shrubs upon which most phasmids live and feed. The body and legs are long and slender, and the protective resemblance to twigs is enhanced by the usually brown, gray or green coloration and the immobile attitudes assumed. Some species are further protected by an armature of sharp spines. Many are wingless; when winged the leathery wing covers (tegmina) are short, but the hind wings are often large and beautifully coloured, though usually kept folded over the abdomen. As in the leaf insects (*q.v.*), to which phasmids are closely allied, the eggs closely resemble seeds. Stick insects attain their largest size and greatest profusion of species in the tropics, one Australian species, *Palophus titan,* reaching a length of ten inches. A few smaller forms occur in southern Europe, and about 25 species are known from the United States, among them *Diapheromera femorata,* which sometimes defoliates oak forests. Phasmids freely regenerate amputated legs and antennae; under certain conditions the stump of an antenna will regenerate a leg (*see* REGENERATION). (T. H. HL.)



EDWIN WAY TEALE

AMERICAN STICK INSECT (DIAPHEROMERA FEMORATA), COMMONLY KNOWN AS WALKING STICK, SHOWN FEEDING ON LEAF OF WILD BLACK CHERRY

STICKLEBACK, a group of small fishes which inhabits the fresh and brackish waters as well as the coasts of the temperate zone of the northern hemisphere. The majority have a compressed well-proportioned body, which in the marine species is of a more elongate form. The mouth is of moderate width, oblique and armed with small teeth. The head is protected by hard bone. There is a series of scutes protecting the sides. The first dorsal fin and the ventrals are transformed into pointed formidable spines, and joined to firm bony plates of the endoskeleton. A remarkable degree of variability is shown by many species. Sticklebacks construct a nest for the reception of the spawn, which is jealously guarded by the male until the young are hatched, from 10 to 18 days after oviposition. He also protects them for the first few days of their existence.

Sticklebacks are short-lived animals said to reach an age of three or four years. In spring each male selects a territory, which he fiercely defends and to which he invites all females, until the nest is filled with ova. At this period he assumes a dress of blue and red. The eggs are comparatively large, one female depositing from 50 to 100.

About 12 species of sticklebacks are known. The three-spined stickleback (*Gasterosteus aculeatus*) is found everywhere in the northern part of the northern hemisphere; in North America it occurs as far south as the Virginia coast. The ten-spined stickleback (*G. pungitius*) rarely exceeds two inches in length. Its geographical range nearly coincides with that of the other species. The four-spined stickleback, *Apeltes quadracus,* is common along the North American east coast from New Brunswick to Virginia; though it may reach a length of two and one-half inches, one and one-half inch is commoner. The sea stickleback (*Spinachia spinachia*) attains a length of seven inches, and is armed with 15 short spines. It is common around the British coasts. At suitable localities sheltered from the waves and overgrown with seaweed, especially in rock pools and shallows covered with eelgrass, one or two males establish themselves with their harems. The elaborate nest is constructed by binding together suitable fronds of seaweed.

STIEGLITZ, ALFRED (1864–1946), U.S. photographer, who devoted his life first to a fight for photography as an art, then to introducing modern painting to the United States, was born in Hoboken, N.J., on Jan. 1, 1864. Stieglitz was educated in New York and Germany, where he began photographing at the age of 19. His earlier work (*c.* 1883–1910), characterized by constant innovations believed impossible (*e.g.,* the first successful snow photographs), won over 150 important prizes. Having achieved public fame, his experimentation changed from preoccupation with techniques and wide subject matter to intense concentration on material closest to daily life. Two series, over 400 prints each (1917–37), are often considered his greatest photographs: the portrait of Georgia O'Keeffe (his wife) and the "Equivalents." His photographic prints were the first that U.S. museums acquired and housed with other graphic arts.

Stieglitz believed passionately that America could give rise to great works of art. His fostering of the work of others paralleled his own creative work, and is demonstrated through over 175 exhibitions in his three galleries (1905–46), many of which offered the public its first sight of modern art. Starting with Rodin and Matisse exhibitions in 1908 at his "291" gallery, Stieglitz followed with many other first U.S. exhibitions of equally important works (*e.g.,* African Negro sculpture as art, Cézanne, Picasso, Toulouse-Lautrec and others). When the public accepted modern European art, Stieglitz concentrated on U.S. artists, especially the five he showed most often in later years: Demuth, Dove, Hartley, Marin and O'Keeffe. *Camera Work* (1903–17), edited and published by Stieglitz, documents the history of these stormy years. He died on July 13, 1946, in New York city. (DS. BY.)

STIEGLITZ, JULIUS (1867–1937), U.S. chemist, was a pioneer in recognizing the value of the electron theories of valence to the interpretation of the behaviour and structure of organic compounds, and in applying to organic chemistry the methods and principles of physical chemistry. Born May 26, 1867, in Hoboken, N.J., he received his early schooling in New York, before attending the *Realgymnasium* in Karlsruhe, Ger., from 1881 to 1886. In 1889 he was awarded the Ph.D. in organic chemistry by the University of Berlin, where he had worked under Ferdinand Tiemann. Most of his professional career was spent at The University of Chicago, where he was chairman of the department of chemistry from 1915 to 1933. He was influential in the university as a whole as well as in the American Chemical society, of which he was president in 1917.

He died in Chicago, Jan. 10, 1937.

Stieglitz' research in organic chemistry dealt with four major topics: molecular rearrangements, for which he won the Gibbs medal in 1923; homogeneous catalysis; the theory of indicators; and the stereochemistry of organic nitrogen compounds. He was widely known as an inspiring teacher and, through his text on qualitative analysis (2 vol., 1911–12), the first to base that discipline on the theories of ionization and of chemical equilibrium. (H. I. SR.)

STIERNHIELM, GEORG (1598–1672), Swedish poet and

scholar, often called "the father of Swedish poetry," was born Aug. 7, 1598, at Vika, the son of a miner in Dalecarlia (Dalarne). After studying at Uppsala he spent several years at German universities (Greifswald, Wittenberg, Helmstedt). He returned to Sweden in 1626 and later obtained an official position in Dorpat. In 1631 he was raised to the nobility. From *c.* 1640 onward he was occasionally in Stockholm, as poet in attendance at the court of Queen Christina, although his home was in Estonia until 1656, when he fled before the Russian invaders. Thereafter he lived in Stockholm in straitened circumstances. In 1661 he was appointed councilor of war and in 1667 director of the college of antiquities. He died at Stockholm, April 22, 1672.

Stiernhielm's first poetic works in Swedish appeared during the 1640s. They included verses in celebration of the queen and three court masques adapted from the French, one of which, *Then Fångne Cupido* (1649), contains fine passages. His most important work is the allegorical, didactic epic, *Hercules* (written *c.* 1647; published 1658), a fine example of late Renaissance classicism, written in hexameters. Imbued with the spirit of humanism, it is a high-minded sermon on virtue and honour. The theme is developed with power and originality; the imagery is exuberant; the construction faultless. It had great influence on the development of Swedish poetry. Stiernhielm's poems were collected in *Musae Suethizantes* (1668).

Stiernhielm's scholarship was encyclopaedic, and in numerous writings, many unpublished, he dealt with philological, historical and philosophical problems. A great patriot, he claimed that Swedish was man's original language. In his fragmentary works of natural philosophy he expounded a cosmogony based on the platonic, mystical tradition of such thinkers as Paracelsus, Robert Fludd and Comenius.

Stiernhielm's *Samlade skrifter* were edited by J. Nordström, 12 vol. (1924 *et seq.*).

See also P. W. Wieselgren, *Georg Stiernhielm* (1948). (S. H. L.)

STIGAND (d. 1072), archbishop of Canterbury, is first mentioned in 1020. He was then chaplain to Canute and afterward to his son, Harold Harefoot, and after the death of Canute appears to have acted as the chief adviser of his widow, Emma. In 1043 he was consecrated bishop of Elmham and in 1047 was translated to Winchester; he supported Earl Godwine in his quarrel with Edward the Confessor, and in 1052 arranged the peace between the earl and the king. In this year the archbishop of Canterbury, Robert of Jumièges, having been outlawed, Stigand was appointed to the archbishopric; but Pope Leo IX. and his two successors refused to recognize him. In 1058, however, Benedict X. gave him the pall, but this pope was deposed in the following year. Stigand is said by Norman writers to have crowned Harold in Jan. 1066; but this ceremony was probably performed by Aldred, archbishop of York. Stigand submitted to William, and assisted at his coronation. But the Conqueror was anxious to get rid of him. In 1070 he was deposed by the papal legates and was imprisoned at Winchester, where he died, probably on Feb. 22, 1072. Stigand was an avaricious man and a great pluralist, holding the bishopric of Winchester after he became archbishop of Canterbury, in addition to several abbeys.

See E. A. Freeman, *The Norman Conquest* (1870–76), vols. ii., iii. and iv.; and J. R. Green, *The Conquest of England* (1899), vol. ii.

STIGMATIZATION, the impression of *stigmata.* Originally, the process of marking or branding slaves and criminals.

The apostle Paul is the first to speak of stigmata in a religious sense. He reminds the Galatians of the distinctive "marks (*stigmata*) of the Lord Jesus" in opposition to the old mark of circumcision: "I bear the marks of the Lord Jesus in my body" (Gal. 6:17). Paul's stigmata were scars of wounds received during his apostolate from persons opposed to his preaching (II Cor. 11:24 f.).

Specifically, stigmatization is a phenomenon observed in a number of Christian saints and mystics for which no satisfactory natural explanation has been offered as yet. It consists of the appearance, on the body of a living person, of wounds or scars corresponding to those of the crucified Christ.

The first and most celebrated stigmatization of this kind is that of St. Francis of Assisi (1181–1226). His biographers, Thomas of Celano, St. Bonaventure and other contemporaries tell us that early one morning, Sept. 14, 1224, while Francis was praying on mount La Verna, a six-winged seraph appeared to him bearing between his wings the form of one crucified and clearly the figure of Christ. When the apparition vanished, a great love of God was felt in Francis' soul and in his body there appeared an impression of the wounds of the Lord Jesus. The reproduction of a lance-wound, fresh and bleeding, was seen on his side, whence thereafter blood came out, on many occasions staining his garments. His hands and feet seemed perforated in the middle. Pope Alexander IV and many other witnesses declared that they had seen the stigmata both before and after his death.

According to Dr. A. Imbert-Gourbeyre, up to 341 cases of stigmatization have been recorded after the time of St. Francis. In 300 cases, the stigmatized person was a woman. Better known stigmatists of modern times are: Catherine Emmerich of Muenster, (1774–1824); Mary von Möerl of Kaltern, Tyrol, (1812–68); Louise Lateau of Bois de Haine, (1850–83). Two renowned cases of stigmatization belong to the 20th century, those of Gemma Galgani of Lucca, Italy, (1878–1903) and of Theresa Neumann of Konnersreuth, Germany. Gemma is the last stigmatist to be canonised by the Roman Catholic Church (May 2, 1940). However, she was declared a saint not on account of her stigmata but for her heroic virtue. Gemma received the stigmata on June 8, 1899. In addition to the wounds in her hands, feet and side, she experienced also the stigmata of the crown of thorns and those of the flagellation. Her stigmata were not permanent but recurrent, reappearing ordinarily every Friday.

The case of Theresa Neumann attracted the attention of scholars and scientists for several years. Born on Good Friday, April 8, 1898, she received the first stigmata on the upper surface of her hands and feet on Good Friday, April 2, 1926. On this occasion, her eyes began to bleed profusely, a phenomenon that used to occur whenever she meditated on Christ's suffering. Stigmata on the lower surface of her hands and feet appeared on the following Good Friday, April 15, 1927, and in subsequent years those on her side, shoulder and brow became manifest. They were witnessed by thousands within the first decade of their appearance and became an object of investigation. (P. P. P.; X.)

STILBITE, a mineral of the zeolite group consisting of hydrated calcium aluminum silicate. The colour is usually white, sometimes red, and on the perfect cleavage (parallel to the plane of symmetry) the lustre is markedly pearly; hence the name stilbite (Gr. *stilbein,* "to shine") given by R. J. Haüy in 1796. After the separation of heulandite from this species in 1818, the name desmine (Gr. *desme,* "bundle") was proposed, and this name is employed in Germany. The composition is $CaAl_2Si_7O_{18}.7H_2O$. Usually a small proportion of the calcium is replaced by sodium. Crystals are monoclinic, and are invariably twinned, giving rise to complex groups and characteristic sheaflike aggregates. The hardness is 3.5 and the specific gravity 2.2. Stilbite occurs with other zeolites in the amygdaloidal cavities of basic volcanic rocks; it is sometimes found in granite and gneiss, and exceptionally in metalliferous veins. Beautiful, salmon-pink crystals occur with pale-green apophyllite in the Deccan traps near Bombay and Poona, India; white sheaflike groups encrust the calcite (Iceland spar) of Berufjord near Djupivogr, Ice.; and crystals of a brick-red colour are found at Old Kilpatrick, Dumbartonshire, Scot. Good specimens of stilbite have been found in New Jersey and Nova Scotia.

For discussion of the crystalline make-up of the zeolites, zeolitic dehydration and the property of base exchange, *see* ZEOLITE.

STILES, EZRA (1727–1795), U.S. clergyman and educator, seventh president of Yale college, was born on Nov. 29, 1727, in North Haven, a parish of New Haven, Conn. He graduated at Yale in 1746; was licensed to preach in 1749 and was a tutor at Yale in 1749–55. He also was admitted to the bar in 1753, and practised in New Haven for two years. After pastorates at Newport, R.I., and Portsmouth, N.H., he became in 1778 president of Yale college and professor of ecclesiastical history, becoming for a few years after 1780 professor of divinity and lecturing on many

subjects. He died in New Haven on May 12, 1795. Under his administration as president Yale college soon recovered after the War of Independence; his intellectual and theological breadth helped to secularize and strengthen the college. Stiles published a number of sermons, and his *Literary Diary* (3 vol., 1901) was edited by F. B. Dexter. *Extracts from the Itineraries and Other Miscellanies* with selections from Stiles's correspondence prepared by the same editor appeared in 1916. Both give a valuable picture of New England life and contemporary figures.

STILICHO, FLAVIUS (?–408), Roman general and statesman, was the son of a Vandal who had served as an officer in the army of the emperor Valens (364–378). He himself entered the imperial army at an early age and speedily attained high promotion. He had already become master of the horse when in 383 he was sent by Theodosius (379–395) at the head of an embassy to the Persian king, Sapor III. His mission was successful, and soon after his return he was made count of the domestics and received in marriage Serena, the emperor's niece and adopted daughter. In 385 he was appointed master of the soldiery in Thrace, and shortly afterward directed energetic campaigns in Britain against Picts, Scots and Saxons, and along the Rhine against other barbarians. Stilicho and Serena were named guardians of the youthful Honorius when the latter was created joint emperor in 394. Rivalry had already existed between Stilicho and Rufinus, the praetorian praefect of the east. Consequently in 395, after a successful campaign against the Germans on the Rhine, Stilicho marched east, with the design of displacing Rufinus; and by connivance with the barbarians he procured the assassination of Rufinus at the close of the year, and thereby became virtual master of the empire. In 396 he fought in Greece against the Visigoths, but an arrangement was effected whereby their chieftain Alaric was appointed master of the soldiery in Illyricum (397). In 398 he quelled Gildo's revolt in Africa and married his daughter Maria to Honorius. Two years later he was consul. He thwarted the efforts of Alaric and Radagaisus to seize lands in Italy by his victories on the Danube at Pollentia and Verona in 401–403.

In a second campaign against Radagaisus, who led a large force of various Germanic peoples into Italy in 405, he surrounded the barbarian chieftain on the rocks of Fiesole and starved him into surrender. Early in 408 he married his second daughter Thermantia to Honorius. It was rumoured about this time that Stilicho was plotting with Alaric and with Germans in Gaul and taking other treasonable steps in order to make his own son Eucherius emperor. The facts are doubtful. It is certain, however, that he was suspected by Honorius and abandoned by his own troops, and that he fled to Ravenna, and, having been induced by false promises to quit the church in which he had taken sanctuary, was executed on Aug. 23, 408.

STILL, ANDREW T. (1828–1917), founder of osteopathy (*q.v.*), was born in Jonesville, Va., on Aug. 6, 1828. His family moved in 1837 to Macon county, Mo., and later to the Shawnee reservation near Kansas City, Kan. Still took an active part in the slave question in Kansas, and in 1857 was elected to the Kansas Territorial legislature on the Free State ticket. He began searching for other means than medicine to combat disease and formulated his principles of osteopathy in 1874. He underwent a long period of opposition, but finally, in 1892, founded the American School of Osteopathy at Kirksville, Mo., in conjunction with William Smith, a Scottish physician. In 1922 the Andrew T. Still College of Osteopathy and Surgery was founded and merged with the original American school.

See Ray G. Hulbert, "Andrew T. Still," *Missouri Historical Review*, vol. xix, 1924; E. R. Booth, *History of Osteopathy* (Cincinnati, 1924); M. A. Lane, *Dr. A. T. Still, Founder of Osteopathy* (Chicago, 1918). (F. L. A.)

STILLBIRTH is a birth in which the child shows no sign of life, including heartbeat, respiration or muscular activity, after complete expulsion from the body of the mother, regardless of whether the umbilical cord has been severed. The term sometimes is used incorrectly to designate the child itself. Most public health authorities require registration of stillbirths occurring after the 20th or 22nd week of pregnancy; those occurring earlier are generally designated abortions. Stillbirths may be divided into two groups: those in which the fetus (unborn child) dies before the beginning of labour (antepartum) and those in which death occurs during labour (intrapartum).

The World Health organization in 1950 recommended that the word stillbirth be given up and fetal death substituted, and that such deaths be divided into four groups according to the period of gestation attained before expulsion from the uterus: (1) before 20 weeks; (2) from 20 to 28 weeks; (3) over 28 weeks; and (4) unknown.

The principal causes of stillbirth are malformation, birth injury, interference with circulation of blood through the umbilical cord or the placenta (afterbirth), infection and erythroblastosis fetalis. Many stillbirths result from unknown causes, especially those in which death occurs before the onset of labour. *See* also Abortion; Childbirth; Fetal Disorders. (E. L. Po.)

STILLINGFLEET, EDWARD (1635–1699), English theologian, was born at Cranborne, Dorset, on April 17, 1635. He graduated from St. John's college, Cambridge, in 1652, and in the following year was elected to a fellowship. At Sutton, Beds., of which he was vicar, he published (1659) his *Irenicum* in which he sought to give expression to the prevailing weariness of the faction between Episcopacy and Presbyterianism, and to find some compromise. He looks upon the form of church government as nonessential, but condemns Nonconformity. Although in 1680 he published his *Unreasonableness of Separation*, his willingness to serve on the ecclesiastical commission of 1689, and the interpretation he then proposed of the damnatory clauses of the Athanasian creed, are proof that to the end he leaned toward toleration. His rapid promotion dates from 1662, when he published *Origines sacrae, or a Rational Account of the Christian Faith as to the Truth and Divine Authority of the Scriptures and the Matters therein contained*. In 1665 the earl of Southampton presented him to St. Andrew's, Holborn, two years later he became prebendary of St. Paul's, in 1668 chaplain to Charles II, in 1670 canon residentiary, and in 1678 dean of St. Paul's. He was also preacher at the Rolls Chapel and reader at the Temple. Finally he was consecrated bishop of Worcester on Oct. 13, 1689. During these years he was ceaselessly engaged in controversy with Nonconformists, Romanists, Deists and Socinians. The range of his learning is most clearly seen in his *Bishop's Right to Vote in Parliament in Cases Capital*. His *Origines Britannicae, or Antiquities of the British Church* (1685), is a strange mixture of critical and uncritical research. In his closing years he had some controversy with John Locke, whom he considered to have impugned the doctrine of the Trinity. He died at Westminster on March 28, 1699.

A collected edition of his works, with life by Richard Bentley, was published in London, 6 vol. (1710); and a useful edition of *The Doctrines and Practices of the Church of Rome Truly Represented* was published in 1845 by William Cunningham.

STILL-LIFE PAINTING may be loosely defined as the depiction of inanimate objects for the sake of their qualities of form, colour, texture and composition. Still life has always existed in painting as a subsidiary element; it occurs in Roman wall decorations at Herculaneum in the 1st century A.D. and again in mosaics (Lateran museum, Rome) of the same date. Four centuries later it appeared as an independent decorative element in the ceiling mosaics at the church of Sta. Costanza, Rome. A significant change came in the 6th century when pure still life disappeared and was replaced by symbols with a specific religious connotation, as exemplified by the two fish on a plate in a mosaic of the Last Supper at S. Apollinare Nuovo, Ravenna. It was not until the rise of a wealthy merchant class, coupled with the prolonged effects of the Reformation, that a new demand grew for secular subjects in painting other than portraiture. Finally, the Renaissance artists' fresh interest in the material details of their environment led to a more naturalistic representation of these as well as of the human figure. Thus Raphael included an admirable still-life group of musical instruments in his "St. Cecilia" (Bologna) and Vittore Carpaccio devoted great care to the still life in his "St. Jerome in his Study" (S. Giorgio dei Schiavoni, Venice). Yet it was the northern artists who provided minutely

detailed still life, and the exquisite draftsmanship of Jan van Eyck and of Albrecht Dürer was a part of the northern artistic tradition most suited to the development of still-life painting as a separate art.

However, the first signed and dated pure still life, of a dead partridge, gauntlets and arrow pinned against a wall (Munich), was painted in 1504 by an Italian, Jacopo de' Barbari, working in the north. Among other early still lifes is one attributed to Pieter Brueghel the Elder (Boymans museum, Rotterdam), and it contains all the elements of the genre. On the table lie an earthenware pot, a jug and bowl, a sliced loaf upon a wooden platter, some pieces of cheese, a joint of meat and three fishes; all these are thrust forward in such a way that the spectator seems to be very close to the table and looking at them from only slightly above. In the top right corner a window overlooks a landscape with two small figures and a windmill. The artist has concentrated on rendering the forms of the objects, which are accentuated by a strong light coming from a window on the left, but already the observation of tonal values of light and shade irrespective of local colour was becoming the chief preoccupation of the still-life painter. Brueghel's austere, tightly composed painting may be contrasted with the work of Pieter Aertszen (1507–73), who preferred the market stall and the kitchen larder, with their piles of vegetables and fruits, as his subjects. Figures are sometimes put into the background, as in "The Butcher's Shop" of 1551 (Uppsala), but interest is concentrated upon gory sides of beef, sausages, game and fish spread upon slabs or suspended from hooks. The composition is overcrowded; the picture literal to a point of nausea.

Giuseppe Arcimboldo (*c.* 1530–93) made a curious use of still life. Human heads and even whole figures are ingeniously fashioned out of various fruits, vegetables, fish and animals to symbolize the four elements or the four seasons. Examples of these fantasies are in the Vienna art gallery. One of the few great Italian masters to produce pure still life was Michelangelo Caravaggio (1573–1610), and his still lifes in the Ambrosiana, Milan, and the Kress collection, Washington, D.C., show him to be a formidable rival of the Dutch and Flemish artists. Solidity of form and composition, brilliant colour and precise drawing are united with a superb feeling for light and atmosphere in his work, as in Francisco de Zurbaran's famous still life of 1633. Qualities such as these are the essence of still life; and during the 17th century many Netherlandish painters achieved a competence of this kind rarely surpassed. Mention need only be made of Willem Heda (1594–1680) and Willem Kalf (1622–93) who used goblets, nautilus shells, delicate porcelain and rich fabrics for much of their work. Kalf's technique, in particular, resembles Vermeer's; and, indeed, the interiors with figures painted by Vermeer partake strongly of the quality of still life. Kalf's pictures are usually simpler and less crowded than those of Abraham van Beyeren (1620–75), although each relied upon similar subjects, and the inclusion of a partly peeled lemon was a favourite trick with many of these painters. Cornelis de Heem and Jan de Heem were the ablest of a large family of still-life painters, while Jan van Huysum (1682–1749) became pre-eminent as a flower painter, and often set his still life against a landscape background. Frans Snyders (1579–1657), along with Jan Fyt (?1611–61) and Jan Weenix (1640/44–1719) also did this, but chose to paint dead game, sometimes adding live birds and dogs for contrast. Melchior d'Hondecoeter (1636–95) also worked in this tradition.

Nothing of new artistic importance emerged from the Netherlands during the 18th century, for the country had become severely impoverished. Jean-Baptiste-Siméon Chardin (1699–1779) added peculiarly French qualities of lucidity and penetration to the enhancement of the art. He broke away from the outmoded conventions of A. F. Desportes (1661–1743) and J. B. Oudry (1686–1755), derived from the Netherlands; and the Louvre version of "The Skate" (1728) is notable for its broad treatment and thick, impasted highlights. There is an escape from the airlessness and overcrowding, the emphasis upon the inessential detail to the detriment of the whole, which often mars Dutch and Flemish painting. The figures and objects in a Chardin household seem timeless and inevitable; "La Fontaine" (National gallery, London) and "Le Bénédicité" (Louvre) are paintings that equal, if not surpass, the work of Gerard Ter Borch and Jan Vermeer.

Something of Chardin's ease and simplicity appears in Francisco Goya's still-life paintings; *e.g.*, the "Still-Life With Lamb's Head" in the Louvre; and he, with Chardin, was to inspire 19th-century artists as diverse as I. H. Fantin-Latour and Edouard Manet.

After a momentary eclipse during the neoclassical period, interest in still life was reawakened by Eugène Delacroix's startlingly original "The Stove," painted in the 1830s. Casually composed, the painting shows a corner of the artist's studio with its stove and other bric-a-brac. There is a new quality of high-spirited realism, and even defiance, and it is significant that Paul Cézanne was to paint the same type of composition (Chester-Beatty collection, London) in the fierce, undisciplined style of his youth.

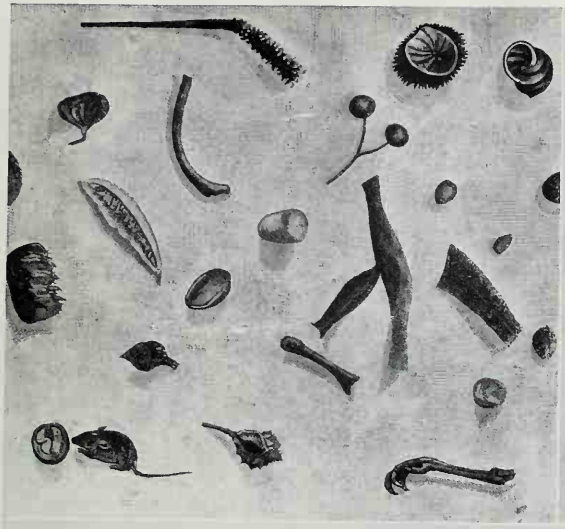
Fantin-Latour's early work is dry and sombre, with the accent upon one bright area, perhaps some roses in a glass, like the Tate gallery painting of 1861. He used a limited palette, and relied upon subtle gradations of tone to obtain plasticity. Influenced by the Impressionists, his colours became richer and his brushwork more lively and impasted, as in the 1881 "Bowl of Flowers" (Lady Lever gallery, Port Sunlight, Cheshire), but at no time did he approach the vigour and immediacy displayed in Gustave Courbet's "Fleurs dans un Panier," 1863 (Cargill collection, London), so reminiscent of Delacroix's painting of 1849, "The Over-turned Flower Basket in a Park" (Wildenstein gallery, New York city). Unhesitatingly, using broad, fluent strokes, Courbet captured the essential features of the objects before him, giving them tangible shape and glowing colour. "Les Grenades," 1871 (Burrell collection, Glasgow), and "The Trout," 1873 (Oeffentliche Kunstsammlung, Zürich), are two of Courbet's most powerful works. Even Manet's "Les Poissons" (Chester-Beatty collection, London), with its cool, brilliant colour, cannot quite rival it.

As might be expected, the still lifes by Claude Monet are flooded with a warm shimmering light. Form and space are revealed more by the acute perception and recording of hitherto unexplored colour harmonies, than by tonal variations and linear perspective.

Cézanne found landscape and still life particularly suited to his own artistic development. Still life acquired a new importance for him, allowing him to concentrate upon the fundamental, recurring problems of form, colour and space, undisturbed by irrelevant anecdotal or literary associations. By 1870, with "Still-Life With Black Marble Clock" (Wildenstein), he showed great originality of conception. Already he was fascinated by the abstract pattern formed by the folds in the tablecloth; the objects on the table, seen at a high eye level, seem to be in the same spatial plane as the clock on the mantelpiece. This spatial ambiguity is even more pronounced in the Louvre "Still-Life With Ginger Jar" (1886–88), where the jar and basket of fruit on the sharply tilted floor seem level with the objects on the table in the foreground. This deliberate use of different eye levels and viewpoints, combined with a slight "flattening" of elliptical shapes (*e.g.*, coffee pot, basket handle), was an attempt to limit the amount of recession into the picture, and to resolve the paradox of representing three-dimensional objects on a two-dimensional surface. Paul Gauguin also used those devices, but emphasized the decorative element, modeling his objects in broad planes of pure, bright colours; a development for which Van Gogh was also partly responsible.

Vincent Van Gogh's intensely personal vision, when applied to still-life painting, could bring a new emotional significance to such mundane objects as a row of peasants' boots, a Bible and a candle, or his "Bedroom Chair" (Tate gallery, London). Here, as in all his work, both the colours and the brush strokes used convey not only shape, but also much of this emotional fire. The contrast between the Tate "Chair" and "Gauguin's Chair" (V. W. Van Gogh collection, Amsterdam), with its fierce clashes of violet, green and red, surely illustrates Van Gogh's extremely wide evocative powers.

At the turn of the century, still-life painting was no longer the province of a few specialists. Between 1907 and 1909, the Cubists, led by Pablo Picasso and Georges Braque, carried Cézanne's ideas much further, and like him found still life to be the subject most suited to their experiments with perspective. Objects were



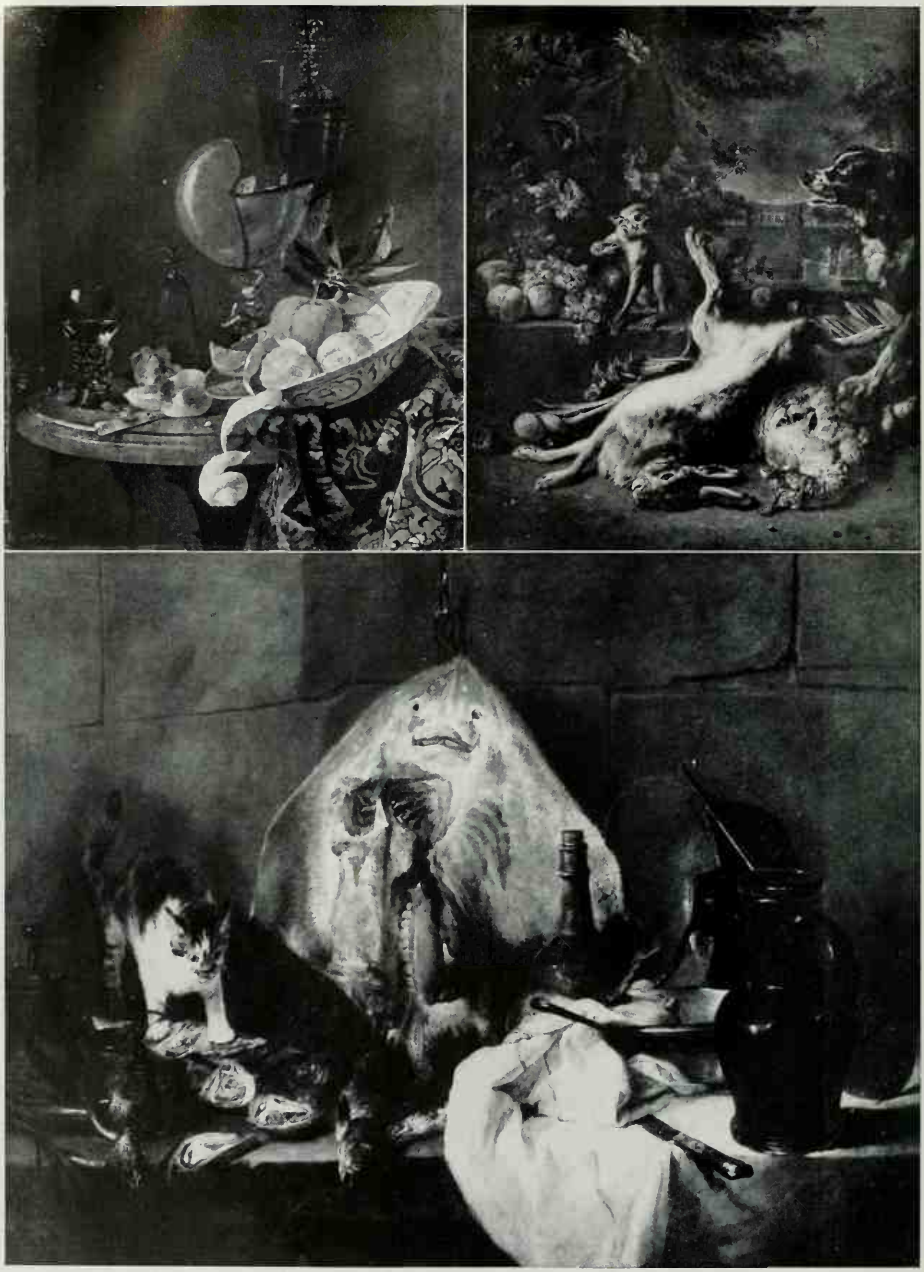
PHOTOGRAPHS, (TOP LEFT, BOTTOM) THE MANSELL COLLECTION, LONDON, (TOP RIGHT) BAYRISCHE STAATSGEMÄLDESAMMLUNGEN, MUNICH

Top left: Mosaic fragment, 1st–2nd century A.D. In the Lateran museum, Rome, It.

Top right: "Perdrix Bird" by Jacopo de Barbari (c. 1440–c. 1516). Bavarian State collection, Munich, Ger.

Bottom: "Fiori e Frutta" by Michelangelo da Caravaggio (1573–1610). In the Ambrosiana gallery, Milan, It.

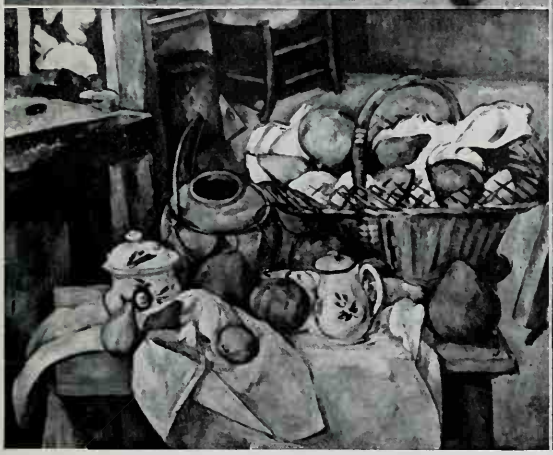
STILL-LIFE PAINTING



BY COURTESY OF (TOP RIGHT) THE RIJKSMUSEUM, PHOTOGRAPHS, (TOP LEFT) HANFSTAENGL, (BOTTOM) THE MANSELL COLLECTION, LONDON

Top left: "Still Life" by Willem Kalf (1622–93)
Top right: "Fruit and Game" by Jan Weenix (1640–1719). In the Rijksmuseum, Amsterdam, Neth.

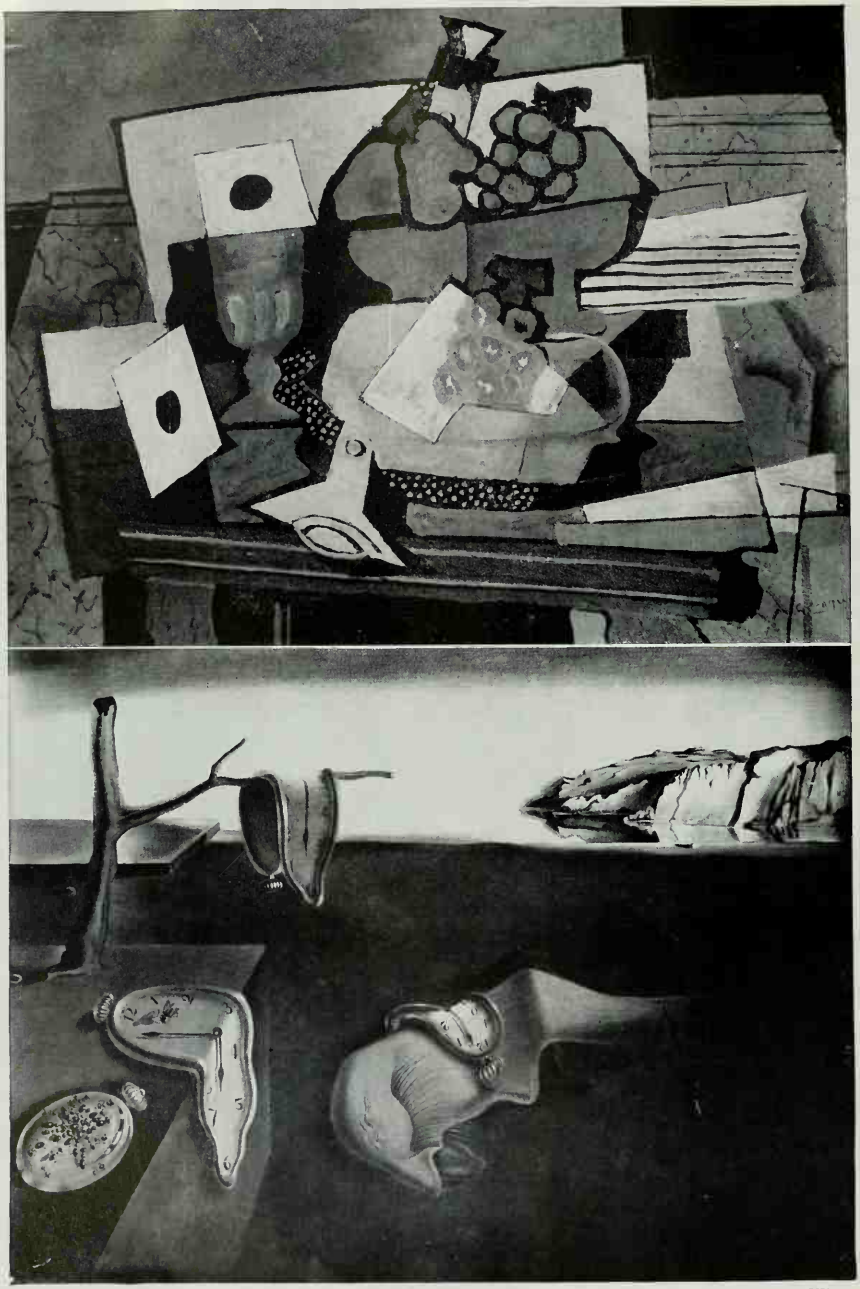
Bottom: "The Skate" by J. S. Chardin (1699–1779). In the Louvre, Paris



BY COURTESY OF (BOTTOM RIGHT) MUSÉE DES BEAUX-ARTS, BASLE; PHOTOGRAPHS, (TOP LEFT) CAISSE NATIONALE DES MONUMENTS HISTORIQUES, PARIS, (TOP RIGHT) GEMEENTE MUSEA, AMSTERDAM, (CENTRE LEFT) KUNSTHAUS, ZÜRICH, (BOTTOM LEFT) THE MANSELL COLLECTION, LONDON

Top left: "Fruits" by Édouard Manet (1832–83). The Louvre, Paris
Top right: "Gauguin's Chair" by Vincent van Gogh (1853–90). Stedelijk museum, Amsterdam, Neth.
Centre left: "The Trout" by Gustave Courbet (1819–77). The Kunsthaus, Zürich, Switz.
Bottom left: "Still-Life With Ginger Jar" by Paul Cézanne (1839–1906). The Louvre, Paris
Bottom right: "Compotier aux Fruits et Pains sur une Table" by Pablo Picasso (1881–). Musée des Beaux-Arts, Basle, Switz.

STILL-LIFE PAINTING



BY COURTESY OF (TOP) THE PHILLIPS COLLECTION, WASHINGTON, (BOTTOM) COLLECTION MUSEUM OF MODERN ART; PHOTOGRAPHS, (TOP) J. H. SCHAEFER & SON, (BOTTOM) SOICHI SUNAMI

Top: "Still Life With Grapes" by Georges Braque (1882–1953). The Phillips collection, Washington, D.C.

Bottom: "The Persistence of Memory" by Salvador Dali (1904–). The Museum of Modern Art, New York city

reduced to their basic shapes, split open, flattened and reorganized into a complex, predominantly two-dimensional composition. The images made by the profile, top, bottom and oblique angles of one object might be superimposed upon each other in one painting. This is the so-called analytical Cubism. The colours were restricted to greens, grays and browns. Synthetic Cubism, or collage, came next, and pieces of material, newspapers, etc., were stuck to the board or canvas, and painted into the picture. These are gayer in colour, and often made rather witty visual paraphrases. Braque's "Still-Life" (Tate) of 1925, although still using Cubist formulas, is slightly more naturalistic. Although he worked in a more rigorous abstract style during the 1930s, unlike Picasso he did not use still life to convey violent changes of mood, such as found in the latter's somewhat disquieting "Tête de Mort au Poireau" (1945).

Contemporary with the Cubists were Henri Matisse, André Derain and Maurice de Vlaminck (*les Fauves,* "the Wild Beasts"), whose work is characterized by radiant colour, great decorative skill and *joie de vivre.* In still life, unusual viewpoints are sometimes chosen, partly for the variety of pattern they afford, and never for the purpose of deliberate distortion. Édouard Vuillard and Pierre Bonnard (*les Nabis,* "the Prophets"), less formalized in style than Matisse, were yet painters of extreme sensitivity and taste.

The Dadaists, and later the Surrealists, stimulated by modern psychoanalytical research, exploited still life in an unusual way. Attempting to explore and record something of the secret working of the subconscious, partially revealed in dreams, they juxtaposed in their paintings animate and inanimate objects which are quite unrelated to each other in normal circumstances. The disconcerting, haunting effect sometimes thus produced is intended as a pictorial approximation to this other (dream) reality. Salvador Dali's "Persistence of Memory," with its melting watches set on a rocky foreshore, is a typical example.

No British artist of note (excepting William Etty), had seriously attempted still life until W. R. Sickert, with the New English Art club, introduced a taste for the genre from France in the 1890s. Matthew Smith, Ben Nicholson, Edward Wadsworth and William Scott are representative modern English exponents.

Still-life painting in the United States stems primarily from the Dutch realist *trompe l'oeil* tradition best exemplified by the work of the Peale family in the early 19th century and continued by William M. Harnett (1848–92). Harnett's reputation was established during the 1880s with works such as "Emblems of Peace" (1890, Springfield [Mass.] museum) in which he depicted commonplace objects with meticulous precision and deceptive realism. Of Harnett's several followers John F. Peto (1854–1907) was the most distinguished, his "Still Life With Lanterns" (Brooklyn [N.Y.] museum) might easily be confused in style with Harnett's "The Old Violin." For nearly 40 years Harnett's work lay forgotten until exhibited in 1939, when American Surrealists soon claimed him as a forerunner, reading into his sly whimsy a more profound psychopathic preoccupation than he could originally have intended. Surrealist overtones are more obvious in the work of Walter Tandy Murch (1907–), who owes much to Harnett but paints more loosely and introduces a stronger feeling for atmosphere. Murch's "The Circle" (*c.* 1948, Brooklyn museum) represents an old gramophone seen cornerwise with a dead beetle pinned inconspicuously to an odd corallike object on the turntable. The paintings of Ivan Le Lorraine Albright (1897–) are very different; he shows an almost overpowering, morbid interest in the minutiae of texture, and in "Self Portrait" (1935, Earle Ludgin collection, U.S.) the flesh seems about to dissolve. Bernard Perlin's "The Jacket" (1951, Whitney museum) represents yet another kind of careful, factual realism.

Bibliography.—H. Furst, *The Art of Still-Life Painting* (1927); M. J. Friedländer, *Landscape, Portrait, Still-Life* (1949); C. Sterling, *La Nature Morte de l'Antiquité à nos jours* (1955). (D. L. Fr.)

STILLMAN, WILLIAM JAMES (1828–1901), U.S. painter and journalist, was born at Schenectady, N.Y., June 1, 1828, and graduated from Union college, Schenectady, in 1848. He studied art under Frederick E. Church and early in 1850 went to England, where he fell under the influence of Rossetti and Millais. He studied art in Paris, returned to the United States and devoted himself to landscape painting on upper Saranac lake in the Adirondacks and in New York city. There also he started the *Crayon;* when it failed for want of funds, Stillman moved to Cambridge, Mass. He returned to England, and afterward painted with Ruskin in Switzerland. He was in Normandy in 1861 when the American Civil War broke out and was appointed U.S. consul in Rome; he resigned in 1865, but immediately afterward was appointed to Crete. When in London he lived with D. G. Rossetti. When the insurrection of 1875 broke out in Hercegovina he went there as a correspondent of the *Times* (London), and his letters from the Balkans aroused so much interest that the British government was induced to lend its countenance to Montenegrin aspirations. In 1877–83 he served as the correspondent of the *Times* at Athens; in 1886–98 at Rome. After his retirement he lived in Surrey, where he died on July 6, 1901. He wrote, among other books, *The Cretan Insurrection, 1866–1868* (1874), *On the Track of Ulysses* (1888), *Billy and Hans* (1897) and *Francesco Crispi* (1899). His *Autobiography of a Journalist* was published in 1901.

STILLWATER, a city of north central Oklahoma, U.S., 65 mi. N. of Oklahoma City; the county seat of Payne county and the seat of Oklahoma State university. The first recorded mention of Stillwater was on Dec. 25, 1884, when an army officer reported that "about 200 men, armed with double-barreled shot guns and Winchesters," had located there on lands not officially opened to settlement. This was the last serious foray into the Indian territory of "boomers," persons from Kansas who for five years had made raids into the territory in an effort to induce the federal government to open the lands; a detachment of 350 cavalrymen forced their surrender in Jan. 1885. A permanent settlement was made in April 1889 in the first great Oklahoma land run. Stillwater was incorporated as a village in 1890 and as a town in 1891; the name is from Stillwater creek, known by that name before any settlement was made. The city gave $10,000 and 200 ac. of land to secure the Oklahoma Agricultural and Mechanical college, which was opened in the Congregational Church on Dec. 14, 1891. It was renamed Oklahoma State university in 1957. The plant, including a 500,000-volume library, was largely a product of the administration of Pres. Henry G. Bennett, 1928–51. Industrial activity in the Stillwater region includes dairying, beef cattle, wheat and oil. For comparative population figures *see* table in Oklahoma: *Population.*

See B. B. Chapman, *The Founding of Stillwater* (1948). (B. B. C.)

STILO PRAECONINUS, LUCIUS AELIUS (*c.* 150–70 B.C.), is the earliest known Roman philologist. He belonged to the equestrian order, was a Stoic by persuasion, and was a friend of Q. Caecilius Metellus Numidicus, the distinguished general and censor, whom he voluntarily accompanied into temporary political exile. At Rome he engaged in teaching, in the writing of speeches for others, and in literary, antiquarian and philological work. His most famous pupils were Varro and Cicero. Only a few fragments of his works remain, in quotations by Varro, Verrius Flaccus, Pliny the Elder and Gellius. He wrote commentaries on the *Carmina Saliorum* (ritual hymns of the Salian priests) and (probably) on the language of the Twelve Tables, made a critical edition of the writings of Metellus Numidicus and perhaps of some others, and wrote a treatise *De proloquiis* ("On Sentences" or "Propositions"), probably under the influence of Chrysippus' work *Peri axiomaton* ("Concerning Axioms"). He investigated the authenticity of the Plautine comedies, of which he recognized 25 as genuine, four more than allowed by Varro.

See J. E. Sandys, *History of Classical Scholarship,* vol. i, 3rd ed. (1921); *Teuffel's History of Roman Literature,* in Eng. trans. by G. C. W. Warr (1900). (R. H. Rs.)

STILT, a wading bird closely related to the avocet (*q.v.*), with which it forms the family Recurvirostridae. The stilt is remarkable for the extreme length of its slender legs. The bill is long and straight; the legs red; the wings and back black; and the head, neck and lower parts white. Its food consists of small arthropods which it obtains by wading in shallow water. The

black-necked stilt, *Himantopus mexicanus,* is found in western United States and also in the southeast. Stilts range over Europe and Asia, not breeding northward of the Danube valley. The nest is a slight structure on the margin of a pool or lake; it contains four eggs which resemble those of the oyster catcher (*q.v.*). Other species occur in Hawaii, South America, Australia and New Zealand.

ALLAN D. CRUICKSHANK FROM NATIONAL AUDUBON SOCIETY

BLACK-NECKED AMERICAN STILT (HIMANTOPUS MEXICANUS), AT NEST

STILTS, poles with footrests fixed at a certain distance above the ground. They were originally designed for use in crossing rivers and marshes. As a means of amusement stilts have been used by all peoples of all ages, as well as by the inhabitants of marshy or flooded districts. The city of Namur in Belgium, which formerly suffered from the overflowing of the Sambre and Meuse rivers, has been celebrated for its stilt walkers for many centuries. Not only the townspeople but also the soldiers used stilts. The governor of Namur having promised the archduke Albert (about 1600) a company of soldiers that should neither ride nor walk, sent a detachment on stilts, which so pleased the archduke that he conferred upon the city perpetual exemption from the beer tax—no small privilege.

The home of stilt walking in modern times is the *département* of Landes in Gascony, where, because of the impermeability of the subsoil, all low-lying districts are converted into marshes. Stilts used by children are very long, the upper half being held under the arms; they are not strapped to the leg.

STILWELL, JOSEPH WARREN (1883–1946), U.S. army officer, was born at Palatka, Fla., March 19, 1883. In 1904 he graduated from the U.S. Military academy. Stilwell's work with the 4th corps in combat intelligence during World War I won him the distinguished service medal. After the war he exploited his talent for languages, making Chinese a specialty. Three tours of duty in China followed, during which his abilities as improviser, tactician and trainer impressed Gen. George C. Marshall, U.S. army chief of staff in World War II. When the United States entered that war in 1941, the war department, fearing that China might make a separate peace, ordered Stilwell, with the concurrence of China, to improve the combat efficiency of the Chinese army, to become chief of an international staff the Chinese then proposed organizing and to command all U.S. army forces in China, Burma and India.

In 1942 Stilwell fought a losing campaign to preserve the Chinese supply line through Burma, the famous Burma road (*q.v.*). Thereafter he urged China to revitalize its army and break the Japanese blockade by retaking Burma, but the Chinese preferred to stress U.S. air power. The Japanese overran U.S. air bases in east China in 1944, as Stilwell was advancing deeply into Burma. The U.S. then proposed that Stilwell be placed in command of all Chinese forces, but Chiang Kai-shek and the Chinese central government, fearing he might upset the domestic balance of power, forced his recall.

At his death, which took place at San Francisco, Calif., on Oct. 12, 1946, Stilwell commanded the 6th army. *The Stilwell Papers* were published in 1948.

BIBLIOGRAPHY.—Claire L. Chennault, *Way of a Fighter* (1949); Charles F. Romanus and Riley Sunderland, *Stilwell's Mission to China* (1953) and *Stilwell's Command Problems* (1956). (R. Sb.)

STIMSON, HENRY LEWIS (1867–1950), U.S. lawyer and statesman, was born in New York city on Sept. 21, 1867, the son of a successful and socially prominent banker and physician. He was educated at Phillips academy (Andover, Mass.), at Yale university (A.B., 1888) and Harvard university (A.M., 1889) and at the Harvard law school. Admitted to the New York bar in 1891, he joined the law firm of Elihu Root two years later.

From 1906 to 1909 he was U.S. attorney for the southern district of New York state; he distinguished himself especially for his devotion to the "trust-busting" policies of his friend and mentor, Pres. Theodore Roosevelt. At Roosevelt's request, he ran in 1910 as the Republican candidate for governor of New York, but was defeated. From 1911 to 1913 he was secretary of war in the cabinet of Pres. William Howard Taft. During World War I he served briefly as colonel of the 31st field artillery in France, and after the war he opposed Pres. Woodrow Wilson's plans for U.S. membership in the League of Nations. Stimson prospered as a lawyer during the 1920s, made some shrewd investments, and became, as he said, a "rich man." After 1927 his law partnership bore the name of Winthrop, Stimson, Putnam & Roberts.

In that year he was recalled to public life by Pres. Calvin Coolidge, who sent him as a special commissioner to Nicaragua to mediate between Pres. Adolfo Díaz and the rebel leader José María Moncada. He arranged an armistice on the understanding that the U.S. should supervise the election of 1928. Appointed governor general of the Philippine Islands in Dec. 1927, he opposed the Philippine independence movement and pursued a program of economic development of the islands with a view to converting them into "self-governing possessions or colonies whose citizens did not participate in our citizenship."

From 1929 to 1933 Stimson served as secretary of state in Pres. Herbert Hoover's cabinet. He led the U.S. delegation to the London naval conference in 1930. When, in 1931, the Japanese began military operations in Manchuria, he resisted efforts of the Chinese government and the League of Nations to obtain U.S. support against Japan. As the Japanese proceeded to overrun Manchuria, however, he took a stronger and stronger stand against Japanese aggression. He sent to Japan and China identic notes (Jan. 7, 1932) stating that the U.S. did not intend to recognize as legally valid any situation, treaty or agreement impairing the treaty rights of the U.S. or brought about by means contrary to the Pact of Paris. This nonrecognition policy afterward became known as the Stimson doctrine. In Jan. 1933 he was very active in conferences on the international situation with President Hoover and president-elect Roosevelt, and advocated close co-operation between the U.S. and the League in censuring and punishing Japan. Later, as an elder statesman, he urged co-operative international action against Italian and German as well as Japanese aggressions.

In 1940 Stimson, though a lifelong Republican, was appointed secretary of war by Pres. Franklin D. Roosevelt. In this capacity he guided the expansion and training of the U.S. army throughout World War II. He also acted as President Roosevelt's and later as Pres. Harry S. Truman's chief adviser on atomic policy. To President Truman he recommended that atom bombs be dropped on Japanese cities that had military importance, and he afterward justified the bombing of Hiroshima and Nagasaki on humanitarian grounds. He wrote that use of the bomb accelerated the surrender of Japan and thus saved many more lives than it cost. He left office in Sept. 1945 and died Oct. 20, 1950, in Huntington, N.Y.

ROCHE

STINKHORN (PHALLUS IMPUDICUS), A BASIDIOMYCETEOUS FUNGUS

In 1948 Stimson, aided by McGeorge Bundy, published an autobiographical work entitled *On Active Service in Peace and War.* Earlier books by Stimson appeared as follows: *American Policy in Nicaragua* (1927), *The United States and Other American Republics* (1931), *Pact of Paris* (1932) and *Far Eastern Crisis* (1936). (R. N. Cr.)

STINKHORN, the popular name of a group of fungi belonging to the Basidiomycetes, so called because of their carrion-like odour. The young plant (egg) resembles a small puffball but soon splits open at the apex

when a stalk emerges with a conical, slimy, evil-smelling head on which the spores are borne. The spores are dispersed by insects attracted by the odour. *See* FUNGI: *Basidiomycetes (Club Fungi)*.

STINKWOOD, a South African tree, known botanically as *Ocotea bullata,* and a member of the family Lauraceae. Other names for it are Cape walnut, stinkhout, Cape laurel and laurel wood. It derives its name from its strong and unpleasant smell when freshly felled. It is used for building in South Africa and is said to be a substitute for teak and equally durable. The wood is dark walnut or reddish brown to black with a yellow sapwood, and the grain extremely fine, close, dense and smooth.

STINNES, HUGO (1870–1924), German industrialist, was born in Mülheim, Feb. 12, 1870. He established a coal-dealing business in 1892 and subsequently expanded his commercial interests to include the ownership of coal mines, barges, river steamers and seafaring vessels. During World War I he commanded a large share of the industrial work required by German authorities. In 1920 he entered the *Reichstag* and a year later was reported to be contemplating the formation of a supertrust to control German industry. After his death in Berlin, April 10, 1924, the business, conducted by his son, rapidly shrank. The trust was investigated in 1925 and bit by bit the great organization was liquidated. (W. H. D.)

STIPPLE AND CRAYON ENGRAVING, two closely related print-making processes that once enjoyed great popularity but became almost obsolete with the advent of lithography in the 19th century. The aim of stipple engraving was to produce tone by a number of dots or flicks, whereas that of crayon engraving was to render the lines of a chalk drawing rather than tonal surfaces. Both were produced on copperplates. In stipple engraving the plate was covered with the usual etching ground, and the outlines and principal shadows were indicated by dotting this ground with the etching needle or a small-toothed wheel called the roulette. The plate was then bitten with acid and the ground removed. The composition was finished by dotting or stippling with the point of the stipple-graver, an instrument especially curved for this purpose. There is a great variety of roulettes, and those used in crayon engraving are made to reproduce the texture of chalk and pencil. With this method, as in stipple engraving, the ground is removed and further elaboration and completion are accomplished by working directly on the plate. Impressions are obtained in both media by putting the plates through a press. The ancestor of these two processes was the "dotted manner" (*manière criblée*) of the 15th century and the dotted indication of lines and shadows found in the engravings of Giulio Campagnola (*c.* 1482–*c.* 1514), Ottavio Leoni (*c.* 1576–*c.* 1628) and Marcello Fogolino (active *c.* 1520–40), but these effects were produced by short flicks of the engraver's burin and did not involve biting the plate with acid. Similar to this was the dotted work employed by goldsmith engravers in Germany and the Netherlands at the end of the 16th and the beginning of the 17th century.

In the 18th century when the crayon and pastel drawings of François Boucher and others were popular in France, the aim of engravers was to reproduce their texture by means of the copperplate. The invention of crayon engraving was claimed separately by Jean Charles François, Gilles Demarteau, who used his own version of it with artistic feeling, and Louis Marin Bonnet. It seems François was the first to work with this method although Bonnet is credited with what is known as the "pastel" manner, a method of printing a crayon engraving in various colours from several plates and using a separate plate for each colour. François taught the process of crayon engraving and the application of colour in printing to William Wynn Ryland (1732–1783), a young English line engraver who had been studying in Paris and who, when funds were lacking after his return to London, bethought him of the new manner of engraving he had learned. Ryland collaborated with Francesco Bartolozzi (1727–1815) and together they modified and developed it as stipple engraving.

The new method proved to be easy, simple and rapid. The outline was etched in a series of dots, and all the shadows were put in with larger or closer dots or tiny groups of dots. When all the biting was over the ground was removed and the finishing was done with drypoint and the stipple-graver. Then the printing was done with black, red or several coloured inks. A rag stump, or *poupée,* was used and the plate freshly cleaned for each impression. Beginning by translating the pseudoclassical designs of Angelica Kauffmann and Cipriani, which had acquired much popularity, Ryland and Bartolozzi found the new method exceedingly profitable. To the ready hand of Bartolozzi it came as a charming gift, with the facilities of the medium linking themselves to his caressing sense of beauty. All the painters of the day were anxious to share in the profits so readily made by the stipple prints, and few native engravers could resist the attraction of the new method. It also brought Italian and French engravers to learn it and work for the English market. There was, up to the war with revolutionary France in 1793, a great trade in English prints on the continent, and Bonnet and other French engravers by adopting the stipple method tried to get a share in this trade by issuing prints with titles in English, often misspelled. The line engraver, Sir Robert Strange, launched an indignant attack against stipple and even denied its claims as engraving at all, though there was no denying its efficacy as a medium for colour-printing. The white of the paper showing beneath the tinted dots afforded a peculiarly luminous quality and thus gave it a superiority over the coloured mezzotint. The popularity of stipple continued as long as the special subjects for which it was used remained in fashion. It waned with the advent of colour lithography and is no longer practised.

BIBLIOGRAPHY.—Julia Frankau, *Eighteenth Century Colour-Prints,* 2nd ed. (1906); Frank Weitenkampf, *American Graphic Art* (1912); A. M. Hind, *Bartolozzi and Other Stipple Engravers . . .* (1912); Malcolm C. Salaman, *Eighteenth Century French Colour-Prints* (1913); Campbell Dodgson, *Old French Colour-Prints* (1924); A. M. Hind, *The Processes and Schools of Engraving* (1952). (H. Es.)

STIRES, ERNEST MILMORE (1866–1951), bishop of Long Island, was born in Norfolk, Va., on May 20, 1866. He received the degree of B. Litt. at the University of Virginia in 1888, and from 1889 to 1891 continued his studies at the Episcopal Theological seminary of Virginia, receiving the degree of D.D. at Trinity college in 1901.

The degree of LL.D. was conferred upon him at Kenyon college in 1903. He received his first appointment as rector at West Point in 1891, and in 1893 went to the Church of the Good Shepherd in Augusta, Ga. In the same year he was called to Grace church, Chicago, where he remained until 1901. He then accepted a call to St. Thomas's church, New York. He was consecrated bishop of Long Island on Nov. 24, 1925, and he served in that capacity until his retirement in 1942. He died in Palm Beach, Fla., on Feb. 12, 1951.

STIRLING, WILLIAM ALEXANDER, EARL OF (*c.* 1567–1640), Scottish poet and statesman, generally known as William Alexander, was born at Menstrie house, near Stirling, about 1567, and probably educated at Stirling grammar school. He may also have been to the universities of Glasgow and Leiden. He followed James VI to England, where he became one of the gentlemen-extraordinary of Prince Henry's chamber. For the prince he wrote his *Paraenesis to the Prince . . .* (1604), a poem in eight-lined stanzas on the theme of princely duty. He was knighted in 1609; in 1614 he was appointed master of requests, and in July 1615 to a seat in the Scottish privy council. In 1613 he began a correspondence with Drummond of Hawthornden. In 1621 he received from James I enormous grants of land in America, including Nova Scotia and New Brunswick, and was appointed lieutenant governor of the new colony which, however, was ceded to the French in 1632. (*See* NOVA SCOTIA: *History.*) From 1626 until his death he was the king's secretary for Scotland, and in 1630 was created Viscount Stirling and Lord Alexander of Tullibody.

In 1631 he edited and published a translation of the Psalms said to be by James I. In 1639 he became earl of Donovan. He died on Feb. 12, 1640, in London.

Alexander's poetical miscellanies and minor verse include *Aurora,* containing the first fancies of the author's youth (1604); *An Elegie on the Death of Prince Henrie,* and shorter pieces. He also wrote four tragedies, *Darius* (1603); *Croesus* (1604); *The*

Alexandrean (1605), and *Julius Caesar* (1607); the first two of which were published together in 1604 as the *Monarchicke Tragedies;* they are didactic poems or dialogues rather than plays but they contain some fine soliloquies. Of Alexander's heroic poem *Jonathan* only the first book was written. *Domesday*, or *The Great Day of the Lord's Judgment* (1614), is a dreary production, in 12 books or "hours," in eight-lined stanzas. A collected edition of his work appeared in 1637, with the title *Recreations with the Muses* (folio), but did not include *Aurora* and the *Elegie*. A complete modern reprint, *The Poetical Works*, etc., was published in 3 vol. (Glasgow, 1870).

His *Encouragement to Colonies* was edited for the Bannatyne Club by David Laing (1867), and by Edmund F. Slafter, in *Sir W. Alexander and Amer. Colonization* (Prince Society, Boston, Mass., 1865). *See* also E. F. Slafter, *The Copper Coinage of the Earl of Stirling, 1632* (1874); *The Earl of Stirling's Register of Royal Letters Relative to the Affairs of Scotland and Nova Scotia from 1615–1635* (edit. C. Rogers, with biographical introduction 1884–85); C. Rogers, *Memorials of the Earl of Stirling* (1877); the introduction to the *Works* (1870) referred to above; the *Register of the Privy Council of Scotland, passim;* and the bibliography for DRUMMOND, WILLIAM, of Hawthornden.

STIRLING, WILLIAM ALEXANDER, titular EARL OF (1726–1783), American soldier, was born in New York city. He served first as commissary and then as aide-de-camp to Governor William Shirley at the beginning of the French and Indian War, and in 1756 he accompanied Shirley to England, where he prosecuted his claim to the earldom of Stirling. In 1759 an Edinburgh jury declared him to be the nearest heir to the last earl of Stirling, and in 1761 he returned to America and assumed the title of Lord Stirling by which he was afterwards known in America. He became a member of the New Jersey provincial council and surveyor general of the colony. In 1775 he sided with the patriot cause and was appointed colonel of a regiment in New Jersey. In March he became brigadier general, and for a time was in command at New York, and supervised the fortification of the city and harbour. At the battle of Long Island he was taken prisoner, but was soon afterwards exchanged, and in Feb. 1777 became a major general. He participated in the battles of Trenton, Princeton, Brandywine and Germantown, and especially distinguished himself at Monmouth. He took an active part in exposing the Conway Cabal, presided over the court-martial of General Charles Lee, and enjoyed the confidence of Washington to an unusual degree. In Oct. 1781 he took command of the northern department at Albany to check an expected invasion from Canada. He died at Albany on Jan. 15, 1783. He was a member of the board of governors of King's college (now Columbia university) and was devoted to the study of mathematics and astronomy.

See W. A. Duer, "Life of William Alexander, Earl of Stirling," in Vol. ii of the *Collections of the New Jersey Hist. Soc.* (1847).

STIRLING, a royal, large and parliamentary burgh, river port and the county town of Stirlingshire, Scot. Pop. (1961) 27,553. It is finely situated on the right bank of the Forth, 37 mi. N.N.W. of Edinburgh and 27 mi. N.N.E. of Glasgow by road. The old town and castle occupy the slopes of a dolerite hill (420 ft. above the sea) terminating on the north and west in a precipice. The modern town was laid out on the level ground at the base, especially toward the south. Remains of a town wall exist at the southern end of the Back walk. Formerly there were two main entrances—the South port and the "auld brig" over the Forth to the north, a high-pitched structure of four arches, now used only by foot passengers. It dates from the end of the 14th century and was once literally "the key to the Highlands." Just below it is the bridge erected in 1829 from designs by Robert Stevenson, and below this again the railway viaduct. The castle crowning the hill is of unknown age, but from the time that Alexander I died within its walls in 1124 until the union of the crowns in 1603, it was intimately associated with the fortunes of the Scottish monarchs. It is approached from the esplanade, on which stands a colossal statue of Robert Bruce. The main gateway, built by James III, gives access to the lower and then to the upper square, on the south side of which stands the palace, begun by James V (1540) and completed by Mary of Guise. The east side of the quadrangle is occupied by the parliament house, a Gothic building of the time of James III, used as a barrack room and stores. On the north side of the square is the chapel royal, founded by Alexander I, rebuilt in the 15th century and again in 1594 by James VI (who was christened in it), and afterward converted into an armoury and finally a storeroom. The former banquet hall contains the regimental museum of the Argyll and Sutherland Highlanders. In the plain to the southwest were the King's gardens, now under grass, with an octagonal turf-covered mound called the King's Knott in the centre. Farther south lies the King's park. On a hill of lower elevation than the castle and separated from the esplanade by a depression styled the Valley—the tilting ground of former times—a cemetery has been laid out where are the Virgin Martyrs' memorial, in memory of Margaret Maclachlan and Agnes Wilson, who were drowned by the rising tide in Wigtown bay for their fidelity to the Covenant (1685), a large pyramid to the memory of the Covenanters, and the Ladies' rock, from which ladies viewed the jousts in the Valley.

Adjoining the cemetery on the south is the Church of the Holy Rude (Rood), portions of which may have formed part of the first church founded by David I. Following the Reformation it was divided into two churches, but was restored to its original condition in 1936–40. The choir (the old East church) was added in 1494 by James IV, and the apse a few years later. At the west end is a battlemented square tower. The nave (the former West church) is a transition between Romanesque and Gothic, with pointed windows. The main entrance to both churches was by the south transept, with its crow-stepped Gothic gable. The choir is in the Decorated and Perpendicular styles. Within its walls Mary Queen of Scots was crowned in 1543, when nine months old, and in the same year the earl of Arran, regent of Scotland, abjured Protestantism; in 1544 an assembly of nobles appointed Mary of Guise queen regent; on July 29, 1567, James VI (then 12 months old) was crowned, John Knox preaching the sermon, and in Aug. 1571 and June 1578 the general assembly of the Church of Scotland met. James Guthrie (1612–61), the martyr, and Ebenezer Erskine (1680–1754), founder of the Scottish Secession church, were two of the most distinguished ministers. To the southwest of the church is Cowane's hospital, founded in 1639 by John Cowane, dean of guild, and now used as a guildhall. Near the principal entrance to the esplanade stands Argyll's lodging, erected about 1630 by the 1st earl of Stirling, Sir William Alexander, the founder of Nova Scotia. On his death in 1640 it passed to the 1st marquess of Argyll and later was made into a military hospital. It is a fine example of a baronial town house. Broad street contains Mar's Work, the unfinished palace begun by John Erskine, 1st earl of Mar, about 1570, according to tradition, out of the stones of Cambuskenneth abbey; the old town house, erected in 1701 to replace that in which John Hamilton, the last Roman Catholic archbishop of St. Andrews, was hanged for alleged complicity in the murders of Darnley and the regent Moray; the town cross, restored in 1891, and the house which was, as a mural tablet says, the "nursery of James VI and his son Prince Henry." The Smith Arts institute in Dumbarton road, founded in 1873 by Thomas Stewart Smith, an artist, contains a picture gallery, museum and reading room.

The Abbey craig, across the Forth, an outlying spur of the Ochils, 1½ mi. N.E. of Stirling, is a thickly wooded hill on the top of which stands the National Wallace monument (1869), a baronial tower, 220 ft. high, with a "Hall of Heroes" containing busts of eminent Scotsmen. Cambuskenneth abbey, on the left bank of the Forth, about 1 mi. E.N.E. of Stirling, was founded by David I in 1147 for monks of the order of St. Augustine. Several Scots parliaments met within its walls. At the Reformation Mary Queen of Scots bestowed it on the 1st earl of Mar (1562), who is said to have used the stones for his palace in Stirling. All that remains of the abbey is the massive, four-storied tower, the west doorway and the foundations of some of the walls. The remains of James III and his queen, Margaret of Denmark, who were buried within the precincts, were discovered in 1864 and reinterred the next year under a tomb at the high altar at the expense of Queen Victoria.

Stirling was also known as Snowdoun, which became the official

title of the Scots heralds. The Romans probably had a station there. The Picts held the hill until ejected in 843 by Kenneth MacAlpine. In 1119 it was a royal burgh and under Alexander I was one of the Court of Four Burghs (superseded under James III by the Convention of Royal Burghs). In 1174 it was handed over to the English in security for the treaty of Falaise, being restored to the Scots by Richard I. The earliest known charter was that granted in 1226 by Alexander II who made the castle a royal residence. The fortress was repeatedly besieged during the wars of the Scottish Independence and fell to the English in 1296. In 1304 it fell with the town to Edward I, after William Wallace had taken it in 1297. The English held it for ten years, and it was in order to raise the Scottish siege in 1314 that Edward II risked the battle at Bannockburn, 2½ mi. S. Edward Baliol surrendered it in 1334 in terms of his compact with Edward III, but the Scots regained it in 1339. From this time until the collapse of Queen Mary's fortunes in 1568, Stirling almost shared with Edinburgh the rank and privileges of capital of the kingdom. It was the birthplace of James II in 1430 and probably of James III and James IV. In 1571 an attempt was made to surprise the castle by Mary's adherents, the regent Lennox being slain in the fray, and seven years later it was captured by James Douglas, 4th earl of Morton, after which a reconciliation took place between the Protestants and Roman Catholics. It was occupied in 1584 by the earls of Angus and Mar, the Protestant leaders, who, however, fled to England on the approach of the king. The next year they returned with a strong force and compelled James VI to open the gates, his personal safety having been guaranteed.

After the union of the crowns (1603) Stirling ceased to play a prominent part on the national stage. The privy council and court of session met in the town in 1637 because of the disturbed state of Edinburgh. In 1641 Charles I gave it its last governing charter, and four years later parliament was held in Stirling because of the plague in the capital, but the outbreak of the pest in Stirling caused the legislators to move to Perth. During the Great Rebellion the Covenanters held the town, to which the committees of church and state adjourned after Oliver Cromwell's victory at Dunbar (1650), but in August of the next year the castle was taken by General George Monk. In 1715 the 3rd duke of Argyll held it to prevent the passage of the Forth by the Jacobites, and in 1746 it was ineffectually besieged by Prince Charles Edward.

Stirling, with Falkirk and Grangemouth, returns one member to parliament.

The most important industries are associated with agriculture: engineering, farm and garden seeds, sawmills, bacon curing and sausage making. Rubber, asbestos and fibre insulating materials are made and there is floorcloth making, coopering and coachbuilding.

Coal mining employs many of the inhabitants. There is some shipping from the small harbour which is accessible only at high water.

STIRLING NUMBERS, in mathematics. In the year 1730 James Stirling, in his *Methodus Differentialis* introduced into analysis two sets of numbers which, because of their uses in various branches of analysis, their properties, and the methods used in their computation, have continued to attract the attention of mathematicians. They have been called "Stirling numbers of the first and second kind" in honour of their discoverer.

Definitions.—The Stirling numbers of order n of the first kind may be defined as the co-efficients in the expansion of

$$(1+x)(1+2x)\cdots(1+nx)=1+{}_nS_1x+{}_nS_2x^2+{}_nS_3x^3+\cdots$$

in ascending powers of x, while the Stirling numbers of the second kind are found in

$$\frac{1}{(1+x)(1+2x)\cdots(1+nx)}=1-{}_nT_1x+{}_nT_2x^2-{}_nT_3x^3+\cdots,$$

also in ascending powers of x.

This definition immediately leads to the following theorems:

(*a*) The Pth Stirling number of order n of the first kind is equal to the sum of the products of the first n integers taken P at a time without repetition.

(*b*) The Pth Stirling number of the second kind of order n is equal to the sum of the products of the first n integers taken P at a time with repetitions.

Stirling used the numbers now named after him as a tool for expressing x^n as a series of factorials.

$$x^2=x(x-1)+x,$$
$$x^3=x(x-1)(x-2)+3x(x-1)+x,$$
$$x^4=x(x-1)(x-2)(x-3)+6x(x-1)(x-2)+7x(x-1)+x,$$
$$x^5=x(x-1)(x-2)(x-3)(x-4)+10x(x-1)(x-2)(x-3)+25x(x-1)(x-2)+15x(x-1)+x.$$

The coefficients of the various factorials (1.1; 1.3.1; 1.6.7.1; 1.10.25.15.1) are the Stirling numbers of the second kind. For about 150 years mathematicians considered it important to be able to express algebraic expressions in form of sequences of factorials. The result was a rich literature consisting of memoirs dealing with the peculiarities of this method of expression.

In the year 1846 Weierstrass proved the general futility of this mode of notation, and since then the subject of "factorial notation" has slowly but surely been eliminated as a topic of discussion in mathematical literature. At present it is used only in certain problems of finite summation, in theories of interpolation, and as an example illustrating some of the uses of the Stirling numbers.

Applications.—George Boole in his treatise on the calculus of finite differences calls attention to the "system of numbers expressed by $\triangle^n o^m$ (differences of nothing)" without indicating whether he realized that these numbers were already used by Stirling. The formula given by Boole for $\triangle^n o^m$ may well be used for the calculation of the above mentioned coefficients T; *i.e.*, the Stirling numbers of the second kind,

$$\frac{1}{n!}\,{}_nT_{m-n}=n^m-\binom{n}{1}(n-1)^m+\binom{n}{2}(n-2)^m-\binom{n}{3}(n-3)^m+\cdots$$

The Stirling numbers of the second kind (T) also appear in the expansion of $(e^x-1)^p$. Thus,

$$\frac{1}{2!}(e^x-1)^2=\frac{x^2}{2!}+\frac{{}_2T_1x^3}{3!}+\frac{{}_2T_2x^4}{4!}+\cdots$$

$$\frac{1}{3!}(e^x-1)^3=\frac{x^3}{3!}+\frac{{}_3T_1}{4!}x^4+\frac{{}_3T_2x^5}{5!}+\cdots$$

$$\frac{1}{p!}(e^x-1)^p=\frac{x^p}{p!}+\frac{{}_pT_1}{(p+1)!}x^{p+1}+\frac{{}_pT_2}{(p+2)!}x^{p+2}+\cdots$$

Formulas.—Of the recurring formulas the most important are

$${}_{n+1}S_p={}_nS_p+n\cdot{}_nS_{p-1},$$
$${}_{n+1}T_p={}_nT_p+n\cdot{}_{n+1}T_{p-1}.$$

Of the independent formulas developed by various writers, none is suitable for practical purposes; but theoretically the following formulas which appear in a work by Ettinghausen (1826) are important:

$${}_nS_p=\Sigma\frac{(n+1)!}{a!\,b!\,c!\cdots 1^a\,2^b\,3^c\cdots}$$

for all integral values of $a,b,c\ldots$ satisfying the equations:

$$a+b+c+\cdots=p+1,$$
$$a+2b+3c+\cdots=n+1,$$

and

$${}_nT_p=\Sigma\frac{(n+p)!}{a!\,b!\,c!\cdots 1^a(2!)^b(3!)^c\cdots}$$

for all integers satisfying the equations

$$a+b+c+\cdots=p,$$
$$a+2b+3c+\cdots=n+p.$$

Relations between the S's and the T's are given by

$$ {}_nT_p = \begin{vmatrix} {}_nS_1 & 1 & \cdots & 0 \\ {}_nS_2 & {}_nS_1 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & 1 \\ {}_nS_p & {}_nS_{p-1} & \cdots & {}_nS_1 \end{vmatrix}, \text{ and } {}_nS_p = \begin{vmatrix} {}_nT_1 & 1 & \cdots & 0 \\ {}_nT_2 & {}_nT_1 & \cdots & 0 \\ \cdot & \cdot & \cdots & 0 \\ \cdot & \cdot & \cdots & 1 \\ {}_nT_p & {}_nT_{p-1} & \cdots & {}_nT_1 \end{vmatrix} $$

The second formula is evidently obtained from the first by interchanging the T's and the S's. It has also been observed (*Amer. Math. Monthly*, 1928) that the Stirling numbers of the first kind can be obtained by performing the algebraic divisions,

$$ \frac{2+x}{(1-x)^5}, \frac{6+8x+x^2}{(1-x)^7}, \frac{24+58x+22x^2+x^3}{(1-x)^9} \cdots $$

while the Stirling numbers of the second kind appear in the quotients of

$$ \frac{1+2x}{(1-x)^5}, \frac{1+8x+6x^2}{(1-x)^7}, \frac{1+22x+58x^2+24x^3}{(1-x)^9} \cdots $$

The Stirling numbers have many important relationships to the Bernoulli numbers (*q.v.*), Euler numbers (*q.v.*), and the tangential coefficients, and there are numerous formulas connecting them. In fact, there is a way of regarding the Bernoulli numbers as a species of Stirling numbers; that is $(e^x-1)^n$ always gives, when expanded, Stirling numbers; and when we let $n=-1$, we obtain a well-known expansion giving the Bernoulli numbers.

The consideration of $(e^x-1)^{-n}$ leads to what may be called ultra-Stirling numbers, a subject which, like the ultra-Bernoulli and ultra-Euler numbers has been very little studied.

BIBLIOGRAPHY.—Chr. Kramp, *Analyse des refractions astronomiques et terrestres* (Leipzig, 1799); v. Ettinghausen, *Die combinatorische Analysis* (Vienna, 1826); Schläfli, "Sur les coefficients du développement du product $(1+x)(1+2x)\ldots[1+(n-1)x]$ suivant les puissances ascendantes de x," *Journal für die reine und angewandte Mathematik* (1852); N. Nielsen, *Handbuch der Theorie der Gammafunktion* (Leipzig, 1906); N. Nielsen, *Traité élémentaire des nombres de Bernoulli* (1923); E. Netto, *Lehrbuch der Combinatonik* (Leipzig, 1927); J. F. Steffenson, *Interpolation* (1927). (J. GI.)

STIRLINGSHIRE, a midland county of Scotland, bounded north by Perthshire, northeast by Clackmannanshire and the Firth of Forth, southeast by Linlithgowshire, south by Lanarkshire and the detached part of Dunbartonshire and southwest and west by Dunbartonshire. Pop. (1961) 194,858. Area 450.5 sq.mi. In the northwest the Grampians end in Ben Lomond (3,192 ft.); the centre is occupied by the Lennox hills. The oldest rocks are the Dalradian schists in the northwest beyond a great fault which runs from near the southern end of Loch Lomond in a northeasterly direction across Scotland; it is the geographical boundary of the Highlands. On the southeastern side of the fault are the conglomerates and sandstones of the Lower Old Red Sandstone age, followed by the Upper Old Red series. Then follows the Carboniferous system occupying the rest of the county. The Coal Measures are developed in the southeast. Intrusive sheets of basalt have penetrated the Carboniferous rocks and are quarried for road metal; Abbey craig and Stirling Castle hill are formed of one of the more important of these intrusions. Much boulder clay covers the older rocks and an interesting blue marine clay is found beneath it in the Endrick valley. The Carse of Stirling is overlaid by the muds and sands of the 50-ft. raised beach; and traces of the 100-ft. beach are also found.

The chief river is the many-looped Forth, which forms most of the northern boundary. The other large streams are the Carron, rising in Campsie fells and flowing to the Forth at Grangemouth; the Endrick, rising in Fintry hills and emptying itself into Loch Lomond; the Kelvin, which, from its source in Kilsyth hills, flows southwestward to the Clyde at Glasgow after a run of 22 mi.; and the Avon, rising in the detached portion of Dunbartonshire, and flowing to the Forth.

The principal lochs include the greater part of the eastern waters of Loch Lomond, a small portion of the northern end of Loch Katrine and Loch Arklet, in the northwestern area, which, like Loch Katrine, provides part of the water supply of Glasgow. The Forth and Clyde canal crosses the southeastern corner of the county from Grangemouth to Castlecary.

History and Antiquities.—The wall of Antoninus, built by Lollius Urbicus in A.D. 142 and connecting the Forth and Clyde, passed through the southeast of the county, in which it is locally known as Graham's dyke. At Castlecary (where the Romans used coal in the fort) and Camelon many relics have been found. The Camelon causeway ran eastward from Castlecary and crossed the rampart at Camelon, then proceeded northward to Stirling and the Forth, where there was a station near the present bridge of Drip. Then it crossed the river to Keir and Dunblane in Perthshire. After the withdrawal of the Romans the native inhabitants gradually retired before the advance of the Saxons and Scots. By the time of Malcolm Canmore (d. 1093) the lowland area had become settled, but the highland tract remained a disturbed and disturbing region until the pacification following the Jacobite rising of 1745–46. The county played a conspicuous part in the struggle for Scottish independence, being particularly associated with many of the exploits of William Wallace and Robert Bruce. The three great battles of the independence were fought in the shire—Stirling Bridge (1297), Falkirk (1298), Bannockburn (1314). James III was stabbed to death in a cottage in the village of Milton after the battle of Sauchieburn (1488), but apart from the disastrous defeat of the Covenanters at Kilsyth (1645) and the transitory triumph which Prince Charles Edward won at Falkirk (1746), the history of the shire practically centres in that of the county town. (*See* STIRLING.)

Agriculture and Industries.—Because of varying rainfall and different soils, Stirlingshire has two types of agriculture. In the west, farms are largely used for dairying, and cropping is mainly for farm consumption. In the east, and particularly on the (alluvial) Carse in the northeast, a system of arable and stock feeding is practised. Wheat and beans grow well but the outstanding crop is timothy grass which gives heavy yields as hay or produces seed of very good quality. The pastures in this part are used to fatten cattle during summer.

Hardy sheep are grazed on the hills. Small holdings in the vicinity of Glasgow are used for fattening pigs on swill, and tomatoes are grown under glass in the west to supply the Glasgow market. Stirling is the market centre for a wide area and has large sales of lambs, sheep and store cattle.

Birches grow on the lower slopes of the mountains in Buchanan and Drymen, and oaks on the banks of Loch Lomond. Larch and Scots fir are the leading trees in modern plantations.

The coalfield of the southeast has been worked out and coal is chiefly produced in the central area near Bannockburn, Denny and Kilsyth. Fireclay is also obtained in large quantities and there are ironworks at Carron, Falkirk, Bonnybridge and Denny.

Nails are made at Lennoxtown and elsewhere; carpets are manufactured at Bannockburn; Grangemouth is one of the biggest centres for dyes and chemicals and has a modern oil refinery. Tanning, iron founding, papermaking, brewing and distilling are carried on and shipbuilding is done at Grangemouth, the chief port.

Population and Administration.—The population in 1951 of the administrative county was 123,030; Gaelic and English were spoken by 725 persons. Falkirk (pop., 1951, 37,535) and Stirling (the only royal burgh, 26,962) are large burghs, the small burghs being Grangemouth (15,432), Kilsyth (9,922), Denny and Dunipace (6,756) and Bridge of Allan (3,173).

The shire, with Clackmannanshire, sends three members to parliament. It forms a sheriffdom with the counties of Dunbarton and Clackmannan; there are resident sheriff-substitutes at Stirling and Falkirk.

STOA, in Greek architecture, a long open building with the roof supported by one or more rows of columns parallel to the back wall. A row of rooms may open from the rear of the colonnade; a second story is often added. Stoae surrounded the market places and later the sanctuaries and served for nearly all public and private business, commercial and legal, as well as for public promenade. Athens, Corinth, Delos and the Asia Minor cities offer good examples. (L. T. SE.)

STOAT, a common name for the small northern weasel *Mustela erminea,* especially when it displays its summer pelage of brown upper parts and whitish under parts. These slender, agile, blood-thirsty mammals measure 7½ to 13½ in., approximately 40% of which is tail. In colder climates the winter coat is white, except for the black tail-tip; the animal in this colour phase is popularly called ermine. In moderately cold climates the fur becomes only partially white. (The colour change is the result of molting and is an inherited hormonal response to the daily amount of light.)

Stoats occur from the arctic southward in eastern America to Pennsylvania and in the west to northern New Mexico. They also extend from the British Isles to Japan and as far south as Algeria. They are versatile as to habitat but are most abundant in thickets, woodland and damp, grassy, semi-timbered areas. Females and southern races are smaller than males and northern races, respectively. Their diet consists of small mammals, birds, eggs, frogs and occasional invertebrates. The prey is seized usually at the base of the skull, the sharp teeth penetrating the brain. Larger prey is seized by the throat and held with fierce tenacity. Stoats are in turn preyed upon by larger carnivores and birds of prey. Mating takes place in late spring or early summer with gestation prolonged as much as 10 months because of delayed implantation. The 4 to 13 blind young are born in a burrow. They reach old age in 7 to 8 years.

JOHN MARKHAM

STOAT OR ERMINE (MUSTELA ERMINEA) IN BROWN SUMMER COAT

The winter-taken pelts, prized for fineness and pure colour, are among the most valuable of commercial furs and are obtained mainly in northern Eurasia. (K. R. Kn.)

STOBAEUS, JOANNES, so called from his native place Stobi in Macedonia, the compiler of a valuable series of extracts from Greek authors. Of his life nothing is known, but he probably belongs to the latter half of the 5th century A.D.

The extracts were intended by Stobaeus for his son Septimius, and were preceded by a letter briefly explaining the purpose of the work and giving a summary of the contents. From this summary (preserved in Photius's *Bibliotheca*) we learn that Stobaeus divided his work into four books and two volumes. In most of our manuscripts the work is divided into three books, of which the first and second are generally called "Physical and Moral Extracts," and the third *Florilegium* or *Sermones.* As each of the four books is sometimes called *Anthologion,* it is probable that this name originally belonged to the entire work; the full title, as we know from Photius, was *Four Books of Extracts, Sayings and Precepts.*

The introduction to the whole work, treating of the value of philosophy and of philosophical sects, was lost with the exception of the concluding portion; the second book is little more than a fragment, and the third and fourth have been amalgamated by altering the original sections. The first book teaches physics by means of extracts. It is often untrustworthy; Stobaeus betrays a tendency to confound the dogmas of the early Ionic philosophers, and he occasionally confused Platonism with Pythagoreanism. For part of this book and much of the second book, he depended on the works of Aëtius, a peripatetic philosopher, and Didymus.

The third and fourth books, like the larger part of the second, are concerned with ethics; the third, with virtues and vices in pairs; the fourth deals with more general ethical and political subjects.

In all Stobaeus quotes more than 500 writers.

STOCK, FREDERICK AUGUST (1872–1942), U.S. conductor and director of the Chicago Symphony orchestra for 37 years, was born on Nov. 11, 1872, in Jülich, Ger., and studied violin and composition at the Cologne conservatory. Theodore Thomas, founder and conductor of the Chicago Symphony orchestra, discovered Stock when he was a violinist with the Cologne orchestra and invited him to Chicago to join the new orchestra in 1895. He made his debut there as a viola player and six years later became Thomas' assistant and associate conductor. Shortly after the Chicago symphony moved into its new quarters at Orchestra hall in 1905, Thomas died and Stock became his successor. Stock composed three overtures, two symphonies, a violin concerto, a 'cello concerto, a *Festival March* and *Hymn to Liberty.* He died in Chicago on Oct. 20, 1942.

STOCK, popular greenhouse and garden flowers, all but one of which belong to the genus *Matthiola* of the mustard family (Cruciferae, *q.v.*). These plants comprise single and double flowering forms in a variety of colours ranging from white through pink and yellowish to purple and scarlet. *M. incana,* also called queen or Brompton stock and gilliflower, the more popular species for both the florist and the gardener, is a perennial or biennial grown principally for late summer or autumn bloom in the open or under glass where flowering may occur at other seasons. The variety *annua,* known as ten-weeks or intermediate stock, is usually grown as an annual. *Matthiola bicornis,* evening or Grecian stock, is a low spreading annual grown for its small, fragrant, lilac flowers, which open at evening.

The Virginian or Malcolm stock (*Malcomia maritima*), a member of the same family, is not a regular florists' flower but is often grown as a garden plant for its small fragrant blossoms.

STOCK (Shares). The money invested by individuals in business corporations is divided into units called stock in the United States or shares in Great Britain. Such stock represents a permanent investment in the business. The investor receives a certificate as evidence of his ownership of a part of the corporation. The detailed contractual relationship between the corporation and its stockholders sets forth the division of the risk, income and control of the business. Stockholders have the right to share in the management of the corporation, to share proportionately in any dividends declared and in the assets of the business in case of liquidation, and to approve or disapprove proposed changes in the original contract. The stock certificate itself, which has been called both a negotiable and quasi-negotiable instrument, may be transferred from person to person. In Great Britain the discussion of the rights of shareholders is relatively simple because the Companies acts apply to the entire nation. In contrast, there are great diversities in the rights of stockholders in the United States because each state has its own corporation law.

Origins.—As business units grew in size and as a continuing existence beyond the lives of their owners became essential, new legal forms were devised. A distinct forerunner of the modern business corporation that can be identified by the mid-16th century was the joint-stock company. Both the London company, which founded the first permanent English settlement in the new world, and the East India company were joint-stock companies. The modern business corporation evolved during the latter part of the 17th century in both England and in the United States. The history of the attempts to define the rights of corporation shareholders may be traced for about 300 years. (*See* Corporation.)

By the middle of the 20th century virtually all large privately owned businesses and a large proportion of smaller businesses were organized as corporations and had issues of stock outstanding. The fact that the ownership of both large and small corporations is divided into shares makes any simple description of classes of stock and the rights of stockholders necessarily vague. Only by unreasonably stretching the legal fabric can the shares of the very large and very small businesses be included under the same cloak.

The Companies acts in Great Britain, the Securities act (1933) and Securities Exchange act (1934) of the United States, and the new corporation laws of the individual states have tended to give more publicity to corporate affairs for the protection of the shareholders of public companies. The rights of minority interests are also protected by law. The rights of the shareholders of private or closely held corporations may be quite different from those of public companies. The right of the shareholder to trans-

fer his shares is often restricted by agreement among the shareholders. Profits may be divided on the basis of private agreements. Shareholders' rights may become indistinguishable from those of partners. In the United States even one-man corporations are not uncommon. By the 1960s the laws governing the establishment of corporations in the United States and the rights of stockholders were beginning to recognize the differences between the problems of small businesses organized as corporations and large publicly owned businesses.

Dividends.—One of the basic rights of stockholders is to share in the profits of the corporation. These profits may be paid out to the investors in the form of cash dividends or retained in the business to help finance its growth. Stockholders have no right to receive dividends until they are declared by the board of directors. Dividends are usually declared as a certain amount per share by class of stockholder.

Cash dividends, with the exception of liquidating dividends, generally cannot be legally authorized if the effect of such a dividend payment would be to reduce the owners' equity in the business below their original contribution. However, such a limitation is of little importance to large corporations that have grown partly or wholly by "plowing profits back into the business." It is usually recognized that profits retained in the business in fact, if not legally, become a part of the permanent investment of the shareholders. Cash dividends generally cannot be paid if the corporation is insolvent or if the payment of the dividend would result in insolvency.

Dividends are sometimes paid in the form of additional common stock of the corporation rather than in cash. Such stock dividends, as they are called, give recognition to the fact that the profits retained in the business represent a permanent investment. Many rapidly growing, publicly held corporations have a policy of paying dividends partly in stock as well as in cash, thus conserving the firm's cash so that it may be used to help finance expansion. The stockholder may, if he wishes, sell the stock received as dividends.

Small, closely held, rapidly expanding businesses usually pay little if any dividends until such time as they prepare for their first public issue of stock. At that time the retained earnings are transferred to the common stock account and the number of shares is increased or split so that the market price of the new shares will be in the popular trading range when they are sold. Stock splits of more than 1,000–1 are not uncommon in these circumstances. In one unusual case, the shares of Biophysical Electronics, Inc. were split 62,500–1 when the firm sold common stock to the public for the first time in 1959.

The decision of a corporation to raise funds or finance by the use of equity shares rather than by borrowing is extremely complex. Some of the factors that influence the decision are: size and rate of growth of the business, distribution of shares of the business (whether closely held or publicly held), maintenance of voting control, stability of earnings, personal and corporate income tax rates, desire to avoid fixed contractual obligations for the payment of interest and principal on a debt, and availability and relative cost of debt and equity funds. *See* also Stock Exchange.

(R. M. S.)

UNITED STATES

As noted above, a share of stock is an ownership interest in a corporation: Each stockholder or shareholder has the fractional interest in the corporation that may be measured by comparing the number of shares he holds with the total number of the corporation's shares outstanding. This interest includes the right to share in the control of the organization's activities, to share in dividends as declared, to purchase additional shares when and if issued and to share in assets remaining after paying creditors upon termination of the corporation's existence. One or more of these rights, however, may be withheld or qualified by particular state laws or by the articles of incorporation.

As compared with partners in an unincorporated business, corporation stockholders have the advantage of limited liability but their freedom to withdraw assets from the firm is restricted. The restriction on owners' withdrawals creates the concept of "stated capital"—the portion of owners' interests which must not be withdrawn. The initial investments of stockholders usually exceed the amount designated in the articles of incorporation as stated capital. The legal term for this excess is "paid-in surplus" or "capital surplus." Profits accumulated as the corporation carries on its functions add to the owners' interests, while dividends paid by the corporation reduce the owners' interests in the corporate fund of assets as these assets are transferred to them. The excess of accumulated profits over dividends declared is called retained earnings or earned surplus. Thus, the owners' interests in corporate assets commonly arise from two sources—original investments and retained earnings—but the legal concept of stated capital requires a distinction between any part of the owners' original contributions that may be returned to them and the part that must be preserved.

Dividends and a rise in the price of the shares are the basic incentives for investing in an incorporated business by purchasing its stock. The payment of dividends entails the transfer of assets from the corporation to its stockholders, thus reducing both the total assets of the corporation and the owners' interests therein. The legal category of owners' interests that are reduced by dividends is partly regulated by law and partly at the discretion of the corporation's board of directors who have the responsibility for declaring dividends. Unless the corporation has covenanted with a specific group of creditors to restrict its dividends in some manner, the directors normally have the power to declare dividends which reduce retained earnings. Paid-in surplus is, in most states, available for reduction by dividends under certain circumstances. Stated capital, however, may not be reduced by dividends, although some states permit dividends not in excess of profits earned in the most recent fiscal period even if the stated capital of the corporation has been impaired by earlier losses. In such a case the earned surplus is negative and the total of the owners' interests is less than stated capital.

Classes of Stock.—Many corporations have only one class of stock, frequently referred to as capital stock or common stock in the United States or ordinary shares in Great Britain. Each share has full voting and dividend rights as well as the right to share in assets upon dissolution. The pre-emptive right to buy new issues of the corporation's stock is provided by law in some U.S. states or by articles of incorporation in others. The existence of two classes of stock means that one or more of these fundamental rights does not attach equally to the shares of all classes.

Voting and Nonvoting Shares.—Shares whose holders may not vote at all or may not vote on certain issues, such as the election of directors, may be referred to as nonvoting shares. The two classes are often officially designated as class A and class B shares.

Preferred Stock.—To appeal to investors who wish to be sure of receiving dividends regularly, many corporations issue a class of stock that gives its holders a preferential claim to dividends. The most common provision of this type is that no dividends may be declared on the common stock unless dividends of a stipulated amount per year have been declared on the preferred shares. Usually this preferred dividend is cumulative, meaning that any dividend omitted is added to future dividends payable on the preferred shares before the common shares may receive a dividend. Noncumulative preferred shares have a weak position legally because the board of directors is under no obligation to declare dividends on any class of stock even if earnings are ample.

Another type of preference that may be granted relates to the distribution of assets upon dissolution of the corporation. The articles of incorporation may provide that one class of stock shall rank before another class when assets are distributed at the end of the corporation's life. The amount of this preference must be specified; it may vary depending upon whether the liquidation is voluntary or involuntary.

Convertible Preferred Stock.—The major disadvantage of preferred stock is that the holder does not share in exceptionally good fortunes of the issuing corporation, because preferred shares rarely receive dividends in excess of the preferential rate. This shortcoming can be remedied by giving stockholders the right to ex-

change preferred shares for common shares in a stated ratio so that conversion may become attractive if the market price of the common stock increases by a modest percentage. For example, when the common shares are selling at $21, preferred shares convertible at the rate of four common for one preferred share may be issued at $100. If the company prospers and the price of the common shares rises to $30, the preferred stockholder may share in the prosperity by converting a share of preferred into four shares of common with an aggregate value of $120. Under such circumstances the market price of the preferred stock will also rise.

Stock Values.—Articles of incorporation must specify the maximum number of shares of stock to be issued by a corporation. The total stated capital to be attributed to this number of shares may also be set, in which case the total stated capital divided by the number of shares determines the par value of each share. Alternatively, the incorporators may exclude from the articles any reference to a maximum stated capital or par value per share, instead designating the shares as no-par shares. In such cases, either the stockholders as a group or the board of directors is charged with the responsibility for declaring what portion, if any, of the proceeds from the issuance of the shares is to be designated as paid-in surplus. Thus, no-par shares outstanding are said to have a stated value per share equal to the increase in stated capital recorded upon their issue divided by the number of such shares issued. The significance of the par or stated value of a share of stock is tied to the concept of stated capital and the dividend limitations connected therewith; par or stated value is unlikely to be related to the stock's market value.

The "book value" of a share of stock is the shareholder's equity in the corporation as indicated by the balance sheet. If only one class of stock is outstanding, the entire net worth—including stated capital, paid-in surplus and retained earnings—divided by the number of shares outstanding, is the book value per share. If both preferred and common shares are outstanding, the total liquidation preference of the preferred shares is subtracted from net worth before computing the book value of the common shares.

Shares of stock are assignable personal property and are frequently bought and sold in the stock market. The market value of a share subsequent to its issue date may be substantially different from its book value. The general reason for the purchase of stock is to obtain cash receipts in the future in excess of the purchase price. The anticipated receipts either take the form of dividends or occur upon resale to another party who purchases it to obtain dividends. The corporation's capacity to pay dividends over a long period is affected by factors other than the capital funds it employs; hence the market value of shares may be higher or lower than their book value.

Changes in Capital Structure.—Stock dividends can be declared by the board of directors if sufficient authorized but unissued shares are available. A 5% stock dividend involves the issuance for no consideration of one share for each 20 shares held. The net worth of the corporation is not changed by this step, which is sometimes characterized as dividing the pie into more pieces. Unlike a cash dividend, a stock dividend paid to a holder of common shares is not considered income. The stated capital of the corporation is increased by a transfer from retained earnings or paid-in surplus.

A stock split is similar to a stock dividend in that the net worth is unchanged but the number of shares is increased. However, stated capital is not increased by a stock split. An example is the splitting of 1,000,000 shares of $10 par value into 2,000,000 shares of $5 par value. Occasionally the number of shares is reduced and the stated capital relating to each increased by a reverse stock split.

When a corporation whose stockholders enjoy the pre-emptive right wishes to raise capital funds through the issue of additional stock it first offers the stock to its shareholders in proportion to their holdings. Thus, if the new shares will increase the number of shares outstanding by 20%, a stockholder who owns 100 shares is entitled to purchase 20 shares. In order to assure the success of the rights offering, the subscription price on the new shares is usually set below the current market price. Accordingly, the right to purchase the new shares is valuable in itself, and these rights are frequently bought and sold.

Treasury shares are shares acquired by their issuer and held pending reissue or retirement. The acquisition by a corporation of its own shares is strictly regulated by law in order to prevent the circumvention of the stated capital restriction by returning assets to owners through the purchase of their shares to the detriment of creditors. (G. J. St.)

GREAT BRITAIN AND THE COMMONWEALTH

Great Britain.—In Great Britain the fundamental difference between stock and shares is a technical one only, though the two words have acquired overtones of meaning which tend to differentiate them in general usage. Since 1862, the term "stock" has signified a subscribed fund resolvable into parts as small as pounds, shillings and pence for use as corporate capital or a loan. Stock is therefore divisible and transferable in any amount desired not lower than a fraction specified by practical considerations. The term "share" on the other hand is used when the subdivision of the capital is of uniform amount. Thus a holding of stock is described as "£100 stock" or "£100 worth of stock," while a holding of shares is called "100 £1 shares." Furthermore, stock may be bought or transferred in odd amounts, *e.g.*, £10 8*s.* 5*d.* stock, whereas shares must always be subscribed or exchanged in multiples of the specified face value, be it £1, 10*s.*, 5*s.* or even 1*s.* Thus a sale of fractions of a share is not permitted.

Dividends payable on British shares are always expressed in one of two forms: "a dividend of 6*d.* (or so much) per share" or "a dividend of five (or so much) per cent," the latter meaning a dividend equivalent to the named percentage on the face value of each share. If the shares are £1 shares, 5% is equivalent to 1*s.* per share, but if they are 10*s.* shares, it is equivalent to 6*d.* per share. Dividends are normally paid annually or twice-yearly, those in the latter case being called interim and final payments.

The market price of stock or shares, which is related to the dividend payments, is normally expressed either in shillings and pence (*e.g.*, 11*s.* 6*d.* or 27*s.* 4½*d.*) or in pounds and fractions of pounds (*e.g.*, 13¼, meaning £13 5*s.* 0*d.*, or, in the case of stock, 87⅝, meaning £87 12*s.* 6*d.* per £100 stock).

In addition to the annual return the price of stock and shares depends on the degree of risk attached to the holding. Stock is a term normally applied in Great Britain to a loan or debenture repayable within or at a certain specified time and repaying a fixed rate of interest until that date. Shares, however, usually represent permanent capital as opposed to loans and debentures and fall into three main categories, each of which implies a different degree of risk to shareholders. These are preference, ordinary and deferred.

Preference shares, like loans and debentures, usually receive a fixed return only, although occasionally they participate with the ordinary shares in any further profits. They are then called participating preference shares. In all cases preference shares, as their name implies, have a prior claim to profits earned and rank ahead of ordinary and deferred shares, both as to the dividend and as to repayment of capital in the event of liquidation. Thus, if a company earns £100,000 and the preference dividend (say 5% on £2,000,000 of £1 preference shares) requires the full £100,000, then the holders of ordinary shares get nothing. If the preference shares are cumulative, any arrears of dividends on them are carried forward until a sufficient profit is earned to pay off the accumulated arrears; and this must be done before any payment is made on the ordinary shares. If preference shares are not cumulative, once a preference dividend is not paid in any one year, the shareholders have no claim in respect of that dividend on any future profits.

Ordinary shares rank after preference shares but, whereas the preference shareholder is entitled to a fixed rate of dividend only if profits are made and to the return of the nominal value of his shares only in the event of a liquidation, in both cases the ordinary shareholder receives what is left. If the company has also issued deferred shares, however, it is the ordinary shareholder who receives a fixed return only (sometimes with participating rights),

while the deferred shareholder receives what is left. A company seldom issues all three classes of shares; normally two only will be issued and, if these should be the first and the third, they are frequently called preferred ordinary and deferred ordinary.

In no case are the directors of a company bound to pay a dividend, however, and in practice a company will normally pay only a portion of its earned profit in any one year, frequently putting the surplus into reserve funds or development expenditures. The profits made in any one financial year are often expressed as a percentage of the issued capital, this percentage being known as earnings.

A company increases its capital by issuing additional shares. These may be issued either free as a bonus to existing shareholders, often termed a scrip bonus issue (*e.g.*, two shares for every three held, frequently abbreviated to "two-for-three scrip"), or at a price below their current market price as a partial bonus to shareholders. In the latter case, for example, if the market price of £1 shares already in issue is 85*s.*, new shares may be offered at 45*s.*, a holder of two shares being entitled to buy one new one at the lower price, or to sell his right in the market to anyone who will buy it. This type of issue is known as a rights issue and the above example would be "a one-for-two rights."

Throughout the British Commonwealth the concept of stock and shares resembles closely that current in Great Britain both in terminology and practice. The one country providing a striking exception to this is Canada, which has in many respects followed the U.S. pattern.

Canada.—In Canada the term "stock" frequently takes the place of the English term "share." Thus, although preference shares form part of the capital structure of Canadian companies, the ordinary shares are known collectively as "stock" and investors in that company are "stockholders" and not "shareholders." Furthermore, the terms "ordinary" and "deferred" are not normally used, the equity being known as "common" stock. This common stock is frequently "bearer" stock, as opposed to the "registered" shares common in Great Britain; *i.e.*, ownership is vested simply in the bearer and no nominal registration of specific ownership is made in the company's books. Unlike British shares, Canadian stock does not always have a specific nominal or face value; it is frequently designated "of no par value" (commonly abbreviated to NPV). Common stock of Canadian companies may be divided into two sorts, class A and class B, the latter only carrying voting rights and the former carrying either a degree of preferential treatment with regard to dividends and liquidation, or participating rights, or sometimes both. Dividends are expressed in Canada as so many dollars (or cents) per share and not as a percentage, and are frequently issued together with trading information quarterly, not yearly or half yearly as in Great Britain. The most important Canadian stock exchanges are those of Toronto and Montreal.

Australia and New Zealand.—Australian and New Zealand shares bear a marked similarity to those in Great Britain, although the deferred share is not normally found. Australian companies, however, unlike their British counterparts, sometimes issue unsecured registered notes. These fulfill the same function as a debenture stock, representing a loan repayable at or within a specified period; they are, however, issued in units not of £100 but of small denominations. Unsecured notes normally pay a fixed rate of interest but sometimes carry participating rights also. Occasionally they are convertible into ordinary shares. The leading stock exchanges of Australia are those of Melbourne and Sydney, while the main centre of share dealing in New Zealand is at Wellington.

Malaya and Hong Kong.—Australian shares enjoy a sizable following in the stock exchanges at Singapore and Hong Kong. Both these markets follow closely the British pattern and the former is the home market for Malayan rubber shares in which active dealings take place in London.

OTHER COUNTRIES

Stocks and shares have formed an integral part of the economic development of almost every country in the world, both in assisting their various governments to raise the finances necessary for their administrative program and in providing the capital required by private concerns in each country. Owing to the different rates of growth in economic and industrial development, however, the degree to which comprehensive trading of stocks and shares has been centralized and developed varies widely. In this respect the stock exchanges in London and New York had, by 1900, established a predominant position both in breadth and volume of trading and this position has since been consolidated, but the following cities also have stock exchanges which have gained international recognition: Tokyo, Cairo, Amsterdam, Paris, Brussels, Milan, Geneva, Frankfurt am Main, Düsseldorf, Munich, Lisbon, Stockholm, Oslo and Vienna.

On all these exchanges dealings take place in governmental and provincial securities and other private bonds and shares of a national character and the concept of a share in each case is fundamentally the same as that in Great Britain, although there are many minor technical and terminological differences in the case of European and Asian countries. The capital structure of many European companies, as well as that of Japan, shows a marked preponderance of ordinary shares over preference shares, the latter forming a much smaller proportion of the total share capital than is customary in Great Britain. As against this, however, it is not unusual to find among companies of other countries different denominations of both ordinary and preference shares, a feature not found in Great Britain and alien to the British definition of a share. For example, Philips' Incandescent Lamp works, the Netherlands company whose shares, together with those of Unilever and Royal Dutch, had by 1958 achieved perhaps the widest international circulation of any non-British or U.S. industrial share, has issued both preference and ordinary shares in denominations of 1,000*f*, 100*f* and 50*f*. In certain countries, notably Belgium, it is common to find shares with no par value, or nominal face value, as with Canada.

As in Great Britain, shares issued in other countries normally entitle their holders to vote in matters concerning the administration of the company, but in some European countries, particularly Switzerland and the Netherlands, comprehensive steps are frequently taken to ensure that voting control does not fall into the hands of non-nationals. This goal is achieved in spite of the fact that European companies tend to have a larger number of small shareholders than their counterparts in Great Britain and the United States.

In almost all European countries with the exception of Great Britain and Italy, shares are "bearer" shares and not "registered" shares. A "bearer" share certificate is, as its name suggests, a certificate vesting ownership simply in the bearer (as with an English pound note), whereas a registered certificate specifically designates the owner or subscriber by name and implies his specific registration by name in the books of the company. A bearer share certificate normally has dividend coupons attached to it; *i.e.*, perforated slips entitling the bearer to his dividend on presentation at a bank. Partly because of this and partly because of the smaller scope and less highly developed organization of many European and Asian stock exchanges, banks in Europe and Asia tend to take a bigger and closer part in investment and share dealing than they do in Great Britain or the United States.

South Africa is the home of many of the leading gold mining companies whose shares have an active following on the London stock exchange (where they are known as Kaffirs). The home market for these issues, together with many other shares of a national and local character, is Johannesburg, where the concept of stock and shares is very similar to that in London. Dividends, however, are always expressed in shillings and pence per share and not as a percentage. These are normally paid twice yearly, while gold mining companies also issue quarterly reports to shareholders.

See also references under "Stock" in the Index volume.

BIBLIOGRAPHY.—A. A. Berle and G. C. Means, *Modern Corporation and Private Property* (1948); J. B. Lindon *et al.*, *Buckley on the Companies Acts*, 13th ed. (1957); H. G. Guthmann and H. E. Dougall, *Corporate Financial Policy*, 2nd ed. (1948); R. H. Code Holland,

Poley's Law and Practice of the Stock Exchange, 5th ed. (1932).
(B. P. W.)

STOCKBRIDGE, a town of Berkshire county, Mass., U.S., on the Housatonic river, is located 116 mi. W. of Boston and 138 mi. N.E. of New York city. A summer and autumn resort and a region of historic interest, there are, within its area, the Stockbridge Bowl, a lake that provides opportunities for boating and fishing; Ice Glen (where caverns are lined with ice even in midsummer); Monument mountain (1,750 ft.); and most of Tanglewood, the home of the Boston Symphony orchestra's summer centre and the Berkshire Symphonic festival. Also located in the town is the Riggs Centre, a clinic for psychiatric research and treatment.

In 1734 a plantation 6 mi. square was laid out for the Mohican Indians who had come from the Hudson valley when white settlements encroached upon their territory, and among whom Rev. John Sergeant had established a mission in 1734. In 1739 this Indian town was incorporated as the town of Stockbridge; Jonathan Edwards was in charge of the Indian church and school from 1750 to 1758. By 1800 nearly all the Indians had moved west.

The resident population of Stockbridge is over 2,000.
(R. C. L. S.)

STOCKBROKER. A stockbroker is an agent who deals, either on or off an organized exchange, in stocks, bonds and other securities. Legally the relationship between the broker and customer or client is not only that of agent and principal but becomes that of creditor and debtor and that of pledgee and pledgor when the customer purchases on margin (*q.v.*) and the broker furnishes the required additional funds secured by collateral.

Apart from the technical knowledge of buying and selling securities for clients, the stockbroker must have a wide fund of general experience. He must keep in close touch with the affairs of industry and finance, and with events abroad likely to affect such investments as foreign bonds or overseas commercial securities. The importance of such general knowledge is shown by the fact that many brokers never actually go onto the floor of the stock exchange to deal themselves. They concern themselves solely with giving investment advice and receiving instructions from clients. The mechanics of the transactions are handled by other members of the same firm who specialize in dealing. With the steady growth in the volume of institutional business—business from insurance companies and pension funds—a wide knowledge of financial and industrial matters becomes increasingly more important for brokers dealing with such sophisticated clients, and most large firms maintain statistical departments that produce detailed reports on securities and industries, as well as handle clients' portfolios. (E. Mt.; K. L. T.)

United States.—Relationships between brokers and their customers in the U.S. are governed by various and differing state customs, business policies and the constitutions and rules of the organized exchanges. Before 1953 the New York Stock exchange did not permit corporations as member firms; hence brokerage houses with memberships on the exchange were partnerships.

When a person acts as broker, in either buying or selling securities, he is an agent of his customer and charges a commission for his services. Usually the buying or selling of securities listed on an exchange is done on a brokerage basis. Unlisted securities and most bonds are usually bought or sold in the over-the-counter market—that is, not on an exchange. The purchase or sale of a security off an exchange is sometimes made on a brokerage basis, but often through a dealer acting for himself. Unlike the broker, who receives a commission from his principal, the dealer receives his compensation in the "spread"—the difference between what he pays for the security and what he sells it for. The same person or firm may serve as broker in one transaction and as dealer in another, but strict rules and numerous court decisions prevent a broker from acting as a broker and dealer in the same transaction.

Members of the New York Stock exchange may be classified, according to the special nature of their business, into five groups: (1) commission brokers, who act for the public at a commission fixed by the exchange; (2) "two-dollar" brokers, who act as agents for other member brokers on the exchange at a rate per 100 shares fixed by the exchange (though the old rate of $2 was changed in 1919, the name still prevails); (3) floor traders or dealers, who buy and sell exclusively for themselves and for their own profit; (4) specialists, each of whom confines his trading to a few specific stocks, performing the functions of both commission broker and dealer since he trades for his own account as well as buying and selling stocks on a commission basis for other brokers and dealers; and (5) the odd-lot dealers, who supply or buy from commission brokers a sufficient number of odd lots to equal the trading unit (usually 100 shares), which is the minimum amount that may be purchased or sold on the exchange. While few houses deal in odd lots, probably one-fourth of all business done on the exchange originates in odd lots. The odd-lot dealer enables the small investor to trade in a few shares at a price only slightly higher than that of large lots.

Members of the New York Stock exchange are controlled rigidly by the constitution, bylaws, rules and customs of the exchange in an effort to secure fair competition and the proper relationship between broker and customer. Other exchanges in the United States are modeled after the New York Stock exchange, and all registered exchanges and members thereof are regulated by the laws of their respective states and by the federal Securities and Exchange commission (SEC).

Almost all over-the-counter brokers and dealers are members of the National Association of Securities Dealers Inc., a nonprofit association organized in 1939 and registered with the Securities and Exchange commission. The association, under the supervision of the SEC, makes and enforces rules of conduct consistent with equitable principles of trade within the requirements of the Securities Exchange act. (K. L. T.)

Great Britain.—In Great Britain a stockbroker must be a member of a recognized stock exchange or licensed by the board of trade to deal in securities. Most brokers are stock exchange members. The biggest exchange in the British Isles is in London, but there are also provincial exchanges in other large towns, such as Glasgow, Liverpool, Manchester, Birmingham and Dublin. There is, in addition, a Provincial Brokers Stock exchange, which has no exchange in which its members can meet but which has members in nearly 150 of the smaller towns.

A person may become a member of the London stock exchange in two ways. Either he may purchase his membership outright (the more expensive course) and in addition buy a nomination from a retiring member, or he may qualify for membership through service. All candidates for membership, who must be over 21 and British subjects, are governed by very strict rules aimed at the protection of the investing public, and must be annually re-elected by the council. For conditions of membership *see* STOCK EXCHANGE: *Membership*.

There is an important distinction between the two classes of stock exchange members, brokers and jobbers. The broker is simply an agent, who buys and sells shares on behalf of his clients, the members of the public wishing to invest. Unlike the jobber, who is a specialist, he must be an all-round man for his clients will expect him to give an opinion on all sorts of investments. The jobber, on the other hand, is a principal. He keeps what is known as a "book" in a certain group of shares, buying and selling on his own account. The broker is the intermediary between the investor wishing to deal in certain shares and the jobber who holds a stock of such shares, or, to be more correct, "makes a book" in them. (Though the broker comes into direct contact with the public, he is not allowed to advertise for business; his clients come through personal introductions.) The broker receives a commission on each transaction. Commissions are fixed by the stock exchange and are known as the "minimum scale of commissions." (*See* also STOCK EXCHANGE.) (E. Mt.)

STOCK EXCHANGE, an auction market for the purchase and sale of securities such as shares, stocks and bonds. Stock exchanges, also known as bourses, play important roles in the economic life of all major industrial nations. Their primary function is to provide a convenient means for persons to invest money in business enterprises and to liquefy their investments at any time by selling their securities. Without stock exchanges, capital

would become immobilized, for once it was invested it could not readily be liquefied. (*See* also BROKER; CORPORATION; INVESTMENTS; INDIVIDUAL; STOCK.)

This article is organized according to the following outline:

I. United States
  A. Historical Background
  B. Organization of the New York Stock Exchange
    1. Admission of New Members
    2. Board of Arbitration
    3. Committee on Member Firms
    4. Committee on Floor Procedure
    5. Committee on Public Relations
    6. Executive Committee
    7. The Trading Floor
  C. Classes of Members
    1. Commission Brokers
    2. Specialists
    3. Two-dollar Brokers
    4. Odd-lot Dealers
    5. Traders
  D. Types of Orders
    1. The Quotation
    2. Market Orders
    3. Limited Orders
    4. Stop-loss Orders
    5. Stop Orders
    6. Time Limits
II. Great Britain
    1. Historical Background
    2. Organization
    3. Membership
    4. Business
III. Other Countries
    1. Paris
    2. Amsterdam
    3. Basel, Geneva and Zürich
    4. Milan
IV. Glossary of Stock Exchange Terms

I. UNITED STATES

There are three national stock exchanges in the United States: the New York Stock exchange, the American Stock exchange (formerly the New York Curb exchange) and the National Stock exchange (opened in 1962), all located in the city of New York. These exchanges account for about 90% of the total trading on organized exchanges in the United States. In addition, there are regional exchanges located in many of the large cities, although their number has diminished due to consolidations. Securities are also bought and sold in the over-the-counter market, which consists of brokers and investment bankers in all parts of the country who arrange their transactions by telephone or telegraph. (*See* OVER-THE-COUNTER MARKET.)

A. Historical Background

The first informal organization of the New York Stock exchange took place in May 1792, although there had been trading in government securities some time before. A group of 24 merchants and brokers decided to charge a commission for acting as agents for other persons and to give preference to each other in the purchase and sale of securities. They operated under a buttonwood tree on Wall street except during inclement weather, when they went into a coffee shop.

From 1792 to 1817, interest in stocks and bonds rose as a result of the increase in government obligations and the growth of banks and insurance companies. During this period there was no restriction as to the time of operation or the number of brokers who could transact business. When the number of brokers became large, a group of them decided to organize formally. They adopted a simple constitution comparable to that of a small-town social club and called themselves the New York Stock and Exchange board. They agreed to accept the regular brokers as members, but each new member had to be sponsored by one of the existing members and then approved by the entire membership.

The crowded conditions in the street influenced the members to take permanent quarters inside. They rented a room at 40 Wall street for $200 a year. New members were required to pay an initiation fee of not less than $25, the amount varying with the location of the seat in the room. Each member was assigned a seat and the chairman called roll each day. An absence without excuse resulted in a fine. Continued absence resulted in expulsion. If a member did not attend regularly, it became uneconomical for him to hold his seat. (The use of the phrase, "seat on the exchange," stems from this time, when members were assigned particular seats when admitted. There are no seats on the present exchange, yet the term continues to denote membership.) During the day the chairman would call the name of each stock once, and any bids or offers would be heard for that stock and transactions effected. If a member called for a stock that had already been called, he was fined. The procedure was orderly and unexciting.

The American Civil War created much interest in speculation and the number of brokers interested in gaining seats increased. Since the number of members was limited, an outside group of brokers organized a separate exchange called the Open Board of Brokers. This group operated in uptown New York.

Because of the increased activity, a need arose for more adequate quarters for the New York Stock and Exchange board. Through contributions from members, the first building was erected. The construction of the building raised the problem of the transferability of memberships. Before this time, memberships were not transferable. Seats of deceased members were sold at auction and the proceeds given to charity. But since members now had an equity in the building, seats were made transferable in 1868. The price of a seat at that time reflected the equity of a member in the real estate as well as the value of the trading privileges.

With the advent of transferable memberships it was necessary to limit the number to protect each member's equity. The maximum number of seats was set at 533. In the meantime, however, competition was being felt from the younger members of the Open Board of Brokers, who operated at night as well as during the day. The older group had previously decided to close at 3 o'clock each afternoon. In 1869, the two groups consolidated to form the present New York Stock exchange, with members of the Open Board of Brokers admitted without cost. This brought the total membership to 887. At about the same time, 173 new memberships were sold for $1,000 each, bringing the total seats to 1,060. By 1879 the building had become inadequate and 40 new memberships were sold at current prices to pay for enlarging the building. This brought total membership to 1,100. Later admissions increased it to 1,375.

B. Organization of the New York Stock Exchange

The New York Stock exchange is organized as a nonprofit unincorporated association. The government of the exchange is vested in a board of governors consisting of members, the chairman, the president of the exchange, and representatives of the public. The principal powers of the board are to elect officers, establish rules governing trading, impose penalties for violations of rules, supervise member firms and consider applications for listing. The board has a chairman and vice-chairman, but the chief officer of the exchange is the president. He is a paid employee of the association and must not be a member of the exchange. Subject to the approval of the board of governors, he selects officers to whom he delegates authority vested in him by the board.

The organization of the exchange was completely revised in 1938 in substantial accordance with the recommendations of the Conway committee, which had been formed for that purpose in 1937. Previously, the administrative problems were handled by the members. Since 1938, in addition to the paid president, the exchange has had several paid vice-presidents and a total payroll of about 1,800. This staff, however, continued to be subject to the direction of the board of governors and its standing committees. They may not hold memberships on the exchange. There is a department headed by an officer of the exchange for each standing committee. These departments, at the direction of their committees, perform the functions necessary to the proper administration of the exchange. Final responsibility and authority for decisions, however, rest with the committees.

1. Admission of New Members.—The first requirement for acquiring membership on the New York Stock exchange is to be in

a position to meet the existing offering price. This price varies from time to time, depending on the amount of trading activity. In 1929, for example, a seat was sold for $625,000; in 1942 a seat was sold for $17,000. Once a potential member agrees to meet the offering price he must appear before a subcommittee on admissions with two members who act as his sponsors. By the time he appears before the subcommittee the secretary's office has gathered as much information about him as possible. This office usually attempts to get information even before an applicant submits a bid for a seat in order to avoid embarrassment in case he fails to qualify for membership. He must have a record of honesty in business affairs, a knowledge of the functions of the exchange, financial stability and a definite purpose for acquiring membership other than for speculation on his own account. The exchange guards carefully against speculation in the price of memberships. If the application is approved by the subcommittee it is posted on the bulletin board of the exchange. If within two weeks no member submits evidence to indicate that the applicant should not be admitted, the board of governors takes final action.

If the applicant borrows funds to purchase the seat, the lender must agree to subordinate his claims against the membership or proceeds of its transfer to all claims of the New York Stock exchange or any of its members. If payment is made possible through proceeds of a gift, the donor must provide the applicant with evidence that the latter is under no obligation to make payment to him. Where a member firm provides funds for the purchase of a seat for one of its partners, the firm must sign an agreement subordinating its claims to those of the exchange.

The exchange also provides for an unlimited number of allied members. They must pledge themselves to abide by the constitution and the rules and become general partners in member firms. An allied member may not go upon the floor of the exchange unless he has permission of the board to serve as an alternate for a partner who is a regular member.

2. Board of Arbitration.—To handle disputes that arise between members, between members and allied members, and between members and nonmembers, the board of governors appoints a board of arbitration. In disputes involving nonmembers, the board of arbitration may be composed of some members who do not own seats on the exchange. There is a panel of about 100 members and 100 nonmembers who may be drawn upon to serve as arbitrators in disputes. The secretary's office handles the technical aspects of the dispute. Certain procedures are followed in filing the complaint, selecting arbitrators, covering costs of arbitration, and selecting time and place for the hearing.

As the exchange has designed its arbitration procedure to conform with New York arbitration laws, the decision of the board of arbitration has the force of law. The decision has really more force than the law, for the exchange will not allow an appeal to be made to the board of governors from the decision of the arbitrators, unless fraud can be proved or unless additional information is subsequently obtained which might reveal the unfairness of the arbitrators' decision. Counsel is not necessary at the hearings. However, if counsel is desired, the secretary's office must be notified. If one side is granted permission to obtain counsel, the other side must be granted that privilege. A member may not be represented by counsel in a hearing with a nonmember where the latter is not so represented.

3. Committee on Member Firms.—The function of this committee is to ensure that member firms observe the regulations of the exchange. Even though a firm finances the purchase of a seat for a partner, the membership represents a personal franchise of the partner. Should the member sell his seat either voluntarily or involuntarily while in debt to the exchange or one of its members, the proceeds of the sale will be used first to pay those debts. On the other hand, the partnership must sign an agreement to the effect that it will assume the risks of fluctuations in the value of the membership where the firm is financing, in part or in whole, its purchase. All members have a right to retain their memberships, subject to making agreed payments to their firms for their seats. These are but a few of the regulations governing member firms.

4. Committee on Floor Procedure.—Trading on the floor of the exchange is governed by rules of the exchange and by rules of the Securities and Exchange commission (SEC). It is the duty of the committee on floor procedure to see that these rules are observed. For example, specialists (see *Classes of Members*, below) are required to file with the exchange a report of all their transactions. These reports are checked for infractions of trading rules. Sometimes the SEC may question activities of members which the committee feels are not justified. A regular check on members' trading strengthens its position. For example, the SEC believes that specialists should always take a position against the market. The committee on floor procedure believes that this is not always possible. By having full information on their members' activities, the committee frequently may be able to convince the SEC that some trading with the market is justified.

5. Committee on Public Relations.—As with many private corporations, the New York Stock exchange was late in recognizing the importance of its relations with the public. While the exchange has had a public relations department for some time, it originally was concerned only with internal public relations. With the reorganization which took place in 1938, the exchange showed concern about its relations with the public. It launched an advertising campaign to urge investors to be cautious. It told the public about the purpose of the exchange and how the public could invest in corporate enterprise through careful employment of savings.

6. Executive Committee.—The principal function of this committee is to co-ordinate the functions of all other committees. Its members participate in the important policy decisions of the various departments. Through the eyes of this committee is seen in perspective the results of the administrative efforts of the exchange.

7. The Trading Floor.—All trading takes place on the floor of the exchange where there are several large U-shaped posts. Each stock is traded at only one post. The selection of a post for a new issue is based on the relative activity around the several posts as well as the existing location of issues in the same industry as the new issue. In selecting this post, the exchange is interested primarily in preventing congestion on the floor. When, for example, a new aircraft issue is listed, a post is selected around which the activity is relatively light and where no other aircraft issue is traded.

Members who are interested in buying or selling a particular issue will congregate around the post where it is traded. These members constitute what is called a "crowd." The Chrysler crowd, for example, includes all members who are standing at the post where Chrysler stock is bought and sold, and who are participating in the auction process. Some will be bidding for Chrysler stock; others will be offering it for sale. When there is a meeting of minds, a transaction is effected.

Pages are the only other persons allowed on the floor. These pages record transactions as they are effected in the crowd. If the members involved agree that the transaction as repeated by the page is correct, the page walks to the open end of the post and sends the report, via pneumatic tube, to another floor in the building where teletype machines send the report to other cities. In this manner, reports of the latest transactions are flashed to brokerage houses throughout the country by ticker tape minutes after they are effected.

Along the sides of the floor are several hundred telephones over which clerks receive orders from member firms. Since the floor of the exchange is large, two huge boards on the walls provide a medium of communication between a clerk and his broker. Each member has a number assigned to him. When a clerk receives an order on the telephone he flashes the number of his broker on this board. His broker, observing the signal, goes to his booth, picks up the order, and takes it out into the appropriate crowd. Telephone clerks are not allowed on the floor, for they would add to the confusion.

C. Classes of Members

1. Commission Brokers.—The largest group among the mem-

bers of the New York Stock exchange is commission brokers. They are partners of member firms and own seats on the exchange in order to execute orders for their customers. They are their customers' agents; through them customers from all over the world can meet at one place and trade in securities. For their services they charge commissions which constitute the principal source of income for member firms.

2. Specialists.—Specialists are another group of members. For each stock listed there is a specialist. There may be more than one for one stock and several stocks may be handled by one specialist. For each stock the specialist has a book in which he enters all limited (see *Types of Orders*, below) orders given him by commission brokers as well as the time they were received. He stands at his post all day; therefore, if he specializes in more than one issue, they must all be located at his post. Whenever he executes an order for another broker, the specialist charges a commission, but a customer whose order is executed by a specialist does not pay this commission. Instead, the commission broker must pay it. Therefore, the latter will give orders to specialists only when he cannot himself gain an execution in a reasonable length of time. The origin of the function of the specialist dates back to the days when trading took place in the street. Some of the older brokers who could not stand the physical strain of hurrying from one place to another decided to stay in one place and specialize in the trading of one or more issues.

3. Two-dollar Brokers.—A third group of brokers is the so-called two-dollar brokers. They are not associated with member firms; they make their livelihood by executing orders for commission brokers who may become too busy to handle all their orders quickly. Suppose a commission broker receives two orders at the same time, one to buy 500 shares of General Motors and the other to buy 100 shares of Curtiss-Wright. Since the posts at which these two stocks are traded are located at some distance from one another, he elects to execute the order for General Motors himself and gives the smaller order for Curtiss-Wright to a two-dollar broker to execute. The latter charges a commission for his services equal to those charged by the specialist. The two-dollar broker also executes orders for members while the latter take time out for lunch or while they are away from the floor due to vacations or illnesses. Several members of the board of governors are two-dollar brokers.

4. Odd-lot Dealers.—The fourth group includes odd-lot dealers and brokers. The dealers act as principals in making up so-called round lots—100 shares or multiples thereof—to trade in the auction market from the odd-lot orders—less than 100 shares—which are received by member firms. While they act as principals, their primary function is to service the odd-lot customer and their main source of income is the commissions they charge for this service.

5. Traders.—The traders represent the final group of members. They buy and sell stock for their own account with the hope of making speculative profits. Contrary to prevailing opinion, they are small in number and account for only a small percentage of the total transactions.

D. Types of Orders

The market for every security listed on an organized exchange is an auction market. At all times there is a quotation for these securities which consists of the best bid and the best offer. The best bid comes from that broker in the crowd who is willing to pay as much or more than any other person for that stock and the best offer comes from the broker who is willing to sell stock at a price below that of any other offer. These brokers are buying and selling stock for their customers, and the prices which they bid or ask are those specified by the customer. All stocks are quoted in terms of dollars per share. Except for very low-priced stocks and those traded on the National Stock exchange, the smallest change in price is one-eighth of a dollar. Thus, a stock could sell for 50, 50⅛ or 50¼. All orders in the auction market are for 100 shares or multiples thereof, except for some inactive stocks traded at post 30 where the unit of trading is less than 100 shares. Every stock has a symbol which identifies it and which speeds the handling of orders and reports of transactions.

1. The Quotation.—Let us suppose that the last price for GM (symbol for General Motors) was 50 and that several brokers in the GM crowd want to buy it. The highest bid is made by one broker who is willing to pay 49¾. Since no other broker is willing to pay as much, this broker's bid constitutes part of the quotation, even though there may be others willing to pay 49⅝ or below. There are other brokers in the crowd who want to sell GM. The one who is willing to sell at the lowest price constitutes the best offer. Suppose he offers GM at 50¼. Others may be offering to sell at 50⅜ or above. The quotation for GM, however, is 49¾–50¼, or the best bid and best offer received from brokers in the crowd. This quotation appears above the post where GM is traded and is kept up to date by the specialist or his clerk inside the post. If the spread between the bid and ask prices for GM is too great to constitute a good market for the stock, it is the duty of the specialist in GM to bid higher than the best bid or to offer to sell at a price below the best offer. It is possible that several brokers may be bidding for GM at 49¾. In this case, the broker whose bid came first is the one who will have the opportunity to purchase the stock if anyone offers to sell at that price.

A complete quotation includes the size or the amount of stock available at the best bid and at the best offer. If there were three brokers each offering to buy 100 shares of GM at 49¾ and two brokers willing to sell 100 shares at 50¼, then the complete quotation would be 49¾–50¼, 300–200. The size of the orders does not appear on the post where the bid and offer are shown. A broker may wish to know the size, however, and he will ask the specialist for that information. The specialist will reply: "49¾ for 300, 200 at 50¼." The size is often an important factor in measuring the strength or weakness of the market at any one time. If there are considerably more shares available at the best bid than at the best offer, the market is stronger than if the reverse were the case.

The quotation may change even before the next execution. One of the brokers in the crowd may better the best offer by asking 50⅛. The quotation now becomes 49¾ for 100, 100 at 50⅛, which narrows the spread between the bid and ask and brings an execution nearer to realization. The broker who offered to sell at 50⅛ may now decide that he may be able to get a better price and cancels his offer. Since there is still one broker in the crowd who is willing to sell at 50¼, his offer now becomes the best offer and the quotation is changed to 49¾–50¼. These are the two principal methods by which a quotation changes before an execution is realized—through a better bid or offer or through a cancellation. Most of these bids and offers are made by commission brokers on behalf of their customers. If a customer changes his mind and bids a higher price he tells his broker to change his bid and this order is sent to the floor of the exchange where the commission broker receives the instructions to change the bid.

Suppose a broker who is bidding for GM now decides to meet the best offer. If the quotation is 49¾–50¼, the broker says: "50¼ for 100," meaning he is willing to buy at that price. An execution is effected at 50¼, and the quotation now is changed to include the next best bid and offer. If another broker is willing to bid 49⅞ while another offers at 50⅜, and these constitute the best bid and offer, the quote is changed to 49⅞–50⅜. It is possible, however, that an execution will leave the quotation unchanged, for there may be more bids and offers at the same price.

2. Market Orders.—The simplest type of order which a customer can enter is a market order. If a customer enters a market order to buy 100 shares of a certain stock the commission broker will take that order into the crowd and announce his willingness to buy at the best offer. If the quote is 49⅞–50⅜, he will buy stock for his customer at 50⅜, for he holds a market order and must meet the best offer. If a customer wishes to sell 100 shares at the market, the broker enters the crowd and sells 100 shares at 49⅞, which is the best bid. This is the only course of action a broker may take when executing a market order. A customer who enters a market order receives the worst possible price, from his point of view, within the range of the quotation. The broker, on the other hand, dares not try to get a better price for his

customer through the auction process without specific instructions to do so. If he decided not to execute the market order at 50⅜, but instead bid something less, trying to get a better price for his customer, it is possible that no one would offer to sell at less than 50⅜ and that, while he is holding the order, the broker who is offering to sell at that price cancels that offer and offers stock at 50½ instead. If the offering price continues to advance without an execution, the customer's broker may be forced finally to meet an offering price higher than the one that existed when he entered the crowd. If this happened, the broker would be forced to account for the difference to his customer out of his own pocket. For example, if he is forced to buy at 50⅝ because he withheld a market order which could have been executed at 50⅜, he must make up ¼ of a dollar per share from his own funds. All orders are time-stamped when they reach the floor, and it is a simple procedure to check the quotation at the time the order was received. For this reason, a market order should be executed immediately. The customer who enters the market order should be aware, however, that he must meet the best offer, which is the highest price in the quotation range, if he wishes to buy stock and that he must meet the best bid, which is the lowest price in the quotation range, if he wishes to sell.

3. Limited Orders.—Any order other than a market order is called a limited order. Referring again to the hypothetical quotation of 49⅞–50⅜, a customer may wish to buy 100 shares but does not want to pay 50⅜, the price he would be forced to pay if he entered a market order. He is willing to pay no more than 49½, so he enters a limited buy order at that price. When the telephone clerk receives this order on the floor, he may decide not to bother his broker, who may be executing market orders in other crowds, so he sends the order to the specialist via pneumatic tube. The specialist receives the order, time-stamps it, and enters it in his book. The order will not be executed until someone offers to sell at 49½. This may take several hours or even days. The customer will be notified as soon as he gets an execution, but he cannot expect to get confirmation within a few minutes after he enters the order, as is the case for a market order. If he has decided that he is unwilling to pay more than 49½ he must be willing to wait until someone offers stock at that price. It may be possible that he will get an execution below that price, for a limited order becomes a market order as soon as an execution of 49½ is reached. It is a modified type of market order, however, for it cannot be executed above 49½ but can be executed below that price. The procedure for executing a limited order to sell is exactly the same as that for limited buy orders. The former, of course, is entered above the market price, for the customer wants to receive more for his stock than the best bid. A customer who wants to buy stock but does not want to meet the best offer may nevertheless enter a limited order which betters the best bid, such as a limited buy order at 50. In this case, the quote would be changed to 50–50⅜, with the bid representing the customer's order. Under such circumstances, the broker might take the order into the crowd and bid 50, rather than give it to the specialist, in order to save the commission he would have to pay the specialist for executing it. Where a customer enters a limited order to buy at a price which is above the best offer, because he did not know the quotation at the time, the order will be executed as a market order.

4. Stop-loss Orders.—Another type of limited order is called a stop-loss order. Suppose a customer buys 100 shares of a certain stock at 50 but is unwilling to take a loss of more than three dollars per share. He may protect himself by entering a stop-loss order at 47. This order is entered on the specialist's book. If the stock falls to 47 the specialist tries to sell it. This type of order, however, does not guarantee a loss of more than three points, for it becomes a market order as soon as the stock drops to 47. If the stock continues to fall in price, it is likely that the best bid will be something less than 47 after an execution at that price. Since the stop-loss has become a market order, it will be executed at the best bid, which may be 46½.

The stop-loss order is also used to protect profits. If stock bought at 48 goes to 53, the owner may wish to hold the stock and still protect part of his paper profit. In this case, he might enter a stop-loss order at 50. Should the stock fall back to 50, it will be sold at or near that price and the owner will realize some profit. If, on the other hand, it continues to rise, he can enter stop-loss orders at progressively higher prices, thus protecting more of his profits while letting them increase.

It is also possible to enter a buy stop-loss order. This type is used to limit losses or protect gains when a customer has sold short. A customer sells short when he sells something he does not own. Of course, he must deliver stock when he sells it, and since he does not own it, he borrows some stock from someone who does and delivers this stock to the buyer. At some time in the future he will be forced to buy the same number of shares from someone else in order to repay the lender. An individual sells short when he believes a particular stock will fall in price. He sells short at a relatively high price and hopes to cover his short sale—buy the stock later to repay the lender—at a lower price.

Suppose an investor sells a certain stock short at 50. He decides to limit his losses to three dollars per share, so he enters a buy stop-loss order at 53. Should the stock rise to 53, his order will be executed. Should it drop to 45, however, he could enter a buy stop-loss order at 48, thus protecting two points' profit per share while letting his profits build up if the stock continues to fall.

5. Stop Orders.—Stop orders, of which stop-loss orders are a type, are often employed by chartists or persons who believe that a stock has resistance levels. An investor may feel that if a certain stock does not reach 50 it will probably go down to 40. He may also feel that if it does reach 50 or above it may go to 60. If the stock is selling at 48 the investor may enter a buy stop order at 51. This order is entered on the specialist's book. As soon as it reaches 51 the specialist will buy 100 shares. According to the analysis the stock will now probably go up to 60. If the stock does not reach 50, the specialist will not buy it and the stock, if it moves as anticipated, will go down to 40. The buy stop order differs from the limited buy order in that it is entered above the market while the latter is entered below the market. It differs from a buy stop-loss order only in its use. The former is used when one has no position in the stock. The latter is used when one is short in the stock. A sell stop order, on the other hand, is entered below the market and the procedure is reversed.

6. Time Limits.—A customer may wish to limit the length of time on a limited order. He may do so by telling his broker that the order is good for today only, good this week (G.T.W.), good this month (G.T.M.) or good until canceled (G.T.C.). If the order is not executed within the time limits, it will be canceled. A customer may cancel an order at any time, however, even though he did not so designate at the time it was entered. If orders do not carry a specific time limit, they are considered open orders. One danger of an open limited order is that the customer may forget about having entered it. It may be executed for him later, to his surprise. Most brokerage houses guard against this possibility by sending monthly notices to customers concerning open orders which they are holding for them.

The price of certain types of limited orders may be changed by the specialists when a stock sells "ex-dividend" (see *Glossary*, below.) If a stock is paying $1.25 every quarter, it will drop in price one and one-quarter points when it sells ex-dividend. This drop is exclusively due to the stock going ex-dividend, not to any fundamental or technical factors in the market. Therefore, on limited open buy orders and on limited open sell stop orders the specialist will reduce the price by one and one-quarter points. This accounts for the drop in the price of the stock due to its selling ex-dividend, and leaves the order, therefore, in exactly the same position as it was previously. Stop orders to buy or limited sell orders are not reduced. Where the dividend is in a fraction not divisible by the minimum trading range, the price is reduced to the nearest higher fraction.

See also LISTED SECURITIES; CURB MARKET.

See George L. Leffler, *The Stock Market* (1951); C. A. Dice and W. J. Eiteman, *The Stock Market* (3rd ed. 1952). (F. GR.)

II. GREAT BRITAIN

1. Historical Background.—London was once the greatest of the world's stock exchanges in the range and volume of its transactions; and even after New York's rise to rival importance in the 20th century the range of London's business covered every corner of the earth and every class and type of security. It was not until the latter part of the 18th century that the London stockbrokers definitely formed themselves into a stock exchange with premises of their own. They had been in the habit of meeting and doing business in the Royal exchange and neighbouring places, but in 1773 those who had hitherto met at Jonathan's (a coffeehouse in Change alley) moved to a room in Sweeting's alley to which the name of the stock exchange was formally given, the building becoming known as the Stock Exchange Coffeehouse or tavern. Business grew apace, and in 1801 a group of members raised £20,000 of capital in 400 shares of £50 each for the purpose of providing an adequate building and acquired a site in Capel court, Bartholomew lane.

The new building was opened in 1802. At that time the members of the stock exchange numbered about 500. Rules were drawn up for the conduct of business and for regulating membership of the house. It was provided that all future members should be admitted by ballot, and members were required to pay a subscription of ten guineas each for themselves and five guineas for their clerks. It was the general intention that the shares should be held by members, but there were always a large number of members who did not hold shares. The first extension was made in 1823 when the floor space was nearly doubled by the addition of a second room for dealings in foreign stock. In 1854 the two rooms with some small additions were rebuilt as one in the form of a dome with two transepts. Several subsequent extensions of this building were made, the principal one being in 1885 when another dome was added. In 1953 a public gallery was constructed to help dispel some of the ignorance about the true functions of the house. The stock exchange building occupies the greater portion of the triangular area formed by Throgmorton street, Bartholomew lane and Old Broad street.

2. Organization.—The control and administration of the assets and activities of the London Stock exchange have from June 1945 been in the hands of a body called the council of the stock exchange. After June 24, 1954, by which date the last ex officio members dating back to an earlier regime were compelled to retire, all members of the council (numbering from 30 to 36) were elected annually from and by the general membership. The introduction of the council to replace the earlier administrative bodies was accompanied by other revisions. Among these was the cessation of the earlier practice of distributing profits to the proprietors. In Dec. 1947, paid-up capital was reduced from £720,000 to £1,000 by writing down the paid-up capital value of each share from £36 to a nominal value of 1*s*. In compensation for the reduction of capital and surrender of rights to dividends on these 1*s*. shares, members received two £4 cumulative redeemable annuities for each share held. These annuities are redeemable by purchase at any time and any price, or by drawings after March 1, 1958, at £100 for each annuity; and they may be reissued any time. One important result of this change was that the purchase of a share qualification for membership was abolished. It was provided that each member should hold at least one share for voting purposes, to be obtained on application to the council.

The discipline of the stock exchange is very strict, and the council deals severely with members guilty of improper conduct. It has power to suspend or expel a member. The council also settles disputes between members and between members and their clients.

Securities cannot be quoted on the stock exchange except by permission of the council, and conditions have to be fulfilled before that permission is granted; this regulation marks a state of affairs sharply different from, and very much more satisfactory than, the ruling prior to World War I by which any security could be dealt in without permission and grave abuses occasionally arose.

3. Membership.—The annual income of the stock exchange in the early 1960s was about £1,170,000, derived from the annual subscriptions of members and their clerks, entrance fees paid by new members and quotation fees. The fees have been raised from time to time. In the early 1960s the entrance fees and annual subscriptions were: for a new member with three sureties, 1,000 guineas and 100 guineas; for a new member with two sureties (*i.e.*, a clerk who had served four years in the house), 500 guineas and 50 guineas; for an authorized clerk (*i.e.*, a clerk authorized to deal), 50 guineas and 100 guineas; and for an unauthorized clerk, 15 guineas and 30 guineas.

Since 1904, the only method by which a person may ordinarily become a member is to obtain a nomination (by purchase) from an existing member who must retire in his favour. Nominations may be obtained from a former member, or, if he has not disposed of his nomination, from the legal personal representative of a deceased member. Nominations, which are personal and non-transferable, are usually purchased, and the cost may range between £40 and £2,000. A candidate for membership must be recommended by three members, who become sureties for him during the first four years for £500 each, or £1,500 altogether. A certain number of clerks who have completed four years' service are admitted each year without nomination; for these, only two sureties of £300 each are needed, their liability extending for four years. Membership is for 12 months only, so that everyone who wishes to remain a member must apply to the committee for re-election. By the early 1960s there were approximately 3,400 members of the stock exchange, employing more than 2,000 clerks (of whom about 1,200 had right of entry to the floor of the house).

4. Business.—By custom, each market has a special place allotted to it; for instance, the space immediately in front of the World War memorial is the market for British government securities and other gilt-edged stocks. Dealers or jobbers in these securities take up their posts in this space every day, and dealers in other classes of securities similarly occupy the floor space allotted to their particular markets. In this way the locality of a market can be readily and easily found by the brokers who go to the house to buy or sell securities for their clients. There is an important distinction between brokers and dealers. Brokers are not allowed to deal on their own account, nor can dealers act as brokers—this is the vital difference (*see* also STOCKBROKER). The dealer or jobber occupies a position similar to that of the wholesale dealer in commodities. He deals in a particular group of securities, say oil shares, and his function is to buy oil shares when they are offered to him and to sell them if they are wanted. He quotes two prices for a share, his buying price, which is the lower, and his selling price, which is the higher, and the difference represents his profit. A broker acts for the public which desires to buy or sell securities. Ordinarily, both for buying and selling securities, though in certain circumstances only for buying, a commission is charged by the broker.

When a broker has bought or sold shares he usually enters on a printed slip the name of the shares and the price at which the business was done; he then signs the slip and drops it into one of the boxes provided for the purpose inside the house. These slips are all collected, and the prices are printed that evening in the official list. This is what is known as the marking of bargains.

Except in the gilt-edged market, every bargain that a broker executes for his client is "for the account," unless otherwise specified. This means that the bargain will be completed (*i.e.*, the stock handed over and paid for) on the next settling day. The usual period of the account is a fortnight, but there are four 21-day accounts in the course of a year. Settling days are usually fixed for a Tuesday; the passing of tickets—ticket day—takes place on a Wednesday; and making-up day is Tuesday, prices for the settlement of current securities being fixed in the morning. A broker on executing an order for a client sends him a contract note which sets forth the details of the transaction, the price obtained or to be paid, the amount of brokers' commission, stamp and transfer fee and the amount of the contract stamp. Later, as settlement day approaches, a further note is sent to a client who has bought stock, requesting him to forward the purchase money. The bargain may be "closed" in the same account, that is, stock purchased may be sold (and vice versa) before settlement is due; in this case the client is not required to pay the full

consideration (or deliver the stock), but merely receives or pays the difference between the buying and the selling price.

This latter procedure is one of the commoner methods of speculation. Speculation differs from genuine investment in that the genuine investor normally buys for income purposes or steady capital appreciation, whereas the speculator operates in anticipation of a quick movement in the share price. He may buy in anticipation of a rise or sell in anticipation of a fall. He may do this in the hope of closing the bargain at a profit before he is called upon to pay for or to deliver the stock. Before 1930 facilities for buying and holding stock without paying for it (contango) and for lending shares to speculators who had sold them without possession (backwardation) were commonly employed. Since then, although these facilities are available, such transactions are less common. In 1958 reintroduction of options offered alternative facilities that largely replaced contango and backwardation.

The hours of business on the London Stock exchange are nominally from 9:30 A.M. to 3:30 P.M., after which time the streets around the house are crowded with brokers and jobbers returning to their offices. Since World War I no sessions have been held on Saturdays. (C. J. Mt.; E. Mt.; B. P. W.)

III. OTHER COUNTRIES

1. Paris.—There are three types of markets on the Paris bourse: the official market or *parquet;* the curb market or *marché des courtiers;* and the free market or *marché hors-cote.* The official stockbroker or *agent de change* on the *parquet* has a transferable monopoly (now subject to governmental sanction), must act only as an agent and must not recommend the purchase or sale of securities to his clients. He is a sworn government official, subject to close supervision. Business is divided into cash (*au comptant*) transactions and transactions for the term or for future delivery and settlement (*à terme*). Both investors and speculators use the more advantageous term market. At the end of each month, all open accounts are settled through the central liquidation service. On the curb market and the free market the *courtiers,* like the *agents de change,* have a transferable monopoly, which, however, covers securities not quoted on the official market. They are not government officials but only commission agents. The free market comprises all shares quoted neither on the *parquet* nor on the *marché des courtiers.* A new share is generally first quoted on the free market and then, as it steadies, rises to the *parquet.*

2. Amsterdam.—In this exchange, stockjobbing is separated from the trade in commodities. During their occupation of the Netherlands in World War II the Germans tried to disrupt business on the stock exchange as much as they could; not until 1953 were the old regulations re-established. The Stock Exchange union, though a private association, is subordinate to the minister of finance insofar as the latter has the right to regulate the exchange's opening and closing and to approve the listing of new securities. The Amsterdam banks are members of the stock exchange, so that many transactions are effectuated not on the exchange itself but in the banks; such internal clearings, however, are based on the official quotations and governed by the regulations. Most dealings are "for money"; all purchases must be fully paid in cash. Bear transactions are prohibited, though in practice a bear seller may borrow or hire the securities he has to deliver.

3. Basel, Geneva and Zürich.—These are the most important Swiss stock exchanges and are supervised by the cantonal authorities. In the other centres, the markets rank as private organizations. The Vereinigung Schweizerischer Effektenboersen endeavours to secure the maximum possible uniformity. In the three principal markets only firms approved by the cantonal authorities and belonging to the Boersenkammer or the Effekten Boersenverein can transact business. Dealing on these stock exchanges is restricted to authorized banks, and only securities officially introduced can be dealt in. Banks work on commission and technically as responsible principals even when representing third parties. Business is transacted daily between members by direct negotiation (no jobbers or equivalents being employed).

4. Milan.—This is the most important Italian stock exchange. The rules governing Italian stock exchanges are fixed by law; they are established and supervised by the government and the *agente di cambio* is a public official appointed by presidential decree, with a rate of commission fixed by the treasury. Direct negotiations by word of mouth can be conducted only by these *agenti* or their representatives. Methods of trading include: transactions *a contanti* (for cash), almost entirely confined to securities of fixed value; transactions *a termine* (for the term), concerned with shares, *i.e.,* the majority of business; and transactions *a premio,* of which option dealing alone is fairly widespread. Settlement takes place at the end of the month, through clearing houses.

IV. GLOSSARY OF STOCK EXCHANGE TERMS

Accounts.—In British usage, accounts are the periods into which the stock exchange year is divided. Usually 24 in number, they are normally of 14 days' duration; sometimes they are of 21 days because of holidays.

Arbitrage.—The adjustment of anomalies in values between two different markets, *e.g.,* between London and Johannesburg. (*See* ARBITRAGE.)

Bear.—Originally a term limited in its application to a person who sold shares which he did not own, in the belief that he could buy them at a lower price when the time came for delivery. It has since been extended to cover any market operator who anticipates that prices will fall.

Bearer Stock or Bond.—Securities for which registration of ownership by the company concerned is unnecessary for a good title to pass from one person to another.

Bid.—Term indicating the price one is willing to pay for shares. Also the price which a buyer has declared that he will pay; commonly, therefore, the lower of the two prices quoted by a jobber or dealer.

Blue Chip.—An American expression also widely used in Britain to describe ordinary shares of the highest investment calibre.

Bonus Issue.—See *Capitalization Issue.*

Bourse.—A European term for a stock exchange, derived from French usage and first used for the Paris exchange.

Bucket Shop.—This term refers to a type of brokerage house that once flourished but is now illegal. It was not a brokerage firm in the true sense, for its operators merely accepted orders from customers without placing them on the market. The bucket shop operator gambled that customers were wrong in their judgment of the market, and that he would later be able to buy or sell the stock at more favourable prices.

Bull.—An investor who expects that prices will rise. The term was formerly used chiefly to describe speculators who bought shares which they hoped to sell at a profit before settlement day, when they would have to pay for them. A bull market is one in which prices are rising.

Buying Limit.—An order to buy shares at a price below the current market price. It is given to a broker to enable the would-be purchaser to take immediate advantage of a fall in price—which may be short-lived—without the necessity of the broker's referring to the client.

Buy or Cancel.—A customer may give his broker an order to buy today or cancel. If the order is not completed that day, the order is automatically canceled.

Capital Gain.—A profit from the sale of securities or other capital assets. *See* CAPITAL GAINS, TAXATION OF.

Capitalization Issue.—The official description in Great Britain of a scrip issue; also known as a bonus issue or, in the U.S., as a stock dividend. This expression means the issue by a company, to existing shareholders, of shares for which payment is not required. It represents the capitalization of reserves; to the extent that these have been accumulated from annual profits, it is a formal acknowledgment that the ploughing-back of profits in past years is the equivalent of the reinvestment of those profits in the company.

Chart Reader.—Some investors and speculators employ charts which depict such data as the daily market averages, the market prices of individual stocks, or the volume of sales. They think they perceive patterns in formation, the completion of which can be anticipated profitably.

Close to Close.—A price made by British jobbers indicating that they are prepared to deal $\frac{1}{8}$ of a £ (*i.e.,* 3¾*d.*) either side of a specified price. For example, "seven-eighths close to close" = 17*s.* 2⅛*d.* to 17*s.* 9⅞*d.*

Closing.—Buying or selling to close a bargain opened in the same account.

Cum.—Prefix meaning "with." "Cum all" means "with all advantages"; cum cap., "with capitalization rights"; cum div., "with dividend"; cum rights, "with rights."

Customers' Man.—The customers' man is the "contact man" for the brokerage firm by whom he is employed. He accepts and solicits orders from customers, advises them in their trading, and comments on the market outlook.

Daylight Trading.—Daylight trading refers to the practice of disposing of all stocks purchased during a single session.

Dealing for New Time.—A British term for the purchase or sale of securities during the two market days that precede the first day of an

account. In "dealing for new time" the same stamp duty and commission concessionary rules apply as for "deals for the account." The jobber will, however, quote a higher price.

Discretion.—Concession of freedom by a client to a broker to act as he thinks fit, generally limited to a stated margin.

Either Side.—Term used to indicate that a quotation is $\frac{1}{32}$ (or 7½*d.*) either side of a specified price. For example, "either side of the seven" means either side of $\frac{7}{16}$ and would indicate a range of from $\frac{13}{32}$ to $\frac{15}{32}$. If the "big figure" which (by implication) precedes it were 3, then the price quoted would be 68*s.* to 69*s.* 4½*d.*

Ex-dividend.—When a company announces a dividend to stockholders of record as of a certain day, the books of the corporation are closed on that day for the payment of the dividend. Purchasers of the stock after the books are closed do not receive the dividend, and the stock is said to sell "ex-dividend."

Flat.—Bonds are usually quoted and sold "and interest." Bonds in default, however, are sold "flat," since no interest is being paid.

Floating Supply.—The floating supply is that portion of the outstanding stock which is usually held for speculative purposes, and which is available for trading purposes at or near the current market price.

Gilt-edged.—Originally, writing paper of the finest quality, and hence first-class bills of exchange. Now most commonly applied to British government securities, dominion and colonial government stocks, United Kingdom corporation issues, etc.

Growth Stock.—The stock of a company which gains in value over a period of time more than the general market. The company usually has sales and earnings growth prospects which are greater than those of the economy.

G.T.C.—This is an abbreviation for "good 'til canceled" and refers to a limited price offer which remains open until canceled.

Kaffirs.—Familiar name given to South African gold mining shares. The market in which they are dealt—the particular spot inside the stock exchange premises—is sometimes called the Kaffir circus.

Loan Crowd.—The loan crowd is composed of exchange members who are willing to borrow or lend stocks.

Long.—A buyer or holder of securities is said to be "long" of the market or "long" of stock if he has more of a particular security than he has contracts to deliver. The longs are usually those who have bought expecting a rise in prices.

Margin.—When a customer borrows money from his broker to buy stocks, he is said to have bought stock on margin. In the United States the federal reserve board sets limits as to how much a broker can lend on stock transactions. *See* MARGIN.

Mixer.—A security which, although too speculative to justify a large holding in relation to the size of the total portfolio, is recommended for "mixing purposes," *i.e.*, for raising the average yield.

New Time.—This British expression describes the account, also called "new account," following the one in which dealings are taking place. For the two days preceding the new account, brokers may deal for "new time" so that settlement does not take place until the settlement day of the following account.

Not to Press (*NTP*).—Short for "not to press for delivery"; expression used when a jobber sells shares which he has not got and which he believes may take him some time to acquire.

Offer.—A term indicating the price one is willing to sell his shares for. It is synonymous with "ask," and is distinguished from "bid," the price one is willing to pay for shares.

Option.—Literally, a right to choose. In stock exchange matters it refers usually to "single options" and "double options." Single option means the right (purchasable for a consideration) to choose whether or not at a given future date to buy ("call option") a given security for a given price; or, similarly, the right to choose whether or not to sell ("put option") a given security for a given price. In both cases the right must be exercised within a given period, normally three months. A double option, sometimes known as a straddle, gives the right to decide at the end of the period whether or not to buy or sell.

Rights Issue.—An issue of shares to existing shareholders in a fixed proportion to those they already hold, to be purchased at a given price, which is generally below the market price of the existing shares.

Selling Against the Box.—When the holder of securities, possessing stocks which he is holding "long," sells an equal amount of stock "short," he is said to be "selling against the box."

Shop.—Term to describe a finance house which sponsors issues and subsequently deals in the shares. Hence such expressions as "The 'shop' is buying so-and-so."

Short.—A buyer or holder of securities is said to be "short" of the market or "short" of stock if he has contracted to deliver at some future date more of a particular stock than he holds. The shorts are principally those who have "sold short" in anticipation of a drop in prices, in which event they would buy at the lower price in time for delivery.

Split.—Frequently when the price of a stock rises to levels which discourage the average investor, the company may split its shares to reduce the price. For example, a stock selling at 200 may be split four shares for one and the price would drop to 50. The old stockholder now has four times as many shares but they are worth no more in total than before the stock split.

Spread.—The spread is the difference between the bid and asked prices.

Stop-loss Order.—The stop-loss order or stop order is an order to sell at the market if and when a transaction occurs at or below a stated price. By using "stops" holders of stocks limit their losses in a declining market.

Ticker.—A term used to describe the transmission of stock prices to brokerage offices all over the country. Appearing on the ticker tape are prices of securities which represent transactions completed only minutes before.

Warrant.—A certificate giving the holder the right to purchase a fixed number of shares of stock at some future time at a fixed price. It is properly known as a stock purchase warrant.

Wash Sale.—When two individuals agree to buy from or sell to each other securities to stimulate market activity, and where no intention to exchange goods or money is present, a wash sale is said to have been effected. This is an illegal transaction on the exchange floor and members are liable to expulsion when discovered engaging in it.

Yield.—Interest on bonds or dividends on stock expressed as a percentage of the amount invested on an annual basis.

See also references under "Stock Exchange" in the Index volume.

(B. P. W.; F. GR.)

STOCKHOLM, the capital of Sweden, on the east coast, not far south of the junction of the Baltic sea and the Gulf of Bothnia. The population of Stockholm in 1751 was 61,040; in 1850, 93,070; in 1880, 176,875; in 1900, 300,624; and in 1960, 808,484.

Origin.—Before Stockholm arose, Birka, Sigtuna and Uppsala were places of great importance. Birka, on the island of Björkö (the isle of birches), was a kind of capital where the king lived occasionally at least; history speaks of its relations with Dorestad in the Netherlands, and the extensive refuse heaps of the old city, as well as the numerous sepulchral monuments, show that the population must have been large. But, though situated at a central point on Lake Mälar (Mälaren), it was destroyed, apparently before the beginning of the 11th century (exactly when or by whom is uncertain), and it never recovered. Sigtuna, lying on the shore of a far-reaching northern arm of Lake Mälar, also a royal residence and the seat of the first mint in Sweden, where English workmen were employed by King Olof at the beginning of the 11th century, was destroyed in the 12th century. Stockholm was founded by Birger Jarl, it is said, in or about 1255, at a time when pirate fleets were less common than they had been, and the government was anxious to establish commercial relations with the towns which were now beginning to flourish on the southern coast of the Baltic. The city was originally founded as a fortress on the island of Stadholm. The castle was erected at the northeastern corner, and the city was surrounded with walls having fortified towers on the north and south. It came to be called Stockholm (the isle of the log, Latin *Holmia*, German *Holm*); the true explanation of the name is not known. During the middle ages the city developed steadily, and grew to command all the foreign commerce of the midlands and north, but it was not until modern times that Stockholm became the capital of Sweden. The mediaeval kings visited year by year different parts of the kingdom.

Geographical Situation.—Stockholm is famed for the beauty and physical characteristics of its situation. The coast there is thickly fringed with islands (the *skärgård*), through which a main channel, the Saltsjö, penetrates from the open sea, which is nearly 40 mi. from the mainland. A short stream, with a fall normally so slight as to be sometimes reversed by the tide, drains the great Lake Mälar into the Saltsjö. The city stands at the junction of the lake and the sea, occupying both shores and the small islands intervening. From the presence of these islands a fanciful appellation for this city is derived—"the Venice of the North"; but actually only a small part is insular. There are five main divisions, Gamla Staden, (old town), ancient nucleus of the city, properly confined to Stadholmen (the city island) which divides the stream from Lake Mälar into two arms, Norrström and Söderström; Norrmalm on the north shore of the channel, including the business section; Östermalm, eastern residential section; Kungsholmen, to the west, manufacturing and municipal centre; and Södermalm, to the south, manufacturing and residential area.

Gamla Staden.—Ancient origins are apparent in the narrow, winding streets, though the individual houses are not very old, because of the ravages of frequent fires. A few, however, preserve antique narrow fronts with gables, as in some of the north German

towns. The old market, still called Stortorget (great market), is now one of the smallest in Stockholm. At the north angle of the island is the royal palace (Slottet). The original building was destroyed by fire in 1697, the body of Charles XI being rescued with difficulty from the flames. A new palace after designs of Nicodemus Tessin the younger (d. 1728) was not completed, because of wars and general distress, until 1754; a restoration carried out in 1901 included many ornamental details devised by the architect and executed at the expense of Oscar II.

West of the palace are the offices of the majority of the ministries, some of them in the former buildings of the royal mint. Beyond these, on the west side of the island, is a square named from the palace on its northern side, the Riddarhustorget. The Riddarhuset (house of the nobility) was the meeting place of the Council of the Nobles until 1866, and its hall is adorned with the armorial bearings of noble families. The north forecourt has a statue (1890) of Axel Oxenstjerna, the chancellor, by J. Börjeson. In Riddarhustorget is a statue of Gustavus Vasa, unveiled in 1774 on the 250th anniversary of his accession. Southwest of the royal palace is the Storkyrkan (great church), dedicated to St. Nicholas, the oldest church of Stockholm. The date of its foundation is *c.* 1260, but it was practically rebuilt in 1736–42. Within it is richly adorned with paintings and wood carving, notably the equestrian St. George and the Dragon by Bernt Notke (late 15th century). Gamla Staden is the commercial centre of the city. At the broad shipping quay (Skeppsbron) which flanks the palace on the north and east, most of the seagoing steamers lie; and the exchange, customhouse, numerous banks and merchants' offices are in the immediate vicinity. Riddarholmen (nobles' island), lying immediately west of Stadholmen, contains the old Franciscan church (Riddarholmskyrka), no longer used for regular service, which since the time of Gustavus Adolphus has been the burial place of the royal family. It was restored, and contains mediaeval paintings and many trophies of the European wars of Sweden. On one side of it stands the old house of parliament; on the other a statue of Birger Jarl, the reputed founder of the city.

Norrmalm.—Staden is connected with Norrmalm by the Norrbro (north bridge) and Vasabron, the first crossing Helgeandsholmen (the island of the Holy Ghost), on which are situated the new houses of parliament and the Bank of Sweden. A third bridge connects with the main thoroughfare of Norrmalm, Drottninggatan (Queen street). The Norrbro connects with Gustaf Adolfs Torg, where a statue of that king stands between the royal opera house and the foreign office. Norrmalm is the finest quarter of the city, with broad straight streets, several open spaces with gardens, and handsome buildings. East and north of the royal opera, the Karl den Tolftes Torg and Kungsträdgården (royal garden) form a favourite summer meeting place, where open-air concerts and arts-and-crafts exhibitions take place. There are a statue of Charles XII and a fountain with allegorical figures, by J. P. Molin, also a statue of Charles XIII, and in the small Berzelii park close at hand one of the chemist J. J. Berzelius. Near Drottninggatan is the Klara church, the burial place of the poet K. M. Bellman, and west of this, occupying one side of a square, is the central railway station.

Other Districts.—On the island of Kungsholmen, to the west, is the town hall, designed by Ragnor Östberg, which, combining Swedish traditional styles with Italian influences, is one of the country's most magnificent buildings. There are also several hospitals, including the Serafimerlasarettet (1752), the royal mint and factories. Östermalm, lying east, that is, on the seaward side of Norrmalm, is a good residential quarter, containing no public buildings of note, save the barracks of the Swedish guards and the fine royal library, which is entitled to receive a copy of every work printed in Sweden. The library stands in the beautiful park of Humlegården (hop garden), in which is also a statue of Linnaeus. South of Östermalm, and east of the Kungsträdgården and Gamla Staden (Old Town), lies the peninsula of Blasieholmen (formerly an island) and, connected by bridges, the islands of Skeppsholmen and Kastellholmen, the three forming the foreground in the beautiful seaward view from the Norrbro. On the first is the National museum (1866), with ancient and modern sculptures, picture gallery and engravings. The National Historical museum contains the numismatic collection which is notable for its series of Anglo-Saxon coins. About 11,000 pieces came from the island of Gotland, some dating from 901–924, but the majority are later. In front of the museum is a bronze cast of the famous group of J. P. Molin (1859), the "Bältespännarna" (belt-bucklers), representing an early form of duel in Scandinavia, in which the combatants were bound together by their belts.

East of Skeppsholmen an inlet, Ladugårdslandsviken, so named from the proximity of the former royal farmyard (*ladugård*), and bordered on the mainland by a quay with handsome houses called Strandvägen, throws off a narrow branch (Djurgårdsbrunnsviken) and separates from the mainland an island about 2 mi. in length by ¾ mi. broad. This is mainly occupied by Djurgården (the deer park), a beautiful park containing the buildings of the Northern museum, a collection of Scandinavian costumes and domestic and agricultural utensils, and a biological museum housed in a wooden building imitating the early Norwegian timber churches. There also is Skansen, an ingenious reproduction in miniature of the salient physical features of Sweden with its flora, fauna and characteristic dwellings inhabited by peasants in the picturesque costumes of the various districts. Both the Northern museum and Skansen were founded by Arthur Hazelius (1833–1901). There is a bust of the poet K. M. Bellman, whose festival is held on July 26. Södermalm, the southern quarter, is principally residential. Among the many suburbs, Vällingby is particularly notable for its advanced planning.

To the north of the city, accessible by rail and water, are the residential suburbs of Haga and Ulriksdal, with royal *chateaux*, and Djursholm. Saltsjöbaden, 9 mi. E. of Stockholm, on Baggensfjärden, is the nearest and most favoured seaside resort, but Dalarö (20 mi. S.E.) and Nynäshamn (39 mi. S.) are much frequented. Vaxholm, 12 mi. N.E. by water, is a pleasant fishing village where numerous villas have been built.

Institutions.—Stockholm is the centre of government and the usual residence of the king; in summer he usually occupies one of the neighbouring country palaces. The city is the seat of the high court of justice (*Högsta Domstolen*) and of the court of appeal for the northern and midland districts (*Svea Hovrätt*). As regards local government, Stockholm is a *län* (administrative district) in itself, distinct from the rural *län* of the same name, under a high governor (*överståthållare*) and deputy, with departments for secretarial work, taxation and police. The city is in the diocese of Uppsala, but has a separate consistory, composed of the rectors of the city parishes, the president of which is the rector of St. Nicholas (Storkyrkan). Stockholm has no state university. A private university (Högskola) was founded in 1877, and was brought under state control in 1904. The president of the governing body is appointed by the government, while the appointment of the remaining members is shared by the Swedish academy, the Academy of Sciences and the city council. The faculties are three—philosophy, history and philology, mathematics and natural sciences and jurisprudence. The Royal Caroline Medico-Surgical institute (*Kungliga Karolinska Mediko-Kirurgiska Institutet*) is a medical foundation dating from 1810, which ranks after 1874 with the state universities of Uppsala and Lund in the right to hold examinations and confer degrees in its special faculty. Special and secondary education is highly developed; there are schools of agriculture, mining and forestry, technical schools, notably the Royal Institute of Technology (*Kungliga Tekniska Högskolan*), a veterinary school, a school of pharmacy, etc. Among the public schools under state control, one, the Nya Elementarskolan, was founded experimentally in 1828, after the education committee of 1825–28, among the members of which were E. Tegnér and J. J. Berzelius, had reported on the want of such schools. This school retains its separate governing board, whereas others of the class are under a central board.

Stockholm is the seat of the principal learned societies and royal academies. There are schools of painting, sculpture and architecture under the direction of the Royal Academy of Arts; a conservatory of music under that of the Royal Academy of Music; and experimental gardens and laboratories under the Royal Society of Agriculture. Outside the city is the Thiel gallery, with a rich

collection of the works of Bruno Liljefors, Carl Larsson and Anders Zorn. On Lidingö nearby, the home of Carl Milles contains a great collection of the works of the sculptor. The Natural History museum, the observatory and meteorological office, and the botanical gardens are under the supervision of the Royal Academy of Sciences.

Other collections deserving mention are the State Historical museum, the Nautical History museum and the Technical museum, all three excellent examples of modern display techniques, and the archives in the record office (*Riksarkivet*).

Among places of entertainment, the royal opera and the royal dramatic theatre (*Kungliga Dramatiska Teatern*), at Nybroplan, receive state subsidies and are known for excellent presentations. The Vasa, Oscars, Intima and Nya theatres may also be mentioned. In summer, music-hall shows are performed at the China theatre and open-air plays at Skansen and elsewhere. In early June a great attraction is the Stockholm Festival of Music, Drama and Ballet. Several of the leading sporting clubs have their headquarters in Stockholm.

An annual regatta is held in July by the Royal Swedish Yacht club (*Kungliga Svenska Segelsällskapet*).

Industry and Commerce.—Stockholm's industries are varied, including iron and steel, engineering, shipbuilding, and the manufacture of a great number of consumption goods such as flour, beer, tobacco, clothing, leather, etc. About 60,000 workers are employed in the city's industries. Government minting, printing, etc., is done there. Although not an industrial city in aspect, Stockholm is Sweden's foremost manufacturing centre. It is the first port of the country for imports, but in exports about equals Malmö and falls below Göteborg. About 50% of the shipping tonnage is Swedish. The channel is 32-ft. deep, and vessels of 23-ft. draught can tie up at Skeppsbron and Blasieholmen.

There is an outport at Värtan on Lilla Värtan channel to the northeast. The Hammarby channel takes vessels up to 3,500 tons. There is an intricate transportation centre in and over Slussen, connecting Lake Mälar with the Baltic by water and also Gamla Staden with Södermalm by land.

(*See* SWEDEN.)

BIBLIOGRAPHY.—T. Nordberg, *Stockholm i tolv vandringar* (Stockholm, 1946); G. Johansson, *Trettiotaleto Stockholm* (Stockholm, 1942); G. Selling *et al*, *Stockholm i närbild* (Stockholm, 1943); Svenske Turistföreningen, *Stockholm med omgivningar* (Stockholm, 1948); S. Trenter and K. W. Gullers, *The Old Town of Stockholm* (Stockholm, 1953); Burnett Anderson, *Stockholm, Capital and Crossroads* (Stockholm, 1953). (G. HN.)

STOCKJOBBER: *see* STOCK EXCHANGE.

STOCKMAR, CHRISTIAN FRIEDRICH, BARON VON (1787–1863), Anglo-Belgian statesman, who came of a Swedish family, was born at Coburg on Aug. 22, 1787. He was educated as a physician, and became attached in 1816 to Prince Leopold of Saxe-Coburg-Gotha on his marriage to Princess Charlotte of England. When she died next year he remained Leopold's private secretary, controller of the household and political agent, until the prince became in 1831 king of the Belgians. His disinterestedness and profound acquaintance with English and European social and political questions impressed themselves on all who were associated with him. In 1831 he retired to Coburg, in order not to excite Belgian jealousies by residing at court as confidential adviser, but he continued to be Leopold's right-hand man. In 1837 Leopold sent him to England as adviser to the young Queen Victoria, and in the next year he accompanied Prince Albert (afterward Prince Consort) on his tour in Italy. He won the complete confidence of the prince as well as of the queen, and on their marriage in 1840 he became their trusted though unofficial counsellor, dividing his time between England and the continent. He died at Coburg on July 9, 1863.

See the articles on VICTORIA, QUEEN; and ALBERT, PRINCE CONSORT OF ENGLAND. Selections from Stockmar's papers were published by his son Ernest in 1872, and a biography by Juste appeared at Brussels in 1873; *see* also *The Letters of Queen Victoria* (1907).

STOCKPORT, a municipal (1835), county (1888) and parliamentary borough, partly in Lancashire and partly in Cheshire, Eng., 7 mi. S.E. of Manchester and 38½ mi. N.E. of Chester by road. Pop. (1961) 142,469. Area 12.5 sq.mi. The ancient town stood on the southern slope of a narrow gorge where the Tame and Goyt rivers meet to form the Mersey, but the modern town has spread over the more level ground above. The town hall (1908) and the war memorial building opposite, which is the art gallery, stand on high ground and are built of Portland stone. Close by is Stockport college (1887–89). The chief of the 48 parks and open spaces are Vernon and Woodbank Memorial parks, which lie along the banks of the Goyt.

It is recorded that in the rebellion of 1172 the castle of Stockport (later demolished) was held against Henry II by his son, Geoffrey. Sir Robert de Stokeport granted a charter to the town about 1220, making it a free borough. In 1260 an annual fair and weekly market was granted, to be held on the site of the present market place. Stockport's grammar school was founded under the will of Sir Edmund Shaa, dated 1487. The 14th-century church of St. Mary, reconstructed in the 19th century, was built on an earlier foundation. During the rebellion of 1745 the defeated forces of Prince Charles Edward descended on the town whose bridge had been demolished. Stockport figures largely in the industrial disturbances of the early 19th century. Trade unionism increased as cotton spinning and weaving were transformed from cottage to factory industries. By 1800 local weavers were persistently pressing parliament for the regulation of their wages, and in 1808 demanded a fixed minimum price for their goods. Distress was increased by the bankruptcy of the employers, and by 1812 serious disorders broke out after which two weavers were hanged. Later, during the Chartist demonstrations, a procession marched on Stockport workhouse and took forcible possession. By 1822 cotton spinning was already Stockport's staple industry. Its importance as a hat-making town dates from the 18th century, although a document of 1585 indicates that even then hatting was a local industry. Today, engineering and the manufacture of chemicals have been added to these earlier industries.

STOCKS, a wooden structure formerly in use both on the continent of Europe and in Great Britain as a method of punishment for petty offenses. The culprit sat on a wooden bench with his ankles, and sometimes his wrists or even neck, thrust through holes in movable boards, generally for at least several hours. That stocks were used by the Anglo-Saxons is proved by their often figuring in drawings of the time (*see* Harleian mss. No. 65). The second Statute of Labourers (1351) ordered the punishment for unruly artisans. It further enjoined that stocks (ceppes) should be made in every town between the passing of the act and the following Pentecost. The act appears to have been ill observed, for in 1376 the commons prayed Edward III that stocks should be set up in every village. Though never expressly abolished, the punishment of the stocks began to die out in England during the early part of the 19th century, though there is a recorded case of its use so late as 1865 at Rugby. In many villages may still be seen well-preserved examples of stocks, in some cases with whipping posts attached. In the United States stocks were of frequent use in the 18th century, more particularly in the New England states; while in the southern states they were employed for punishing slaves (*see* PILLORY).

STOCKTAKING. To prepare a balance sheet it is necessary for a trader to arrive at a valuation of his stock-in-trade; this process is called stocktaking or, in the U.S. inventory taking. When proper stock records are available, they are an invaluable check upon the accuracy of the stocktaking; but in their absence it is essential that the process of stocktaking should be performed with the utmost care. If a wrong value be placed upon stock, the accounts embodying that valuation will to a corresponding extent mis-state the profits of the undertaking and its current financial position, and to a corresponding extent the accounts of the ensuing period will be affected in the opposite direction. The lower the rate of turnover the more serious will be any error in the periodical stocktaking. In many cases, the only independent check that the trader has upon the accuracy of his stocktaking is the inherent probability of the percentage of profit shown by the trading account. (*See* also FINANCIAL STATEMENT.)

The generally accepted basis for the correct valuation of stock-

in-trade, assuming the goods to be in a readily saleable condition, is either the price they actually cost or the price at which they could then be bought in the open market, whichever of the two be the lower.

But although it is unsound to take into account any rise that there may have been in market values since the goods were bought, it would be equally unwise for a trader to dispose of his stock upon a rising market at normal prices, where he can only replace the goods sold at a greatly enhanced figure. Usually, therefore, although rising market prices are ignored for stock-taking purposes, selling prices are advanced as buying prices increase; and this is equalized by the fact that as buying prices fall, it commonly becomes necessary for the trader to reduce his selling prices, though he still has stock purchased at the higher rate.

Responsibility.—In most businesses it is usual to make the general manager responsible for the general accuracy of the stock-taking. The managers of departments are responsible to him, and the actual work is split up among the various employees, each of whom is held responsible for his section of the work. Those employed in counting, measuring or weighing goods are usually those familiar with the handling of these goods, and thus able to identify them as being correctly described. The cost prices are then entered on the list by clerks from actual invoices. The managers decide what (if any) percentage is to be deducted from goods that are old-fashioned, or otherwise out of condition, and the remaining calculations are then completed in the office, after which the summary is certified by the general manager.

The actual process of taking stock is often rendered more difficult and uncertain because it has to go on while trading operations are in full swing, and goods are accordingly constantly coming in and going out. It is very necessary to take precautions to ensure that no errors arise from this cause. For this reason some business houses prefer to close their premises during stocktaking, even though it means a corresponding loss of trade. Further, it is important to see that the invoices relating to all goods which have in fact been included in stock are also included in the books as purchases, even though (as is sometimes the case) the actual invoices do not come to hand until later; conversely, invoices already to hand relating to goods in transit may, of course, be excluded from purchases.

When the trade is a seasonal one, it is usual to fix a date for stocktaking that comes between the two seasons, as it is naturally more convenient to take stock when the stock is at its lowest. In such cases it sometimes happens that deliveries of new season's goods have already been accepted, the invoices for which are "dated forward," *i.e.*, into the new accounting period. In such cases it seems admissible to exclude the value of these goods both from the purchases of the past period and also from the stocktaking. (L. R. D.)

STOCKTON, FRANK (Francis) **RICHARD** (1834–1902), U.S. writer best known for *Rudder Grange* (1879) and the tantalizingly enigmatic story "The Lady or the Tiger?" (1882). Stockton was born in Philadelphia, Pa., April 5, 1834. Soon after graduating from high school he turned to journalism and contributed prolifically to periodicals. He was variously on the staffs of *Hearth and Home*, *Century Magazine* and *St. Nicholas Magazine*. After 1881 he devoted himself entirely to writing. He began with juveniles but after the success of the absurdly fanciful *Rudder Grange* wrote mainly for adults. Stockton's many novels and stories are typically full of farcical and fantastic incidents in a fairy-tale type of setting, although he also wrote conventional romances and adventure tales. Probably his best work in the romantic vein is *The Casting Away of Mrs. Lecks and Mrs. Aleshine* (1886). An admirer of Charles Dickens and Daniel Defoe, Stockton lacked their realism and moral earnestness, remaining aloof from contemporary concerns. His short stories are generally considered more successful than his novels. He died in Washington, D.C., April 20, 1902.

His work was collected in *The Novels and Stories of Frank R. Stockton* (23 vol., 1899–1904)..

See Martin I. Griffin, Jr., *Frank R. Stockton; a Critical Biography* (1939).

STOCKTON, RICHARD (1730–1781), signer of the Declaration of Independence, was born Oct. 1, 1730, in Princeton, N.J. He was the grandson of another Richard Stockton, who was one of the original settlers of the city and the owner of the land which became the centre of modern Princeton. He was educated at the College of New Jersey at Newark which later, largely through the efforts of his father, was moved to Princeton and became Princeton university. After his graduation in 1748, young Stockton studied law in Newark, and his admission to the bar in 1754 was the beginning of a long and successful career as an attorney. In 1766 he traveled to Europe as the representative of the College of New Jersey and was able to persuade John Witherspoon to accept the presidency of the college, a position the latter had previously declined. From 1768–76 Stockton served on the executive council of his native colony and from 1774–76 as an associate justice of the New Jersey supreme court. In 1776 he became a member of the Continental Congress.

Stockton's attitude on the relation of the colonies to Great Britain underwent a considerable change in the years before the outbreak of the Revolutionary War. The first record of his opinion is dated 1764, at which time he felt that any difficulties between the colonies and England could be settled by the inclusion in parliament of some U.S. members. Ten years later, however, his letter to Lord Dartmouth entitled *An Expedient for the Settlement of the American Disputes* advocated that the colonies be allowed to be completely self-governing and responsible only to the king. By the time of his election to congress, he was convinced of the advisability of independence for the colonies, and he not only signed the Declaration of Independence but argued in favour of it before its passage.

He was re-elected to congress in 1776 and in the same year declined the position of chief justice of the New Jersey supreme court, to which he was elected unanimously, having previously been defeated for governor. In Nov. 1776, he was captured by the British at the home of a friend in Monmouth, N.J., and imprisoned in New York city. Before his release, which was secured several months later by Washington himself, his health had suffered so badly that he never recovered. In addition, most of his fortune was gone, his lands and buildings having been looted by the British in his absence. He did not return to public life, but spent the years until his death, Feb. 28, 1781, in Princeton.

STOCKTON, ROBERT FIELD (1795–1866), U.S. naval officer and public official who played an important role in events leading to the cession of California to the U.S. He was born in Princeton, N.J., Aug. 20, 1795, the grandson of Richard Stockton (*q.v.*), signer of the Declaration of Independence. Robert Stockton attended Princeton college for a time, but in 1811 joined the navy as a midshipman, shortly becoming attached to the frigate "President," under the command of Commodore John Rodgers. He served under Rodgers during the War of 1812 except for a brief period in Washington as aide to the secretary of the navy. At the conclusion of hostilities, Stockton was transferred to the Mediterranean, where he remained until 1821, participating (1815) in the war against the Barbary pirates. In 1821 he was placed in charge of the schooner "Alligator" and dispatched by the American Colonization society to the west coast of Africa where he was able to obtain for the society the land which later became Liberia.

From 1826 to 1838 he lived in Princeton, devoting his attention to the foundation of the New Jersey Colonization society, of which he was first president, and to the construction of the Delaware and Raritan canal, which owed its existence to him. He was responsible for obtaining the charter for the canal, for financing it with his own fortune and for directing it as president during the last years of his life.

In 1838 he went to sea again, as commander of the flagship of the Mediterranean fleet. Pres. John Tyler asked him to join his cabinet as secretary of the navy, but Stockton refused, preferring to remain on active duty. With John Ericsson, later designer of the "Monitor," he drew up plans for the "Princeton," the first warship to be driven with a screw propeller, designing himself one of its guns, the largest in the U.S. fleet. On one of the ship's first voyages, Stockton's gun was fired and exploded, killing the

secretaries of state and the navy and wounding numerous other passengers, but Stockton was absolved of all blame.

In 1845 he was chosen by Tyler to sail in the "Princeton" to Texas to deliver to the state government the annexation resolution of the U.S. government, and later in the same year he commanded the "Congress" in its voyage from Norfolk to the Pacific, where he had been ordered to take command of the fleet from Commodore John Sloat. War with Mexico having broken out while he was en route, Stockton immediately took command of U.S. land and sea forces, capturing Los Angeles, the Mexican stronghold, on Aug. 13, 1846.

Stockton set up a civil government, first naming himself as governor and later John C. Frémont (*q.v.*), who had been associated with him in the fighting. He subsequently fought and defeated the Mexicans at San Diego, San Gabriel and La Mesa, after which Los Angeles, which the Mexicans had recaptured, again fell to the U.S. forces. The entire province of California was then ceded to the United States.

Stockton returned east in 1847 and resigned from the navy three years later, on May 28, 1850. He served in the U.S. senate from 1851 to 1853 and after 1853 was president of the Delaware and Raritan Canal company. He died in Princeton on Oct. 7, 1866.

STOCKTON, a city of California, U.S., the seat of San Joaquin county, is on the San Joaquin river, 78 mi. E. of San Francisco, approximately in the centre of the great central valley of California. Although situated 88 nautical miles from the sea, Stockton is a deepwater port and has become a major shipping point for the produce of the rich interior valley. The city had a population of 86,321 in 1960, and is the centre of a standard metropolitan statistical area (San Joaquin county) with a population of 249,989 in 1960. (For comparative population figures *see* table in CALIFORNIA: *Population.*)

Stockton traces its origins to 1843 when Capt. Charles Weber, a German immigrant, endeavoured to develop a grant of 48,747 ac. of rich valley land which he and a Mexican partner had secured from the government of Mexico. The first settlement, named Tuleburg after the characteristic tules or bulrushes of the region, failed to attract settlers until the discovery of gold in the famous Mother Lode of the Sierra Nevada in 1848. As its position at the head of navigation on the San Joaquin river made it the most advantageous supply point for the mines, the community grew rapidly. By the summer of 1849 the one permanent building of 1848 was surrounded by a tent city of about 1,000 inhabitants, the great majority of whom catered to the needs of the miners. During the height of the gold rush the little port of Tuleburg was visited by ships from Europe, South America and Asia, contemporary accounts telling of over 100 ships at anchor at one time.

In 1850 after fires virtually destroyed the original tent city and its ramshackle wooden successor, a planned community was begun by Captain Weber. Surveys were made, areas were set aside for schools, parks and churches, and the name was changed from Tuleburg to Stockton in honour of Commodore Robert Stockton, whom Weber had met during the course of the war with Mexico. By 1853 the city had reached a population of 5,000.

With the coming of the Central Pacific railroad in 1869 and the rapid settlement of the fertile interior valley, the economy of the area soon changed. Modern Stockton is a thriving trade centre but the wealth of the mines has long been superseded by the much greater wealth of the surrounding agricultural region. San Joaquin is a leading county in acreage value for vegetable crops and in acreage for Tokay grapes, asparagus and tomatoes. Agriculture is highly diversified, with approximately 100 different crops.

As large steamships supplanted the small ocean-going craft of the mining days, the natural channel of the San Joaquin river soon proved inadequate and the growing commerce of the area was lost to San Francisco. As early as 1871 plans had been made for the construction of a deepwater channel to serve the city, but despite a growing realization of its desirability, the financial burden of the undertaking was such that nothing could be done until 1930 when federal assistance was secured. Since its completion in 1933 the Port of Stockton has been greatly enlarged and improved and is now one of the largest general cargo terminals on the Pacific coast. The 32-ft. channel and a 33-ac. turning basin make it possible for any ship in the U.S. merchant marine to dock at Stockton. Among the facilities for handling the diversified cargoes of the port are the 3,000,000-bu. capacity grain elevators and a fully automatic bulk ore terminal.

The mid-20th century was a period of increasing industrialization of the Stockton area, particularly in those industries relating to agriculture and food processing. Several military establishments are located nearby, notably the Sharpe general depot with its atomic food processing plant; and a naval supply annex and west coast communications centre.

Stockton is the home of the University of the Pacific, a Methodist liberal arts institution chartered in 1851. It includes a conservatory of music and schools of education, pharmacy and engineering; its facilities include, in addition to the Stockton campus, the Pacific Marine station at Dillon beach and the Pacific summer theatre at Columbia. The Stockton public-school system extends through junior college.

The city's program of cultural activities and recreational facilities includes the Haggin Art gallery and Pioneer Historical museum, a symphony orchestra, a successful civic theatre and an extensive park system. Stockton has had a council-manager form of government since 1923. (M. H. Mo.)

STOCKTON-ON-TEES, a municipal and parliamentary borough and port in County Durham, England, on the left bank of the river Tees, 11½ mi. above its mouth and opposite Thornaby-on-Tees. Pop. (1951) 74,155. Area 8.5 sq.mi. Stockton's first market charter was granted in 1310 and it is still a big market centre for cattle and agricultural produce. Because of the presence of iron ore in the Cleveland hills to the south (which has long been mined) and the Durham coal field to the north, it grew into a big industrial centre and the leading Tees port until superseded by Middlesbrough. The port has 980 ft. of quayage, a minimum depth of 12 ft. at low water ordinary spring tide and 22½ ft. at high water ordinary spring tide and trades principally with Scandinavia and the Netherlands. The historic Stockton-Darlington railway (opened 1825) was the first passenger-carrying railway. Stockton is interested in the development of diesel and electric locomotives. Many light industries are located in the town and partly within the borough is the Imperial Chemical Industries establishment at Billingham.

Stockton grew up around the manor house and later castle of the bishops of Durham, to whom the town belonged. It is said that King John granted a charter in 1214, but no trace of this remains. The borough is first mentioned in 1283. Since 1867 Stockton has sent one member to parliament.

At Norton, 1 mi. north, St. Mary's church, once collegiate, shows Saxon and Norman work and has the remains of a fine three-decker pulpit.

STOCKYARDS: *see* SLAUGHTERHOUSE.

STODDARD, AMOS (1762–1813), U.S. soldier, was born in Woodbury, Conn., Oct. 26, 1762. He enlisted in the colonial army in 1779, serving until the close of the Revolutionary War, when he became a clerk of the Massachusetts supreme court. He spent the years from 1793–98 as a lawyer in Hallowell, Me., serving for one year (1797) in the legislature of Massachusetts, of which Maine was at that time a district. In 1798 he rejoined the army, receiving a captain's commission on his entrance and becoming a major in 1807 and deputy quartermaster in 1812. When the territory of Louisiana was purchased by the United States, Stoddard was chosen to accept the territory from Spain in the name of France and to accept it the following day from France in the name of the United States. He was then named civil and military governor of the territory, a post he held for the few months in 1804 before congress set up a government. He was assigned to Fort Meigs, Ohio, during the War of 1812 and during the siege of that post received a fatal wound. He died May 11, 1813.

Stoddard was the author of *The Political Crisis* (1791), written and published in London during a visit he made in his youth, and of *Sketches, Historical and Descriptive, of Louisiana* (1812).

STODDARD, RICHARD HENRY (1825–1903), U.S. editor, poet and critic, was born in Hingham, Mass., July 2, 1825.

His formal education was limited but as a boy he read widely and he was befriended by literary figures in New York. Soon Stoddard was contributing frequently to magazines and saving money enough to publish a collection of poems entitled *Footprints* (1848). However, he burned the whole edition except the one copy which was sold. He later published *Poems*, which was favourably reviewed, and in 1853 Nathaniel Hawthorne helped him to secure a position in the New York customhouse. Later he was private secretary to Gen. George B. McClellan in the York York dock department (1870–73); city librarian of New York (1874–75); literary editor of the *New York World* (1860–70) and the *Mail and Express* (1880–1903); and editor of the *Aldine*, a journal (1869–74). He died in New York, May 12, 1903. Stoddard's poems were popular in his day but modern critics do not regard them as important contributions.

STODDERT, BENJAMIN (1751–1813), first U.S. secretary of the navy, was born in Charles county, Md., in 1751. He planned to be a merchant, but at the outbreak of the Revolutionary War joined the army and served for a time as a cavalry officer. From 1779–81 he was secretary of the board of war and from 1781–98 he conducted a highly successful shipping business in Georgetown, Md. He was one of the men whom George Washington commissioned to purchase tracts of land in what is now the District of Columbia, to be repurchased later by the government and used as the site of the national capital.

In May 1798, Stoddert joined the cabinet of John Adams as first secretary of the new navy department, although Adams had previously offered the position to George Cabot, who refused. Stoddert's experience stood him in good stead in his new post. When he took office the U.S. navy consisted of little more than three frigates. When he left office, in 1801, he had enlarged it to more than 50 ships, which were able to defeat the French in the short and undeclared naval war between the two countries (1798–1800). He also authorized the purchase of land in six locations on the eastern seaboard for the construction of navy yards and was responsible for organizing the marine corps on the basis on which it still operates. During his secretaryship of the navy, Stoddert also served as head of the war department for a brief period between incumbents. He left office with Adams and returned to his business but it never prospered again and he died in Bladensburg, Md., Dec. 18, 1813, a poor man.

STOESSEL, ALBERT (1894–1943), U.S. conductor and composer, was born in St. Louis, Mo., Oct. 11, 1894. He studied in Berlin, Ger., and made his debut as a violinist there in 1914, returning afterward to the United States. He became conductor of the New York Oratorio society in 1921 and head of the music department of New York university in 1923. He conducted the Worcester (Mass.) music festivals from 1925 to 1942. His opera *Garrick*, on the life of the famous actor, was produced in New York city on Feb. 24, 1937; he also composed several orchestral works. He died in New York city on May 12, 1943, while conducting Walter Damrosch's symphonic poem *Dunkirk*. (N. Sy.)

STOFFLET, JEAN NICOLAS (1751–1796), Vendéan general, was born at Lunéville, the son of a miller. Long a private soldier in a Swiss regiment in France, and afterward gamekeeper to the comte de Colbert-Maulevrier, he joined the Vendéans when they rose against the Revolution to defend their religious and royalist principles. During the war in La Vendée he served first under Gigot d'Elbée, fought at Fontenay, Cholet and Saumur, and distinguished himself at the battles of Beaupréau, Laval and Antrain. He was appointed major-general of the royalist army and in 1794 succeeded La Rochejaquelein as commander-in-chief. He accepted the terms of the treaty of La Jaunaie (May 2, 1795) but he soon violated them, and at the instigation of royalist agents took arms in Dec. 1795 on behalf of the count of Provence (the future Louis XVIII), from whom he had received the rank of *maréchal-de-camp*. This last attempt of Stofflet's failed completely. He was taken prisoner by the republicans, condemned to death by a military commission, and shot at Angers on Feb. 23, 1796.

See General d'Andigné, *Mémoires*, edited by E. Biré (1900–01); C. Loyer, "Cholet sous la domination de Stofflet," in *L'Anjou historique*, vol. iii (1902–03).

STOICHEIOMETRY, in chemistry, is a term which, strictly, denotes the determination of the proportions in which elements or compounds react with one another, and is occasionally extended to include the determination of atomic and molecular weights (Gr. *στοιχεῖα*, fundamental parts or elements; *μέτρον*, measure). The term, as used in analytical chemistry, refers to that branch of chemistry concerned with the weight relationships in gravimetric analysis, the volume relationships in volumetric analysis and the calculations based on these relationships (*see* CHEMISTRY: *Analytical Chemistry*). The important stoicheiometric relationships are expressed by the laws of constant composition, multiple proportions and combining volumes. (G. W. Wd.)

STOICS, a school of philosophers founded at the close of the 4th century B.C. by Zeno of Citium and so called from the place where he delivered his lectures; viz., the Stoa or painted corridor on the north side of the market place at Athens. Stoicism, however, though it arose on Hellenic soil, is scarcely to be considered a product of purely Greek intellect, but rather as the first fruits of that interaction between west and east which followed the conquests of Alexander the Great. Zeno himself was of Phoenician descent; and hardly a single Stoic of eminence was a citizen of any city in the heart of Greece, unless we make Aristo of Chios, Cleanthes of Assus and Panaetius of Rhodes exceptions. Such lands as Cyprus, Cilicia and Syria, such cities as Citium, Soli, Heraclea in Pontus, Sidon, Carthage, Seleucia on the Tigris, Apamea by the Orontes, furnished the school with its scholars and presidents; Tarsus, Rhodes and Alexandria became famous as centres of its teaching. Moreover, Stoicism did not achieve its crowning triumph until it was brought to Rome, where the grave earnestness of the national character could appreciate its doctrine and where for two centuries or more it was the creed, if not the philosophy, of numbers of great men. It therefore stands in marked antithesis to the Academy, which was the typical school of Greece and of the Greek intellect.

The history of the Stoic school may conveniently be divided in the usual threefold manner: the old Stoa; the middle or transitional period (Diogenes of Seleucia, Boethus of Sidon, Panaetius, Posidonius); and the Stoicism of the Roman empire.

THE OLD STOA

By the old Stoa is meant the period (*c.* 304–205 B.C.) down to the death of Chrysippus, the second founder; then was laid the foundation of theory, to which hardly anything of importance was afterward added. Confined almost to Athens, the school made its way slowly among many rivals. Aristo of Chios and Herillus of Carthage, Zeno's heterodox pupils, Persaeus, his favourite disciple and housemate, the poet Aratus and Sphaerus, adviser of the Spartan king Cleomenes, are noteworthy minor names; but the chief interest centres about Zeno, Cleanthes and Chrysippus (*qq.v.*).

Zeno.—Zeno's residence at Athens fell at a time when the great movement which Socrates originated had spent itself in the second generation of his spiritual descendants. Neither Theophrastus at the Lyceum, nor Xenocrates and Polemo at the Academy, nor Stilpo, who was drawing crowds to hear him at Megara, could be said to have inherited much of the great reformer's intellectual vigour, to say nothing of his moral earnestness. Zeno visited all the schools in turn, but seems to have attached himself definitely to the Cynics; as a Cynic he composed at least one of his more important works, "the much admired *Republic*," which was a stumbling block to the school later on. In the Cynic school he found the practical spirit which he divined to be the great need of that troubled age. For a while his motto must have been "back to Socrates," or at least "back to Antisthenes." The Stoics always counted themselves among the Socratic schools and canonized Antisthenes and Diogenes; while reverence for Socrates was the tie which united to them such an accomplished writer upon lighter ethical topics as the versatile Persaeus, the protégé of Antigonus Gonatas. Zeno began, then, as a Cynic; and in the developed system we can point to a kernel of Cynic doctrine to which various philosophemes of other thinkers (more especially Heraclitus and

Aristotle, but also Diogenes of Apollonia, the Pythagoreans and the medical school of Hippocrates in a lesser degree) were added. Thus, quite apart from the general similarity of their ethical doctrine, the Cynics were materialists; they were also nominalists, combating the Platonic theory of ideas; in their theory of knowledge they made use of "reason" (*logos*), which was also one of their leading ethical conceptions.

In all these particulars Zeno followed them, and the last is the more important, because, Chrysippus having adopted a new criterion of truth—a clear and distinct perception of sense—it is only from casual notices we learn that the elder Stoics had approximated to Cynicism in making right reason the standard. At the same time, it is certain that the main outlines of the characteristic physical doctrine, which is after all the foundation of their ethics and logic, were the work of Zeno. "Reason," which had been an ethical or psychological principle to the Cynics, received at his hands an extension throughout the natural world, in which Heraclitean influence is unmistakable.

Cleanthes.—If the recognition of physics and logic as two studies co-ordinate with ethics is sufficient to differentiate the mature Zeno from the Cynic author of the *Republic*, no less than from his own heterodox disciple Aristo, the elaboration on all sides of Stoic natural philosophy belongs to Cleanthes, who certainly was not the merely docile and receptive intelligence he is sometimes represented as being. He carried on and completed the assimilation of Heraclitean doctrine; but his own contributions were distinctive and original. He was able to transform Zeno's seeming dualism of God (or force) and formless matter into the lofty pantheism which breathes in every line of his famous hymn to Zeus. Heraclitus had indeed declared all to be in flux, but we ask in vain what is the cause for the unceasing process of his ever-living fire. It was left for Cleanthes to discover this motive cause in a conception familiar to Zeno, as to the Cynics before him, but restricted to the region of ethics—the conception of tension or effort. The soul of the sage, thought the Cynics, should be strained and braced for judgment and action; his first need is firmness and Socratic strength. But the mind is a corporeal thing. Then followed the flash of genius: this varying tension of the one substance everywhere present, a purely physical fact, accounts for the diverse destinies of all innumerable particular things; it is the veritable cause of the flux and process of the universe. Herein lies the key to the entire system of the Stoics, as Cleanthes' epoch-making discovery continually received fresh applications to physics, ethics and epistemology.

Chrysippus.—Zeno had caught the practical spirit of his age—the desire for a popular philosophy to meet individual needs. But there was another tendency in post-Aristotelian thought—to lean upon authority and substitute learning for independent research—which grew stronger just in proportion as the fresh interest in the problems of the universe and the zeal for discovery declined. The representative of this tendency, Chrysippus, addressed himself to the congenial task of assimilating, developing, systematizing the doctrines bequeathed to him and, above all, securing them in their stereotyped and final form, not simply from the assaults of the past, but, as after a long and successful career of controversy and polemical authorship he fondly hoped, from all possible attack in the future. To his personal characteristics can be traced the hairsplitting and formal pedantry which ever afterward marked the activity of the school, the technical procedure of the "dialecticians," as they were called. He created their formal logic and contributed much that was of value to their psychology and epistemology; but he did not hesitate also to relabel and rearrange in every department; and he lavished care and attention on the logical terminology and on the refutation of fallacies, or, as his opponents declared, on the excogitation of fallacies which even he could not refute.

Stoic Conception of Philosophy.—What is philosophy? No idle gratification of curiosity, no theory divorced from practice, no pursuit of science for its own sake, but knowledge realized in virtuous action, the learning of virtue by exercise and effort and training. So absolutely is the "rare and priceless wisdom" identical with virtue itself that the three main divisions of philosophy current at the time and accepted by Zeno—logic, physics and ethics—are defined as the most generic or comprehensive virtues. Accordingly Aristo rejected two of these parts of philosophy as useless and out of reach—a divergence which excluded him from the school, but strictly consistent with his view that ethics alone is scientific knowledge. Of the three divisions logic is the least important; ethics is the outcome of the whole and historically the all-important vital element; but the foundations of the whole system are best discerned in the science of nature, which deals preeminently with the macrocosm and the microcosm, the universe and man, including natural theology and anthropology or psychology, the latter forming the direct introduction to ethics.

Physics.—The materialism of the Stoics was not open to some of the objections that had been leveled at the doctrine of the Cynics and of the atomists. In Plato's *Sophist* (246 C *et seq.*), the Eleatic stranger had argued as follows against those "who drag everything down to the corporeal": "They admit the existence of an animate body. Is soul then something existent? Yes. And the qualities of soul, as justice and wisdom—are they visible and tangible? No. Do they then exist? They are in a dilemma." Now, this reasoning is ineffective against the Stoics, who take boldly the one horn of the dilemma. That qualities of bodies (and therefore of the corporeal soul) exist they do not deny; but they assert most uncompromisingly that they are one and all (wisdom, justice, etc.) corporeal. And they strengthen their position by taking Plato's own definition (247 D), namely "being is that which has the power to act or be acted upon," and turning it against him. For this is only true of body; action, except by contact, is inconceivable; and they reduce every form of causation to the efficient cause, which implies the communication of motion from one body to another. Again and again, therefore, only body exists. The most real realities to Plato and to Aristotle had been thought and the objects of thought, *nous* and *noeta*, whether abstracted from sensibles or inherent in "matter," as the incognizable basis of all concrete existence. But this was too great an effort to last long. Such spiritualistic theories were nowhere really maintained after Aristotle and outside the circle of his immediate followers. The reaction came and left nothing of it all; for five centuries the dominant tone of the older and the newer schools alike was frankly materialistic. "If," says Aristotle, "there is no other substance but the organic substances of nature, physics will be the highest of the sciences," a conclusion which passed for axiomatic until the rise of Neoplatonism. The analogues therefore of metaphysical problems must be sought in physics; particularly that problem of the causes of things for which the Platonic idea and the Peripatetic "constitutive form" had been, each in its turn, received solutions.

Yet the doctrine that all existence is confined within the limits of the sensible universe—that there is no being save corporeal being or body—does not suffice to characterize the Stoic system; it is no less a doctrine of the Epicureans. It is the idea of tension or tonicity as the essential attribute of body, in contradistinction to passive inert matter, which is distinctively Stoic. The Epicureans leave unexplained the primary constitution and first movements of their atoms or elemental solids; chance or declination may account for them. Now, to the Stoics nothing passes unexplained; there is a reason (*logos*) for everything in nature. Everything which exists is at once capable of acting and being acted upon. In everything that exists, therefore, even the smallest particle, there are these two principles. By virtue of the passive principle the thing is susceptible of motion and modification; it is matter which determines substance. The active principle makes the matter a given determinate thing, characterizing and qualifying it, whence it is termed quality. For all that is or happens there is an immediate cause or antecedent; and as "cause" means "cause of motion," and only body can act upon body, it follows that this antecedent cause is itself as truly corporeal as the matter upon which it acts. Thus we are led to regard the active principle "force" as everywhere coextensive with "matter," as pervading and permeating it and together with it occupying and filling space. This is that famous doctrine of universal permeation, by which the axiom that two bodies cannot occupy the same space is practically denied. Thus that harmony of separate doctrines

which contributes to the impressive simplicity of the Stoic physics is only attained at the cost of offending healthy common sense, for body itself is robbed of a characteristic attribute. A thing is no longer, as Plato once thought, hot or hard or bright by partaking in abstract heat or hardness or brightness, but by containing within its own substance the material of these qualities, conceived as air currents in various degrees of tension. We hear, too, of corporeal days and years, of corporeal virtues and of actions (like walking) which are bodies. Obviously, again, the Stoic quality corresponds to Aristotle's essential form; in both systems the active principle, "the cause of all that matter becomes," is that which accounts for the existence of a given concrete thing. Only here, instead of assuming something immaterial (and therefore unverifiable), we fall back upon a current of air or gas, pneuma; the essential reason of the thing is itself material, standing to it in the relation of a gaseous to a solid body. Here, too, the reason of things—that which accounts for them—is no longer some external end to which they are tending; it is something acting within them, germinating and developing as from a seed in the heart of each separate thing that exists. By its prompting the thing grows, develops and decays, while this "germinal reason," the element of quality in the thing, remains constant through all its changes.

Psychology and Theory of Knowledge.—In the rational creatures—man and the gods—pneuma is manifested in a high degree of purity and intensity as an emanation from the world soul, itself an emanation from the primary substance of purest ether—a spark of the celestial fire, or, more accurately, fiery breath, which is a mean between fire and air, characterized by vital warmth more than by dryness. The physical basis of Stoic psychology deserves the closest attention. On the one hand, soul is corporeal, else it would have no real existence, would be incapable of extension in three dimensions (and therefore of equable diffusion all over the body), incapable of holding the body together, as the Stoics contended that it does, herein presenting a sharp contrast to the Epicurean tenet that it is the body which confines and shelters the light vagrant atoms of soul. On the other hand, this corporeal thing is veritably and identically reason, mind and ruling principle; in virtue of its divine origin Cleanthes can say to Zeus "We too are thy offspring," and Seneca can calmly insist that, if man and God are not on perfect equality, the superiority rests rather on our side. What God is for the world, that the soul is for man. The cosmos must be conceived as a single whole, its variety being referred to varying stages of condensation in pneuma. So, too, the human soul must possess absolute simplicity, its varying functions being conditioned by the degrees or species of its tension. It follows that of "parts" of the soul, as previous thinkers imagined, there can be no question; all that can consistently be maintained is that from the centre of the body—the heart—seven distinct air currents are discharged to various organs, which are so many modes of the one soul's activity. The ethical consequences of this position will be seen at a later stage.

With this psychology is intimately connected the Stoic theory of knowledge. From the unity of soul it follows that all psychical processes—sensation, assent, impulse—proceed from reason, the ruling part; that is to say, there is no strife or division: the one rational soul alone has sensations, assents to judgments, is impelled toward objects of desire just as much as it thinks or reasons.

Formal Logic.—The dynamic materialism of the Stoics led them to develop a nominalist two-valued logic of propositions which was of considerable importance and differed radically from the Aristotelian logic of terms. Admitting the reality only of the individual, subject to a series of changes, the Stoics made central the singular proposition and interpreted the relation between subject and predicate, not as that of inherence or inclusion, but as that of concomitance or sequence, thus moving away from the Aristotelian notion of essence toward the concept of law. A detailed account of Stoic logic, both as distinct from the syllogistic of Aristotle and as related to the logic of the Megarian school of philosophy, is given in the article Logic, History of, where its technical aspects are discussed at length. It may be claimed that the Stoics, insisting on the logical primacy of words, not meanings, were pioneers of formalistic logic.

Cosmopolitanism.—In their view of man's social relations the Stoics were greatly in advance of preceding schools. We saw that virtue is a law which governs the universe: that which reason and God ordain must be accepted as binding upon the particle of reason which is in each one of us. Human law comes into existence when men recognize this obligation; justice is therefore natural and not something merely conventional. The opposite tendencies, to allow to the individual responsibility and freedom and to demand of him obedience to law, are both features of the system; but in virtue even of the freedom which belongs to him *qua* rational, he must recognize the society of rational beings of which he is a member and subordinate his own ends to the ends and needs of this society. Those who own one law are citizens of one state, the city of Zeus, in which men and gods have their dwelling. In that city all is ordained by reason working intelligently, and the members exist for the sake of one another; there is an intimate connection between them which makes all the wise and virtuous friends, even if personally unknown, and leads them to contribute to one another's good. Their intercourse should find expression in justice, in friendship, in family and political life. But practically the Stoic philosopher always had some good excuse for withdrawing from the narrow political life of the city in which he found himself. The circumstances of the time, such as the decay of Greek city life, the foundation of large territorial states under absolute Greek rulers and afterward the rise of the world empire of Rome, aided in developing the leading idea of Zeno's *Republic*. There he had anticipated a state without family life, without law courts or coins, without schools or temples, in which all differences of nationality would be merged in the common brotherhood of man. This cosmopolitan citizenship remained all through a distinctive Stoic dogma; when first announced it must have had a powerful influence upon the minds of men, diverting them from the distractions of almost parochial politics to a boundless vista.

Religion.—The religious problem had peculiar interest for the school which discerned God everywhere as the ruler and upholder, and at the same time the law, of the world that He had evolved from Himself. The physical groundwork lends a religious sanction to all moral duties, and Cleanthes' hymn is evidence how far a system of natural religion could go in providing satisfaction for the cravings of the religious temper:

> Most glorious of immortals, O Zeus of many names, almighty and everlasting, sovereign of nature, directing all in accordance with law, thee it is fitting that all mortals should address. . . . Thee all this universe, as it rolls circling round the earth, obeys wheresoever thou dost guide, and gladly owns thy sway. Such a minister thou holdest in thy invincible hands—the two-edged, fiery, everliving thunderbolt, under whose stroke all nature shudders. No work upon earth is wrought apart from thee, lord, nor through the divine ethereal sphere, nor upon the sea; save only whatsoever deeds wicked men do in their own foolishness. Nay, thou knowest how to make even the rough smooth, and to bring order out of disorder; and things not friendly are friendly in thy sight. For so hast thou fitted all things together, the good with the evil, that there might be one eternal law over all. . . . Deliver men from fell ignorance. Banish it, father, from their soul, and grant them to obtain wisdom, whereon relying thou rulest all things with justice.

To the orthodox theology of Greece and Rome the system stood in a twofold relation, as criticism and as rationalism. That the popular religion was open to criticism hardly needed to be pointed out: the forms of worship could be described as trivial or as mischievous, the myths as immoral. But Zeno declared images, shrines, temples, sacrifices, prayers and worship to be of no avail. A really acceptable prayer, he taught, can only have reference to a virtuous and devout mind: God is best worshiped in the shrine of the heart by the desire to know and obey Him.

At the same time the Stoics felt at liberty to defend and uphold the truth in polytheism. Not only is the primitive substance God, the one supreme being, but divinity must be ascribed to His manifestations—to the heavenly bodies, which are conceived, like Plato's created gods, as the highest of rational beings, to the forces of nature, even to deified men; and thus the world was peopled with divine agencies. Moreover, the myths were rationalized and allegorized. The search for a hidden meaning besides the literal one had been begun by Democritus, Empedocles, the Sophists and the Cynics. It remained for Zeno to carry this to a much greater

extent and to seek out or invent "natural principles" (*logoi physikoi*) and moral ideas in all the legends and in the poetry of Homer and Hesiod.

THE MIDDLE STOA AND ROME

The principal representatives of the middle period of Stoicism were Panaetius (*q.v.*) and Posidonius (Poseidonius, *q.v.*); and the introduction of Stoicism to Rome, which took place during this period, was an event of great moment both in the history of the school and in that of the city. After the first sharp collision with the jealousy of the national authorities it found a ready acceptance and made rapid progress among the noblest families. It has been well said that the old heroes of the republic were unconscious Stoics, fitted by their narrowness, their stern simplicity and devotion to duty for the earnestness of the new doctrine. In Greece its insensibility to art and the cultivation of life was a fatal defect; not so with the shrewd men of the world, desirous of qualifying as advocates or jurists. It supplied them with an incentive to scientific research in archaeology and grammar; it penetrated jurisprudence until the belief in the ultimate identity of the *jus gentium* (*see* ROMAN LAW) with the law of nature modified the praetor's edicts for centuries. Even to the prosaic religion of old Rome it became in some sort a support. Q. Mucius Scaevola, following Panaetius, explained that statesmen had established this public institution in the service of order midway between superstition and the barren truths of philosophy. Soon the influence of the pupils reacted upon the doctrines taught. Of speculative interest the ordinary Roman had little; for abstract discussion and controversy he cared nothing. All had to be simplified and disengaged from technical subtleties. To attract his Roman pupils Panaetius would naturally choose simple topics susceptible of rhetorical treatment or of application to individual details. He was the representative, not merely of Stoicism, but of Greece and Greek literature and would feel pride in introducing its greatest masterpieces. Among all that he studied, he valued most the writings of Plato. He admired the classic style, the exquisite purity of language, the flights of imagination, but he admired above all the philosophy. He marks a reaction of the Hellenic spirit against the austerity of the first Stoics.

His pupil, Posidonius, a man of great natural austerity, carried still further the eclectic tendencies evident in Panaetius. His wide learning in history and in natural science and his moral seriousness appealed to the cultivated Roman mind. His religious and psychological teaching, deeply influenced by Plato, prepared the way for the late Stoa with its dominantly religious note. The thought of such men as Cicero, Seneca, Manilius and Plutarch is greatly indebted to him.

No account of Stoicism in Rome can be complete which fails to mention the part played by Cicero who, though not a professing Stoic, was sympathetic to much in Stoicism. His philosophical writings contain exposition and discussion of Stoic doctrine in all branches of philosophy, and without them our knowledge of Stoicism would be much less adequate than it is.

THE LATER STOICS

The writings of the later Stoics have come down to us, if not entire, in great part, so that L. Annaeus Seneca, L. Annaeus Cornutus, Aulus Persius, Lucan, Epictetus and Marcus Aurelius (*qq.v.*) are known at first hand. They do not profess to give a scientific exposition of doctrine, and may therefore be dismissed somewhat briefly. We learn much more about the Stoic system from the scanty fragments of the first founders, or even from the epitomes of Diogenes Laërtius and Stobaeus, than from these writers. They testify to the restriction of philosophy to the practical side and to the increasing tendency, ever since Panaetius, toward a relaxation of the rigorous ethical doctrine and its approximation to the form of religious conviction.

This finds most marked expression in the doctrines of submission to Providence and universal philanthropy. Only in this way could they hold their ground, however insecurely, in face of the revival of religion inaugurated by the emperor Augustus. In passing to Rome, Stoicism had quitted the school for actual life; and the fall of the republic was indeed a gain, for it released so much intellectual activity from civic duties. The life and death of Cato, moreover, fired the imagination of a degenerate age in which he stood out both as a Roman and a Stoic. To a long line of illustrious successors, including men like Thrasea Paetus and Helvidius Priscus, Cato bequeathed his resolute opposition to the dominant power of the times; unsympathetic, unpractical, but fearless in demeanour, they were a standing reproach to the corruption and tyranny of their age. In the early years of the empire, however, when Augustus had restored order, philosophy had become bolder and addressed every class in society, public lectures and spiritual direction being the two forms in which it mainly showed activity. Books of direction were written in Greek by Q. Sextius (as afterward in Latin by Seneca), almost the only Roman who had the ambition to found a sect, though in ethics he mainly followed Stoicism. His contemporary Papirius Fabianus was the popular lecturer of that day, producing a powerful effect by his denunciations of the manners of the time. Under Tiberius, Sotion and Attalus were attended by crowds of hearers.

Seneca.—Seneca is the most prominent leader in the direction which Roman Stoicism now took. His penetrating intellect had mastered the subtleties of the system of Chrysippus, but they seldom appear in his works, at least without apology. Incidentally we meet there with the doctrines of pneuma and of tension, of the corporeal nature of the virtues and the affections, and much more to the same effect. But his attention is claimed for physics chiefly as a means of elevating the mind and as making known the wisdom of Providence. Chrysippus had argued plausibly that natural evil was a thing indifferent—that even moral evil was required in the divine economy as a foil to set off good. The prosperity of the wicked and the calamity of the just he had explained in various ways: sometimes he alleged the forgetfulness of higher powers; sometimes he fell back upon the necessity of these contrasts and grotesque passages in the comedy of human life. Seneca gives the true Stoic answer in his treatise *On Providence:* the wise man cannot really meet with misfortune; all outward calamity is a divine instrument of training, designed to exercise his powers and teach the world the indifference of external conditions. In the soul Seneca recognizes an effluence of the divine spirit, a god in the human frame; in virtue of this he maintains the essential dignity and internal freedom of man in every human being. Yet, in striking contrast with this orthodox tenet, he has a vivid concept of the weakness and misery of men, the hopelessness of the struggle with evil, whether in society or in the individual. Thus he describes the body (which, after Epicurus, he calls the flesh) as a mere husk or fetter or prison of the soul; with its departure begins the soul's true life. Sometimes, too, he writes as if he accepted an irrational as well as a rational part of the soul. In ethics, if there is no novelty of doctrine, there is a surprising change in the mode of its application. The ideal sage has receded; philosophy comes as a physician, not to the whole but to the sick. We learn that there are various classes of patients in "progress"; *i.e.*, on their way to virtue, making painful efforts toward it. The first stage is the eradication of vicious habits: evil tendencies are to be corrected and a guard kept on the corrupt propensities of the reason. Suppose this achieved, we have yet to struggle with single attacks of the passions: irascibility may be cured, but we may succumb to a fit of rage. To achieve this second stage the impulses must be trained in such a way that the fitness of things indifferent may be the guide of conduct. Even then it remains to give the will that property of rigid infallibility without which we are always liable to err, and this must be effected by the training of the judgment. Other peculiarities of the later Stoic ethics are due to the condition of the times. In a time of moral corruption and oppressive rule, a feeling of insecurity led the student of philosophy to seek in it a refuge against the vicissitudes of fortune.

Musonius.—From Seneca we turn, not without satisfaction, to men of sterner mold, such as Musonius Rufus, who certainly deserves a place beside his more illustrious disciple, Epictetus. As a teacher he commanded universal respect, and wherever we catch a glimpse of his activity he appears to advantage. His philosophy, however, is yet more concentrated upon practice than Seneca's,

and in ethics he is almost at the position of Aristo. Epictetus testifies to the hold that Musonius acquired upon his pupils.

Epictetus.—Stoicism bore abundant fruit in the life and teaching of Epictetus. His character, which rose superior to weakness, poverty and slave's estate, is presented to us in the *Discourses* edited by his disciple Arrian as a model of religious resignation, of forbearance and love toward our brethren, that is, toward all men, since God is our common father. With him even the "physical basis" of ethics takes the form of a religious dogma—the providence of God and the perfection of the world. We learn that he regards the *daimon* or "guardian angel" as the divine part in each man; sometimes it is more nearly conscience, at other times reason. His ethics, too, have a religious character. He begins with human weakness and man's need of God: whoso would become good must first be convinced that he is evil. Submission is enforced by an argument which almost amounts to a retractation of the difference between things natural and things contrary to nature, as understood by Zeno. Would you be cut off from the universe? he asks. Go to, grow healthy and rich. But if not, if you are a part of it, then be resigned to your lot.

Marcus Aurelius.—Epictetus is marked out by his renunciation of the world. He is followed by a Stoic emperor, Marcus Aurelius Antoninus, who, though in the world, was not of it. The *Meditations* give no systematic exposition of belief, but there are many indications of the religious spirit, together with an almost Platonic psychology. Following Epictetus, he speaks of man as a corpse bearing about a soul; at another time he has a threefold division—(1) body, (2) soul, the seat of impulse, and (3) intelligence, the proper ego. In all he writes there is a vein of sadness: the flux of all things, the vanity of life, are thoughts which perpetually recur, along with resignation to the will of God, forbearance toward others and the religious longing to be rid of the burden and to depart to God. These peculiarities in Marcus Aurelius may perhaps be explained in harmony with the older teaching; but, when taken in connection with the rise of Neoplatonism and the revival of superstition, they are certainly significant. A generation after Marcus Aurelius there is hardly a Stoic professor to be named.

See also ETHICS, HISTORY OF; LOGIC, HISTORY OF.

BIBLIOGRAPHY.—E. Zeller, *Die Philosophie der Griechen*, iii, 1, 5th ed. (Leipzig, 1923); R. Hirzel, *Untersuchungen zu Ciceros philosophischen Schriften*, 3 vol. (Leipzig, 1876–83); F. Ogereau, *Essai sur le système des Stoïciens* (Paris, 1885); A. C. Pearson (ed.), *Fragments of Zeno and Cleanthes* (London, 1891); A. Schmekel, *Die Philosophie der mittleren Stoa* (Berlin, 1892); A. Bonhöffer, *Epiktet und die Stoa* (Stuttgart, 1890), *Die Ethik des Stoikers Epiktet* (Stuttgart, 1894); E. Bréhier, *Chrysippe* (Paris, 1910); E. V. Arnold, *Roman Stoicism* (Oxford, 1911); R. D. Hicks, *Stoic and Epicurean* (London, New York, 1910); J. von Arnim (ed.), *Stoicorum veterum fragmenta*, 3 vol. (Leipzig, 1903–05); E. R. Bevan, *Stoics and Sceptics* (Oxford, 1913); V. Brochard, *Études de philosophie ancienne et moderne* (Paris, 1912); I. Heinemann, *Poseidonios' metaphysische Schriften* (Breslau, 1921–28); K. Reinhardt, *Poseidonios* (Munich, 1921), *Kosmos und Sympathie* (Munich, 1926); E. Grumach, *Physis und Agathon in der alten Stoa* (Berlin, 1932); P. Barth, *Die Stoa*, 6th ed. (Stuttgart, 1946); J. Lukasiewicz, "Zur Geschichte des Aussagenskalküls," *Erkenntnis*, v (Leipzig, 1935); M. Pohlenz, *Grundfragen der stoischen Philosophie* (Göttingen, 1940), *Die Stoa*, 2 vol. (Göttingen, 1948–49), *Stoa und Stoiker* (Zürich, 1950); M. van Straaten, *Panétius* (Amsterdam, 1946); A. Virieux-Reymond, *La Logique et l'épistémologie des Stoïciens* (Chambéry, 1950); B. Mates, *Stoic Logic* (Berkeley, Calif., 1953).

(R. D. HI.; A. R. C. D.)

STOKE NEWINGTON, a metropolitan borough (1899) of London, Eng., which under the London Government Act 1963 was to be amalgamated on April 1, 1965, with Hackney and Shoreditch to form the new London borough of Hackney (*see* LONDON). The metropolitan borough of Stoke Newington is bounded by Tottenham, Hornsey, Islington and Hackney. Pop. (1961) 52,280. Area 1.4 sq.mi. Stoke Newington and Hackney North form a parliamentary constituency. Referred to as Neutone in Domesday Book, the district has always been largely residential, and since World War II many blocks of council flats have been built. Its industries are chiefly concerned with clothing and furniture. In 1613 Sir Hugh Myddelton's New river (a canal to supply London with water) was finished; it now ends in the waterworks and reservoirs north of Clissold park. The old parish church of St. Mary has a brick 16th-century tower and a medieval nave. The new parish church, designed by Sir Giles Gilbert Scott and consecrated in 1858, is in the Early English style. There are some fine early 18th-century houses in the High street.

Stoke Newington became a stronghold of nonconformity in the 17th century, and Daniel Defoe studied and in later life lived there. Isaac Watts, the theologian and hymn writer, who also studied there, was a tutor at Hartopp house and later lived and died at Abney house. Samuel Rogers, the poet, was born and educated there, and Edgar Allan Poe, the American writer, was at the Manor House school for three years. In the 18th century well-to-do Jewish families, including the D'Israelis, and Quaker families, such as the Hoares and Listers, settled in Stoke Newington.

STOKE-ON-TRENT, a city and a county and parliamentary borough of Staffordshire, Eng., 16 mi. N.N.W. of Stafford and comprising the district known as the Potteries (*q.v.*). Pop. (1961) 265,506. Area 33.1 sq.mi. The county borough was constituted in 1910 from a federation of the Potteries towns—Tunstall, Burslem, Hanley, Stoke-upon-Trent, Fenton and Longton. In 1922 it was considerably extended, in 1925 it became a city, in 1928 the chief magistrate became lord mayor and in 1929 the boundaries were again extended. Of its six pre-federation town halls, that of Stoke-upon-Trent is now the meeting place of the city council and the chief administrative centre.

The North Staffordshire Technical college contains the Solon Ceramic library, and the British Ceramic Research laboratories were opened in 1951. Stoke co-operated with the county council and Burton-on-Trent to establish the University College of North Staffordshire, founded in 1949 at Keele, about 5 mi. W. of Stoke. Next in importance to the potteries, which make almost every kind of ware, using local coarse clay for bricks, tiles, etc., and Cornish kaolin for finer work, come the coal mines. A statue of the great potter, Josiah Wedgwood (*q.v.*; 1730–95), faces the railway station. Other names made famous by local potteries are Spode, Minton and Copeland. Arnold Bennett, who pictured the area in his novels of "the Five Towns" (there are in fact six), was born in Hanley. Stoke returns three members to parliament.

STOKE POGES, parish and village in Buckinghamshire, Eng., about 2 mi. N. of Slough. Pop. (1951) 2,470. Area 4 sq.mi. It is a residential area. Humbert de Pugeis held the manor in 1255. St. Giles's church is of mixed architectural styles; Thomas Gray is buried in the churchyard, held to be the original of his *Elegy*. William Penn (*q.v.*) built the manor house, later a golf club.

STOKES, SIR (FREDERICK) WILFRID SCOTT K.B.E. (1860–1927), engineer and inventor, son of Scott Nasmyth Stokes, was born in 1860 at Liverpool. After some years in railway work he joined Ransomes and Rapier of Ipswich, becoming managing director and chairman. He was distinguished in many branches of engineering. His inventions, for which some fifty patents were granted, included improvements in sluice gates, breakdown cranes, railway, hydraulic, refrigerating and cement making machinery, ordnance and projectiles. He had a share in many notable engineering achievements at home and abroad, including the Manchester ship canal and the irrigation works in India, Egypt and the Sudan. He was responsible for the design of the sluice and lock gates, and he was present both at the opening of the Assuan dam 1901 and of the Sennar dam 1926.

From 1915 until 1918 Stokes served on Ministry of Munitions inventions committees, and he was also chairman of the East Anglian and other munitions committees. In 1915 he acquired fame by the timely invention of the Stokes trench mortar and ammunition.

At great personal risk, without previous technical knowledge of explosives he perfected the shells of which more than twenty million rounds were used with deadly effect in the trenches and at Zeebrugge and elsewhere. He also promoted the invention of effective anti-aircraft shells. For his public services he was created K.B.E. in 1917. Stokes was president of the British Engineers' Association 1916–1918.

See *Handbook of Stokes' Trench Mortar Equipment* (1920); *History of the Ministry of Munitions* (1922); T. E. Lawrence, *Revolt in the*

Desert (1927); W. Stokes, "Sluices and Lock Gates of the Nile Reservoir, Assuan," *Proc. Inst. Civil Engineers* (1902–03); "The Stokes Gun and Shell and their development," *Engineering*, June 28, 1918, "The Sennar Dam," *Engineering*, Jan. 22, 1926, "The late Sir W. Stokes," *Engineering*, Feb. 11, 1927; "The Stokes Gun," *Whitehall Gazette*, Oct. 1919; "Sir W. Stokes," *The Engineer*, Feb. 11, 1927.

STOKES, SIR GEORGE GABRIEL, 1ST BART. (1819–1903), British mathematician and physicist, who with James Clerk Maxwell and Lord Kelvin especially contributed to the fame of the Cambridge school of mathematical physics in the middle of the 19th century. Born on Aug. 13, 1819, at County Sligo, Ire., in 1837 he entered Pembroke college, Cambridge, where he became a fellow in 1841. He lost his fellowship when he married in 1857, but 12 years later, under new statutes, he was re-elected. In 1902 he was elected master. As Lucasian professor of mathematics (1849–1903), secretary (1854–85) and president (1885–90) of the Royal society, he held three offices which had only once before been held by one man, Sir Isaac Newton. He was a member of parliament for the university from 1887 to 1892, was created baronet in 1889 and died on Feb. 1, 1903.

The greater part of Stokes's work was concerned with waves and the transformations imposed on them during their passage through various media. Stokes's first published papers, in 1842 and 1843, were on the steady motion of incompressible fluids and some cases of fluid motion; these were followed in 1845 by one on the friction of fluids in motion and the equilibrium and motion of elastic solids, and in 1850 by another on the effects of the internal friction of fluids on the motion of pendulums. To the theory of sound he made several contributions, including a discussion of the effect of wind on the intensity of sound and an explanation of how the intensity is influenced by the nature of the gas in which the sound is produced. These inquiries together put the science of hydrodynamics on a new footing and provided a key not only to the explanation of many natural phenomena, such as the suspension of clouds in air and the subsidence of ripples and waves in water, but also to the solution of practical problems, such as the flow of water in rivers and channels and the skin resistance of ships.

His best-known researches are perhaps those on the undulatory theory of light. His first papers on the aberration of light appeared in 1845 and 1846 and were followed in 1848 by one on the theory of certain bands seen in the spectrum. In 1849 his paper on the dynamical theory of diffraction showed that the plane of polarization must be perpendicular to the direction of vibration. Two years later he discussed the colours of thick plates; and in 1852 in his famous paper on the change of refrangibility of light, he described the phenomenon of fluorescence, as exhibited by fluorspar and uranium glass, materials which he viewed as having the power to convert invisible ultraviolet rays into rays of wave lengths which are visible (*see* LUMINESCENCE: *Early History: Fluorescence and Luminescence*). A mechanical model illustrating the dynamical principle of Stoke's explanation was shown in 1883, during a lecture at the Royal institution, by Lord Kelvin, who said he had heard an account of it from Stokes many years before and had repeatedly but vainly begged him to publish it. In the same year, 1852, there appeared the paper on the composition and resolution of streams of polarized light from different sources, and in 1853 an investigation of the metallic reflection exhibited by certain nonmetallic substances. About 1860 he was engaged in an inquiry on the intensity of light reflected from, or transmitted through, a pile of plates; and in 1862 he prepared for the British association a valuable report on double refraction, which marks a period in the history of the subject in England. A paper on the long spectrum of the electric light bears the same date and was followed by an inquiry into the absorption spectrum of blood.

The discrimination of organic bodies by their optical properties was treated in 1864; and later, in conjunction with the Rev. W. Vernon Harcourt, he investigated the relation between the chemical constitution and the optical properties of various glasses, with reference to the conditions of transparency and the improvement of achromatic telescopes. A still later paper connected with the construction of optical instruments discussed the theoretical limits to the aperture of microscopical objectives. In other departments of physics may be mentioned his paper on the conduction of heat in crystals (1851) and his inquiries in connection with the radiometer; his explanation of the light border frequently noticed in photographs just outside the outline of a dark body seen against the sky (1883); and still later his theory of the Röntgen rays, which he suggested might be transverse waves traveling as innumerable solitary waves, not in regular trains. Two long papers published in 1849—one on attractions and Clairaut's theorem and the other on the variation of gravity at the surface of the earth—also demand notice, as do his mathematical memoirs on the critical values of the sums of periodic series (1847) and on the numerical calculation of a class of definite integrals and infinite series (1850) and his valuable discussion of a differential equation relating to the strains, stresses and other factors involved in the breaking of railway bridges (1849).

Many of his discoveries were only touched upon in lectures. An instance is his work in the theory of spectrum analysis. Some of Stokes's friends and pupils claimed that he had anticipated the work of G. R. Kirchhoff but Stokes maintained that he had failed to see an essential step in the argument and disclaimed priority. As Lucasian professor, Stokes announced that he wished to help any member of the university in his mathematical studies, and pupils were glad to consult him, even after they had become colleagues. During the 30 years of his secretaryship to the Royal society he advanced the cause of mathematical and physical science, not only by his own investigations, but by suggesting problems for inquiry and inciting men to attack them.

He received the Rumford medal in 1852 and the Copley medal in 1893. His numerous other honours included the Prussian order of merit.

Sir George Stokes's mathematical and physical papers were published in a collected form in five volumes; the first three (1880, 1883 and 1901) under his own editorship and the two last (1904 and 1905) under that of Sir Joseph Larmor, who also selected and arranged the *Memoir and Scientific Correspondence* of Stokes published at Cambridge in 1907. Stokes was the author of *Light* (1884–87) and *Natural Theology* (1891).

STOKES, WHITLEY (1830–1909), British lawyer and Celtic scholar, was educated at Trinity college, Dublin. He became a barrister and in 1862 went out to India. In 1877 he became legal member of the viceroy's council, and drafted the codes of civil and criminal procedure. After his return to England in 1882 he devoted himself to the Celtic studies which made him famous. He studied Irish, Breton and Cornish texts, and among his numerous works are editions of *Three Irish Glossaries* (1862); *Three Middle-Irish Homilies* (1877); and *Old Irish Glosses at Würzburg and Carlsruhe* (1887). He was one of the editors of the *Irische Texte* published at Leipzig (1880–1900); and he edited and translated *Lives of Saints From the Book of Lismore* (1890). His principal legal work was *The Anglo-Indian Codes* (1887). He died in London, April 13, 1909.

STOKESAY, a fortified manor house in Shropshire, Eng., 6½ mi. N.W. of Ludlow. Stokesay castle is one of the most remarkable moated manor houses surviving in England. It was built in 1240 by John de Verdon with banqueting hall, north tower and solar (paneled in the 17th century). The south tower was added in 1291 when the Ludlows bought Stokesay and fortified it. The 16th-century half-timbered gatehouse replaces the old entrance tower. From 1620 to 1869 (when it was thoroughly repaired) Stokesay belonged to the lords Craven.

STOKESIA, an American genus of the aster family (Compositae, *q.v.*), occasionally cultivated for ornament, often called Stokes's aster. The single species, *Stokesia laevis*, is a fibrous-rooted perennial, one to two feet high, with blue or purplish flower heads an inch or two wide in wild plants, or up to three or four inches wide in cultivated forms. White, pink, lilac and creamy yellow forms are also known in cultivation. Well-developed plants are loosely branched, with a flower head terminating each of the several branches. The basal leaves are long-stalked, usually with a rather narrow blade; the middle and upper ones are shorter and often relatively wider, sometimes with a broad, clasping base. The bracts beneath the flower heads are weakly but rather conspicuously spinose-ciliate. The plant is anomalous in its tribe (Ver-

nonieae) in having the marginal corollas of the head enlarged and five-cleft, like those of the chicory tribe. Native to the coastal plain from South Carolina to Florida and Louisiana, it is hardy as far north as Massachusetts and central New York. It grows best in well-drained, sandy loam and blooms from August until frost. It is propagated by seed and by division. (A. Ct.)

STOLBERG, FRIEDRICH LEOPOLD, Graf zu (1750–1819), German poet, the younger son of Count Christian Stolberg, was born at Bramstedt in Holstein on Nov. 7, 1750. He studied in Göttingen and was a member of the famous Göttinger Hain or Dichterbund. After leaving the university he made a journey to Switzerland with his brother Christian, in company with Goethe. In 1777 he was appointed envoy of the prince-bishop of Lübeck at the court of Copenhagen, but often stayed at Eutin, where he was the intimate associate of J. H. Voss. In 1782 he married Agnes von Witzleben, whom he celebrated in his poems. After her early death in 1788, he became Danish envoy at the court of Berlin, and contracted a second marriage with the countess Sophie von Redern in 1789. In 1791 he was appointed president of the Lübeck episcopal court at Eutin; he resigned this office in 1800 and, retiring to Münster in Westphalia, there joined, with his whole family, the eldest daughter only excepted, the Roman Catholic Church. For this step he was severely attacked by his former friend Voss (*Wie ward Fritz Stolberg ein Unfreier?* 1819). After living for a while (from 1812) in the neighbourhood of Bielefeld, Stolberg moved to his estate of Sondermühlen near Osnabrück, where he died on Dec. 5, 1819. He wrote many odes, ballads, satires and dramas—among the last the tragedy *Timoleon* (1784), translations of the *Iliad* (1778), of Plato (1796–97), Aeschylus (1802) and Ossian (1806); he published, in 1815, a *Leben Alfreds des Grossen*, and a voluminous *Geschichte der Religion Jesu Christi* (17 vol., 1806–18).

Stolberg's brother, Christian, Graf zu Stolberg (1748–1821), was also a poet. Born at Hamburg on Oct. 15, 1748, he became a magistrate at Tremsbüttel in Holstein in 1777, and died on Jan. 18, 1821. Of the two brothers Friedrich was undoubtedly the more talented, but Christian, though not a poet of high originality, excelled in the utterance of gentle sentiment. They published together a volume of poems, *Gedichte* (edited by H. C. Boie, 1779); *Schauspiele mit Chören* (1787), their object in the latter work being to revive a love for the Greek drama; and a collection of patriotic poems *Vaterländische Gedichte* (1815). Christian von Stolberg was the sole author of *Gedichte aus dem Griechischen* (1782), a translation of the works of Sophocles (1787), and of a poem in seven ballads, *Die weisse Frau* (1814), which last attained considerable popularity.

The collected works of Christian and Friedrich Leopold zu Stolberg were published in 20 vol. (1820–25; 2nd ed., 1827). Friedrich's correspondence with F. H. Jacobi appeared in Jacobi's *Briefwechsel* (1825–27).

Selections from the poetry of the two brothers were published in A. Sauer, *Der Göttinger Dichterbund*, iii (Kürschner's *Deutsche Nationalliteratur*, vol. 50, 1896).

STOLBERG, a town of Germany formerly in the Prussian Rhine province and after 1945 in North Rhine-Westphalia, about 6 mi. E. of Aachen. The population of Stolberg in 1961 was 37,462. Its prosperity began in the 17th century, when French refugees introduced brass founding. A castle was built on the site of a church said to have been used by Charlemagne. Leading industries developed there include metalworking in zinc, brass and iron.

STOLE, a liturgical vestment of the Roman Catholic Church, peculiar to the higher orders; *i.e.*, deacons, priests and bishops. Its equivalents in the Eastern Churches are the epitrachelion of priests and the orarion of deacons. *See* Vestments, Ecclesiastical.

STOLP (Slupsk), a town formerly in Pomerania, Ger., and after 1945 in Koszalin province, Pol. Pop. (1960) 53,000.

Stolp received town rights in 1273. From the 14th to the 16th century it was a member of the Hanseatic league. Until 1637, when it passed to Brandenburg, the town usually belonged to the dukes of Pomerania. The large church of St. Mary, with a 14th-century tower, the 16th-century Renaissance castle which became a prison, one of the ancient towngates restored in 1872 and the church of St. John (13th century) are the oldest buildings.

STOLYPIN, PETER ARCADIEVICH (1863–1911), Russian statesman, was born at Baden-Baden in 1863. He was educated at St. Petersburg, and in 1902 he became governor of Grodno, and in 1903 of Saratov, where he was a firm administrator. In 1906 Stolypin was minister of internal affairs, and in July succeeded Goremykin as minister-president. (*See* Russia.) His relentless policy led to attempts upon his life. In Aug. 1906 a bomb was exploded at his summer residence, injuring a daughter, and on Sept. 14, 1911, he was shot in a theatre at Kiev, before the eyes of the imperial family, by a Jew, Mordka Bogrov. He died on Sept. 18, 1911.

STOMACH, the baglike digestive organ which in man lies in the upper left part of the abdomen between the esophagus and the duodenum. For anatomical details *see* Gastrointestinal Tract. For the diseases of the stomach in general *see* Gastrointestinal Tract, Diseases of; and for special forms, Gastric and Duodenal Ulcer; Gastritis; Indigestion; etc.

See Digestion; *see* also references under "Stomach" in the Index.

STOMACH, SURGERY OF. The first successful major operations upon the stomach were performed by Theodor Billroth (*q.v.*) of Vienna. His procedures included two separate approaches for excisional therapy of cancer of the stomach: (1) a partial removal of the stomach with re-establishment of direct continuity between the cut end of the stomach and the duodenum (the Billroth I gastric resection, first performed successfully in 1881), and (2) a similar partial removal of the stomach but with the construction of an indirect continuity between the stomach and the jejunum (Billroth II gastric resection, 1885). Surgery of the stomach was not possible before Billroth chiefly because the required laparotomy (*q.v.*) was not feasible until general anesthesia and antisepsis were developed.

Stomach operations are performed mainly for (1) cancer of the stomach, (2) ulcer of the stomach or duodenum and (3) hypertrophic pyloric stenosis.

Cancer of the Stomach.—Surgical treatment for this condition involves a subtotal or total removal of the stomach, reconstitution of the continuity of the intestinal tract so that food can again pass into the small intestine, and removal of the greater omentum, spleen and surrounding lymph node areas. If there are obvious extensions into other organs that can be sacrificed (colon, left lobe of liver, pancreas, etc.) these are removed as well. When all the apparently cancerous tissue is removed, about one-third of the patients are still alive after five years. Surgery for cancer of the stomach has about a 10% mortality; without such treatment, less than 5% of the patients survive for three years.

Ulcer of the Stomach or Duodenum.—The aim of surgery for gastric or duodenal ulcers is different from that for cancer. The aim is not primarily to remove the ulcer but instead to reduce the secretion of acid by the parietal cells in the central portion (corpus) of the stomach so that the ulcer will heal. This may be accomplished by removing most of the corpus (*i.e.*, by subtotal gastric resection) or by cutting down on the strongest stimulators of acid secretion. These stimulators include nervous stimuli transmitted by the two abdominal branches of the vagus nerve, which may be cut bilaterally at the level of the diaphragm or selectively to the stomach only, and hormonal stimuli that arise from the lower portion (antrum) of the stomach; the hormonal stimuli may be eliminated by removing the antrum. Surgery for ulcer, therefore, usually involves a combination of one or more of the following items: cutting the vagal nerves, subtotal resection of the corpus, removal of the antrum, and often a side-tracking procedure to permit more easy exit of food from the stomach. Surgery for ulcer has about a 1%–2% mortality and gives good results in 90% of cases.

While these principles apply to ulcers of both the stomach and duodenum, treatment of duodenal ulcer differs in that it is directed at the adjoining organ, the stomach, where the acid is secreted, rather than at the ulcerated duodenum itself. (*See* also Gastric

AND DUODENAL ULCER.)

Hypertrophic Pyloric Stenosis.—This condition represents a hypertrophy, or overgrowth, of the muscle that surrounds the pylorus (outlet) of the stomach. Its highest incidence is in first-born males, in whom it occurs at a rate of 1 in 300 at about the third week of life. If the condition is not treated surgically, the stomach contents cannot go through the narrowed ring and the mortality is 50%–75%. The preferred method of surgery involves cutting the thickened muscular ring without cutting into the interior of the pyloric canal (Fredet-Ramstedt operation). This operation has about a 0.5% mortality and gives good results in 99% of cases.

See also ABDOMEN, SURGERY OF: *Gastrointestinal Surgery;* GASTROINTESTINAL TRACT, DISEASES OF: *Stomach.*

BIBLIOGRAPHY.—L. R. Dragstedt, "Vagotomy for Gastroduodenal Ulcer," *Ann. Surg.,* 122:973–989 (Dec. 1945); Henry N. Harkins and Lloyd M. Nyhus (eds.), *Surgery of the Stomach and Duodenum* (1962); N. C. Tanner, "The Indications for Surgery in Peptic Ulcer," *Edinburgh Medical Journal,* 58:261–278 (June 1951); O. H. Wangensteen, *Cancer of the Esophagus and the Stomach* (1951); R. Maingot, *Abdominal Operations,* 4th ed. (1961); R. W. Postlethwait *et al.* (eds.), *Results of Surgery for Peptic Ulcer* (1963). (H. N. Hs.)

STONE, EDWARD JAMES (1831–1897), British astronomer, who improved the fundamental constants of astronomy, was born in London on Feb. 28, 1831. Educated at King's college, London, and Queen's college, Cambridge, in 1860 he became chief assistant to the astronomer royal at the Royal observatory, Greenwich. Stone deduced the solar parallax, first from observations of Mars, obtaining 8.932″ and 8.945″, and then from the transit of Venus in 1869, which yielded 8.91″. He received the gold medal of the Royal Astronomical society in 1869 and was appointed astronomer at the Cape observatory, South Africa, in 1870. He was appointed Radcliffe observer at Oxford in 1878 and served as president of the Royal Astronomical society from 1882 to 1884. Stone died at Oxford on May 9, 1897. (O. J. E.)

STONE, HARLAN FISKE (1872–1946), 12th chief justice of the United States, participated in making more of his dissenting positions become law than any other justice in history. He was born in Chesterfield, N.H., on Oct. 11, 1872, and died in Washington, D.C., on April 22, 1946. He was graduated from Amherst college in 1894 and received his law degree from Columbia university law school in 1898. He taught law in Columbia law school and practised in New York from 1899 until 1924. He was dean of Columbia law school from 1910 until 1923. In 1924 Pres. Calvin Coolidge appointed him attorney general of the United States, and in 1925 associate justice of the United States supreme court. In 1941, Pres. Franklin D. Roosevelt appointed him chief justice.

His tenure on the supreme court was marked by three rather distinct areas of constitutional debate in addition to the perennial constitutional controversy concerning the authority of some one state to regulate or tax persons or institutions connected with several states. During the first half of his tenure the great constitutional controversy was state experimentation in the power of a state to make changes in its economic or legal order by regulations, prohibitions or taxes. During most of this period he was associated with justices Oliver Wendell Holmes and Louis Brandeis, and part of the time with Justice Cardozo as one of three "great dissenters." A second period of his judicial career involved a constitutional controversy over the power of the national government to make changes in the economic or legal order of the nation. This was the period of the New Deal legislation. During his chief justiceship the great constitutional controversy concerned the power of the popular majority to control matters touching affairs of conscience and expression.

In each of these areas, first as a dissenter and later writing for the majority, Stone made contributions to constitutional jurisprudence. The culmination of his views on the power of a state to regulate interstate commerce came in an opinion shortly before his death, *Southern Pacific Company* v. *Arizona,* 325 U.S. 761 (1945). Therein he restated the historic position of the supreme court that the commerce clause of the constitution gave the court some power to invalidate local legislation affecting persons connected with several states. He rejected any mechanistic formula to determine how the court executed its power, such as the formula that if the burden of the regulation on interstate commerce was "direct," it was unconstitutional, but if "only indirect," the regulation was valid. The court, said Stone, must look to all facts and circumstances and determine for itself, without any help from congress, whether state regulation imperiled the national interest in the free flow of commerce. This required the court openly to make a value judgment.

When the problem was the state's power under the due process and equal protection clause of the 14th amendment to the constitution to regulate economic enterprise, Stone adhered to the position developed earlier by Holmes and Brandeis that the court should abstain from invalidating the legislation, whatever the individual members thought about the policy of the legislation, as long as anyone could find a reasonable basis for supporting the statute. His unique contribution to the constitutional debate concerning the 14th amendment came in a footnote to *United States* v. *Carolene Products Co.,* 304 U.S. 144 (1938), wherein he suggested that while the court should abstain from expressing judgment concerning economic regulation, the court in matters concerning personal freedom and conscience and concerning the processes of democracy, such as voting, should make a more searching inquiry concerning the reasonableness of the legislation. In application of this principle he was the lone dissenter in *Minersville District* v. *Gobitis,* 310 U.S. 586 (1940) wherein the majority held that the state could compel a member of Jehovah's Witnesses actively to participate in the flag salute ceremony. In 1943 the majority of the court switched to Stone's position and held in *West Virginia Board of Education* v. *Barnette,* 319 U.S. 624, that such a compulsory flag salute was unconstitutional.

As chief justice during World War II, Stone presided over cases of martial law, military courts and treason, many of which involved constitutional matters presented to the supreme court almost for the first time in U.S. history. The court was convened in special session that the German saboteurs might have their conviction reviewed without delay and in accordance with standards of due process. Later the court passed on the validity of the "war trial" of the Japanese General Yamashita; the validity of the curfew for reasons of defense of the west coast; and the question whether military tribunals could supersede civilian courts in Hawaii.

During his 22 years on the court, Stone wrote more than 600 opinions and dissents. While they cover the entire range of judicial business, the accidents of history projected him into the limelight for his work in the field of constitutional law.

In one of his more famous dissents, *U.S.* v. *Butler,* 297 U.S. 1 (1936), involving the constitutionality of the first Agricultural Adjustment act of the New Deal, he expressed one of the great problems of judicial review of executive and legislative action when he said that "while unconstitutional exercise of power by the executive and legislative branches of the Government is subject to judicial restraint, the only check upon our own exercise of power is our own sense of restraint." Stone's conduct on the court indicated how thoroughly he subordinated self-interest and applied this principle. As Stone's biographers point out, some of Stone's passionate dissents, such as the one in the case cited above, found him upholding legislation which, if he had been a legislator, he would have passionately denounced.

Stone was stricken on the bench while reading a dissent in the case of *Girouard* v. *U.S.,* 328 U.S. 61 (1946), involving the question whether unwillingness to bear arms because of pacifistic religious belief was a bar to naturalization. In 1931, when this question had first come to the supreme court, Stone joined in a dissent declaring that the act of congress should not be construed to bar naturalization. In 1946 the court changed its position to adopt Stone's original dissenting position, but he again dissented on the ground that subsequent consideration by congress had confirmed the interpretation of the statute that earlier he had thought an erroneous one. Thus his last act on the supreme court was a dramatic illustration of his deference to the legislative will as he saw it.

See Alpheus Mason, *Harlan Fiske Stone: Pillar of the Law* (1956); Allison Dunham and Philip B. Kurland (eds.), *Mr. Justice* (1956). (A. Dm.)

STONE, LUCY (1818–1893), U.S. reformer and women's rights leader, was born in West Brookfield, Mass., Aug. 13, 1818. As a child, indignant at the way in which women were treated, she determined to go to college, study Greek and Hebrew and try to determine whether the Scripture texts regarding subjection for women were correctly translated. She enrolled at Oberlin, O., the only college admitting women at that time. She was graduated in 1847, and in 1848 was engaged as lecturer for the Anti-Slavery society. It was agreed that she should lecture for the Anti-Slavery society on Saturday evenings and Sundays—times regarded as too sacred for a women's rights meeting. The rest of the week she might speak for women's rights on her own responsibility. In time she won wide popularity. She headed the call for the first National Woman's Rights convention, held at Worcester, Mass., in 1850—a call signed by many distinguished men and women. A report of this convention reached England and attracted the attention of Harriet Taylor, afterward the wife of John Stuart Mill. Mrs. Taylor wrote an article about it for the *Westminster Review*, which was the beginning of the modern movement for women's rights in England.

In 1855 Miss Stone married Henry B. Blackwell (1824–1909), an antislavery worker. They issued a joint protest against the inequalities in the marriage laws, and with full approval she kept her own name. In Boston, in 1870, she founded the *Woman's Journal* and served on its editorial staff, 1872–93, aided by her husband and later by her daughter. She died Oct. 18, 1893. (A. S. B.)

STONE, MELVILLE ELIJAH (1848–1929), U.S. journalist, who was general manager of the Associated Press, was born in Hudson, Ill., on Aug. 22, 1848, and attended school in Chicago, Ill. He worked for various newspapers there and was one of the founders of the evening *Chicago Daily News;* the city's first one-cent paper, it began publication on Dec. 25, 1875. With Victor F. Lawson, who had become his partner, Stone established the *Chicago Morning News* (later the *Record*) in 1881. In 1888 he sold his newspaper interests to Lawson and later became a banker.

Stone was general manager of the Associated Press from 1893 to 1921. During that 28 years the A.P. budget grew from about $500,000 to $6,000,000 and it came to furnish more than half the news printed in U.S. papers.

He published his autobiography, *Fifty Years a Journalist,* in 1921. Stone died in New York city on Feb. 15, 1929.

STONE, NICHOLAS (1586?–1647), the most important English mason-sculptor of the early 17th century. A native of Woodbury, near Exeter, Devon, he was apprenticed in London to Isaac James, a Flemish refugee, and after 1606 spent seven years as a journeyman in Amsterdam with Hendrik de Keyser, the leading Dutch architect and sculptor, whose daughter he married. On his return he set up a workshop in London and quickly acquired a large practice as a tomb sculptor and mason. About 80 monuments are recorded in his notebook in the Soane museum, London. There is also an account book covering the years 1631–42, which throws interesting light on the organization of a big mason's yard.

Stone's earliest monuments—the Bodley monument at Merton college, Oxford (1615); Lady Carey (1617) at Church Stowe, Northamptonshire; or Sir Charles Morrison (1619) at Watford, Hertfordshire—show links with the style of the Low Countries and with late Elizabethan work. The architectural settings are florid, much use is often made of coloured alabasters and of brightly blazoned shields. The figure sculpture is, however, naturalistic and sensitively cut.

In the next decade Stone began to be aware of Italian art. From 1619 to 1622 he served as master mason under Inigo Jones at the Banqueting house at Whitehall and became master mason to the crown in 1632. He was concerned with the repairs of the antique sculpture bought by Charles I, and the knowledge he thus gained of classical art undoubtedly affected his style. Much of his later work, *e.g.,* the second Morrison tomb (1630) at Watford or the Lyttelton monument (1634) at Magdalen college, Oxford, is more classical both in the drapery treatment and in the architectural settings. Because of the extensive use of assistants, however, his later work is uneven in quality.

He also undertook a large number of masonry contracts for royal palaces. He could design as well as build—the gateway to the Botanic garden (1632), and probably the porch of St. Mary the Virgin, at Oxford are his, as were Goldsmiths' hall, London (now destroyed), and parts of Cornbury house, Oxfordshire. He built York Water gate, London, but whether he designed it is uncertain.

He was an excellent craftsman and his best sculpture has some distinction, but he cannot be regarded as an artist of high rank. He died at his house in Long Acre, London, on Aug. 24, 1647.

The Note-Book and Account Book of Nicholas Stone was edited by W. L. Spiers (Walpole society, vii, 1918). (M. D. Wy.)

STONE, as the word is most generally used, is a piece of rock (*see* Petrology) or of the solid crust of the earth, and hence of natural origin and generally of inorganic composition.

By their resemblance to rock fragments, however, many objects that originate artificially are also called stone. Examples are commercial grindstones and whetstones (commonly made of crushed natural mineral substances cemented together artificially). A frequent but even more divergent use of the word is the medical one: gallstones and the like are physiologically secreted materials, also known as concretions (*see* Urinary System; Liver and Gall-Bladder, Diseases of). Because of their hardness, the seeds of some vegetables and fruits (*e.g.,* peaches) are called stones. Kinds of stones or rocks are distinguished by prefixes, *e.g.,* limestone, sandstone. To describe its use by man the word stone may have a prefix, *e.g.,* gemstone, millstone. The term is applied to markers used in surveying or building (*e.g.,* cornerstones).

Finally, stone is used somewhat archaically for units of weight, especially in Great Britain—the legal unit of measure, the stone (14 lb. avoirdupois), the stone of the wool weighers (formerly, and still in Scotland, 24; later usually 14 lb.), the Smithfield or butcher's stone (8 lb.; unofficial but still in use). The word has special meanings in certain trades, *e.g.,* a defect in glass, generally caused by foreign material, the glassmakers' stone, and a flat surface for assembling type known to printers as a stone for historical reasons.

GENERAL USES OF STONE

When used for a fragment of rock, stone is widely applied to building material—for example, a house of stone, a stone wall. The stone may be crushed (crushed rock, crushed stone) or ground. Or the rock may be broken, quarried, cut or blasted, along natural lines of weakness. The resulting blocks, if large but irregular, are riprap; if more or less polygonal, thin slabs, they are flagging or flagstones. If sized to fairly definite dimensions, they are dimension stones, and if cut, cut stones. Boulders, pebbles or gravel are terms reserved, respectively, for the individual, generally large rounded pieces of stone, for such pieces when smaller and for masses of such boulders or pebbles, frequently embedded in sand or clay. Aggregate is used in the United States for pebbles, crushed stone or their natural or artificial substitutes which, when mixed with cement as binder, give concrete. Much stone is raw material for industrial processes. Important instances are in making glass and various refractories, in smelting several ores especially iron ore to make pig iron, in preparing cements notably portland cement, and in the chemical industries, where limestone is an inexpensive, widely available basic chemical material.

BUILDING STONE

Structural Features.—Most stone will split along pre-existing or incipient planes of weakness. If originally laid down as a sediment, stone will part (after consolidation or cementation) parallel to bedding planes or stratification, though such planes may heal or be otherwise obliterated. The partings separating the beds may be closely or widely spaced, ranging from fractions of a millimetre to thicknesses of several feet. If the rock has been compressed in mountain building or folding, with or without heating, it may be dynamically metamorphosed and recrystallized (*see* Metamorphism). In such cases, cleavage planes are developed

parallel to the most numerous and most flat surfaces of the constituent minerals. Examples are many schists and slates. Such parting planes, perhaps only millimetres apart, may transect the bedding at almost any angle. Slate also may have a much less well-defined tendency to fracture normal to the cleavage—the sculp or grain of quarrying. If once molten and later solidified in its present position—igneous rock, technically—stone may show some tendency to part along two ill-defined directions, rift and grain, normally about at right angles to each other and to the surface of the ground; further, it may have a tendency to part along nearly horizontal but gently curved planes (sheeting).

Almost all rocks have been slightly compressed or have been shrunken by drying or cooling and therefore show real or incipient parting, distinct from the planes already described. This may follow a regular pattern, giving jointing, which yields rectangular or rhombohedral blocks—forms obviously desirable for use in foundations, walls or pavements.

Quarrying and Preparation.—If the stone is to be used in bulk (as for ballast or chemical raw material), the quarry operation may be preceded by stripping useless material, followed by drilling holes into the rock and blasting it to size suitable for transport. Wherever possible, such operations are mechanized. They may be followed by crushing the rock in a gyratory, a heavy vertical spindle rotating within a hard metal collar, against which the rock is crushed. Or crushing may be by two metal jaws that close upon the rock as it is fed between them. Further stages of crushing or grinding, grading by screens or washing or a similar process may follow. If the stone must meet dimensions, the parting planes previously mentioned may be exploited in quarrying. Holes are drilled from the exposed surface to the level of an observed or inferred parting plane. Through these are introduced a gas or water under pressure, or explosives to make the rock break along the desired plane of weakness. Other surfaces are separated from the solid rock in like manner, until an approximately regular block of desired size is freed. Alternative methods of quarrying are by channeler, a chisellike blade rapidly forced against the rock, cutting a groove; by drilling a line of holes close together and broaching (breaking) the rock between; or by wire saw, a continuous wire which moves under tension dragging sand across the stone, thus sawing its way through. A block is next usually transported to a shed where it is sawed on a movable table or with a movable saw, toothed with hard alloys. Or the large block is cut, as in much marble production, with one or several parallel bands of hard steel swinging across one surface, sprayed with water and fed with abrasive sand. Such methods reduce the block to pieces approximating the desired thinness for possible use as facing (veneer). Blocks also may be trimmed to size and shape by hand or by smaller carborundum or other hard alloy saws, drilled if they are to be bolted in place, and polished. Polishing is on a sand bed, a rotating, horizontal disk, moving under the stone to be polished and carrying sand as abrasive. Another polishing device consists of one or several polishing heads that rotate while pressing against the slab from above and are operated manually or, more commonly, mechanically. Further processing depends upon special requirements—the more processing, the higher the unit value. (*See* also QUARRYING.)

Natural Features Affecting Use.—Factors affecting the use of stone for roadbeds, ballast, dam and embankment facing, construction of bridge piers, supporting walls, roofing or facing of buildings, surfacing of areas of much wear (like stairs) and of little wear (like shower stalls), decoration of tables and chairs, veneering of interiors of buildings and for statuary are as varied as the uses to which the stone is put. Special uses are mentioned below in discussing the appropriate stone.

Such properties as resistance to abrasion or sudden shock, and the tendency to fracture or spall when heated are normally considered in selecting the stone to be used for ballast or concrete aggregate. Examples of the tests applied are the Dorry hardness test, the Los Angeles test for measuring resistance to surface abrasion and the Page test for determining relative resistance to impact. Crushing strength tests are also used, especially with structural stone for foundations of heavy dams, bridges and large buildings.

Porous rock may absorb water and, on the freezing of the water, will spall or chip with changes in temperature. Accelerated freezing tests, using some soluble compound that expands as it crystallizes and is introduced into the rock in solution, have proved useful in determining freezing resistance.

For decoration, critical properties are colour, texture and resistance to disfiguring developments (such as pitting, staining and spalling). Colour varies with the mineral constituents. Pink, red, brown and yellow are usually due to iron minerals, especially hematite and limonite. But granitelike stone is pinkish because of potassium feldspars. Grays may be due to carbon or dark minerals like amphiboles, pyroxenes and micas—silicates usually containing some iron. Green is usually from magnesium or iron silicates. Bluish gray and rare blue are commonly fundamentally black, but if iridescent, may indicate calcic or sodic feldspars, as in some of the iridescent peacock-coloured dark rocks from Scandinavia. White and light gray chiefly result from quartz, feldspars, calcite and clay minerals. Colours do not necessarily reflect major constituents; stones containing 99% white quartz may be rusty brown because of 1% iron oxide, a strong tinting agent.

Texture is important in many ways. Some rocks, such as certain lavas are extremely porous. Stone of mixed grain sizes does not polish as readily as one more uniform in grain. Marble too coarsely granular cannot be worked easily for statuary. Nevertheless, highly porous travertine with irregularly wavy pattern is at a premium as a decorative stone, and rocks with conspicuous veins, crystals, fossils or fragments (breccias) are valued for interior decoration. Similarly, sea-green and light-gray wavelike patterns in certain marbles are much in demand and so are matched panels, cut and set into patterned designs.

Kinds and Sources of Stones.—As suggested, building stones fall into three geologic classes: sedimentary, igneous and metamorphic. The first includes limestone (and its magnesium-bearing variety, dolomite), sandstone and the rarely used shale. Some marbles are properly limestones. Igneous building stones include granite and related light-coloured (acid) coarsely crystalline igneous rocks, basalt or trap and other dark-coloured (basic) igneous rocks, largely fine-grained but partly coarsely crystalline, like diorite and gabbro and fine-grained, light-coloured lavas and volcanic tuffs. Among the metamorphic rocks, marble, slate, gneiss, schist and serpentine are most used.

Of all types, limestone and dolomite, or magnesian limestone, find the largest and most versatile use. Being sedimentary rock, with the beds in many places essentially horizontal, these are commonly easily quarried. The uses of limestone and dolomite are highly varied—as a flux in metallurgy, particularly ironmaking, as a basic chemical raw material (if not too high in magnesium), as a source of coarse crushed stone, or road metal, and aggregate in concrete wherever available and as a satisfactory foundation, building and facing stone. Thus, limestone and dolomite yield more tons of stone in the United States than any other. Generally white to gray and fine textured, it may grade into the coarsely crystalline texture of marble. A major producing region is near Bedford, Ind. Characteristic of the Bedford limestones and those of southern England and northwestern France are the small pellets, or oölites, suggesting fish roe, composing them (*see* OÖLITE). Limestone is also available on tropical coasts in the form of coralline lime if of corallike origin, and of coquina if composed of shell fragments.

Second in quantity used as building stone is sandstone. It ranges in colour from white through buff, red and dark brown. It consists of sand grains cemented with silica, calcium carbonate or clay. The porosity of certain sandstones makes them subject to water ingress and spalling. As coarseness increases sandstone grades into grit and finally conglomerate (cemented gravel). Sandstone is widely quarried for building, especially in western Europe and the United States (Ohio, New York, Wisconsin, Pennsylvania and New Jersey). Unlike limestone, it serves little use in industry except in glass and refractories.

Similar to sandstone in quantity used is granite (*q.v.*), a term loosely but inaccurately applied to most light-coloured, coarsely

grained igneous stone. Granite is highly useful for crushed and building stone and for monuments and decoration. Its great variety of colours—from white through gray, pink, greenish and almost to black—and of coarseness of grain make it much in demand. Productive localities include especially the Scandinavian countries, Scotland, Canada and New England.

Many dark-coloured igneous rocks (gabbro, andesite, basalt, diabase) are collectively sold as trap and used chiefly in crushed stone for roads and aggregate. Typical localities of production are central France and western Scotland, and Washington, Oregon and New Jersey in the United States.

Marble is a recrystallized and commonly beautifully banded limestone. Its original layers may be greatly contorted, or shattered and recemented, resulting in strikingly contrasting colours and patterns. Its chief use is for decoration—mainly indoors—and for statuary. Famous sources are Italy, Greece, Egypt and Tennessee and Vermont.

Slate is a gray to red, green, mottled or black stone marked by closely spaced and splitting cleavage planes. It is used chiefly for roofing, blackboards, switchboards and partitions, but also as structural stone. Outstanding producing regions are Wales, France, Germany, Switzerland, Belgium, and the eastern United States (Pennsylvania, New York, Vermont and Maine). Similar foliated rocks with wavy cleavage, called schists, are locally used.

Serpentine and related types veined with white calcite (ophicalcite, verd antique) are massive and greenish, generally ranging from almost chartreuse to olive. They are used for counters and panels. Though widespread, large deposits are rare. Travertine, a hot-spring deposit, usually porous, and onyx, generally a precipitate from underground waters in caves, are other decorative stones.

See also separate articles on varieties of stone, as GRANITE; LIMESTONE; SLATE, etc.

BIBLIOGRAPHY.—Oliver Bowles, *The Stone Industries* (1934), general, origin, technology, economics; American Institute of Mining and Metallurgical Engineers, Committee on the industrial minerals volume, *Industrial Minerals and Rocks*, 3rd ed. (1959), sections dealing with cement materials, crushed stone, dimension stone, granules, sand and gravel, slate; G. P. Merrill, *Stones for Building and Decoration* (1910), thorough but somewhat outmoded treatment; Heinrich Ries, *Economic Geology*, pp. 102–130 (1942), brief treatment of occurrences and distribution in U.S. and Canada; J. Hirschwald, *Handbuch der Bautechnischen Gesteinspruefung* (1921), numerous tables of extensive tests.
(C. H. BE.)

STONE AGE: *see* ARCHAEOLOGY: *Prehistory.*

STONE CARVING: *see* SCULPTURE; SCULPTURE TECHNIQUE.

STONECHAT, a small thrush, *Saxicola torquata*, ranging over Europe, Asia and Africa. The male is black, white and rufous in colour. It is an active, restless bird, of open brushlands where it sits on tops of bushes flicking its wings and tail, and flying down to the ground to pick up its insect food. Other related species are the wheatear and the whinchat (*qq.v.*) (A. L. RD.)

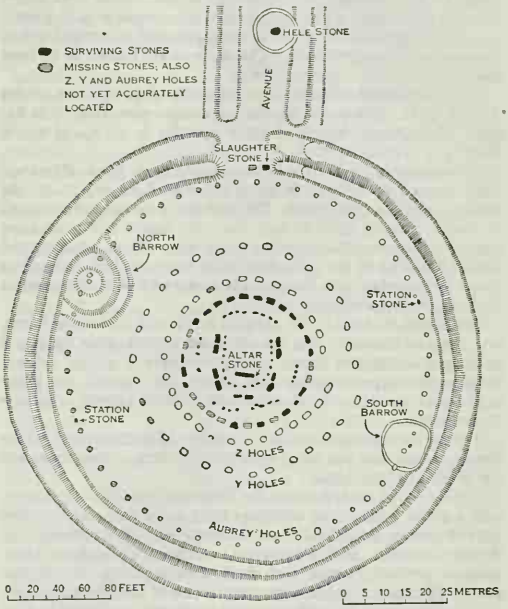
STONE FLY, a member of the insect order Plecoptera (*q.v.*), a group of wide distribution and containing about 800 species. The immature stages are aquatic, confined chiefly to streams. (H. H. R.)

STONEHENGE, a circular setting of large standing stones surrounded by an earthwork, situated about 8 mi. N. of Salisbury, Wiltshire, England. Among the earliest references to it is that of Geoffrey of Monmouth, who relates in his *Historia Regum Britanniae* (*c.* 1136) the legend that the stones were magically transported from Ireland by Merlin. This legend perhaps enshrines a folk memory of the bringing from Pembrokeshire of the "bluestones" which form part of the monument. The supposed connection of Stonehenge with the Druids, which, though without foundation, has held the public imagination since the 17th century, originated with John Aubrey (1626–97), and was more particularly elaborated by William Stukeley in 1740. The modern interpretation of the monument is based chiefly on excavations carried out by the Society of Antiquaries of London since 1919.

The monument consists of a number of structural elements, mostly circular in plan. On the outside is a circular ditch, broken by an entrance gap on the northeast, with a bank immediately within it.

Inside the bank is a ring of 56 pits, known after their discoverer as the Aubrey holes. Between these and the stones in the centre are two further rings of pits, now invisible on the surface, known as the Z and Y holes. The stone setting consisted of two circles and two horseshoes of uprights, the first and third being capped by stone lintels. Additional stones include the Altar stone, lying on the axis southwest of the centre; the Slaughter stone inside the entrance of the earthwork; two Station stones just within the bank on the northwest and southeast; and the Hele stone, standing on the Avenue outside the entrance and surrounded by a narrow circular ditch.

Similar ditches enclose two flat areas on the inner edge of the



ADAPTED, BY PERMISSION, FROM MINISTRY OF WORKS PLAN

STONEHENGE PLAN SHOWING STONES AND PROGRESS OF EXCAVATIONS. THE AUBREY, Z AND Y HOLES ACTUALLY FORM COMPLETE CIRCLES BUT ONLY THOSE EXCAVATED ARE SHOWN ON THE PLAN

bank, known as the North and South barrows, with empty stoneholes at their centres.

These structures can be divided into four groups, each representing a separate phase of building. To the first belong the bank, ditch and Aubrey holes, and the Hele (Heel) stone with its ditch.

This earliest monument is thought to have incorporated at least one block of rhyolite, one of the varieties of bluestone imported from southern Wales which figure prominently in a later stage of the history of Stonehenge.

The ditch is very irregular in outline and was dug in a series of interconnected pits; it averages 13 ft. in width, 5 ft. in depth, and 350 ft. in diameter. The sides are steep and the bottom flat. The bank stands on its inner edge, and is of only moderate height. It is made of chalk rubble, laid directly on the underlying chalk rock after removal of the ancient turf.

The Aubrey holes form a circle about 290 ft. in diameter and are spaced 16 ft. apart. They are circular holes varying in diameter from 30 in. to 70 in. and in depth from 24 in. to 48 in., and were deliberately filled with burned chalk and chalk rubble, soil and ashes; many of them show signs of subsequent disturbance in early times. Although two-thirds of those excavated contained cremated human bones, their purpose is not necessarily sepulchral, nor is there any evidence that they ever held uprights of stone or wood. With them should be associated a number of other cre-

mated burials in the bank and the silting of the ditch.

These structures together form an open-air sanctuary of the kind known as a Henge monument, belonging to the late Neolithic period. This date is confirmed at Stonehenge not only by pottery and by objects associated with the cremations, but also by a radioactive carbon determination made on charcoal from the Aubrey holes (1848 B.C. ± 275 years).

To the second phase belong the outer ring and horseshoe of uprights, the Station stones and barrows, and the Slaughter stone. The stones are all sarsens, a form of sandstone whose nearest probable source is the surface of the downs near Avebury, about 18 mi. to the northwest. The centre of these structures differs slightly from the earlier centre, and lies at the intersection of lines drawn between the Station stones and between the centres of the barrows, which thus appear to represent preliminary markers erected to find the centre of the earthwork. The barrows are known to be later than the bank and the Aubrey holes.

The outer sarsen ring consisted originally of 30 uprights, set on a circle 98 ft. in diameter, supporting a continuous ring of 30 lintels. The average length of these uprights is 18 ft., of which about 4 ft. is below ground.

In the outer horseshoe the five sarsen trilithons (each of two uprights capped by a lintel) increase in height from 16½ ft. at the outside to 22 ft. at the centre. The uprights were erected by means of ramps entering their sockets from the inside, and must have been raised before those of the outer ring. Both uprights and lintels, here and in the outer ring, have been carefully dressed to shape by pounding with stone mauls, many of which were used as packing-stones round their bases.

The lintels are fixed to the uprights by mortice-and-tenon joints, and those of the outer ring are also fixed to each other by toggle or tongue-and-groove joints. This technique is copied from a wooden prototype, probably not directly but through the intermediary of an earlier local monument built of bluestone as suggested below.

The sarsen structure exhibits a number of architectural refinements. Great care was taken with the alignment of the stones and the levelling of the lintels.

The inner and outer faces of the lintels are worked to the appropriate curve, and on the trilithons their upper faces are broader than the lower, so that when viewed from below their sides do not appear to recede. Such refinements are unknown elsewhere in prehistoric northern Europe, and imply influence from the contemporary Mycenaean and Minoan architecture of the Mediterranean. The probability of such influence was startlingly confirmed in 1953 by the discovery of carvings of Bronze Age weapons on three of the sarsen stones. These comprise over 30 representations of bronze axeheads, of a pattern current in Britain (and of Irish manufacture) between 1600 and 1400 B.C., and a single carving of a hilted dagger, which can with some confidence be identified as a type used at Mycenae between 1600 and 1500 B.C., but not found elsewhere in northwestern Europe. The erection of the sarsen stones can thus be dated to around 1500 B.C.

The order and absolute date of the two remaining phases, represented by the Z and Y holes and by the bluestones, is not known, though both are certainly later than the sarsen stones. The Z and Y holes average 4 ft. by 6 ft. in diameter and 40 in. in depth, and though roughly concentric with the sarsens are irregularly aligned.

After the original excavation of these holes they appear to have silted up naturally over a period of many centuries, during which pottery of the pre-Roman Iron Age found its way into their filling. There is, however, no reason for supposing that the holes are of this date, or for associating them with the Druids. Though their date and purpose is unknown, it is perhaps significant that they are of a suitable size and shape to serve as emplacements for the bluestones, and that their number is the same as the best estimate of the original number of the bluestones in their present setting. It may be, therefore, that the Z and Y holes represent an unfinished project for the erection of the bluestones at Stonehenge, which for some reason unknown was abandoned in favour of the present setting of these stones.

The bluestone setting, consisting of the inner circle and horseshoe and the Altar stone, is made up of at least four varieties of igneous rock foreign to the region. The inner horseshoe is of spotted dolerite, and the inner circle mainly of the same rock with four specimens each of rhyolite and volcanic ash; the Altar stone alone is of micaceous sandstone. The first three of these rocks are found in association only in the Prescelly mountains in Pembrokeshire, whence they were brought to Wiltshire.

The occurrence of micaceous sandstone near Milford Haven suggests transport of the stones at least in part by seagoing boats.

The bluestones are all carefully dressed to shape, and taper gradually from the ground to flat tops. They vary in height from 6 ft. to 8 ft. It is clear that they were erected after the sarsens, since some of them stand in the filling of the ramps of the sarsens' sockets. Three of them, all of dolerite, retain the remains of a tenon and two pairs of mortices respectively. This suggests that they, and probably the remainder of the dolerites, originally stood elsewhere as a separate monument incorporating at least two trilithons, and it is doubtless from this that the mortice-and-tenon joints of the sarsen stones were derived.

Recent work suggests that another structure, of bluestones other than dolerite, stood about a mile northwest of Stonehenge, and that the bluestones were brought to Wiltshire in the late Neolithic period, some centuries before their erection in their present position at Stonehenge.

Stonehenge is approached from the northeast by an avenue, 75 ft. in width, bounded by a slight bank and ditch on each side, which runs by three successive alignments to the bank of the River Avon 1¼ mi. to the southeast. In date it appears to be intermediate between the first and second phases. The section adjoining Stonehenge, and the axis of symmetry of Stonehenge itself, is aligned approximately on the point of midsummer sunrise (though not, as is often stated, on the Hele stone). This fact has occasioned much fruitless conjecture.

BIBLIOGRAPHY.—For a bibliography complete to 1900 see *Wiltshire Archaeological and Natural History Magazine*, xxxii (Devizes, 1901). A later book is E. H. Stone, *The Stones of Stonehenge* (London, 1924). *See* also the following articles: R. C. C. Clay, R. S. Newall and O. G. S. Crawford in *Antiquity*, i, iii, xxviii (Gloucester, 1927–29; Newbury, 1954); S. Piggott, *Aspects of Archaeology in Britain and Beyond*, ed. W. F. Grimes (London, 1951); R. J. C. Atkinson in *Proceedings of the Prehistoric Society*, xviii (London, 1953). For excavations *see* reports by: W. Hawley, and by R. J. C. Atkinson, S. Piggott and J. F. S. Stone in *Antiquaries Journal*, i–vi, viii, xxxii (London, 1920 ff.); W. Gowland in *Archaeologia*, lviii (London, 1902); J. F. S. Stone in *Archaeological Journal*, civ (London, 1947); and R. J. C. Atkinson in *Nature*, clxxii (London, 1953). (R. J. C. A.)

STONE MOUNTAIN MEMORIAL. Stone mountain, about 15 mi. N.E. of Atlanta, Ga., is a gray monadnock, shaped like a half parabola, 1,683 ft. high, covered with growths of mosses and lichens and various plants, including rare varieties.

In 1916 the sculptor Gutzon Borglum (*q.v.*; 1871–1941) conceived the plan of carving upon the side a huge memorial commemorating the valour of the south and its loyalty to the lost Confederate cause. The design, 200 ft. high and 1,300 ft. long, was to represent the army of the Confederacy with Robert E. Lee and other southern leaders at its head. Work was begun by Borglum in 1917, interrupted by World War I, and continued in 1923–25 when a dispute with the Stone Mountain Monumental association, which raised the funds, resulted in his dismissal.

In 1926 Augustus Lukeman was employed as Borglum's successor, whereupon he destroyed all Borglum's models for the work except that of Jefferson Davis, and demolished Borglum's Lee. A section of Lukeman's work was unveiled in 1928. Work ceased on the monument after 1930 but later a commission was appointed to make plans for its completion, to buy the mountain from the private owners and to create a state park which would include the mountain and a nearby lake. (A. K. McC.)

STONE RIVER, BATTLE OF, a battle of the American Civil War, called the battle of Murfreesboro by the Confederates, fought on Dec. 31, 1862, and Jan. 2, 1863. After his appointment in October to command the Army of the Cumberland, Gen. W. S. Rosecrans, with Chattanooga as his objective, moved from Nashville upon Gen. Braxton Bragg, who left the winter quarters he had established at Murfreesboro and met the Union army on

Stone river immediately north of Murfreesboro, on the last day of December. The plan of attack on each side was to crush the enemy's right. Bragg's left, commanded by Lieut. Gen. W. J. Hardee, overlapped and bore back the Union right under Maj.-gen. A. McD. McCook, and Maj.-gen. T. L. Crittenden, commanding the Union left, was hurriedly called back from his attack on the Confederate right to support McCook. The Union right was crumpled up on the centre, where Maj.-gen. G. H. Thomas's corps checked the Confederate attack. There was practically no fighting on Jan. 1, but on the 2nd the Confederates renewed the attack, Maj.-gen. J. C. Breckinridge with Bragg's right attempting in vain to displace one of Crittenden's divisions which had established itself during the 1st on high ground across the river. On the night of the 3rd Bragg withdrew and the Union army occupied Murfreesboro. Tactically a drawn battle, Stone River was strategically a Union victory. The losses on both sides were heavy.

STONEWORT, the common name for algae of the class Charophyceae, so called from the deposit of lime in their cells. Most stoneworts grow submerged in fresh water, but a few grow in brackish water. *See* ALGAE; *Classification of Algae: Chlorophyta.*

STONY POINT, a town of Rockland county, N.Y., U.S., lies in the scenic highlands along the west shore of the Hudson river and about 38 mi. from midtown New York city. Stony Point derives its name from the rocky promontory jutting into the Hudson. A part of Palisades interstate park lies within the town limits. At the town's northern end was anchored one end of the heavy chain stretched across the Hudson during the American Revolution to prevent the British fleet from ascending further, while at the southern end stood the Treason (Joshua Hett Smith) house where Gen. Benedict Arnold and Maj. John Andre met on Sept. 22, 1780, to arrange for Arnold's betrayal of West Point to the British. At midnight on July 16, 1779, occurred the successful and daring storming of the strong British fortifications at Stony Point by Gen. Anthony Wayne and about 1,350 gallant, hand-picked American troops. Wayne's brilliant feat, with a loss of only 15 killed and 83 wounded as against 63 British killed and 70 wounded, and the capture of 543 of the garrison, was a severe setback to the British. As the fort was undefendable against river bombardment, however, it was speedily dismantled and evacuated on July 18; the British thereupon reoccupied it a second time but abandoned it late in Oct. 1779, and the upper Hudson was safe.

Major industries include a chemical company, a trap rock company, a Rockland county power plant and a gypsum plant. Stony Point is governed by a town board consisting of a supervisor, two justices of the peace and two councilmen.

For comparative population figures *see* table in NEW YORK: *Population.* (M. D. HH.)

STOOLBALL. Of the ancient game of stoolball very little indeed is known, and it has nothing whatever in common with the modern game which, after 1916, became popular in Great Britain and elsewhere. During World War I a simple but not strenuous outdoor game was very badly needed for wounded officers, soldiers and sailors; and the idea occurred to Maj. W. W. Grantham, who was stationed at Brighton, that the old game of stoolball adapted to modern requirements, would be suitable for the purpose. The game was soon played in scores of hospitals and convalescent homes, in England and France. He made a few simple rules adapted from cricket, and the first public match took place on the County cricket ground of Hove in 1917.

In 1924 a Stoolball association was formed at Lord's cricket ground, London, headquarters of the Marylebone Cricket club. In Sussex alone the game is played in hundreds of villages, and by large numbers of schools of all kinds, women's institutions and girl guides associations. It was later introduced into Iceland, Japan, Siberia, Switzerland and Finland.

The materials required for the game are very few and very simple; and although with the exception of the ball they can be easily and quickly made by any carpenter they may now be obtained from most of the outfitters. Two bats, a ball and two wickets comprise the materials. The bats are wooden, made like a racket, the diameter of the round part being not more than 7½ in., and, including the handle, the bat must be not more than 18 in. in length. The ball is usually known as "best tennis no. 3," but a hard lawn tennis ball will serve the purpose. Each of the two wickets consists of a board 1 ft. square fastened to a stout post which must be firmly fixed in the ground in such a way that the top of the board is 4 ft. 8 in. above the ground; a tripod is often substituted for the post.

The wickets must be fixed at a distance of 16 yd. from each other, and a bowling crease not more than 1 yd. in length must be marked at a spot 10 yd. from each wicket, for this is where the bowler stands. The bowling is underhand, and the number of balls to an over, originally ten, has been reduced to eight. A batsman may be bowled, caught, or run out (the ball hitting the face of the board), or may be given out "body before"; runs are made just as in cricket. In order that the batsman shall be out, "bowled," the ball must hit the face or edge of the wicket (but not the stump), without having previously touched the ground—in other words the ball must be a full pitch. On all points for which there are no special rules, the laws of cricket, as far as possible, hold good for stoolball.

The game can be played on any piece of ground which is fairly level and not so rough as to be dangerous; the minimum space required would be about half the size of an ordinary cricket ground.

The ancient form of the game was once very popular in England, and was commonly considered to be the ancestor of cricket. Writing in 1801, Joseph Strutt gives a description of it.

See W. W. Grantham, *Stoolball Illustrated and How to Play It.*

(W. A. BE.)

STORACE, STEPHEN (1763–1796), English composer, was born in London but studied (under his father, Stefano Storace, an Italian contrabassist) at the Conservatorio di Sant' Onofrio, at Naples. His first opera, *Gli Sposi malcontenti,* was produced at Vienna, in 1785. There he made the acquaintance of Mozart, in whose *Nozze di Figaro* his sister, Anna Selina Storace, first sang the part of Susanna. There also he produced a second opera, *Gli Equivoci,* founded on Shakespeare's *Comedy of Errors,* and a Singspiel entitled *Der Doctor und der Apotheker.* In England, after creating a favourable impression by bringing out his Singspiel at Drury Lane, under the title of *The Doctor and the Apothecary,* Storace attained his first great success in 1789, in *The Haunted Tower,* an opera which ran for 50 nights in succession. The music of *The Pirates* (1792) affords an early instance of the introduction of a grand finale into an English opera.

Storace's music is pre-eminently English. He doubtless learned something from his sister Anna (1766–1817), a brilliant singer who died leaving £50,000. Storace himself died on March 19, 1796.

STORAGE BATTERY: *see* BATTERY.

STORK, the name for various large to very large long-legged, long-necked, large-billed birds related to the herons and ibis (*qq.v.*) and placed with them in the order Ciconiiformes. The 17 species of typical storks form the family Ciconiidae. They frequent open fields and shallow water, where they stalk about looking for their food of insects, fish, frogs or almost any other small animal. Some storks, such as the adjutant, *Leptoptilos dubius,* of India, and the marabou, *L. crumeniferus,* of Africa feed on carrion.

During the nonbreeding season storks often travel in flocks, and at times soar majestically high in the air. Their voices are little more than grunts or peeps, but storks commonly make a rattling noise by snapping their great mandibles together. While stately, or impressively grotesque in everyday activity, at mating time storks perform grotesque dances, hopping about with spread wings and clattering mandibles.

The nests are bulky affairs, usually placed in trees, sometimes in colonies. The eggs are white and unmarked; the young are down-covered and are cared for in the nest by the parents. In plumage storks may be mostly glossy black, or black and white, the head feathered or naked and with brightly coloured lapets on the bill, depending on the species.

Most storks live in the old world where the group is widespread,

only three occurring in tropical America, including the jabiru (*q.v.*), *Jabiru mycteria,* and wood ibis, *Mycteria americana.*

The white stork, *Ciconia ciconia,* of Europe is one of the best-known birds. It stands more than three feet high and is white with black in the wings. It nests on houses where platforms are sometimes put up for it, and it is thought to bring good luck. The hammerkop stork, *Scopus umbretta,* of Africa is the only species in the related family Scopidae and the whale-headed stork, *Balaeniceps rex,* of Africa is the only species in another related family, Balaenicipitidae. (A. L. RD.)

STORM, THEODOR WOLDSEN (1817–1888), German poet and novelist, was born at Husum, Schleswig, on Sept. 14, 1817, and studied jurisprudence at Kiel and Berlin, where he formed a close friendship with the brothers Theodor and Tycho Mommsen. Storm is hardly less remarkable as a lyric poet than as a novelist. As the former, he made his debut, with the two Mommsens, with *Liederbuch dreier Freunde* (1843); but his *Gedichte* (1852), in the wake of Eichendorff and Mörike, first obtained for him general recognition. As a novelist he gained his first great success with *Immensee* (1852); and this was followed by numerous other short stories. He is at his best when treating, with retrospective resignation, episodes and incidents from his own earlier life. Later he passed to psychological problems with *Aquis submersus* (1877) and *Zur Chronik von Grieshuus* (1884), and made a deep impression with his fantastic *Schimmelreiter* (1888).

BIBLIOGRAPHY.—T. Storm, *Gesammelte Schriften,* 19 vol. (Brunswick, 1868–89); *Sämtliche Werke,* ed. by A. Köster, 8 vol. (Leipzig, 1919–20); *Werke,* ed. by T. Herzel and F. Böhme, 9 vol. (Leipzig, 1936); *Briefe,* ed. by G. Storm, 4 vol. (Brunswick, 1915–17); G. Storm, *Theodor Storm, Ein Bild seines Lebens,* 2nd ed. (Berlin, 1912–13); A. Biese, *Theodor Storms Leben und Werke* (Leipzig, 1917); E. Steiner, *Theodor Storm* (Basle, 1921); F. Stuckert, *Theodor Storm* (Halle, 1940).

STORM. This term is sometimes applied to any disturbed atmospheric condition, but more often only to the violent disturbances. Windstorm, thunderstorm, snowstorm, sandstorm, dust-storm, etc., are more specific terms. Ordinary cyclones (*q.v.*) are frequently referred to as storms. In hydrology, the name is applied to an area of rapidly accumulated rainfall whether caused by one meteorological storm, a series of them or only a portion of one. In the Beaufort scale (*q.v.*) a wind of force 11 (64 to 75 m.p.h.) is called a storm wind. For magnetic storms *see* GEOMAGNETISM. (H. R. B.)

STORMBERG, ACTION OF (Dec. 10, 1899): *see* SOUTH AFRICAN WAR.

STÖRMER, FREDRIK CARL MÜLERTZ (1874–1957), Norwegian mathematician and geophysicist, was born at Skien, Nor., Sept. 3, 1874. Professor of mathematics at the University of Oslo, 1903–46, Störmer began his scientific work in pure mathematics, particularly number theory. In 1904 he became interested in experiments of the physicist Kristian Birkeland, in which aurora effects were obtained by electron bombardment of a magnetized terrella in vacuum. Störmer calculated the trajectories of electrical particles in the field of a magnetic dipole and explained many of the peculiarities of the aurora borealis. The results are also applicable to cosmic rays.

In 1910 he spent a winter at Bossekop, in the northernmost Norwegian county of Finnmark, observing auroras; later he organized a permanent network of stations in Norway to photograph and measure the heights of auroras and special types of clouds. Störmer was a member of numerous geophysical commissions and received a great number of scientific distinctions. He died at Oslo on Aug. 13, 1957.

See Carl Störmer, *The Polar Aurora* (1955). (O. OE.)

STORNOWAY, a small burgh on the island of Lewis-and-Harris (*q.v.*), Outer Hebrides, Scot., 53 mi. N. of Portree, Isle of Skye, by steamer. Pop. (1961) 5,229. It is the principal town in the county of Ross and Cromarty and is a natural harbour. There is about ¾ mi. of quayage and wharfage. The herring fishery, formerly much larger, declined when the fish migrated. Stornoway's chief occupation is now the manufacture of Harris tweed. There is daily steamship connection with Kyle of Lochalsh and Mallaig, air services with Glasgow and Inverness, and weekly cargo services to Aberdeen, Dundee, Leith, Glasgow and Liverpool.

STORY, JOHN (*c.* 1510–1571), English Roman Catholic martyr, twice abjured his faith but twice returned to it. He was educated at Oxford, where he became lecturer on civil law in 1535, being made later principal of Broadgates hall, afterward Pembroke college. He appears to have disavowed his Roman Catholic opinions just after the accession of Edward VI, but having been chosen a member of parliament in 1547 he gained notoriety by his opposition to the Act of Uniformity in 1548. For crying out during a session of parliament "Woe unto thee, O land, when thy king is a child," he was imprisoned by the house of commons, but he was soon released and went into exile. Returning to England in 1553, he resigned his position at Oxford, which was now that of regius professor of civil law, and was made chancellor of the dioceses of London and of Oxford and dean of arches. Story was one of Queen Mary's most active agents in prosecuting heretics, and was proctor at the trial of Archbishop Thomas Cranmer in 1555. Under Elizabeth I he was again returned to parliament, but in 1560 he underwent a short imprisonment for boasting about his work in the former reign. In 1563 he was again arrested, but managed to escape to Flanders, where he became a pensioner of Philip II of Spain. The duke of Alva authorized him to exclude English books from the Netherlands and, in 1570, while engaged in this work, he was decoyed onto a ship at Antwerp and conveyed to Yarmouth. In spite of his claim that he was a Spanish subject, he was tried for high treason and executed at Tyburn on June 1, 1571. In 1886 Story was beatified by papal decree.

STORY, JOSEPH (1779–1845), U.S. legal writer and educator, associate justice of the U.S. supreme court and a principal juristic supporter, with Chief Justice John Marshall (*q.v.*), of American nationalism, was born on Sept. 18, 1779, at Marblehead, Mass. One of the great figures in U.S. legal history, his influence on the development of U.S. law rivaled that of Marshall. His preparatory training was begun at the Marblehead academy. He entered Harvard college in Jan. 1795, joining the class admitted the previous term and unsuccessfully contesting with William Ellery Channing (*q.v.*) for first rank in the class of 1798. Upon graduation, Story returned to Marblehead and began reading law with Samuel Sewall, who later became chief justice of the supreme judicial court of Massachusetts. In 1801, he moved to Salem and entered the office of Samuel Putnam, where he read law until July of that year when he was admitted to the bar.

From 1801 to 1811 Story practised law in Salem and occasionally argued cases in New Hampshire. In Massachusetts he became prominent in the Jeffersonian Republican (afterward called the Democratic) party. In 1805, he was elected to the state legislature. He also served part of one session (1808–09) in the U.S. congress. In 1810, he was counsel in the celebrated case of *Fletcher* v. *Peck,* before the U.S. supreme court. He was returned in 1810 to the Massachusetts legislature, becoming speaker of the state house of representatives in 1811.

Supreme Court Appointment.—Appointed an associate justice of the U.S. supreme court by Pres. James Madison in Nov. 1811 to fill the vacancy created by the death of William Cushing, Story was then 32 years old, the youngest justice ever to sit upon that court, and lacking in judicial experience. A leader in the Massachusetts Jeffersonian Republican party, he wrote shortly after receiving the life appointment, "I am no longer a political man." Although he was raised to the bench by Madison for the express purpose of contesting the nationalism of Chief Justice John Marshall, Story himself became a confirmed nationalist and a supporter of the Federalist views of the chief justice.

The supreme court was then composed of seven justices. One term of the court was held each year at Washington in a basement chamber of the old capitol building; in addition, each justice was assigned a "circuit" in which two terms of the court were held. Story's circuit consisted of Maine, Massachusetts, Rhode Island and New Hampshire. His circuit court decisions (reported by Gallison, Mason, Sumner and Story) in admiralty, prize, patent and copyright matters helped make United States law on these subjects.

Notable Cases.—The supreme court, which eight years previously had first declared a law of congress unconstitutional (*Marbury* v. *Madison*), was just entering upon its real importance when Story came to the bench, and many of the cases decided during his tenure established basic principles of U.S. public and constitutional law. Story's supreme court opinions, found in the reports of Cranch, Wheaton, Peters and Howard (1812–45), number more than 200.

In *Martin* v. *Hunter's Lessee* (1816), a decision that has been called "the keystone of the whole arch of Federal judicial power" (Charles Warren, *The Supreme Court in U.S. History*, p. 449, Little, Brown and Company, Boston, 1922), Story held that the supreme court's appellate civil jurisdiction embraces every case not otherwise to be decided initially and directly by way of its original jurisdiction; that there is no constitutional restraint upon the exercise of this appellate jurisdiction over state tribunals; and that sec. 25 of the Judiciary act of 1789, insofar as it granted such appellate jurisdiction, was constitutional. This decision established the supremacy of the supreme court over the highest court in any state in all civil cases where the federal constitution, laws and treaties were involved. (Five years later, Marshall's opinion in *Cohens* v. *Virginia* did likewise for state criminal cases.)

Although Marshall wrote the opinion of the court in *Dartmouth College* v. *Woodward* (1819), holding that charters of private corporations, as then issued, were contracts which could not constitutionally be impaired by the states granting them, Story concurred in a separate opinion—one of his most learned—which reviews English and U.S. authorities and reveals his devotion to higher law principles. This decision gave stability to the rights of corporations and influenced subsequent economic developments in the United States.

In *Ogden* v. *Saunders* (1827), the first case in which Marshall was defeated on a constitutional issue, Story joined in a dissenting opinion written by the chief justice. They both took the position that the contract clause of art. i, sec. 10 of the constitution denounced state insolvency legislation concerning the prospective, as well as the retrospective, operation of contracts. This case marked the beginning of a trend away from the principles of economic conservatism which Marshall and Story had sought to entrench in U.S. law.

John Marshall died in July 1835. Story presided over the court while a successor was being chosen, but he was never chief justice. "I have never for a moment imagined that I should be thought of . . . ," he wrote on July 24, 1835. Roger Brooke Taney (*q.v.*) was commissioned chief justice in March 1836, and Story served under him until 1845.

Left over from the Marshall period were three cases, involving the constitutionality of state laws, that had been argued orally before the supreme court, but that had gone undecided: *City of New York* v. *Miln* (1837), *Briscoe* v. *Bank of Kentucky* (1837) and *Charles River Bridge* v. *Warren Bridge* (1837). In all of these cases, the Taney court upheld the state laws, and Story delivered dissenting opinions which lamented the death of Marshall and the passing of the "old order." He now referred to himself as the last of the "old race of judges," and he decided to resign. Friends, however, persuaded him to remain on the bench. Henceforth, on important constitutional issues, he was forced to dissent from the judgments of the Taney court, or to appear to agree in silence with doctrines which seemed to him at variance with his previous decisions.

Two cases decided during this final period, in which Story wrote the opinions of the court, were *Prigg* v. *the Commonwealth of Pennsylvania* (1842) and *Swift* v. *Tyson* (1842). In the Prigg opinion, Story held that since the constitution gave the federal government the power to deal with fugitive slaves, this power could not be exercised by the states, even if no conflict with federal laws was involved. Further, he expressed doubt that the federal government could compel state officers to enforce federal laws. In the Swift case, Story held that sec. 34 of the federal Judiciary act of 1789 did not require that local interpretations of general commercial law be followed in diversity cases in the federal courts. In 1938, the supreme court reversed this decision, holding that the law to be followed in such cases was the law of the state as interpreted by state courts.

Story and Marshall Compared.—Marshall, who was 24 years older than Story and versed mainly in constitutional law, disliked the labour of investigating legal authorities to support his judgments. Story, who was the master of several fields of law (including constitutional law), loved legal research, and his opinions drew heavily upon past authorities. If Marshall's opinions were businesslike, Story's were scholarlike; if the chief justice was concise and to the point in his judgments, his younger colleague was eager, luminous and had a tendency to be expansive. The difference between their judicial minds is seen in the Dartmouth college case, where opinions were written by both. Marshall served on the court just over 34 years; Story's tenure lacked only months of equaling that of the chief justice. In the field of constitutional law, they laboured together for nearly a quarter of a century in upholding the same legal doctrines.

Harvard Professorship.—In 1829 Nathan Dane founded a professorship of law at Harvard college, the duties of which were to prepare, deliver and revise for publication a course of lectures on the law of nature, the law of nations, commercial and maritime law, federal law and federal equity. By an express condition of the $10,000 endowment, Story became the first Dane professor of law at Harvard, which position he held, together with that of associate justice of the supreme court, until his death in 1845. Story's lectures on the designated subjects led to the publication of his nine great commentaries on the law, the first of which appeared in 1832. His first associate at the law school was John Hooker Ashmun, Royall professor of law, who died in 1833 and was replaced by Simon Greenleaf. The Story-Greenleaf period marked the beginning of the law school's growth to greatness. Story was exceptional as a teacher, and he devoted a great deal of time to the practical presentation of moot-court cases and arguments by his students. His annual salary at Harvard was $1,000.

In 1845 Story resolved to resign from the supreme court and devote his life to teaching and writing at Harvard. Early in September, however, before his letter of resignation was sent to Washington, he became ill, and on Sept. 10, 1845, he died at his home in Cambridge, Mass.

Story was married in 1804 to Mary Lynde Fitch Oliver, who died without issue; and in 1808 to Sarah Waldo Wetmore (1784–1855), the daughter of a Boston jurist, who bore him seven children, only two of whom survived childhood. A Calvinist, Story became a Unitarian and was several times president of the American Unitarian association.

Writings.—Story is best known as an author for a series of commentaries on *Bailments* (1832), *Constitution of the United States*, 3 vol. (1833), *Conflict of Laws* (1834), *Equity Jurisprudence*, 2 vol. (1836), *Equity Pleading* (1838), *Agency* (1839), *Partnership* (1841), *Bills of Exchange* (1843) and *Promissory Notes* (1845). His works on equity were particularly important; with Chancellor James Kent (*q.v.*), he helped found equity jurisprudence in the United States. He also prepared *Commentaries on the Constitution*, 1-vol. abridgment (1833) and *Constitutional Classbook* (1834). He wrote scholarly articles for *The American Jurist* and *The North American Review* and published biographical sketches of many prominent men. He drew up the bylaws of a bank, the *Report* of the Massachusetts Commission on Codification and the *Report on Salaries of the Judiciary* and prepared elaborate notes (for Wheaton's *Reports*) on prize, admiralty, etc. He drafted a bankruptcy law, a crimes act and a judiciary act. He published a volume of poems, contributed law articles to Francis Lieber's *Encyclopedia Americana*, constructed a portion of a legal digest and prepared an edition of the laws of the United States. He delivered many public addresses, some on literary, scientific and historical subjects, and participated in the convention called to revise the Massachusetts constitution.

BIBLIOGRAPHY.—W. W. Story, *Life and Letters of Joseph Story*, 2 vol. (1851) and *Miscellaneous Writings of Joseph Story* (1835 and 1852); Roscoe Pound, "The Place of Judge Story in the Making of American Law," *American Law Review*, 48:676–697 (Sept.–Oct. 1914); W. Schofield, "Joseph Story, 1779–1845," *Great American Lawyers*, ed. by W. D. Lewis, 3:123–185 (1908). Constitutional developments during the

period are evaluated in: Carl B. Swisher, *American Constitutional Development* (1943); and W. W. Crosskey, *Politics and the Constitution in the History of the United States*, 2 vol. (1953). (JN. C. H.)

STORY, WILLIAM WETMORE (1819–1895), U.S. sculptor and author, who is most famous for his statue "Cleopatra," was born in Salem, Mass., on Feb. 12, 1819. He graduated from Harvard college in 1838 and studied law under his father Justice Joseph Story. After his graduation from Harvard law school in 1840 he decided while practising law in Boston to become a sculptor. In 1856 he settled in Rome and remained there for the rest of his life. He lived in the Palazzo Barberini, where he entertained famous literary, theatrical and social figures of the day.

Story's "Cleopatra" was exhibited at the London exposition in 1862 (a replica of it is in the Metropolitan Museum of Art, New York city). Nathaniel Hawthorne's description of this work in his novel *The Marble Faun* (1860) contributed greatly to its wide popularity in the United States and Great Britain.

In addition to his sculpture, Story wrote several volumes of poetry, essays, a biography of his father and various legal works.

He died at Vallombrosa, Italy, on Oct. 7, 1895. Story's fame has been kept alive by the interest in Henry James's book about him titled *William Wetmore Story and His Friends* (1903; reprinted 1957).

See A. Gardner, *Yankee Stonecutters* (1945). (A. T. G.)

STORY, SHORT: *see* SHORT STORY.

STOSS, VEIT (1438? or 1447?–1533), German sculptor, was born in Swabia but reared in Nürnberg. From 1477 to 1496 he worked mainly in Poland, Bohemia and Hungary. His principal works are the majestic high altar, carved in limewood and painted, of the German church, St. Mary's, in Cracow (1477–89) and the sculptured tombs of King Casimir IV and Archbishop Zbigniew Olesnicki in the cathedrals of Cracow and Gniezno. When he returned to Nürnberg, he was defrauded of his savings, attempted to regain them by forgery, was discovered and branded, and passed an embittered old age encumbered with civic disabilities. His work of this period includes important wood and stone sculpture in the churches of St. Sebaldus (1499, 1520) and St. Lorenz (1513, 1518) in Nürnberg, and a carved altar in Bamberg cathedral (1523). His style, developed from that of Nicholas Gerhaert, astonishes by its dramatic realism and emotional force no less than by the technical virtuosity of the carving. He was active also as a painter and engraver. He died at Nürnberg in 1533.

See Eberhard Lutze, *Veit Stoss* (1938); Tadeusz Dobrowolski, *Wit Stwosz: Ostarz Krakowski* (1951). (D. Kg.)

STOTHARD, THOMAS (1755–1834), English subject painter, designer and book illustrator, best known for his work in illustration, was born in London on Aug. 17, 1755. He was apprenticed in Spitalfields to a draftsman of patterns for silks and in 1777 became a student of the Royal Academy. He was elected associate in 1791 and full academician in 1794, being appointed librarian in 1812. He died in London on April 27, 1834.

In 1779 Stothard became a contributor to *Novelist's Magazine* and to Bell's *Poets*. Because he was so ill-paid he supplemented his income by designing plates for pocket books, tickets for concerts, illustrations for almanacs, portraits of popular actors—and into even the slightest sketches he infused grace and distinction. He is at his best in domestic or ideal subjects; the heroic and tragic were beyond his powers. He began to exhibit pictures in oil at the Royal Academy in 1778 and continued to do so until his death. His oils are usually small in size and free in handling, but their colouring is often rich and glowing. His best-known painting is the "Canterbury Pilgrims," in the Tate gallery, the engraving from which attained immense popularity and heralds the methods of Pre-Raphaelitism.

Stothard illustrated many books, including *The Adventures of Peregrine Pickle*, *Clarissa Harlowe*, *Tristram Shandy*, *Robinson Crusoe*, *The Pilgrim's Progress*, *The Vicar of Wakefield*, *The Rape of the Lock* and the works of Shakespeare, Byron, Milton and other poets. He also decorated the mansions of many of his wealthy patrons. (D. C. T. T.; X.)

STOUR, the name of several English rivers, an ancient word of doubtful etymology. (1) The East-Anglian Stour rises in southeast Cambridgeshire and flows 47 mi. (76 km.) to the North sea at Harwich. (2) The Kentish or Great Stour rises on the south of the North Downs, one branch, the East Stour, rising northwest of Hythe but flowing away from the sea, while the western branch rises near Lenham and the two unite at Ashford. Passing Canterbury, the Stour again divides: the Great Stour, joined by the Little Stour, flows east to Pegwell bay; the small branch enters the North sea near Reculver. It is navigable to Fordwich. Its length is about 40 mi. (64 km.) Other rivers named Stour are: (3) a tributary of the Hampshire Avon; (4) a left-bank tributary of the Severn, which it joins at Stourport; and (5) a small tributary of the Warwickshire Avon, rising west of Banbury.

STOURBRIDGE, a municipal borough (1914) in the Dudley parliamentary division of Worcestershire, Eng., 12 mi. (19 km.) W. of Birmingham by road. Pop. (1961) 42,631. Stourbridge glass, dating from 1556, is still made in the district, and the college of further education teaches glassmaking. At the Council house in Mary Stevens park is a permanent glass exhibition. King Edward school (1430) was attended by Samuel Johnson; Oldswinford Hospital (bluecoat) school was founded in 1667. Manufactures include hollow ware, refractories, chains, tools and leathers.

STOURPORT-ON-SEVERN, an urban district in the Kidderminster parliamentary division of Worcestershire, Eng., 12 mi. (19 km.) N.N.W. of Worcester by road. Pop. (1961) 11,748. Area 5.0 sq.mi. (8.1 sq.km.) It was built (1765–71) where Stour joins Severn at the south end of the Staffordshire and Worcestershire canal. Industries include carpets, insulators, chains, hollow ware and engineering. There are also an electric generating station, an oil depot, warehouses and wharves.

STOUT, GEORGE FREDERICK (1860–1944), British psychologist and philosopher, provided a fresh philosophical basis for British psychology in place of the older materialistic doctrine. He was born Jan. 6, 1860, at South Shields, Eng., and educated at Cambridge. He was appointed reader in mental philosophy at Oxford in 1898 and professor of logic and metaphysics at St. Andrews in 1903. He spent the last eight years of his life in Australia, and died in Sydney, New South Wales, on Aug. 18, 1944.

Stout approached the problems of the mind primarily from a philosophical standpoint. His three best-known works, *Analytic Psychology* (1896), *A Manual of Psychology* (1898) and *Mind and Matter* (1931), together provided a systematic theory of the mind and its nature which entirely supplanted the primitive associationism of the 19th century. He firmly rejected the Cartesian dualistic conception of body and mind and with it all the alternative hypotheses of interactionism, parallelism and materialistic behaviourism to which that notion gave rise; instead he reverted to a biological concept of an embodied self, not unlike Aristotle's idea of the soul as the form of the body. He fully endorsed the view that mind is essentially purposive or conative.

In his *Analytic Psychology* Stout anticipated many of the principles later developed by the Gestalt school and by psychoanalytic writers such as Sigmund Freud. His idea of relative suggestion provided Charles Spearman with his noegenetic law of the eduction of correlates. His interpretation of apperception in terms of noetic synthesis fitted well with the later neurological theory of the integrative action of the nervous system and with the doctrine that the mind, like the nervous system, has a hierarchical structure. Finally, his account of physical objects was intended to combine all that was sound in both the idealistic and the realistic theories. Because of their eclectic and synthetic character, his writings supplied a coherent and systematic justification for the empirical position to which later British psychologists, working in the newer and narrower fields of biological, physiological, experimental and statistical psychology, were led. He was by far the most influential British psychologist of the early 20th century. (Cy. B.)

STOUT, SIR ROBERT (1844–1930), New Zealand judge and statesman, was born on Sept. 28, 1844, at Lerwick, Shetland Islands. In 1863 he went to New Zealand and in 1867 began to study law, being admitted a barrister and solicitor of the supreme court in 1871. After serving in the Otago provincial council during 1872–75 and as provincial solicitor, 1873–75, he became Liberal

M.P. for Caversham in 1875 and for Dunedin in 1876, and played an important part in the formation of the Liberal party. In 1878 he was made attorney general and minister for lands and immigration. In 1884 he formed a ministry with Sir Julius Vogel in which he was premier, attorney general and minister of education. He lost his seat at the general election in 1887 but was elected as an independent Liberal at a by-election in 1893 and at the general elections of 1893 and 1896. He resigned in 1898 and was chief justice, 1899–1926, and a member of the legislative council, 1926–30.

Stout's principal measures were the Land act of 1877; the first Land Tax act (which he drafted in co-operation with John Ballance and which became law in 1878); the Hospitals and Charitable Institutions act of 1885; and the Civil Service Reform act of 1886. He was a member of a royal commission appointed in 1881 which formed the code of civil procedure. Enacted in 1882, it made the supreme court procedure of New Zealand one of the simplest in the world. Stout died in Wellington on July 19, 1930.

Stout's works include *Leasing State Lands* (Dunedin, 1875); *Notes on the Progress of New Zealand, 1864–84* (Wellington, 1886); "The Hon. John Ballance," *Review of Reviews* (Australian edition) (Melbourne, 1893); with Sir W. A. Sim, *The Practice of the Supreme Court and Court of Appeal of New Zealand*, 9th ed. (Wellington, 1955).

(A. T. CL.)

STOUT: *see* BEER; BREWING.

STOVES: *see* HOME EQUIPMENT; FURNACE, ELECTRIC; HEATING AND VENTILATION.

STOW, JOHN (*c.* 1525–1605), English historian and antiquary, was born about 1525 in London. He learned the trade of his father, a tailor, but possibly did not practise it much after he grew up. In 1561 he published his first work, *The Woorkes of Geffrey Chaucer, Newly Printed With Divers Addicions Whiche Were Never in Printe Before*. This was followed in 1565 by his *Summarie of Englyshe Chronicles*, which was frequently reprinted, with slight variations, during his lifetime. In 1580 Stow published his *Annales, or a Generale Chronicle of England from Brute Until the Present Yeare of Christ 1580*.

The work by which Stow is best known is his *Survey of London*, published in 1598, not only interesting from the quaint simplicity of its style and its amusing descriptions and anecdotes, but of unique value from its minute account of the buildings, social condition and customs of London in the time of Elizabeth I. Through the patronage of Archbishop Parker, Stow was enabled to print the *Flores historiarum* of Matthew of Westminster in 1567, the *Chronicle* of Matthew Paris in 1571, and the *Historia brevis* of Thomas Walsingham in 1574. At the request of Parker he had himself compiled a "farre larger volume," *An history of this island*, but the manuscript is lost. Stow remained poor all his life, and was authorized by James I to appeal for alms in 1604. He died on April 6, 1605, and was buried in the church of St. Andrew Undershaft, where is his monument.

A number of Stow's manuscripts are in the Harleian collection in the British museum. Some are in the Lambeth library (No. 306); and from the volume which includes them were published by the Camden society, edited by James Gairdner, *Three Fifteenth-Century Chronicles, with Historical Memoranda by John Stowe the Antiquary, and Contemporary Notes of Occurrences written by him* (1880).

Of the many editions of Stow's *Survey of London, see* that with notes by C. L. Kingsford (Oxford, 1908).

STOWE, HARRIET ELIZABETH BEECHER (1811–1896), U.S. writer and philanthropist, is famous as the author of *Uncle Tom's Cabin*. She was born at Litchfield, Conn., on June 14, 1811, the daughter of the clergyman Lyman Beecher (*q.v.*). After her mother's death in 1815, she came most directly under the influence of her eldest sister Catharine, who a few years later set up a school in Hartford, where Harriet was first a pupil and afterward a teacher. In 1832 her father, who had for six years been the pastor of a church in Boston, accepted the presidency of the newly founded Lane Theological seminary at Cincinnati, O. Catharine went with him and established a pioneer college for women, the Western Female institute. Harriet was an assistant at the school, taking an active part in the literary and school life, contributing stories and sketches to local journals, and compiling a school geography, until the school closed in 1837. She was married Jan. 6, 1836, to one of the professors in the seminary, Calvin Ellis Stowe. In the midst of privation and anxiety, due largely to her husband's precarious health, she wrote continually, and in 1843 published *The Mayflower, or Sketches of Scenes and Characters among the Descendants of the Pilgrims*.

She lived 18 years in Cincinnati, separated only by the Ohio river from a slave-holding community, coming in contact with fugitive slaves and learning from friends and her own visits the life of the South. In 1850 her husband was elected to a professorship in Bowdoin college, Brunswick, Me., and moved his family there.

In Brunswick, Mrs. Stowe wrote, for serial publication in the *National Era*, an antislavery paper of Washington, D.C., the story of *Uncle Tom's Cabin; or, Life Among the Lowly*. The publication in book form (1852) was a factor which must be reckoned in summing up the causes of the Civil War. Her name was anathema in the South. The book sprang into unexampled popularity, and was translated into at least 23 languages; the dramatic adaptation was also immensely popular. Mrs. Stowe reinforced her story with *A Key to Uncle Tom's Cabin* (1853), in which she accumulated a large number of documents and testimonies against slavery and in defense of criticism of her accuracy. In 1853 when she made a journey to Europe she was lionized in England by people from all walks of life. (A magazine article published in 1869, "The True Story of Lady Byron's Life," detailing her charge that the poet had had an incestuous love for his sister, later turned British public opinion against her.) In 1856 she published *Dred: a Tale of the Dismal Swamp*, in which she threw the weight of her argument on the deterioration of a society resting on a slave basis. The establishment of the *Atlantic Monthly* in 1857 gave her a constant vehicle for her writings, as did the *Independent* of New York, and later the *Christian Union*, of which papers successively her brother, Henry Ward Beecher, was one of the editors.

She thereafter led the life of a woman of letters, writing novels, of which *The Minister's Wooing* (1859) is best known, and many studies of social life in both fiction and essay. Her *Pearl of Orr's Island* (1862) was credited by the U.S. writer Sarah Orne Jewett with having revealed to her the literary value of the country folk. Mrs. Stowe published also a small volume of religious poems, and toward the end of her career gave some public readings from her writings. In 1852 Calvin Stowe accepted a professorship in the Theological seminary at Andover, Mass., and the family made its home there till 1863, when he retired and moved to Hartford. After the close of the Civil War, Mrs. Stowe bought an estate in Florida, and she spent many winters there. After the death of her husband in 1886, she lived in the seclusion of her Hartford home, where she died on July 1, 1896. She is buried by the side of her husband at Andover.

BIBLIOGRAPHY.—The Riverside edition of Mrs. Stowe's works was published in 1899 and 1906 in 16 volumes (with an additional volume of biography). *See* also Annie Fields (ed.), *Life and Letters of Harriet Beecher Stowe* (1897); C. E. and L. B. Stowe, *Harriet Beecher Stowe: the Story of Her Life* (1911); F. Wilson, *Crusader in Crinoline* (1941); C. H. Foster, *Rungless Ladder: Harriet Beecher Stowe and New England Puritanism* (1954); and J. C. Furnas, *Goodbye to Uncle Tom* (1956).

(H. E. S.; X.)

STOWELL, WILLIAM SCOTT, BARON (1745–1836), English admiralty judge, author of influential doctrines in international law and brother of Lord Chancellor Eldon, was born Oct. 17, 1745, near Newcastle upon Tyne. He was educated at the Newcastle grammar school and at Corpus Christi college, Oxford, from which he received his B.A. degree in 1764, his M.A. in 1767, his bachelor of civil law in 1772 and the doctor of civil law in 1779. He was an intimate friend of Samuel Johnson. He elected to practise in the admiralty and ecclesiastical courts and became advocate general for the office of lord high admiral in 1782 and in 1798 judge of the high court of admiralty. On the coronation of George IV in 1821, he was raised to the peerage as Baron Stowell. He died on Jan. 28, 1836.

He stands in the front rank of English judges with Sir Matthew Hale and Lord Mansfield. His decisions in admiralty were the first decisions of the English admiralty court to be published, and

he was therefore free to be guided by writers on Roman, canon and international law. His decisions were accepted by the United States courts through the work of Justice Joseph Story. The doctrines of international law with which the name of Lord Stowell are identified include: the perfect equality and entire independence of all states; the elementary rules of international law binding even semibarbarous states; the principles that a blockade, to be binding, must be effectual, and that a prize court is a court of the law of nations. (A. Dm.)

STRABANE, an urban district and market town of County Tyrone, Northern Ire., 14 mi. S.S.W. of Londonderry by road. Pop. (1951) 6,619. Area 2.9 sq.mi. The rivers Mourne and Finn join there to form the Foyle. Shirt and collar making is the chief industry, with flax spinning at Sion Mills village.

STRABO of Amasya in Pontus, Greek geographer and historian whose *Geography* is the only general treatise on the subject surviving from antiquity, was born *c.* 63 B.C. Of distinguished family, especially on his mother's side, Strabo was sent abroad at an early age to study under well-known teachers at Nysa in Caria, Rome and perhaps elsewhere. Later he lived for long periods in Alexandria and in Rome. It is not known where he finally settled. He boasts of having traveled more widely than any of his predecessors, "east to west from Armenia to the Tuscan region facing Sardinia, and north to south from the Black sea to the borders of Ethiopia." With the one major exception of a journey up the Nile *c.* 25 B.C. in the retinue of the governor of Egypt, Aelius Gallus, there are practically no details of these journeys and there is no idea of their purpose. Strabo seems to have kept to a few main routes: he had seen little of Italy, and in Greece he had visited Corinth and probably Athens, but not Delphi or Mycenae.

Strabo was a Stoic, and he wrote (as he tells us) for the instruction of educated statesmen and other readers of high station. His belief in the importance of history and geography to such men derives from Polybius, who influenced him profoundly. His main work—a *History* in 47 books, lost except for a handful of small fragments—after four introductory books continued Polybius' *History* at least down to Caesar's death. The *Geography,* probably a long appendix to this work, survives almost in full. The first two books form a general introduction, the next eight deal with Europe, the following six with Asia, the last with Africa (its shape and size totally misconceived). Since little is said about events after 7 B.C., the work was probably completed about that date. A few notes were added about A.D. 18–19, and there is one reference (added by an editor?) to an event of *c.* A.D. 23. But the work was never systematically revised.

Although Strabo defends the orthodox picture of a spherical universe centred in a spherical earth, both divided into five zones, and although, in discussing his predecessors' theories, he has much philosophical argument, he attaches little importance to such general matters and defines the geographer's proper task as "to describe the known parts of the *oikoumene*" (inhabited portion of the earth). He stresses the geographer's need for mathematical and scientific knowledge and makes extensive use of Hipparchus; but his own inclination is clearly toward descriptive and historical geography, though with no stylistic pretensions. His knowledge is chiefly gathered from books: his reading—mainly Greek, but some Latin—was wide and critical. He loves controverting famous predecessors, often with common sense and more recent information, sometimes with irritating puerility. His work shows little sense of order, except that imposed by the subject, and is full of historical and mythological excursuses, which are often of interest. His information is at best as good as that of his sources and can be far from recent: full of admiration for Homer, he bases his three books on Greece on Apollodorus' commentary on Homer's "Ships' catalogue."

The *Geography* had no influence on 1st-century writers. By the 6th century it was already a classic, and its reputation endured for centuries. To us its chief importance lies in the valuable source material (from Eratosthenes, Polybius, Posidonius and others) that it transmits: without Strabo we could hardly begin to follow the development of ancient geography. The first good edition was by I. Casaubon (1587, revised 1620), by whose pages the text is normally cited. The Loeb edition (1917–32) by H. L. Jones with bibliography is very useful. The first volume (a general commentary) of a projected German edition by W. Aly, appeared in 1957. (E. Ba.)

STRACHAN, JOHN (1778–1867), educator and first Anglican bishop of Toronto, whose principal achievement was to organize the church in Canada as a self-governing denomination within the Anglican communion, was born in Aberdeen, Scotland, April 12, 1778. Educated at King's college, Aberdeen, he emigrated to Canada in 1799, was ordained (1803), and was appointed successively missionary at Cornwall 1803, missionary at York (later Toronto) 1812, archdeacon of York, 1827, and bishop of Toronto, 1839. A determined champion of the church-state connection, he fought a steady rear guard action to retain the privileged position of the church as a branch of the English establishment in the colony. As that position was gradually lost under colonial pressure, he created the first Anglican synod of clergy and laity in the British colonial empire (Toronto, 1851), raised endowment funds, subdivided his diocese, and presided over the first episcopal election by clergy and laity outside of the United States of America (1857).

During his middle years he was a controversial political figure, as member of the executive council of Upper Canada after 1815, and as the power behind the ruling Tory clique known as the "family compact," most of whose members had been educated at his hands. His political power waned with the granting of responsible government to the colony.

In 1827 he secured a royal charter for the University of King's college, Toronto, and served as its first president. When the foundation was secularized as the University of Toronto, he founded the University of Trinity college in 1852 as a church institution. His published works, about 30 in number, are largely polemical. He died in Toronto on Nov. 1, 1867. (J. L. H. H.)

STRACHAN, JOHN (1862–1907), Scottish philologist, noted for his classical and Celtic studies, was born near Keith, Banffshire, Scot., on Jan. 31, 1862. After a brilliant record as a student (Aberdeen, Cambridge, Göttingen and Jena universities), he held professorships of Greek and comparative philology at Manchester university, then the new chair of Celtic. In this field he won fame notably through the monumental *Thesaurus Palaeohibernicus* (with W. Stokes; 2 vol., 1901–03), by scores of articles in German and British journals and by the posthumous *Introduction to Early Welsh* (1909). He died on Sept. 25, 1907, at Prestwich. (R. A. Fs.)

STRACHEY, GILES LYTTON (1880–1932), British critic and biographer, son of Sir Richard Strachey, the Indian administrator, was born on March 1, 1880, and educated at Trinity college, Cambridge. He was a leading figure in the remarkable "Bloomsbury" circle which included his Cambridge friends J. M. Keynes, Desmond MacCarthy, Leonard Woolf and Clive Bell and also the two daughters of Sir Leslie Stephen, Vanessa Bell and Virginia Woolf. In 1912 he published *Landmarks in French Literature* and throughout his life did much by his articles to make his countrymen appreciate the precision and delicacy of French poetry and prose. His greatest achievement, however, was in biography. He made his reputation with *Eminent Victorians* (1918) and *Queen Victoria* (1921) and added to it with *Elizabeth and Essex* (1928) and *Portraits in Miniature* (1931). Strachey made a revolution in the writing of biography. Treating his subjects from a highly idiosyncratic point of view, he determined to create a work of art rather than to record activities and events. He was fascinated by personality and motive and delighted in pricking the pretensions of the great and reducing them to life-size in a prose style that was at once elegant, sinuous and singularly ironical.

His aim was to paint a portrait; and though this led to caricature and sometimes, through tendentious selection of material, to inaccuracy, he taught biographers a sense of form and of background as well as sharpening their critical acumen. His many imitators also hastened to "debunk" the past, but they lacked Strachey's detachment, taste and Voltairean assurance; nor could they rival his best work, *Queen Victoria,* in which Strachey combined his gift

for narrative with a power of analyzing characters with whom he had nothing in common. His defects as a biographer arose mainly from his clear but limited vision of life. He did not love the past for its own sake; and he saw politics largely as intrigue, religion as a ludicrous anachronism and personal relations as the supremely important facet of life. Though bitterly attacked during and after his lifetime, Strachey remains a phenomenon in English letters and a pre-eminent English humorist and wit.

Two collections of essays, *Books and Characters* (1922) and *Characters and Commentaries* (posthumously in 1933) were also published. Strachey died unmarried on Jan. 21, 1932.

(N. G. A.)

STRACHWITZ, MORITZ KARL WILHELM ANTON, Graf Von (1822–1847), German poet, was born on March 13, 1822, at Peterwitz near Frankenstein in Silesia. After studying in Breslau and Berlin he settled on his estate in Moravia, where he devoted himself to literary pursuits. He died on Dec. 11, 1847, at Vienna in his 25th year. He had already revealed a lyric genius of remarkable force and originality. His heroic ballads are notable and had much influence upon Theodor Fontane and Detlev von Liliencron. His first collection of poems, *Lieder eines Erwachenden*, appeared in 1842 and went through several editions. *Neue Gedichte* were published in 1848.

Bibliography.—Strachwitz's collected *Gedichte* appeared first in 1850 (8th ed., Breslau, 1891); a convenient reprint will be found in Reclam's *Universalbibliothek* (Leipzig). H. M. Elster (ed.), *Sämtliche Lieder und Balladen* (Berlin, 1912) has a biography. *See* also T. Baumstark, *Die Dichtung des Grafen Moritz von Strachwitz* (Münster, 1935).

STRADELLA, ALESSANDRO (?1645–1682), Italian composer, one of the most accomplished musicians of the 17th century, was probably born at Naples. The generally accepted story of his life was first circumstantially narrated in Bonnet-Bourdelot's *Histoire de la musique et de ses effets* (1715). According to this account, Stradella was engaged by a Venetian nobleman to instruct his mistress, Ortensia, in singing. Stradella eloped with Ortensia to Rome, where the outraged Venetian sent two paid *bravi* to put him to death. On their arrival in Rome the assassins learned that Stradella had just completed a new oratorio, over the performance of which he was to preside on the following day at S. Giovanni in Laterano. They determined to kill him as he left the church; but the beauty of the music conquered them, and they warned the composer of his danger. Thereupon Stradella fled with Ortensia to Turin, where, notwithstanding the favour shown to him by the regent of Savoy, he was attacked one night by another band of assassins, who, headed by Ortensia's father, left him on the ramparts for dead. Through the connivance of the French ambassador the ruffians escaped; and Stradella, recovering from his wounds, married Ortensia, by consent of the regent, and removed with her to Genoa. There he believed himself safe; but a year later he and Ortensia were murdered in their house by a third party of assassins in the pay of the implacable Venetian. Research has, however, driven several holes in this picturesque story.

The first certain date in Stradella's life in 1672, in which year he composed a prologue for the performance of Cesti's opera *La Dori* at Rome; he probably spent a considerable time at Rome about this period, since his cantatas and other compositions contain frequent allusions to Rome and noble Roman families. There is, however, no proof that he ever performed the oratorio *S. Giovanni Battista* in the Lateran. Documents in the archives at Turin relate that in 1677 he arrived there with the mistress of Alvise Contarini, with whom he had eloped from Venice. We hear of Stradella last at Genoa. An opera by him, *La Forza dell' amor paterno*, was given there in 1678, and his last composition, *Il Barcheggio* (*i.e.*, a "Water-Music"), was performed on June 16, 1681 in honour of the marriage of Carlo Spinola and Paola Brignole. That he died at Genoa in Feb. 1682 is established by documents in the archives at Modena.

Stradella's best operas are *Il Floridoro*, also known as *Il Moro per amore*, and *Il Trespolo tutore*, a comic opera in three acts which worthily carried on the best traditions of Florentine and Roman comic opera in the 17th century. The oratorio *S. Giovanni Battista* displays the same skill in construction and orchestration (so far as the limited means at his disposal permitted) as the operas. A serenata for voices and two orchestras, *Qual prodigo ch'io miri*, was used by Handel as the basis of several numbers in *Israel in Egypt*, and was printed by Chrysander (Leipzig 1888); the manuscript, however, formerly in the possession of Victor Schoelcher, from which Chrysander made his copy, has entirely disappeared. The well-known aria *Pietà, signore*, also sung to the words *Se i miei sospiri*, cannot possibly be a work of Stradella.

STRADIVARI, ANTONIO (1644–1737), Italian violin-maker, was associated throughout his life with Cremona, where he brought the craft of violin-making to its highest pitch of perfection. The obscure details of his life were thoroughly worked out in the monograph on him by W. H. Hill, A. F. Hill and Alfred Hill (1902). He was still a pupil of Nicolas Amati in 1666, when he had begun to insert his own label on violins of his making, which at first follow the smaller Amati model, solidly constructed, with a thick yellow varnish. It was not till 1684, that he began to produce larger models, using a deeper coloured varnish, and beautifying them in various details, his "long" patterns (from 1690) representing a complete innovation in the proportions of the instrument; while from 1700, after for a few years returning to an earlier style, he again broadened and otherwise improved his model. He also made some beautiful violoncellos and violas. The most famous instruments by him are:—*Violins:* the "Hellier" (1679), the "Sellière" (before 1680), the "Tuscan" (1690), the "Betts" (1704), the "Ernst" (1709), "La Pucelle" (1709), the "Viotti" (1709), the "Vieuxtemps" (1710), the "Parke" (1711), the "Boissier" (1713), the "Dolphin" (1714), the "Gillot" (1715), the "Alard," the finest of all (1715), the "Cessot" (1716), the "Messiah" (1716), the "Sasserno" (1717), the "Maurin" (1718), the "Lauterbach" (1719), the "Blunt" (1721), the "Sarasate" (1724), the "Rode" (1722), the "Deurbroucq" (1727), the "Kiesewetter" (1731), the "Habeneck" (1736), the "Muntz" (1736). *Violas:* the "Tuscan" (1690), two of 1696 formerly belonging to the king of Spain, the "Archinto" (1696), the "Macdonald" (1701), and the "Paganini" (1731). *Violoncellos:* the "Archinto" (1689), the "Tuscan" (1690), the "Aylesford" (1696), the "Cristiani" (1700), the "Servais" (1701), the "Gore-Booth" (1710), the "Duport" (1711), the "Adam" (1713), the "Batta" (1714), the "Piatti," the finest of all (1720), the "Bandiot" (1725), the "Gallay" (1725). Antonio Stradivari's sons Francesco (1671–1743) and Omobono (1679–1742) were also violin-makers, who assisted their father, together with Carlo Bergonzi, who appears to have succeeded to the possession of Antonio's stock-in-trade. The Stradivari method of violin-making created a standard for subsequent times; but the secret of his varnish, soft in texture, and shading from orange to red, though much debated, has never been discovered.

See E. N. Doring, *How Many Strads?* (Chicago, 1945).

STRAFFORD, THOMAS WENTWORTH, Earl of (1593–1641), English statesman, sponsor of the policy of "Thorough" and victim of the failure of mediation between king and parliament before the Great Rebellion, was born on April 13, 1593, the son of Sir William Wentworth of Wentworth Woodhouse, near Rotherham, and of Anne Atkinson. He was educated at St. John's college, Cambridge, and was admitted to the Inner Temple in 1607. His father married him in 1611 to Margaret Clifford, daughter of Francis, earl of Cumberland, and the Clifford connections were important for Wentworth's early career. In 1614 Wentworth sat for Yorkshire in parliament, and in 1615 when Sir John Savile, an elder statesman of Yorkshire politics, out of favour at court, resigned his office of *custos rotulorum* of the West Riding, Wentworth succeeded him. Savile claimed in 1617 that he had resigned on the understanding that Wentworth would surrender the office when Savile was back in favour. Wentworth denied this and kept the office. The resulting feud between them dominated the politics of Yorkshire for the next decade.

Wentworth represented Yorkshire in the parliament of 1621. He had to use all his influence to get Sir George Calvert, the secretary of state, elected and to defeat Savile. In this parliament Wentworth spoke frequently, though he was not a leader of the commons. His alliance with Calvert symbolized not only his contacts with court circles, but also one abiding attitude, his pro-Spanish views on foreign policy. Wentworth never took a very deep interest in foreign policy; but his instinct to avoid foreign entanglements went with a considerable hostility to Roman Catholic recusants. He regarded the remedying of grievances as one of parliament's main tasks, but one to be performed in close collaboration with the king. In 1621 he was much concerned with law reform and the promotion of the cloth trade, but such business was interrupted by the dispute between king and commons over foreign policy, which led to a dispute about parliamentary privileges. To Wentworth, parliamentary privilege and royal prerogative were complementary, but in 1621 his emphasis was on the former.

Parliament was summoned again in 1624 because the duke of Buckingham wanted revenge on Spain, after the failure of the

match between Charles and a Spanish infanta. Wentworth kept clear of foreign policy in the commons, but expressed his mistrust of Buckingham's policy privately. In 1625 he married (his first wife had died in 1622) Arabella Holles, daughter of John, earl of Clare, whom James, immediately before his death, was thought to have chosen for high office in order to check Buckingham. Wentworth's attitude became more public, when in Charles's first parliament he opposed the granting of further supply, until internal affairs had been dealt with. When parliament was dissolved he and six others were made sheriffs to prevent their re-election to the next parliament. Wentworth's old enemy Savile was now high in Buckingham's favour. In 1626 Wentworth's protégé Sir Christopher Wandesford was one of the managers of Buckingham's impeachment. When Buckingham tried to finance the war without parliament, Wentworth refused to pay the forced loan and was imprisoned.

Wentworth was an acknowledged leader of the commons when they met in 1628. He and the other leaders agreed that their first aim should be to obtain a bill of rights to secure the liberties of the subject. Charles refused to consider such a bill, whereupon Sir Edward Coke produced the expedient of a Petition of Right (*see* ENGLISH HISTORY). Wentworth stood out for a bill, but eventually gave way. After Charles's reluctant acceptance of the Petition, Sir John Eliot began an attack on Buckingham which eventually led to a prorogation. Wentworth did not join the attack, but played the part of a mediator between crown and commons on other subjects. Meanwhile Buckingham was becoming readier to compromise with the parliamentary leaders. As part of this new deal, Wentworth was made a baron in July 1628.

After Buckingham's death (Aug. 23, 1628) the policy of Richard Weston, the lord treasurer (later earl of Portland), was conciliation at home and peace abroad. In December Wentworth became president of the Council of the North. But when parliament met again and Weston found himself under attack by Eliot, he lost his nerve and parliament was dissolved. Wentworth only became a privy councilor eight months later in Nov. 1629.

As president of the north Wentworth was a vigorous administrator, especially active in the collection of fines for knighthood and recusancy, in the supervision of justices of the peace and in asserting the council's jurisdiction as a prerogative court. He was also active as a privy councilor and so began his friendship with William Laud, then bishop of London, who became archbishop of Canterbury in 1633. His success was such that some thought that Weston, his patron in 1628, was becoming jealous and offered him the lord deputyship of Ireland at the end of 1631 in order to get him out of the way. Wentworth accepted the deputyship, but he did not go to Ireland until July 1633. His predecessors had often had their policies thwarted by factions in the Irish privy council intriguing with the English government, sometimes in alliance with the Irish parliament. Wentworth determined to avoid this and to crush all attempts to resist his authority.

He was supremely successful in managing the Irish parliament. His aim was to complete the assimilation of Irish society to the forms of English law and government and to ensure a Protestant hegemony by reviving the policy of large-scale plantation. This meant opposition not only from the native Roman Catholics but from the clique of Protestant settlers which dominated the administration. For, if the government and established church were to play their part, they had to be rescued from the exploitation of profiteers, such as the earl of Cork (*q.v.*). So juries were browbeaten into accepting the plantation of Connaught, while to break the opposition to his will in the Irish privy council Wentworth fined the earl of Cork for misappropriating church lands, court-martialed the vice-treasurer, Lord Mountnorris, and imprisoned the lord chancellor, Adam Loftus. Wentworth himself undertook the farm of the Irish customs at a higher rent than others would offer; but increased trade, one symptom of the prosperity brought by his rule, gave him large profits which excited further jealousies. The increased revenues were used to support a small but efficient army, to suppress piracy and to assist the English treasury. But those who had been thwarted by Wentworth sought a hearing in England, where courtiers, who found Wentworth unco-operative about Irish land grants, were ready to listen. Such opposition centred round the court of Henrietta Maria. To counter it Wentworth relied on Laud's influence with Charles and in their correspondence "Thorough" was a code word for administrative efficiency, the antithesis of corrupt court intrigues. Yet Charles forced Wentworth to dissolve parliament in 1635 and ultimately disregarded his advice about the plantations of Connaught and Londonderry.

When the first Bishops' War in Scotland ended disastrously (June 1639), Wentworth's advice was in demand. He came to England in Sept. 1639 and became one of Charles's inner council. But he was not in charge of affairs; instead Charles enforced a formal reconciliation between Wentworth and his enemies at court. With difficulty Wentworth persuaded Charles to summon parliaments in England and Ireland in order to obtain supplies to crush the Scots. But Henrietta Maria's nominee, Sir Henry Vane, was appointed as a new secretary of state. Charles, however, created Wentworth earl of Strafford (Jan. 1640).

Strafford crossed to Ireland in March, obtained parliamentary subsidies and arranged to increase the Irish army. Returning to England in April, he found the commons unwilling to grant supplies. Strafford persuaded Charles to make concessions by reducing the amount asked for and by allowing ship money to be called in question. But Vane, presumably on Charles's authority, again asked the commons for the original sum of 12 subsidies. This request was refused and in the privy council Strafford reluctantly voted with the majority for a dissolution.

Strafford believed that in order to defend England against the Scots, the king was entitled to use emergency powers to raise and maintain an army. Charles's other councilors had already despaired of a successful defense and were unwilling to risk vigorous action. All the odium fell on Strafford. When a truce had to be made with the Scots and a new parliament summoned, Strafford was a marked man. He wished to retire to Ireland, but Charles persuaded him to come to London. Strafford's enemies at court were busy making their peace with the parliamentary leaders and part of the price paid was the sacrifice of Strafford.

Strafford came to London ready to stake everything on striking first by impeaching the opposition leaders for treasonable correspondence with the Scots. But John Pym (*q.v.*), leader of the commons, forestalled him. On Nov. 11, 1640, Strafford's impeachment was voted and he was committed to custody. The main accusation against him was that of endeavouring to subvert the fundamental laws of the kingdom, notably by preparing to rule by military force, but the detailed charges mainly concerned his administration in Ireland. At his trial Strafford defended himself so successfully that Pym resorted to a bill of attainder. By the opportune revelation of the first army plot Pym got the bill through the lords. Charles had promised to save Strafford's life, but under pressure of riots against Henrietta Maria, he signed Strafford's death warrant. On May 12, 1641, Strafford was executed.

Strafford's consciousness of his own rectitude sometimes made him arrogant, overbearing or revengeful in behaviour. He died in failure, his work in Ireland already undone, yet his attainder was itself a tribute from his enemies to his commanding stature and potentiality as a statesman. Strafford, like Francis Bacon, was a victim of that half-mythical vision of a Tudor harmony of royal prerogative and popular consent which both strove to re-create.

BIBLIOGRAPHY.—William Knowler (ed.), *Letters and Despatches of the Earl of Strafford*, 2 vol. (1739); William Laud, *Works*, vol. vii (1860); T. D. Whitaker, *Life and Letters of Sir George Radcliffe* (1810); H. O'Grady, *Strafford and Ireland*, 2 vol. (1923); R. R. Reid, *The King's Council in the North* (1921); F. Relf, *The Petition of Right* (1917); S. R. Gardiner, article in *Dict. Nat. Biog.*; Lady Burghclere, *Strafford* (1931); C. V. Wedgwood, *Strafford* (1938); Lord Birkenhead, *Strafford* (1938). Strafford's papers are in the Sheffield Central library.
(J. P. Cr.)

STRAIN GAUGE. The strain gauge has many scientific applications, particularly in engineering. In addition to the measurement of strain, the gauge has been applied as a transducer to determine magnitudes of forces, torques, pressures, accelerations and similar physical quantities. This wider application has resulted from the development of a gauge using electric phenomena,

but mechanical and optical magnification are also employed to measure strain. If loads are applied to materials such as metals or glass, the deformations usually are not discernible to the eye. The function of a strain gauge is to magnify the small deformations on the surfaces of mechanical parts or structures to the point where they can be measured.

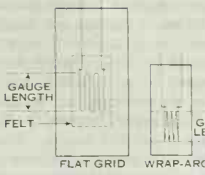


FIG. 1.—STRAIN GAUGES SHOWING CONSTRUCTION

In engineering, normal strain is defined as the deformation in a given direction divided by the gauge length in the same direction. If a 2 in. length, called the gauge length, is marked on a piece of rubber thread, it can be made to increase perceptibly when the thread is stretched. If the forces are such that the original length has increased to $2\frac{1}{4}$ in., the total deformation obviously is $\frac{1}{4}$ in., and the strain is $\frac{1}{8}$ in. per inch. The strain has the same value throughout the length of the rubber strip for a given force, regardless of the gauge length used, and increases as the force becomes greater.

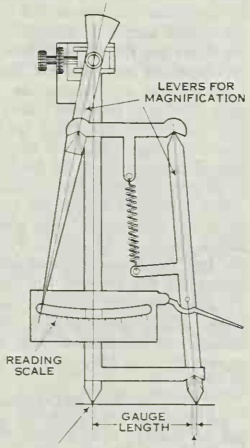
If a block is shortened under a compressive load, compressive strains are developed. In parts of irregular shape which may be subjected to more complicated loading systems, the strain may vary from point to point, and a short gauge length therefore is required.

Engineers are interested in regions of high strain on parts or structures. If strains are known, properties of the material may be used to compute stresses, which are defined as forces per unit area. Conversely, if stresses are known, strains can be calculated. In design, mathematical equations are used to proportion the parts of a machine or structure so that the stresses caused by the applied loads are within the allowable limits of the materials specified. Strain gauges are applied to the actual structure or part to check these mathematical computations or to determine the strains at points when it is difficult or impossible to compute them.

One type of strain gauge widely used for this purpose is the bonded electric resistance strain gauge developed in 1938. This gauge consists of a length of fine wire, 0.001 in. in diameter, formed into a grid as shown in fig. 1 and cemented between two pieces of thin paper. Small and light, the gauge is made in many types, mostly in the size of a postage stamp. Gauge lengths as short as $\frac{1}{32}$ in. are available. The gauge is cemented to the part being tested. Many different cements have been used. As the part is loaded, the wire grid faithfully follows the deformations under it, either in tension or in compression, and the length and cross sectional area of the wire are changed. This causes the electrical resistance of the wire to change; even though this change is small—often a fraction of an ohm—electric circuits and instruments are available to detect the variation. A simple relation called the gauge factor exists between the unit change in resistance and the strain. This relation may be expressed mathematically by

$$\text{gauge factor} = \frac{\Delta R}{R} \Big/ \frac{\Delta L}{L} \qquad (1)$$

where ΔR is the change in resistance caused by strain, R is the original resistance of the gauge, and $\Delta L/L$ is the strain under the grid.



BY COURTESY OF A. U. HUGGENBERGER

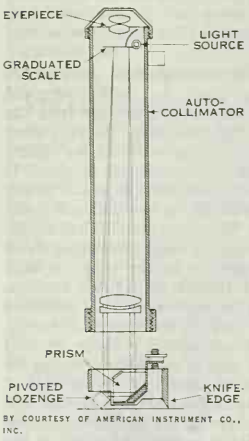
FIG. 2.—A. HUGGENBERGER TENSOMETER LEVER SYSTEM (THE READING SCALE IS APPROXIMATELY 1 IN. LONG)

The manufacturer supplies the value of the gauge factor and the original resistance of each lot of gauges. If ΔR or a voltage change proportional to it is then measured, the strain can easily be calculated. The gauge is sensitive enough to detect strains as low as 0.000001 in. per inch and to measure strains as high as 0.03 in. per inch; special types can measure strains three times this value.

A gauge in which the grid is formed from metallic foil has been used to measure strains at temperatures as high as 1200° F.

Advantages are readily apparent. Many gauges have been bonded to structures and parts at strategic spots and their behaviour under load reported by wires leading to electronic instruments far removed from the test spots. Dynamic or rapidly changing strains occurring for example in freight car couplings or rocket firings have been recorded by oscilloscopes which handle strain variations as high as 50,000 cycles per second.

Strain gauges are used for measuring quantities other than strains. Load cells of various designs have been developed to measure force. Gauges have been bonded to short steel cylinders, the output of the gauges being calibrated in terms of the forces applied to them. Cells with capacities as high as 3,000,000 lb. have been built by the U.S. bureau of standards and calibrated in terms of force with an accuracy of 0.1%. Cantilever load beams have also been designed for force measurements. Gauges have been bonded to diaphragms, and the pressures on the diaphragms—static or variable—have been measured and recorded. Many other applications of the gauge have been made, especially to the measurement of accelerations. Other types of electric gauges—the unbonded resistance, electric inductance and electric capacitance gauges—have been utilized for the purposes described above.



BY COURTESY OF AMERICAN INSTRUMENT CO., INC.

FIG. 3.—TUCKERMAN OPTICAL STRAIN GAUGE

Since 1870, many different types of instruments have been developed using a compound lever system for magnifying strains. The Huggenberger Tensometer shown in fig. 2 is of this type. The gauge length on most models is $\frac{1}{2}$ in. or 1 in., and magnifications of 1,200 to 1 are common. These instruments have been used to measure static strains with reliable accuracy. Their chief disadvantages are large size and inability to respond to dynamic strains. Other gauges employing levers, or levers and dial indicators, are the Whittemore, Berry and Porter-Lipp gauges.

The Tuckerman optical strain gauge was designed to use a beam of light as an optical lever for magnifying deformations. The gauge consists of an extensometer and an autocollimator as shown in fig. 3. The extensometer has a fixed knife-edge and a pivoted lozenge which are brought into contact with the part undergoing deformation. A one-inch gauge length is shown. One side of the Stellite lozenge is ground and lapped to an optically flat mirror surface. The autocollimator houses an optical system which can be adjusted above the gauge and is so designed that its distance above the extensometer is not critical. A beam of light from a lamp in the autocollimator is directed upon a prism in the extensometer, then reflected to the mirror surface of the lozenge and back to a graduated scale in the autocollimator. As the lozenge rotates, the reflected light beam from its mirror surface moves across the scale and may be viewed through an eyepiece in the autocollimator. The smallest deformation measurable is 0.00005 in. and for a gauge length of 1 in. this corresponds to a strain of 0.000002 in. per inch. The accuracy of the gauge is excellent, and it has therefore been regarded as a basic instrument for strain

measurement.

Strain gauges of the types described above have been used in research laboratories and test stands of many companies, large and small, and in the control of manufacturing processes in refineries, steel mills and chemical industries.

See ELASTICITY.

BIBLIOGRAPHY.—M. Hetényi, *Handbook of Experimental Stress Analysis* (1950); C. C. Perry and H. R. Lissner, *The Strain Gage Primer* (1955); *Proceedings of the Society for Experimental Stress Analysis*, vol. i–xv, (1943). (E. O. ST.)

STRAIT JACKET, restraint equipment for violent patients or prisoners, made like a man's jacket, used especially in emergencies and designed to provide safety for both the patient or prisoner and the officer. Since it is subjected to hard usage, the strait jacket is made of some heavy material such as canvas and is securely fastened in the back with several locking buckles and straps. The sleeves are made long so as to extend beyond the fingers. After the strait jacket is put on the patient or prisoner and secured with straps and locks at the back, his arms are folded across his diaphragm and then the ends of the sleeves are fastened at the back. (R. G. CL.)

STRAITS QUESTION. The straits question concerns the international status of the Black sea straits, *i.e.*, the Bosporus and the Dardanelles, which form the only existing passage from the Black sea into the Mediterranean and vice versa. In 1453 these straits became a part of Turkey and thus entered into the eastern question (*q.v.*) until the end of the Ottoman empire in 1920. It is not identical with the problem of Constantinople as the former capital of the Byzantine empire, although the two problems might have been connected in the past. Whatever the political aspirations of the antagonists of the Ottoman empire, including Peter the Great, there was no straits question before 1774, for until then the Black sea was an inland Turkish sea. From 1774 on, with Catherine the Great acquiring the northern shore of the Black sea, the straits question appeared, involving the interests of Russia, Turkey and the non-Black sea naval powers. These interests do not coincide: Russia's aim has been to secure free passage for its warships through the straits but to close them to warships of non-Black sea powers; the latter, and especially Great Britain with its imperial lifeline, have aimed at keeping the Russian navy from entering the Aegean sea, or at least at opening the straits to the warships of all nations. The Ottoman empire resisting both pressures saw in the control of the straits a guarantee of its independence and security, and this point of view remained that of the Turkish republic. From 1947 the control of the straits by Turkey was also considered as a major guarantee of their own security by the non-Black sea powers.

The history of the straits question can be divided into three periods: (1) 1774–1841; (2) 1841–1920; and (3) after 1921.

1774–1841.—This period begins on July 10 (old style; July 21, new style), 1774, with the treaty of Kuchuk Kainarji (a village near Silistra, Bulg.) between Russia and Turkey, which inaugurated an epoch of bilateral Russian-Turkish agreements as to the straits. The treaty opened the Black sea and the straits to Russian commercial navigation, and this privilege was extended to the commercial navigation of other states by the peace treaty of Adrianople in 1829. Thereafter the principle of free commercial navigation in time of peace was not challenged by Turkey. Important for the economic development of south Russia, it did not conflict with the economic interests of the non-Black sea powers either. Russia pressed, however, for naval privileges; and Turkey, yielding in 1833, promised by the treaty of Unkiar Skelessi to close the straits on Russia's demand to warships of all non-Black sea powers. In 1840 the international situation prompted Nicholas I to look for a common action with Great Britain in Turkey; the treaty of 1833 was canceled by the London convention of July 15, 1841, which established a rule that no non-Turkish navy might cross the straits in peacetime.

1841–1920.—The London convention was the first general international convention concerning the straits, and it inaugurated a period in which they became the common concern of all great powers. The regime of the straits agreed upon in the London convention remained without modification until World War I. The prohibition of passage for warships of all nations was confirmed in 1856 (treaty of Paris), in 1871 (treaty of London) and in 1878 (treaty of Berlin). During the Bosnian crisis of 1908–09 Russia attempted to get a straits' regime more favourable to the passage of its warships, but in spite of Russia's political alignment with France and England in European affairs at that time, the attempt ended in a failure. During World War I, Turkey having fought on the side of the Central Powers, the Allied Powers agreed to the inclusion of the straits, after the war, in the Russian empire (treaty of London, 1915). However, the leaders of the October revolution having canceled all the imperial secret treaties and renounced all annexations, the problem of the straits was solved after the war differently and without the participation of Russia. The peace treaty of Sèvres (1920), which never went into force, would have imposed on Turkey the disarmament of the straits and their opening to all foreign warships, while the zone of the straits would have passed under the administration of an international commission.

After 1921.—The proposed regime of the treaty of Sèvres was disrupted by the opposition of the national Turkish government formed by Mustafa Kemal. Thereafter, a forceful Turkish factor was increasingly influencing the straits question. In the beginning, since both Turkey and Soviet Russia were opposed to the treaty of Sèvres, the two powers agreed bilaterally that the problem of passage of warships should be decided by a conference of the Black sea states only (treaty of March 16, 1921). However, no such conference was held, and the regime of the straits was finally decided with the participation of the western powers at Lausanne (convention of July 24, 1923). The convention of Lausanne maintained the demilitarization of the straits and the freedom of passage for warships "under any banner"; also the International Commission of the Straits was maintained under the chairmanship of Turkey—a concession, among others, to Turkey. The convention included a clause that no naval force larger than the largest Black sea fleet might be sent in peacetime into the Black sea by a non-Black sea state.

Because this arrangement satisfied neither Turkey nor Soviet Russia, it was revised at Montreux on July 20, 1936, when a new international situation resulting from a more aggressive German and Italian policy made it appear dangerous also to Great Britain and France. The Montreux conference re-established the full right of Turkey to fortify the straits and disbanded the International commission. Reaffirming the already traditional freedom of navigation for commercial ships of all nations, the Montreux convention agreed to differentiate between the warships of Black sea states and those of non-Black sea states, giving to the former privileges the latter did not receive. With the exception of aircraft carriers and submarines, all warships of Black sea states were allowed to cross the straits unrestrictedly in peacetime; only light warships of non-Black sea states had this right under limitation of tonnage, number of units simultaneously present and duration of cruising time. In time of war the straits would be closed to all participants in the conflict if Turkey was neutral.

The Montreux convention (*q.v.*) was scheduled for revision in 1956, the signatories being unable to agree on the premises of such a revision. In 1945 at Potsdam it was decided that each of the Big Three would communicate its views directly to Turkey concerning this revision, but the U.S.S.R., having raised on June 22, 1946, demands for joint control of the straits with Turkey, Great Britain and the U.S., backed Turkey in rejecting these demands as inconsistent with its independence and endangering its security. On May 30, 1953, the U.S.S.R. withdrew its demands of 1946.

See also TRUMAN DOCTRINE.

BIBLIOGRAPHY.—A. Sorel, *La question d'Orient au XVIIIe siècle*, 2nd ed. (1889); P. P. Graves, *The Question of the Straits* (1930); L. Eisenmann, P. Renouvin (eds.), *Constantinople et les Détroits*, documents, 2 vol. (1930–32); B. H. Sumner, *Russia and the Balkans, 1870–1880* (1937); E. Brüel, *International Straits*, 2 vol. (1947); B. A. Dranov, *Chernomorskiya Prolivy* (1948); J. C. Hurewitz (ed.), *Diplomacy in the Near and Middle East*, 2 vol. (1956). (M. ST.)

STRAITS SETTLEMENTS: *see* MALAYA.

STRALSUND, a seaport in the district of Rostock, Ger., on the Baltic strait of Bodden, opposite Rügen Island; it lies 135 mi. by rail N. of Berlin. Pop. (1950) 58,303. Stralsund was founded early in the 13th century and, though several times destroyed, steadily prospered. It extorted from King Eric of Norway a favourable commercial treaty in 1284–85, and in the 14th century it was second only to Lübeck in the Hanseatic league. Its early Protestant sympathies placed it on the side of Sweden during the Thirty Years' War, and in 1628 it successfully resisted a siege of 11 weeks by Albrecht von Wallenstein. Under the treaty of Westphalia (1648) Stralsund was ceded with the rest of western Pomerania to Sweden, and for more than a century and a half it was exposed to attack and capture. It was taken by France in 1807, by Denmark in 1814, and in 1815 it passed to Prussia. In 1945 Stralsund was taken by Soviet troops.

STRANG, JAMES JESSE (1813–1856), dissident Mormon leader, was born at Scipio, N.Y., March 21, 1813. Admitted to the bar, he also taught school, edited a weekly paper and was a country postmaster, before he followed his wife's family to Burlington, Wis., in 1843. Through his wife's brother-in-law, a Mormon, he visited Joseph Smith (*q.v.*) at Nauvoo, Ill., became a Mormon convert, and was ordained an elder by Smith in Feb. 1844. After Smith's assassination, Strang exhibited a letter, purportedly written by Smith and naming Strang as his successor. He also claimed to have had a vision appointing him "seer, revelator and prophet" of the Mormon Church. He tried to lead the church to settle at Voree, Wis., on the outskirts of Burlington, but was outgeneraled by Brigham Young. With a few of his 2,500 followers, he moved to Beaver Island, in Lake Michigan, where he caused himself to be crowned king after translating—allegedly—*The Book of the Law of the Lord* from golden plates from the Ark of the Covenant. Under its rules polygamy was sanctioned, and he was married to four wives. The Strangites were the object of much persecution, but Strang was able to preserve the sect and to gain acquittal in the several lawsuits brought against him. He had more than 5,000 followers in Michigan and Wisconsin when he was shot June 16, 1856, by two former Strangites. He was taken to Voree, where he died July 9, 1856. The Lake Michigan fishermen drove more than 2,000 of his believers from their homes and all but extinguished the Strangite branch of the Mormon Church.

STRANG, WILLIAM (1859–1921), British engraver and painter, was born at Dumbarton, Scot., on Feb. 13, 1859. He studied at the Slade School of Art, London, 1875–81. Strang worked in etching, mezzotint and engraving and illustrated such books as *The Pilgrim's Progress,* Rudyard Kipling's *Earth Fiend* and *Don Quixote.* In later life he gave more time to painting and his work, both as print maker and painter, is to be found in the Tate gallery, the British museum and elsewhere. Strang died at Bournemouth on April 12, 1921. (H. Es.)

STRANRAER, a royal and small burgh and seaport of Wigtownshire, Scotland. Pop. (1951) 8,618. It is situated at the head of Loch Ryan, an arm of the North channel (Irish sea), 52 mi. S.S.W. of Ayr by road. It lies 39 mi. E. by N. of Larne in County Antrim, Northern Ireland, with which there is regular communication by mail steamer. Grain milling and nursery gardening are carried on, and there is a large trade in farm and dairy produce. Whitefish and herrings are caught and there are oyster beds at Cairn Ryan on the east side of the loch. Stranraer, originally called St. John's Chapel, became a burgh of barony in 1596 and a royal burgh in 1617. In the centre of the town are the ruins of the castle of the 15th century, occupied for a time by John Graham of Claverhouse, Viscount Dundee, when he held the office of sheriff of Galloway (1682).

Three miles east of Stranraer is Lochinch, the residence of the earl of Stair, a modern structure (1870) in the Scots baronial style. The grounds include the White and Black lochs, and the ruins of Castle Kennedy on the isthmus between. This castle was erected in the reign of James VI for the earls of Cassilis, passed into the hands of the Stair family in the 17th century and was burned down in 1716. The gardens, which contain the finest pinetum in Scotland, were restored in the middle of the 19th century.

STRAPWORK, in architecture and the decorative arts, a form of ornament developed in Germany, Flanders and England during the latter half of the 16th and the 17th centuries and consisting of scrolls, straight lines, rectangles and shield forms carved or modeled in flat relief, often with a raised fillet at the edge, and often pierced with circular or oval holes; the whole composition is usually formed of connected units, all upon the same plane, as though made by an elaborately cut out and pierced strap applied to a background. It is a development of certain Italian early baroque motives, such as the multiple shield or cartouche frames, and certain approximations of it occurred in Italian metalwork as early as the silverwork of Benvenuto Cellini (1500–71); these forms are, themselves, Italian interpretations of the flat scrolls so common in Islamic metalwork. Strapwork was used extensively in the Renaissance woodwork of Germany and is occasionally found in stone, as in the Salzhaus at Frankfurt am Main (end of the 16th century). In Flanders and the Netherlands strapwork received an even more complete development, for although in Germany it was usually only a subsidiary motive, in the architectural woodwork and furniture of the Low Countries it often became the only type of ornament used. The form was introduced into England by the flood of Flemish and German woodworkers that came in in the latter part of the 16th and early 17th centuries. (T. F. H.)

STRASBOURG (Strassburg), a town of France, the capital of the *département* of Bas-Rhin, at the junction of the Ill and the Breusch, 2 mi. W. of the Rhine, 88 mi. by rail N. from Basle. Pop. (1954) 193,353. The Ill divides into arms forming an island on which the city grew; it was long a strongly fortified place, and the removal of the fortifications immediately around the city gave opportunities of large expansion after about 1875. The old city contains the cathedral, or Münster (11th to 15th centuries). Part of the crypt dates from 1015; the apse is Transitional; and the nave, finished in 1275, is pure Gothic. The elaborate west façade has a screen of double tracery, numerous sculptures and two towers, the northern with a tall spire. The cathedral has some fine stained glass, a sculptured pulpit and the famous astronomical clock in the south transept; this contains some fragments of the clock built by the mathematician, Conrad Dasypodius, in 1574. The Protestant church of St. Thomas is a Gothic building of the 13th and 14th centuries.

The Palais de Rohan or old episcopal palace, built in 1731–41, was used for university purposes from 1872 to 1895; it is now the municipal museum of art. Other notable buildings are the Frauenhaus, with some interesting sculptures, and the Hôtel du Commerce, the finest Renaissance building in the town. The University of Strasbourg, founded in 1566 and suppressed during the French Revolution as a stronghold of German sentiment, was reopened in 1872 and again as a French institution in 1922.

The chief industries of Strasbourg are tanning, brewing, printing and the manufacture of metal goods, paper and tobacco. To these must be added the fattening of geese for Strasbourg's celebrated *pâté de foie gras.*

History.—Strasbourg has always been a place of great strategical importance, and as such has been strongly fortified. The pentagonal citadel constructed by Vauban in 1682–84 was destroyed during the siege of 1870. The site of the town was originally occupied as a Celtic settlement, which was captured by the Romans, who replaced it by the fortified station of *Argentoratum,* afterwards the headquarters of the eighth legion. In 357 the emperor Julian gained here a decisive victory over the Alamanni, who 50 years later re-conquered the whole of the district. Towards the end of the 5th century the town passed to the Franks, who gave it its present name. The famous "Strassburg oaths" between Charles the Bald and Louis the German were taken there in 842, and in 923, through the homage paid by the duke of Lorraine to the German king Henry I, began the connection of the town with the German kingdom which lasted for over seven centuries. The early history of Strasbourg consists mainly of struggles between the bishop and the citizens. This conflict was finally decided in favour of the citizens by the battle of Oberhausbergen in 1262, and the position of a free imperial city which had been conferred upon Strasbourg by the German king, Philip of Swabia, was not

again disputed. In 1332 there was an internal revolution, which admitted the guilds to a share in the government of the city.

In 1381 the city joined the Städtebund, or league of Swabian towns, and a century later it helped the Swiss confederates at Granson and Nancy. The reformed doctrines were readily accepted in Strasbourg about 1523, and the city was skillfully piloted through the ensuing period of religious dissensions by Jacob Sturm von Sturmeck, who secured for it very favourable terms at the end of the war of the league of Schmalkalden. In the Thirty Years' War the town observed a prudent neutrality. In 1681, during a time of peace, it was suddenly seized by Louis XIV, and this unjustifiable action received formal recognition at the peace of Ryswick in 1697. At the French Revolution the city was deprived of its privileges as a free town. In the war of 1870–71 Strasbourg, with its garrison of 17,000 men, surrendered to the Germans on Sept. 28, 1871. The city and the cathedral suffered considerably from the bombardment.

The bishopric of Strasbourg existed in the days of the Merovingian kings, being probably founded in the 4th century, and embraced a large territory on both banks of the Rhine, which was afterward diminished by the creation of the bishoprics of Spires and Basel. The bishopric was in the archdiocese of Mainz and the bishop was a prince of the empire. The episcopal lands were annexed by France in 1789 and the subsequent Roman Catholic bishops of Strasbourg discharged spiritual duties only.

Strasbourg is the seat of a bishop, of a prefect and of the departmental organization, and of the special government offices for the territories returned to France after 1918. (X.)

Battle of Strasbourg (A.D. 357).—A rising of the Sarmatians, Suevi and Quadi on the Danube compelled the Emperor Constantius to hasten there in 357. Julian, the Caesar, was at Sens with an army 13,000 strong, and as the emperor was unable to assist him he dispatched Barbatio, *magister peditum* and 25,000 troops to Gaul. Julian was to march north, and Barbatio to co-operate with him with Basel as his base of operations. The forces of the barbarians were to be enclosed between them.

Moving to Reims, Julian concentrated his forces. Thence he marched to Zabern. Meanwhile Barbatio, who had refused to obey his orders, was surprised by the Germans and driven back to Basel. Julian's position was now an extremely dangerous one; he was far distant from the emperor, and Barbatio's defeat had raised the morale of the Germans who, under the leadership of Chnodomar, crossed the Rhine and encamped on its left bank near Strasbourg. Undaunted by this mishap Julian set out from Zabern to meet the horde, which was 35,000 strong. The Romans marched in close order and in two columns, the cavalry on the right and the infantry on the left, and as evening was approaching Julian was desirous of deferring the battle until the following day. His men, in spite of the heat and the long march, urged him to lead them on. This Julian agreed to, and took command of the right wing consisting of heavy cavalry and archers. This wing coming into contact with the barbarians was fiercely charged by light horse and driven back. A panic was only stayed by the Caesar riding among his men and exhorting them to rally. The brunt of the attack was borne by the Roman centre and the left, and it was a struggle of footmen against footmen. In spite of the valour of the barbarians, the discipline of the Roman infantry carried the day, the Germans being driven into the Rhine where large numbers perished; 6,000 dead were also left on the field. Chnodomar was captured and sent as a prisoner to Constantius.

This victory resulted in the recovery of the upper Rhine, and the freeing of Gaul from barbarian invasions.

See E. Gibbon, *The Decline and Fall of the Roman Empire*, chap. xix; *The Cambridge Medieval History*, vol. i. (J. F. C. F.)

STRASBURGER, EDUARD (1844–1912), German botanist, one of the first to realize the importance of the nucleus and chromosomes in heredity, was born in Warsaw, then in Russian Poland, on Feb. 1, 1844. He studied in Germany at Bonn and at Jena. He returned briefly to the university at Warsaw in 1868; he was professor of botany at Jena from 1869 to 1880 and thereafter at Bonn till his death.

Strasburger contributed to the advance of many branches of botany. His first work was concerned with the embryology of gymnosperms, then with the more minute details of the life history of angiosperms. Later he turned increasingly to the study of the cell, and made many contributions to the solution of problems of cell and nuclear division in plants. He also studied problems of water transport in plants, and in his last years the inheritance of sex differences in plants.

Strasburger was also the foremost botanical teacher of his day. He wrote a number of textbooks, among which the "Bonner Lehrbuch," *Lehrbuch der Botanik*, first published in 1894 in collaboration with others, established itself as the standard textbook for botanical students the world over and was translated into English, Italian, Russian and Japanese. He died at Bonn on May 18, 1912. (G. H. Be.)

STRATEGUS, a Greek word meaning strictly a general, but frequently in ancient Greece a state officer with much wider functions.

In Athens the board of ten strategi was originally a product of the reorganization of the tribal system under Cleisthenes (*q.v.*), each of the ten tribal units being represented in the army by a *taxis* (regiment) led by a strategus; but there is evidence (Aristotle, *Constitution of Athens*, xxii) against the view that Cleisthenes himself created the board. In any case, in 490 B.C., at Marathon, the head of the army was still the polemarch (*see* Archon), Herodotus' story of successive days of command being based on a misconception of the powers of strategi. After 487, however, when the archons came to be chosen by lot, command of the *taxeis* passed to taxiarchs subordinate to the strategi. In the 5th century, it was usual to appoint a number of strategi for a particular field of operations: while the chief command might be assigned to one of them, they often held co-ordinate powers (*e.g.*, in the Sicilian expedition) or exercised the chief command in turn (*e.g.*, at Arginusae). In crises, however, the whole board could be subordinated to one member (*e.g.*, at the outbreak of the Peloponnesian War, or in the circumstances preceding the recall of Alcibiades). At first each tribe elected a strategus from among its members (Aristotle, *op. cit.*, lxi); but in the second half of the 5th century there are examples of two strategi being elected from a single tribe (*e.g.*, Pericles and Glaucon in 433–432; Alcibiades and Adeimantus in 407–406); and in the later part of the 4th century the strategi were elected from the whole citizen body without regard to tribal divisions. The tribes, however, continued to elect the taxiarchs.

The political powers of the strategi were considerably increased in the 5th century, particularly in foreign affairs. Though not members of the boule, they could attend its meetings, bring motions before it and ask it for emergency meetings of the ecclesia (Thucydides, ii, 59), in which their business regularly took precedence of everything else. In concluding treaties, they acted on the voted resolutions of the ecclesia. The appointment of strategi for specified departments dates from the 4th century.

Two aspects of the office made it important: (1) it was elective, as opposed to the magistracies filled by lot; and (2) strategi were re-eligible indefinitely. A man who commanded popular confidence (*e.g.*, Pericles) might thus not only obtain such commissions as could not be left to chance nominees but also be returned to power in subsequent years to profit from his experience. The board of strategi had no collective responsibility.

Military power prevailed over civil authority in the Greek world after Alexander the Great, and strategi became the supreme magistrates in most leagues and federations. Both in the Achaean and in the Aetolian league the strategus (helped by a commander of the cavalry or *hipparchus*) was the head of the state, being appointed for a year by the appropriate federal assembly: he was at once commander of the army, president of the federal assemblies and foreign minister. The same man could not however be re-elected in consecutive years; and in the Aetolian league the strategus was not allowed to make proposals about war or peace. Alexander the Great appointed Antipater as strategus of Europe when he went to war against Persia. Polysperchon and Cassander filled the same position after Alexander's death; and Antigonus was sometime strategus of Asia. The governors appointed by the

Macedonian kings over regions or cities of Greece were normally called strategi. In the Seleucid kingdom the provinces were governed at least in part by strategi, but details are controversial. There were provincial governors called strategi in other Hellenistic monarchies, and strategi appointed by the king are found in Pergamum. In Egypt the strategi became the civil governors of the nomes during the 3rd century B.C. (they almost completely lost their original military character); a special strategus governed the Thebaid; and there were strategi in the Ptolemaic territories outside Egypt. Under Roman rule strategi remained the heads of the Egyptian nomes until the 4th century A.D. There were officers with powers equivalent to those of strategi (and called strategi by Greek authors) in the Carthaginian state.

BIBLIOGRAPHY.—A. Hauvette-Besnault, *Les Stratèges athéniens* (Paris, 1885); A. H. J. Greenidge, *Handbook of Greek Constitutional History* (London, 1896); J. Sundwall, *Epigraphische Beiträge zur sozialpolitischen Geschichte Athens* (Leipzig, 1906); C. Hignett, *History of the Athenian Constitution* (Oxford, 1952; New York, 1953). For the Greek leagues, *see* G. Busolt, ed. by H. Swoboda, *Griechische Staatskunde* (Munich, 1920–26); and A. Aymard, *Les Assemblées de la confédération achaïenne* (Bordeaux, 1938). For the Hellenistic monarchies, *see* H. Bengtson, *Die Strategie in der hellenistischen Zeit*, 3 vol. (Munich, 1937–52). (A. D. Mo.)

STRATEGY. Strategy is that part of the art of modern warfare and statecraft that deals with the larger aspects of waging war. Narrowly defined, the term means "the art of the general" (from the Greek *strategus*). In this strictly military sense, it first gained currency at the end of the 18th century, when warfare was still relatively simple and limited. In its military phases, the term had to do with stratagems by which a general sought to deceive an enemy, with plans he made for a campaign and with the way he moved and disposed his forces in war. Often defined as the art of projecting and directing campaigns, military strategy came to pre-empt almost the whole field of generalship, short of the battlefield itself. To tactics (*q.v.*), a close partner of strategy, military jargon reserved the art of executing plans and handling troops in battle.

After a consideration of the term "strategy," this article is divided into the following sections and divisions:

I. Fundamentals of Strategy
  1. The Search for Principles
  2. Relation between Strategy and Tactics
  3. Strategic Leadership and War Planning
II. Development of Strategy
  1. 18th-Century Warfare; Frederick the Great
  2. Napoleonic Warfare; Clausewitz and Jomini
  3. American Civil War to World War I
  4. World War I
  5. Between World Wars I and II
  6. World War II
  7. Cold War in the Atomic Age

The term strategy has expanded far beyond its original military meaning. As society and warfare have steadily grown more complex, particularly in western civilization, military factors have become more and more inseparable from the nonmilitary in the conduct of war. Nations have found it necessary to adjust and correlate political, economic, technological and psychological factors, along with military elements, in the management of their national policy. The demarcation between strategy as a purely military phenomenon and national strategy of the broader variety became blurred in the 19th century, particularly in wartime. The distinction became even less clear in the 20th century when nations became more interdependent and the line between war and peace less clearly definable. As a result, the appearance of the term "grand strategy" (or "higher strategy"), meaning the art of employing all the resources of a nation or coalition of nations to achieve the objects of war (and peace), steadily became more popular in the literature of warfare and statecraft of the 20th century.

This broadened scope of strategy has tended to blur distinctions customarily drawn by earlier writers between strategy and statesmanship, and between garden varieties and higher or "grand" forms of strategy. Though there was still no agreed definition of the precise meaning of the term strategy by the middle of the 20th century, few students of the subject any longer accepted the earlier narrow definition. And few contested that strategy, whether in its narrow or broad sense, would, by the very nature of its shifting bases, continue to be a changing art.

I. FUNDAMENTALS OF STRATEGY

1. The Search for Principles.—The starting point of all strategic planning and action is national policy. Once the national aims are set forth by the leaders of the state, the commander sets about drawing up his plans. He must take many matters into account; for example, factors of space and time, the state of his own forces, the enemy's capabilities and intentions, and reactions at home and abroad to his projected moves. As Karl von Clausewitz, the great student of strategy, put it, "In strategy everything is very simple, but not on that account very easy." (Karl von Clausewitz, *On War*, Routledge & Kegan Paul, Ltd., London, 1940.) The strategist deals in many uncertainties and imponderables. Indeed, the art of the strategist is the art of the "calculated risk."

The growing complexity of modern warfare has led some students to take a fresh look at the principles that have traditionally guided military strategists in war. It has long been a favourite occupation of military theorists to seek to distill from the great mass of military experience simple but all-pervasive truths—lists of principles—to guide commanders through the "fog of war." Usually they have derived such principles from a study of campaigns of the great captains of history; occasionally outstanding practitioners have set them down on the basis of personal experience. As far back as 500 B.C. Sun-tzu, a Chinese general, set forth 13 principles. The axioms range from Gen. N. B. Forrest's simple admonition about getting there first with the most men to Napoleon's 115 maxims. The stress varies from list to list. It ranges from the emphasis of the followers and interpreters of Clausewitz on the belief that the battle is all, and defeat of the enemy's armed forces the correct objective and path to victory, to that of exponents of "the strategy of indirect approach," of victory by indirect methods, such as B. H. Liddell Hart, eloquent advocate of the principle of "obliquity" in the 20th century.

In number, the stated principles have varied from writer to writer, from period to period, and from nation to nation. In 1956 the U.S. list contained nine principles; the British and the Russian, ten each. Though there is no complete agreement on the number of principles, most lists include the following: the objective, the offensive, co-operation (unity of command), mass (concentration), economy of force, maneuver, surprise, security and simplicity. The British have added one called "administration"; the Russians another, translated as "annihilation." Despite debate over their precise number and meaning, the principles of war are taught in military schools throughout the world, and most military students continue to accept them as basic concepts.

The individual authors of the lists have almost uniformly claimed the principles to be immutable. They have argued that success in military strategy in the past has been the result of adhering to them; that the advantages of the offensive, the concentration of force, the effort to achieve surprise, the proper movement of forces and their security from attack, sabotage or subversion are in the province of modern as well as ancient warfare. Some authorities have even argued that since war is not the concern of soldiers only, the "principles" deserve a wider application throughout government—in grand as well as military strategy.

Other authorities have argued that the claim of immutability cannot be accepted literally; that there is little agreement as to what the principles are and mean; that they overlap; that they are fluid and require constant re-examination; that they are not comparable with scientific laws since no two military situations are ever completely alike; that the so-called principles are not really principles at all, but merely methods and common-sense procedures adopted by great commanders of the past; and that changes in the conditions of war alter their relative importance.

The debate over principles has been renewed with the coming of the atomic era. Some theorists argue that the new weapons have destroyed whatever infallibility still remained; others contend that the principles are as valid as ever, even more so. To some

extent this is a debate over semantics. Defenders point out that each age must make its own applications of the "fundamental truths" of strategy. Opponents argue that there can be no set rules for the art; the so-called principles must by no means be interpreted as pat formulas for victory to be followed blindly and rigidly; the only sound guide in war and strategy is flexibility.

2. Relation Between Strategy and Tactics.—In the theory of warfare, strategy and tactics have generally been put into separate categories. The two fields have traditionally been defined in terms of different dimensions: strategy dealing with wide spaces, long periods of time and large movements of forces; tactics with the opposite. Strategy is usually understood to be the prelude to the battlefield, and tactics the action on the battlefield itself. As a result, much of the literature and theory of strategy has in the past been preoccupied with the proper approach to the battlefield, the leading of troops up to the time of contact with the enemy. This situation explains the attention to strategic maneuver—aimed at putting one's army into the most favourable position to engage the enemy and compelling the enemy to engage at a disadvantage and depriving him of freedom of movement. Indeed, early writers on strategy dealt heavily in the so-called "geometrical strategy"—the angles formed by lines of movement and supply of opposing armies. Sir Edward Hamley in his *Operations of War* has well described all the various permutations and combinations of strategic maneuver.

Despite distinctions in theory, strategy and tactics cannot always be separated in practice. In fact, the language of strategic maneuver (for example, "envelopment," "penetration," "encirclement") is also largely the language of tactics (*see* TACTICS). Movement begets action, and action results in new movement. The one merges into the other. Strategy gives tactics its mission and wherewithal and seeks to reap the results. But tactics has also become an important conditioning factor of strategy, and as it changes, so does strategy. Battles and fronts are no longer necessarily restricted in space and time, and the distinction between battles and campaigns is no longer so clear-cut, as the experience of World Wars I and II with tridimensional warfare demonstrated. Indeed, in World War II, theatre commanders were as much concerned with the actual fighting of armed forces in battle as they were with larger strategic decisions—with relations to allies, economic problems and political questions on the ground. While in theory strategy continues to occupy a middle ground between national policy and tactics, in practice the line dividing it from the other two fields has become increasingly difficult to draw.

3. Strategic Leadership and War Planning.—Count von Schlieffen, the famous German military leader, once said, "A man is born, and not made, a strategist." (Count Alfred von Schlieffen, *Cannae*, Berlegtbei, Mittler & Sohn, Berlin, 1925.) But it is obvious that even a born strategist—if there be such a natural genius—has much to learn. In the old days strategic leadership was a relatively simple affair. As J. F. C. Fuller, the well-known British student of warfare, has pointed out in *The Foundations of the Science of War*, until relatively recent times the death, capture or wounding of either of two opposing generals normally decided a conflict, "for the general *was* the plan." (J. F. C. Fuller, *The Foundations of the Science of War*, p. 233, Hutchinson, London, 1926.) He could personally devise the plans and direct his troops. By mid-20th century this was rarely possible. As warfare has become complicated, strategic leadership has become more difficult. The art has taken on many more facets and systematic training is required to master them. The strategist has retired from the scene of battle, and large, specialized staffs have grown up to help him. While the responsibility for strategy remains the general's, many of his functions have been delegated to his planning staff. In modern states corporate leadership has become the rule in the management of military strategy, as in the direction of large business enterprises.

The example of an Alexander completing his advance planning and leaping into battle at the head of his troops, as at Arbela, would in modern warfare be considered most unusual. Napoleon was wont to make his plans and then retire with his retinue of trusted advisers to survey the battlefield on horseback from the top of a hill. Generals in World War I were often pictured in their offices in large headquarters—usually in a chateau behind the lines, studying a map on the desk and dispatching orders via the telephone and motor car at hand. In World War II the headquarters staffs of commanders in the theatres of war grew even larger and more elaborate. Tridimensional warfare had enlarged the field of operations far beyond individual battlefields and usually a high commander reached his decisions in a headquarters far removed from the field of battle and months before the battle itself took place. Far from striking the classic pose of the officer on a well-schooled charger, some of the greatest generals issued their orders from desks and fought their most important battles at conference tables. As strategic planning became a highly organized affair, planning committees and conferences in the capital cities of the warring powers made the blueprints for victory in the global, coalition struggle. In their capital command posts military leaders kept in touch with the manifold phases of the national government's war effort and dealt with the world-wide problems transcending those of the individual theatres of war. With the aid of new devices for rapid communication these leaders and their staffs sought to set the patterns of strategy and keep abreast of the movement of armies as the Caesars and Napoleons had done in earlier eras.

As war has become more total, war planning has become an increasingly significant peacetime function of governments. Nations have recognized the vital importance of planning for war long in advance. The manufacture of strategic plans has become a highly specialized industry in modern military establishments. At the same time, more and more governmental agencies have been drawn into the business of planning for national security. The plans they produce may vary from a simple design to shift a small task force to a danger spot to an elaborate plan for the conduct of war in its entirety. To be realistic, strategic plans and estimates must constantly be re-examined and brought into harmony.

Against this general background in the nature of the art, it is now possible to sketch the important contributions made in key periods to modern strategic theory and practice. It is important to remember that the art of strategy has changed from age to age, just as has war itself, and that each is the product of its own society and time.

II. DEVELOPMENT OF STRATEGY

Though the serious and systematic study of modern strategy may be dated from the 18th century, various authorities have identified strategic precedents going back to earliest times. Students of warfare of primitive ages have associated with primitive tribes and clans a stratagem of surprise from darkness or by ambush, and have identified a strategy of hunt and pounce, like that of a lion or tiger. The Bible points to the care with which Moses prepared his operations—an early form of advance planning. The ancient world developed a strategy of mass attack by phalanx, legion or cavalry. Alexander, Hannibal and Caesar, who combined in their own persons political and military direction of the state, planned their famous campaigns far ahead. They have been singled out as forerunners of the modern art of grand strategy. Writers in modern times have used the campaigns of these great captains to illustrate practically every known "principle of war." But important as their attention to strategic considerations in war and especially to strategic approaches to the battlefield may have been, the foundations of the ancient art of warfare were tactics and battles. To a considerable extent battles—often short and furious—also held the centre of the military stage in the middle ages. Strategy was notably absent in the excursions of the Huns, the Moslems or the Crusaders. Far more important from a strategic viewpoint were the campaigns of Genghiz Khan and his general, Sabutai, in the 13th century. Their advance planning and bold strategic maneuvers in broad sweeps from Mongolia across Asia and Europe showed an appreciation of strategic problems most unusual for their age.

In the transition to modern times two other figures that touched on the field of strategy are often mentioned—Niccolò Machiavelli

in the realm of military thought and Gustavus Adolphus in the field of generalship. Machiavelli's *Art of War* (1520) emphasized the larger aspects of war, particularly the close relationship between the civil and military spheres. A century later, Gustavus Adolphus, ruler-general of Sweden, intervened in the Thirty Years' War, and, maneuvering skillfully, drove the imperial armies out of northern Germany.

1. 18th-Century Warfare; Frederick the Great.—After the death of Gustavus, warfare again settled down to a slower pace and a more stable mold. The 17th and 18th centuries saw the growth of professional armies loyal to the king. But the great cost of building and maintaining such armies led to a concern for their safety, a hesitation to risk them in bloody encounters and a preoccupation with defense and fortifications. Strategy was essentially of limited aim and was greatly concerned with the art of siegecraft, for which elaborate rules were prescribed. In Prussia of the mid-18th century, however, circumstances compelled Frederick the Great to try a new and aggressive approach and to break through the accepted military pattern of the day.

Confronted at the outset of the Seven Years' War (1756–1763) by a coalition of Austria, France, Russia and Saxony, Frederick found himself virtually surrounded. His task was to devise a strategy to defend his territory and not to dissipate his outnumbered troops. The strategy he evolved did not follow set rules or recipes. Indeed, never was the definition of strategy as a "system of makeshifts," offered in a later age by Count von Moltke, better demonstrated. In his planning Frederick capitalized on two valuable assets—his army, a superior and highly disciplined instrument of war, and a central position. He sought always to keep the initiative, to attack first one enemy and then another, to assemble at decisive points a force superior to that of his foe and to avoid long, drawn-out wars. Using his central position to concentrate against individual armies of the enemy before they could be reinforced by others, he developed the classical "strategy of interior lines." But even Frederick, the statesman-warrior, could not entirely escape the conditions imposed by the warfare of his times. Indeed, the statesman imposed caution on the warrior. He could not expose his costly armies to the risk of destruction and bloody decision by battle. His battles were not those of annihilation. In the end his wars were decided by reasons of state, and in the final analysis those wars left his nation exhausted.

The age that immediately followed Frederick chose to imitate his caution rather than his aim. Military theory was characterized by ideas of victory without battle, maneuvering for position, a system of lines and angles of operation. Geometric concepts and cunning tricks and artifices replaced his aim to destroy his enemies. Great emphasis was put on terrain and the occupation of key geographic points. The 18th century, it must be remembered, was the era of enlightenment, and warfare conformed to the spirit of the age. Strategy, like all warfare, became "mathematical" and "scientific." Theorists optimistically maintained that a general who knew mathematics and topography could direct campaigns with geometrical precision and win wars without even fighting. But the new mode of warfare ushered in by the French Revolution and the Napoleonic era was soon to challenge these optimistic assumptions.

2. Napoleonic Warfare; Clausewitz and Jomini.—The French revolutionary and the Napoleonic periods witnessed great changes in the methods of war—the revolution in society accompanying and reinforcing the one in warfare. When Napoleon, the first great military strategist of the modern western world, burst upon the European scene, the groundwork for a new age in warfare had already been laid. Fortunately for him, important beginnings had already been made during the French Revolution and the many innovations of his predecessors had put at his disposal new methods and means of war. The French Revolution gave birth (1793) to the "nation in arms" and all Frenchmen became liable for military service. The patriotic citizen-soldier succeeded the mercenary professional. Skirmish tactics, or the loose formation, replaced the straight line; divisional organization came into use, along with lightweight artillery of great range and firing power.

When Napoleon came to reap the benefits of these changes, he completely transformed strategy as well as tactics. He applied the same basic principle to the one as to the other—never to divide his forces, but to concentrate all his might against the enemy forces at the critical point. His emphasis was on careful preparation, on uniting his forces before the action, on overpowering weight of striking power, on shock attack, on great daring and on bloody decision by battle. His methods were simple, direct, overpowering—even brutal; his aim was nothing short of the destruction of the enemy forces. Against such power, neat geometric calculations stood little chance and ordinary stratagems were helpless. Again and again he showed his military genius for bringing a mass to bear against the flanks of his enemy, for selection of battlegrounds advantageous to his forces and for deploying his forces for battle. He gave supreme expression to the idea of victory by battle.

Though as a military leader, operational strategist and tactician of the battlefield he is regarded by many as unparalleled, in the larger field of national or grand strategy he showed shortcomings. Embodying in his own person the leadership of the state and its military affairs, he recognized the value of incorporating political and economic measures, along with military moves, to increase the chances of victory in war. But, in the final analysis, he could not successfully grasp and cope with the challenge put to him by Great Britain and its European allies. British strategy sought to meet the Napoleonic threat to Europe by using naval power to blockade the continent and by conducting a war of exhaustion on land through peripheral warfare, such as Wellington's famous campaign on the Iberian peninsula. Napoleon's reply to the British naval blockade was the continental system prohibiting British goods from entering. But this helped bring his downfall, since he was needed everywhere—to hold the coast, fight in Spain, Holland, against Austria and Russia. His veteran French forces were dissipated and he had to rely on impressed nationalities of Europe. Eventually the coalition of his enemies was to use the methods and means of warfare that the French Revolution had introduced and Napoleon had perfected to reinvigorate their own forces and overwhelm him.

Despite his mistakes, Napoleon's pre-eminent place in the history of strategy is secure. The direct methods of this son of the French Revolution uprooted 18th-century strategy tree and branch. Napoleon's tactics and strategy influenced military leaders for a century after him. His maxims were widely studied and were said to have been carried in the saddlebag of the famous U.S. general, "Stonewall" Jackson. Students of strategy have long pointed to Napoleon's battles and campaigns for classical illustrations of "principles" of war—of surprise, mobility, concentration of force and economy of force. Possibly more than that of any other general, his competent practice oriented modern military theory toward the search for underlying principles. Indeed, the art of strategy as evolved by theorists since 1800 may be traced largely to his operations.

For this development two great interpreters of Napoleonic strategy, Antoine Jomini and Clausewitz (*qq.v.*), were especially responsible. Clausewitz, a Prussian, the first great student of strategy and the father of modern strategical study, lived in the early 19th century. Trained in systematic study of philosophy in the school of Immanuel Kant, it was but natural for him to range widely over the whole field of military knowledge and to reduce Napoleonic warfare to a unified, philosophical conception. His famous work, *On War,* written as an outgrowth of his studies of Napoleon's campaigns, remains the best general study of the art of war. He died with his work unfinished, but his writings published after his death became the standard textbooks on war in Prussia and elsewhere. Their influence was felt profoundly in the Franco-Prussian War of 1870, and leading generals of World Wars I and II were brought up on them and on the works of his followers.

The contributions of Clausewitz to strategic thought are many and diverse. To some his work is the Bible of strategy, and, like that great book, susceptible to many conflicting interpretations. His work set forth fully and clearly for the first time the relation-

ship between political and military leadership. He dwelt on decision by battle as the first rule of war; on seeking the destruction of the enemy's forces; and on achieving superiority at the decisive spot. Rejecting the optimism and rationalism of the 18th century, he held that war was not a scientific game but an act of violence. Mathematical and topographical factors, he held, were important in tactics, but less so in strategy. "We . . . do not hesitate," he asserted, "to regard as an established truth that in strategy more depends on the number and the magnitude of the victorious combats than on the form of the great lines by which they are connected." (Karl von Clausewitz, *On War*, vol. 1, bk. 3, ch. xv, p. 223, Routledge & Kegan Paul, Ltd., London, 1940.) The key to victory was battle, however bloody. Strategy he defined as the employment of battles to gain the end of war.

Clausewitz devoted much of his work to showing that war is both a social development and a political act. He went further and said that, "War is not merely a political act, but also a real political instrument, a continuation of policy carried out by other means." (Maj. Gen. Sir F. Maurice, *British Strategy*, p. 44, Constable & Co., Ltd., London, 1929.) War was therefore not an independent phenomenon unto itself to be handed over to soldiers and sailors. Again and again he asserted that military and political strategy must go hand in hand. "War," he declared, "admittedly has its own grammar, but not its own logic." (Karl von Clausewitz, *On War*, Routledge & Kegan Paul, Ltd., London, 1940.)

Clausewitz' emphasis on the aim of strategy as the destruction of the enemy's forces on the battlefield has had a great influence on subsequent military thinking. His disciples, however, have generally overlooked the fact that he also recognized another strategical form—a strategy of limited aim for limited warfare, of wearing down an opponent. When Clausewitz wrote, warfare was conducted in two dimensions and it was rarely possible for one nation to impose its will without first destroying the opposing army. But Clausewitz recognized clearly what many of his followers in subsequent generations forgot, that the destruction was only a means to enforce policy and not an end in itself.

In the history of military thought, the French general Jomini, one of Napoleon's staff officers and a contemporary of Clausewitz, presents a striking contrast to the Prussian philosopher of war. Lacking the philosophical bent of Clausewitz, Jomini concentrated his thinking on what he regarded as practical issues in war rather than on war as a whole. He became the chief expounder of Napoleonic methods, and out of his studies evolved a theory of strategy. Although he opposed "systems of war" purporting to provide for all contingencies, he nevertheless believed that in the field of strategy certain rules and general principles—eternally true—could and should be formulated "as a compass for the commander-in-chief of an army." To establish these principles, he believed, was the major problem of military science.

The heart of Jomini's theory lay in the theatre of war and the campaign. But he thought primarily of occupying all or part of the enemy's territory rather than of annihilating his army. This occupation was to be achieved by progressive domination of zones of territory. Jomini emphasized throughout his work the proper choice by the general of decisive maneuvering lines and their adaptation to geometrical configurations of zones of operation. Campaigns must be carefully planned in advance. The task of strategy is to make preliminary plans—to establish lines of operation and to bring military means into conformity with geographic realities of the chosen zone of operations. He laid down two basic principles: massing troops against fractions of the enemy by rapid movement and striking in the most decisive direction.

Jomini's great contribution to military thought lay in his definition of the place of strategy in warfare. Probably more than any other work, his *Précis* fixed the major fields of modern military art. Subsequent wars were to cast doubt on much of his work, particularly on his conception of geographical campaigns and of the superiority of interior lines of operation. But, like Clausewitz in German strategic thinking, Jomini had an enduring influence on French military thought. His emphasis on planning for operations and on intelligence took root in military staffs and schools throughout Europe, and his work became the textbook for the conduct of the American Civil War. It has been said that the Civil War was fought with Jomini in the pockets of all the leading generals.

3. American Civil War to World War I.—Often called the first of the really modern wars, the American Civil War marked a transition to a new era in strategy. It gathered up new phenomena that had begun to influence warfare in the middle of the 19th century and whose fuller consequences were to be felt in the half century that followed. It was a period marked by refinement of the old in strategic theory and practice and by the addition of new strands—by such famous figures as Lee, Grant and Sherman on the battlefield; Moltke, Schlieffen, Delbrück and Mahan in the literature of strategy.

The Civil War is significant in the history of strategy in a number of ways. The basis of strategy—particularly factors of time and space—began to change. The use of steam power for land and water military transportation received its first major test. Railroads gave strategy a new speed of movement but tended to make strategy stick to straight lines and fixed routes. The Civil War also tested ironclad ships and heavy naval ordnance. The relation among the combat arms was completely upset by the introduction of the long-range infantry rifle. Artillery could no longer unlimber at case-shot ranges. The accuracy of long-range weapons in the hands of defending infantry shattered the effectiveness of the rapidly concentrated attack in which Napoleonic strategy had culminated. But, as so often has been noted in the history of warfare, armaments and weapons are more readily changed than ideas, and Napoleon's principles continued to be maintained, sometimes with disastrous consequences on the battlefield.

Aside from the effect of new inventions, the Civil War revealed the growing importance of the economy and manpower in war. Industry was called on more and more, and conscription was adopted to provide manpower. The war also revealed the impact of systematic West Point training received by the leading generals on both sides. In fact, it may be said that if the higher officers carried Jomini in their pockets, they carried in their minds the teachings of Dennis Hart Mahan, father of Alfred Thayer Mahan and one of the few outstanding U.S. students of strategy. Finally, the Civil War was long studied for classic examples of maneuver, and of offensive and defensive strategy, and for lessons in the relation among policy, strategy and means of war. Essentially the Civil War demonstrated local or theatre strategy and tactics. Though elements of grand strategy were at hand—political, economic, military and psychological—the art was still not well understood or consistently applied. Despite conscription and the partial mobilization of industry and the railroads, there was no well-worked-out grand design correlating the widely scattered forces and the war industries that supported them.

The strategies of the North and South were rooted in different political objectives. The objective of the North was to prevent the Confederate states from seceding from the Union; that of the South was to attain independence. Since the South was greatly inferior to the North in population and resources it could not hope to conquer the North. The dual purposes of its strategy were: (1) to convince the North that forcing the South to remain in the Union was not worth the cost; and (2) to bring about foreign intervention in favour of the South. Gen. Robert E. Lee, the great Southern leader, believed the best way to realize these objectives was to carry the war into the North and to defeat the Northern armies in their own territory. For a time, therefore, his strategy was essentially offensive. But after his defeat at Gettysburg he no longer had the wherewithal to continue the offensive, and at the same time it became obvious that foreign intervention would not be forthcoming. From that time to the end of the war his strategy was defensive, with the object of wearing down the patience, if not the power, of the North.

To achieve its political object, the North, on the other hand, developed another strategy. The federal design had three main goals: (1) to isolate the Confederacy; (2) to cut it in two; and (3) to strike at Richmond, its capital. The blockade, though not

completely effective, brought virtual commercial isolation to the South. Partition was gained by capturing Vicksburg, in July 1863, and by severing the east-west railroad connections. The capture of Vicksburg cut the South off from its sources of supplies beyond the Mississippi. Only gradually did the North change its design from that of attacking Richmond to that of striking at the main army of the Confederacy and the remaining sources of supply. Grant's elevation to the supreme command of the armies enabled him to put this concept into effect. The famous march of Gen. William T. Sherman through Georgia to the sea in 1864 was an outstanding example of strategic maneuver and surprise. Leaving his supply line, Sherman feinted against one city and attacked another, finally cutting off Lee's army in Virginia from its war resources in the South. The co-operation of the Federal eastern and western armies in a grand converging movement resulted in the evacuation of Richmond and, finally, in the surrender of Lee's army to Grant at Appomattox in April 1865.

The period from the close of the Civil War to the outbreak of World War I saw the further growth of trends already apparent. Space and time factors began to appear in a new light. A nation with a well-developed railway net gained significant advantages in war. The speed of mobilizing and concentrating armies became a basic element in strategic calculations and the timetable based on it the heart of staff plans drawn up in anticipation of war. Increased fire power in the machine gun, universal liability of able-bodied males for military service, rapid mobilization of reserve military units and increased potential of fortifications influenced military planning.

In strategic thinking the half century before World War I showed a remarkable diversity. To the Prussian-German school—Moltke and Schlieffen—the new trends in warfare seemed to reinforce Clausewitz' teachings about battles and the aim of defeating the enemy's armies. Battles, Moltke agreed, were the primary means of breaking the will of the enemy. But Moltke did not believe a strategist could follow a rigid set of rules. To him strategy was a system of "*ad hoc* expedients." It was "the art of action under the pressure of the most difficult conditions." (From Edward M. Earle *et al., Makers of Modern Strategy*, p. 180, Princeton University Press, Princeton, 1948.) No plan of operations, he believed, could look with any assurance beyond the first encounter with the main enemy forces. The offensive, according to Moltke, is "the straight way to the goal," whereas the defensive is "the long way around." (From Hermann Foertsch, *The Art of Modern Warfare*, Veritas Press, New York, 1940. Reprinted by permission of the publishers.) He became famous for his skillful conduct of operations on the outer line leading to encirclement, and Königgratz and Sedan (*see* SEDAN; *Battle of Sedan, 1870*) remain as supreme models of such strategy. In addition to exploiting the altered conditions of space and time created by the railroads and improved highways, he capitalized on the possibilities offered by the telegraph for handling armies of great size. Recognizing that the field of operations had become too vast to be surveyed by the eye of the commander, he introduced a new system of delegating power to subordinate commanders. Broad directives took the place of detailed orders. Moltke always fought with superior forces, and his wars, culminating in that against France in 1870–71, are regarded by some authorities as classical models of conception and execution in military strategy.

Field-Marshal General Count von Schlieffen, chief of the Prussian general staff before World War I and ablest of Moltke's successors, carried the strong line of strategic reasoning running from Napoleon through Clausewitz and Moltke to its logical conclusion in his conception of a "strategy of annihilation." Like Moltke, he stressed the military side of strategy, the concentration on decisive victory by battle. But, unlike Moltke, he could not count on superior forces and had to prepare for war on two fronts. The basis for German strategy before World War I as developed by him became that of the inner line—a conception that was embodied in the famous Schlieffen plan. The plan was extremely simple. The bulk of the German forces were to attack the nearest opponent, the one in the west, and to defeat him in a great battle; meanwhile, in the east the Germans would stand on the defensive. Schlieffen proposed to gain the decision in the great battle by means of an enveloping attack—if possible, by a double envelopment. Once the enemy in the west was defeated, the Germans would attack the foe in the east. This was the essence of the plan with which the Germans entered World War I.

Schlieffen's theories were to have wide influence, largely through his book, *Cannae*. Analyzing Hannibal's great victory over the Romans, he had developed his theory of the battle of annihilation by means of encirclement and double envelopment. The decisive German campaign at Tannenberg in Aug. 1914 was fought in this mold and Schlieffen's theories were studied exhaustively in the higher army schools of the United States and Europe after World War I. As Gen. Walter Bedell Smith, chief of staff to Gen. Dwight D. Eisenhower, supreme commander of the Allied expeditionary forces in World War II, has pointed out, General Eisenhower and many of his staff officers, products of these schools, "were imbued with the idea of this type of wide, bold maneuver for decisive results." (Walter Bedell Smith, *Eisenhower's Six Great Decisions: Europe 1944–1945*, p. 156, Longmans, Green & Co., Inc., New York, 1956.)

Moltke and Schlieffen thought of war as military action—the speediest decisive defeat of the main opponent. But the closing years of the 19th and early years of the 20th centuries witnessed the emergence of new approaches and different emphases in strategy. Two thinkers looking to past history for light on the problems of their times made signal contributions to strategic theory—one, Alfred Thayer Mahan (*q.v.*), an American, in the field of naval strategy; the other, Hans Delbrück, a German, in the area of military strategy. Each recognized an intimate relationship between war and politics in every age; that political and military (or naval) strategy must be in harmony. Each showed an awareness of the growing importance of the economic bases of strategy, of state policy, geographic position and available means as determinants of the mode of strategy; of accommodating strategic action to suit the particular times and needs.

Carrying forward a line of thinking already suggested by Clausewitz, Delbrück presented his theory of the "strategy of exhaustion"—of wearing down an opponent by a variety of means. Clausewitz had merely indicated the existence of two methods of conducting war—one aimed at annihilation of the enemy; the other, limited warfare. Delbrück expounded on the differences. The sole aim of the strategy of annihilation he identified as the decisive battle. The second type he called variously the "strategy of exhaustion" and "two-pole strategy." The commander could move between battle and maneuver; the political object of war could be obtained by other means than battle—by occupying territory, blockade, destroying crops or commerce. In Delbrück's view, Alexander, Caesar and Napoleon had been strategists of annihilation; Pericles, Gustavus Adolphus and Frederick the Great, equally great generals, exponents of the strategy of exhaustion. Holding that the strategy of exhaustion was just as valid as the strategy of annihilation—each depending on the political aims and means at hand—Delbrück's theories ran counter to the military thinking of his day and brought down a storm of criticism about his head. But he persisted in reminding his age, intent on victory by battle, of other important and forgotten aspects of Clausewitz' teachings.

While Delbrück was battling his military critics in Germany, a scholarly navy captain-teacher at the Naval War college in Newport, R.I., was quietly breaking ground in pursuing his brilliant researches in military and naval history and strategy. This pioneer was Alfred Thayer Mahan, indefatigable student of the strategy of Napoleon, Jomini and Nelson. His masterly works, *The Influence of Sea Power upon History, 1660–1783*, published in 1890, and *The Influence of Sea Power upon the French Revolution and Empire, 1793–1812*, published in 1892, marked a revolution in naval thought. While advances in technology were affecting naval architecture and weapons, and steam, armour plate and rifled guns were coming into vogue, Mahan aimed to bring naval strategic thinking up-to-date. The books and articles that poured from his pen down to World War I about warfare of the second dimension —the sea—had a profound influence on the theory of warfare and

on naval policy and strategy in the United States and abroad.

An advocate of a big navy, of overseas bases, of national greatness through sea power, he was the American apostle of "looking outward." Mahan emphasized the significance of commerce in war and of economic warfare through the application of sea power. His researches convinced him that the nation or group of nations that commanded the seas could best draw on the trade, wealth and economic resources of the world and was the more likely to win wars. Strongly influenced by Jomini's teachings, he looked for fundamental truths and formulated "principles" of naval strategy. Naval strategy and sea power, he recognized, were conditioned by a nation's insular or continental situation. To Mahan a central position gave the same great advantages on the sea or on land—interior lines. Concentration of force he viewed as a fundamental principle of land and sea warfare. The backbone of fleet strength, in his opinion, was the battle- or capital-ship. Under Mahan's tutelage, command of the sea approaches took the place of the twin theories of coastal defense and commerce raiding that hitherto held sway in American naval strategy. His concepts of naval strategy and faith in preponderant naval power and the use of the navy as an instrument of national power were accepted by the U.S. navy and Pres. Theodore Roosevelt. His doctrines stimulated the trend toward overseas expansion and growth of the navies of the world between 1898 and 1914. The resultant acquisition of an overseas empire almost overnight in 1898 greatly changed the strategic position and problems of the United States. Its emergence as a world power, beginning in these years, was to have important bearings on the strategic balance of power among the nations of the world, an equilibrium that World War I altered profoundly.

4. World War I.—World War I, the first of the great coalition wars of the 20th century, was an important landmark in the story of the evolution of modern strategy. Never was the phenomenon of cultural lag as applied to warfare more clearly demonstrated. Beginning in the accepted mold of strategic planning popular since 1870, it soon ran head-on into counter trends that were altering the very bases of strategic action and that strategic thinking in the intervening years had not yet fully grasped. Despite the experiences in the Boer and Russo-Japanese wars with the machine gun as a defensive weapon of tremendous fire power, French and German military leaders at the outbreak of the war continued to put their faith in the offensive. In fact, they were convinced that new weapons and methods of control, the radio and telephone, actually improved the offensive capabilities of their mass armies. The universal underestimation of the effect of modern firearms on the defense had important repercussions on strategy both during and after the war.

The first moves in the war began in 1914 as French and German strategists had planned—the Germans in accord with the Schlieffen plan, the French with their Plan XVII. In seven days the Germans concentrated over 3,000,000 men on the eastern and western fronts from mobilization points. In approximately the same time the French assembled 1,200,000 men on the western front. Both sides made heavy use of railroad lines to speed assembly of great masses of troops. Both sides were determined to attack. Out of the movements of mass armies came the first battles on the frontiers. As Schlieffen had planned, the Germans catapulted into Belgium, but the enveloping wing was not so strong as Schlieffen, who had died the previous year, had wished. It was compressed into a smaller corridor by the political decision not to violate Dutch neutrality. The anticipated six-weeks campaign of annihilation against France envisaged by Schlieffen could not be executed. The French attack also soon hit a snag. Though the French army's right wing reached the Rhine, its centre was endangered by a German pincer movement. Only a hasty retreat and a counter-offensive at the Marne saved Paris. The so-called pinwheel strategy—each side attacking and driving the enemy back—had stalled badly.

Meanwhile on the eastern front the German prewar strategy of holding until France had been quickly defeated was compromised by the desire of the Austrian ally to push against the Russians, partners of the French. The German victory at Tannenberg counterbalanced the Austrian defeat at Lemberg. The eastern front became stabilized.

By the close of 1914 the war had become a stalemate on both the eastern and western fronts. The conflict had resolved itself into trench warfare from Switzerland to the English Channel. Machine guns and artillery took over the battlefield. The conflict had settled down into a war of position, and strategic mobility was lost. World War I became a classic case of arrested strategy.

By the close of 1914 the first phase of the war was over. Prewar plans had failed; the war of movement, of mass offensives, had ceased. The big question thenceforth was how to dig the war out of the trenches. In answering that question important elements of grand strategy came into play. The heavy demands upon industry for munitions of war multiplied, and technology was called upon for new means—the tank and poison gas—of breaking the stalemate. Britain's naval blockade to starve Germany took on added significance. The German countermeasures helped bring the United States into the war in 1917. But the United States, as usual in its military history, was not prepared for war. The build-up of its forces across the Atlantic was slow. The Germans, seeking in 1918 to forestall the full impact of U.S. might, put their resources into a great offensive which came close to succeeding. When the Americans, led by Gen. John J. Pershing, finally arrived in force, they played a valuable part in military strategy in reducing the salients within the Allied lines, such as that at St. Mihiel. Eventually the German allies defected, German armies reached a point of exhaustion and the homeland a stage of semistarvation, and Germany asked for an armistice.

Though much has been written on World War I, the strategic lessons of that conflict for coalition warfare had not been fully probed by mid-20th century. Never was the dependence of strategy on statecraft more clearly demonstrated. As political circumstances of the war changed, strategy changed. The Central Powers, led by Germany and Austria-Hungary, never had a common plan of campaign or effective unity of command. The Allied side achieved these only under the pressure of necessity. Along with the military factors, economic and psychological considerations proved important in the conduct of the war and gaining the victory. Though the aim of annihilating the enemy was paramount with both sides—especially in the opening campaigns—the desire to exhaust him also influenced strategy. At the same time fresh confirmation was given to Moltke's description of strategy as a "system of makeshifts."

Military leaders in World War I had to master three basic factors in strategic calculations: masses of men, technological advances and wide areas. The movement of huge masses became an art in itself, for armies had taken on unprecedented dimensions. Millions of men were in action. Railroads and motor transport became important not only for concentration but also for establishing new strategic points on the fronts themselves. The arena of war embraced whole continents. Battles lasted for days and weeks and the fighting continued even after the great battles were over.

New weapons came into play. Aerial reconnaissance enabled a commander to gain some insight into the enemy's intentions and movements. New means of communication—telephone, radio telegraphy, the automobile and the airplane—promoted faster execution of orders and unified command over widely scattered forces. The overwhelming firepower of modern weapons checked the effectiveness of the attack, long considered the ideal path to victory. The tank, however, offered fresh possibilities in redressing the balance between the defensive and the offensive. Tactics became more than ever a prelude and conditioning factor of strategy, since without freedom of movement, strategy was only an academic exercise. Tactics came to mark the beginning rather than the conclusion of an operation.

But there were also larger strategic influences at work. If World War I was a war of masses, it was also a war of matériel. War was becoming increasingly total and cutting deeper into the life of the nation. Some of the foremost leaders and students of World War I—notably Winston S. Churchill and Georges Clemenceau—recognized that military strategy had become but a part of a greater national strategy. Symptomatic of this thinking was

Clemenceau's widely quoted statement that war was too important a business to be left to soldiers. More than ever strategy and politics would have to be correlated. The increasing totality of modern war would have to be matched by a broader national strategy. But the larger impact of the war in the international sphere—the effects of the defeat of Germany, the weakening of England and France, and the rise of the Soviet Union on the strategic balance of power in the world—could not yet be foreseen.

5. Between World Wars I and II.—The period between 1918 and 1939 saw strategy once more in process of flux. As an outgrowth of the experience of World War I, strategy came largely to mean defense. In France, particularly, a Maginot line mentality began to take hold. The belief was strong that field fortifications aided by the machine gun would contain any attack. The huge losses of World War I would thereby be avoided.

But countertrends were soon to dispute this prevalent emphasis in strategic thinking. One strong challenge came from the new school of exponents of air power. In World War I the air arm had had its beginnings. The period between the end of World War I and the beginning of World War II saw it come into its own; air forces and air organization expanded greatly. Theorists began to develop the strategy of warfare of the third dimension. Foremost among these was the Italian Gen. Giulio Douhet. He first presented the doctrine that the air arm alone would decide wars of the future. In his view, land and sea forces would no longer be decisive. On the ground, armies could act henceforth only on the defensive since attack, and with it the decision, could be gained only through the air. Air power could quickly conquer time and space. The air arm could circumvent every kind of ground resistance and nullify fortified positions and obstacles of terrain. It could strike at the enemy's sources of power before his armies could fire a shot. It could strike at his capital, industrial centres and communications. In short, it could so reduce his ability and willingness to resist that he would surrender. Douhet proposed to expand the air arm as much as possible, keep land and sea forces only as support for war in the air, and gain control of the air by defeating enemy air forces in battles or destroying them in their airfields. He made strategic bombing, and the industrial objective —strikes at the opponent's heart—the core of his doctrines.

Douhet's epoch-making ideas found many supporters in other countries, including the United States. This school of thought generally argued that huge armies would no longer be necessary. The opponent's will could be overcome even if his armed forces remained undefeated. Some of Douhet's adherents went further and demanded the abolition of land and sea forces altogether. In any event, the rise of air power accentuated the need of thinking of strategy as dealing with something more than the movements of armies on land, or of ships at sea.

Meanwhile, army leaders began to advocate another solution to break the strategic stalemate of World War I. To overcome the superiority of the defensive, they put their faith in developing a modern cavalry of tanks and armoured, motor-driven vehicles. The best known among the great champions of mechanization and motorization that arose in Great Britain was Maj.-Gen. J. F. C. Fuller (*q.v.*). These advocates saw in the armoured vehicle, combining fire-power, extreme mobility and armoured protection, the best answer to overcoming defensive forces relying on machine guns. This system was particularly suited to needs of an insular country, protected by a strong air force and navy, and of a relatively small army intended primarily for expeditionary purposes in support of continental allies. But this solution on the ground found support in Germany, Russia and the United States. In France, Charles de Gaulle bucked the strong tide of opinion and advocated tank warfare to restore a strategy of mobility and the offensive.

In the late 1930s the Germans combined air power and tanks into a new form of assault that also aimed to overthrow defensive superiority. Developing a highly mobile form of warfare for lightning strikes and mechanized attacks, they were to contribute the art of the blitzkrieg—the spearhead of a conquering, offensive strategy that Hitler unleashed in World War II.

In Germany, too, other influences supporting offensive strategy came to the fore. To overcome strategic stalemate of the World War I variety, Erich Ludendorff contributed his theory of total war. He envisaged total mobilization of a nation's manpower and resources for war. The nation at war would be led by a supreme military commander; strategy would dictate policy. The concept of total war moved geography and economics into prominent positions in nazi thinking. Even before World War I the British geographer, Halford J. Mackinder, had posed the potential threat of a heartland power, in control of Eurasia, to sea power—a counter to Mahan's theory of control of the seas. Led by Karl Haushofer, the German geopoliticians after the war took over the "heartlands" concept. Through their teachings the concept of control of Eurasia became imbedded in nazi statecraft. Their doctrines gave support to the main strands in Hitler's offensive strategy—continental expansion, autarky (national economic self-sufficiency) and *Lebensraum* (living space).

Before war burst upon Europe in 1939, it was apparent that important changes were also brewing in naval strategy. All major sea powers were producing high-speed battleships and the aircraft carrier was becoming a significant and integral member of the fleet. In the crucible of World War II the emerging elements in ground, air, naval and nonmilitary strategy were to take clearer shape.

6. World War II.—Strategy came into its own in World War II. Global and coalition warfare on an unprecedented scale required global and coalition strategy to match it. The new trends in military, air and naval strategy were put to the test. At the same time the world-wide conflict lifted strategy out of the purely military sphere into the field of grand strategy and international relations. The war cut more deeply than ever into every phase of life. It touched all nations, directly or indirectly, all continents and oceans, the air, the land and the sea. Political, economic, technological and psychological factors were drawn into the web of the strategy of all-out effort. The strategy for war grown increasingly total was conditioned by massive armed forces, revolutionary scientific and technological advances, tridimensional warfare and miracles of industrial production and logistics.

World War II began with independent offensive moves of the axis nations, Germany, Italy and Japan. Though these nations formed a war coalition, they never formulated a common blueprint of strategy or achieved the degree of co-operation that the Allied coalition did. Nevertheless, the axis war effort marked an epoch-making development in strategy. The axis concept of total war, particularly as developed in nazi philosophy, revealed more clearly than before that the traditional view of strategy—the art of employing military forces—was too narrow. Diplomacy, propaganda, "fifth columns" and espionage, geography, economics, technology and morale all entered into the nazi concept of strategy. In nazi strategy, the line between war and peace could no longer be clearly defined. The course of Hitler's campaigns, before and after the outbreak of actual war in 1939, showed that military operations and battles were only the last resort against an enemy, to be applied after all other modes of conquest had failed. Hitler early recognized that armed forces are only one of many means available to grand strategy. Indeed, his greatest victories were the bloodless ones before Munich. After war was joined in Sept. 1939, following the invasion of Poland, military strategy perforce came into greater play. Like Caesar and Napoleon, Hitler combined in his own person the two functions of strategy and policy. But his reputation as a military strategist was not to compare with his early triumphs in political warfare.

In the first analysis, Hitler gave the art of offensive strategy a new twist. Like Napoleon, he failed to master all the elements of grand strategy. But in developing and correlating an assortment of modern means of breaking an enemy's will, he showed himself to be far ahead of his opponents.

Despite great German blitzkrieg victories on the continent, successes in the middle east and strategic bombardment of England, the winter of 1941 found Germany still without the victory it sought. Though it had tried to avoid a two-front war, by June 1941 it had become embroiled with the U.S.S.R. In December the Japanese struck at Pearl Harbor and the United States came in to aid England and the U.S.S.R. against the axis powers. The world-

wide strategic contest was then joined in earnest. The axis nations had the advantage of interior lines; the Allies, other than Russia, had to fight on exterior lines, and their lines of support had to fan out from the factories of Britain and the United States all over the world. Inevitably the Allies turned to the strategic weapon of blockade against the axis powers, and Germany and Japan turned to submarine warfare to cripple the Allied lifelines.

The story of Allied strategy in World War II is the search for common denominators among three sovereign powers drawn together by a common bond of danger. That strategy was a product of many minds on both sides of the Atlantic, of diverse pressures and pulls and of changing circumstances in the global war. It was the result of an evolutionary process and a series of compromises, and of a constant struggle to adjust ends and means. Above all, it was fashioned by powers with diverse national interests. If the national objectives sought by each of the participating powers were unchanging, the means and methods used by each to achieve them varied from time to time. In the process of planning and waging the war, moreover, the foundations of the alliance shifted and the relationships among the powers changed. These shifts are part and parcel of the strategy story.

That story falls into at least three phases. The first of these covers the period 1941–42. This phase saw the formation of the Grand alliance and the emergence of divergent British and American concepts of strategic theory.

The Grand alliance was a "marriage of expediency," forged in war and cemented by the common bond of danger. But each member of the alliance expected from the partnership different objectives; each, because of its distinctive traditions, policies, interests and position, looked at the war through different spectacles.

Great Britain, an island dependent upon sea lanes for its very existence and situated precariously on the edge of Hitler's European fortress, had for centuries put its faith in the balance of power. It was natural for it to seek to rally the smaller nations and oppose any strong power that threatened to upset the balance on the continent. It could be expected to intervene actively in the Mediterranean and the middle east through which ran its lifeline to the empire in the east. But in any global war Great Britain's resources were bound to be stretched thin. Its economy, while highly industrialized, was small in comparison with that of the United States. Keenly conscious of its heavy losses in manpower during World War I, Britain would tend to avoid as long as possible large-scale ground battle on the continent. By necessity and choice, therefore, the leaders of Britain put their faith in the navy, the air and mechanized and armoured forces. Experienced in war, diplomacy and empire, Great Britain had a long history of alliances with European powers. Its admirals and generals were accustomed to working closely with the politicians and diplomats, and British strategy usually reflected political objectives.

Across the Atlantic lay the other major western power in the alliance, the United States—young, impatient, rich in resources and highly industrialized. Traditionally opposed to becoming involved in European quarrels, it nevertheless had strong bonds with Europe, especially Britain. To most U.S. citizens war was an aberration; an unwelcome interruption to the normal course of events. Though its tradition in war had been first to declare, then to prepare, the country had between 1939 and 1941 gradually awakened to the dangers from without. It began to mobilize manpower and resources. Aid to Britain became official national policy in 1940, and lend-lease, a strategic weapon, was extended to Britain and other friendly powers in 1941. The strategic planners in Washington, laying aside their earlier academic exercises, began to think in terms of global and coalition warfare. Without committing the administration, the military planners entered into staff conversations with their British opposites on possible ways of using U.S. forces in coalition strategy. For the first time in its history the United States entered a war considerably advanced in its military planning. It had considerable interests in the far east and a tradition of helping China. Throughout World War II the U.S. military staff and the president could not neglect the war in the east any more than the U.S. public could forget the attack on Pearl Harbor. One other factor that distinguished U.S. policy and strategy was the loose relationship between the president and the military. That relationship was in sharp contrast to the closely knit politico-military system of Great Britain.

The third major partner in the alliance against Germany was the Soviet Union, a land power with completely internal lines of communication. Though it possessed an enormous population and great resources, the Soviet Union's industrial facilities were still not fully developed. The U.S.S.R. represented an enigmatic force, devoted to a political and economic ideology different from that of the western partners. Born in revolution and come to power in a civil and foreign war, its policies seemed to combine Russian national socialism, Marxism and tsarist imperialism. Dedicated to the proposition that war was inevitable in capitalistic society until the world revolution ushered in a new millennium, it had developed a strategy of revolution—based on the theories of Marx, Engels and Lenin. In that theory, as in nazi strategic theory, the line of demarcation between war and peace was not clearly drawn. Until attacked by Germany, soviet policy was dominated by the twin notions of security and expansion. The nazi invasion reinforced the soviet desire to strengthen its position in eastern Europe—an objective whose roots lay deep in Russian history. But for almost two years after the German attack the Soviet Union was engaged in a desperate fight for existence. Lacking air and naval traditions, it relied on geography, the loyalty of its people and the huge Red army to see it through. Political and territorial ambitions were subordinated to military necessity. Still fearful of capitalist encirclement, suspicious of friend and foe, the Soviet Union held an uneasy position in what Gen. John R. Deane so fittingly called "the strange alliance."

From the beginning, the inner web of the Grand alliance was the close relationship between the United States and Great Britain. The Soviet Union's part in developing and directing the combined strategy of the war was to be relatively small. Compared with the world-wide demands facing the United States and Great Britain, its strategic problem was simple. Unlike the western partners, the U.S.S.R. would be at war on only one front at a time; it did not enter the conflict with Japan until the closing days of World War II. The Soviet relationship with its western partners consisted of demanding and receiving material aid and pressure against the common enemy. But collaboration was to prove difficult even in these areas. Actually the strategic decisions of the western powers were normally transmitted in general terms to the Soviet Union. The Russians took formal part in strategic decisions only at the international conferences at Moscow (Oct. 1943), Teheran (Nov. 1943), Yalta (Feb. 1945) and Potsdam (July 1945), and even these were called at the initiative of the western powers. Throughout World War II, the Russians remained outside the combined staff system, developed for the coordination of the western effort in the global war.

The basic procedures for formulating Allied strategy emerged in 1941 and 1942. Shortly before Christmas 1941, Prime Minister Winston S. Churchill and his principal military advisers arrived in Washington for the first of their great wartime conferences with Pres. Franklin D. Roosevelt and his staff. Out of this meeting —the Arcadia conference—came the establishment in Washington, in Jan. 1942, of the combined chiefs of staff (C.C.S.), the agency for hammering out Allied strategy and for day-to-day conduct of the war. To supply U.S. opposites for the British in the combined organization, the United States joint chiefs of staff (J.C.S.) and joint staff planning system (army, air and navy) came into being. Since the three principal British military officers normally had to direct the operations of their services from London, they sent senior representatives to function for them in the combined chiefs of staff. As finally constituted, the J.C.S. was composed of the president's chief of staff and the senior officers of the army, navy and army air forces.

The C.C.S. was a unique organization in coalition machinery. Decisions were arrived at by common agreement; no votes were taken. There the leaders concerted policies and plans, outlined strategies, discussed timing of operations, approved programs of allocations and measured requirements against resources. Their

decisions were, of course, subject to the approval of the president and the prime minister.

The prime minister and the president, who were over and above the C.C.S. system, were responsible for all decisions. Each wore two hats—one political, the other military. As political leaders they sometimes had more in common with each other than with their respective staffs. Advising Churchill at the summit of the British system of intragovernmental planning was a war cabinet with which the chiefs of staff committee sat. While President Roosevelt could and did draw on the assistance of the war-born U.S. joint staff system, he never established anything remotely resembling the British war cabinet. The differences between the two politico-military systems—one loosely built, the other closely tied—were sometimes strikingly illustrated at the international war conferences.

The full-dress Anglo-American conferences usually came about when planning had reached a point where top-level decisions on important matters of Allied strategy and policy were necessary. The conferences and the C.C.S. system provided the framework for the important decisions in European and Asiatic strategy. Decisions in the war in the Pacific against Japan were handled somewhat differently. The U.S. joint chiefs of staff who were early given the responsibility for this war submitted their decisions on plans and operations against Japan to the conferences. There the combined chiefs usually gave their stamp of approval.

Beneath the top-level machinery and funneling into it was a whole network of specialized combined and joint planning agencies. At the bottom of the hierarchy lay the planning staffs of the individual services in the capitals. Unique among them on the American side was the war department's operations division, which, under the army chief of staff, Gen. George C. Marshall, developed into the "Washington command post" to control the strategic deployment and activities of the army forces scattered all over the world. As the war progressed, the highly organized planning machinery in the capitals was supplemented by the big theatre headquarters staffs. Such theatre commanders as General Eisenhower in the European theatre and Gen. Douglas MacArthur in the southwest Pacific began to play an increasingly important role in strategic decisions. The vast and speedy network of telecommunications circling the globe enabled strategic plans and decisions to be quickly made and executed. Never had the processes and machinery of strategic planning been so highly organized as in World War II.

Even though much of the Anglo-American discussions at the Arcadia conference centred about the crisis in the Pacific, the western partners agreed that the first and major object of Anglo-American grand strategy must be the defeat of Germany. The principle of "Europe first" had actually been foreshadowed in staff conversations early in 1941. So had the corollary that, should Japan enter the war, military strategy in the far east would be defensive. When war came to the United States in Dec. 1941, this strategy, despite the initial Japanese victories, was adopted.

Political expediency combined with geography and logistics in arriving at this decision—perhaps the single most important one made by the British and U.S. in World War II. British and Soviet power lay close to Germany. Great Britain offered a base for massing western air power; in the Mediterranean, operations against Germany could soon be undertaken. The Soviet Union and Great Britain could not wait for a decision in the war against Japan. Before the west could come to grips with Japan decisively, U.S. naval striking power would have to be restored, ships built and advance bases and lines of communications extended across the far Pacific. It followed, therefore, that defeat of Germany should be the first major objective and that in the meantime the Japanese should be contained until the Allies could assemble enough strength to take the offensive in the Pacific.

The divergent approaches of the Allies were most clearly reflected initially in the conflict between U.S. and British strategy against Germany. The 1941–42 period saw the emergence of what may be called the peripheral theory, espoused by Churchill and the British staff, and the theory of mass and concentration, advocated by General Marshall and his staff. The British stood for hitting the German army at the edges of the continent; the Americans for concentration at the decisive point to defeat the main body of the enemy. In the peripheral concept, emphasis would be on swift campaigns of speed and maneuver, on probing soft spots, on a war of attrition. The cross-channel operation would follow as a last blow against a Germany already in process of collapse. The British concept was in accord with its small-scale economy and limited manpower, its experience in World War I and the prime minister's predilections. The U.S. early thought in terms of coming to grips with the German army and defeating it decisively. This theory reflected U.S. optimism, reliance on the industrial machine to produce the necessary wherewithal, and military faith in a large citizen army built and trained for offensive purposes. Both justified their theories and plans in terms of relieving the pressure upon the Russians. The divergent approaches lay behind the debate in 1942 over the notion of an early cross-channel attack versus the invasion of North Africa—Operation "Torch." The British won out, the president overruling the U.S. staff, and October saw the start of Mediterranean operations.

The "Torch" decision, which so disappointed U.S. military expectations, also complicated relations with the Soviet Union. Stalin, locked in a death struggle on the eastern front, demanded a second front—and soon. Each Anglo-American postponement of this second front was to add fuel to the fire of his resentment.

The cross-channel versus Mediterranean debate of midwar was argued in and out of the big international conferences. North Africa led to Sicily (July 1943); Sicily, to the invasion and surrender of Italy (Sept. 1943). Always the prime minister urged operations in the Mediterranean that would bleed the foe and prepare the way for an eventual cross-channel operation. Always the U.S.—with Marshall as the foremost spokesman—urged the need for a definitely scheduled cross-channel operation in force and for tapering the Mediterranean effort. Both sides agreed on the need for strategic bombardment to pave the way. The series of Anglo-American decisions reached at the 1943 conferences—Casablanca in January, Trident (Washington) in May, Quebec in August and Cairo-Teheran in November and December—reflected the compromises between opportunism and long-range commitments, between a war of attrition and a war of mass and concentration.

The meeting at Teheran in Nov. 1943, where, for the first time in the war, the president, the prime minister and their staffs met with Marshal Stalin and his staff, was the decisive conference for strategy. Churchill made eloquent appeals for operations in Italy, the Aegean and the eastern Mediterranean, even at the expense of a delay in the cross-channel operation ("Overlord"). But Stalin, for reasons of his own, put Soviet weight squarely behind the U.S. concept of strategy. He asserted his full power as an equal member of the coalition and came out strongly in favour of "Overlord." Further operations in the Mediterranean, he insisted, should be limited to the invasion of southern France in support of "Overlord." In turn, he promised to launch an all-out offensive on the eastern front.

Stalin's stand put the capstone on Anglo-American European strategy and in a sense fixed western strategy. Germany was to be crushed by a great pincers—the Anglo-American drive on the west and a Soviet drive from the east. General Eisenhower was appointed commander for "Overlord" and preparations for the cross-channel thrust began. The decision at Teheran marked a subtle but significant change in the relationship of the Allies. Britain was growing relatively weaker, the United States and the Soviet Union stronger. By way of the military doctrine of concentration the strategists of the Pentagon and of the Kremlin had found common ground.

The third phase of the strategy story—from the summer of 1944 to the surrenders of Germany and Japan—was the period of the payoff. In this period the problems of winning the war began to come up against winning the peace. The curtain began to lift on the divergent national objectives of the Allies—objectives hitherto obscured by the common military danger. The western drive unfolded according to plan. The thrust across France coincided with the invasion from the Mediterranean; the Rhine was crossed, the Ruhr encircled and the drive across Germany met up with the

Russian advance. But to Churchill, warily watching the swift Soviet advance into Poland and the Balkans, the war had become more than ever a contest for great political stakes. As the strategy unrolled in the field, the two approaches to the war boiled down to a question of military versus political maneuvers.

While the war in the west was drawing to a close, the strategy for defeating Japan had been taking gradual shape. As the war had advanced, Allied strategy in the Pacific shifted from the defensive to the offensive. Though firm, final plans had to await the defeat of Germany, the momentum of advances in the Pacific was so fast that it almost caught up with the European conflict. Island to island advance was gradually changed to island-hopping. The battle of Midway in 1942 foreshadowed a new trend in naval strategy, in which opposing fleets did not get within gun range of each other, but duelled with their air power and submarines. Submarine warfare, amphibious assaults, task forces built around aircraft carriers, the strengthened American fleet—speeded the pace in midwar. A twofold approach—along the southwest and central Pacific axes—took the Allies to the threshold of Japan by the time the European war closed. Before final plans for invasion were put into effect, however, the Japanese surrendered. A multiplicity of factors led to this result—including the threat of invasion, the attrition of Japanese shipping, air bombardment, the entry of the Soviet Union into the war against Japan and, most important, the dropping of the atomic bomb.

In retrospect, the strategy of World War II represented a curious mixture of old and new strands. Never had strategy had so vast a scope, for this was the first really global war. Never in history had two partners pooled their efforts and ideas so closely as had the Americans and British in the Grand alliance. Once more the scales of war were tipped in favour of the offensive. To the age-old primary aim of seeking the defeat of the enemy's armed forces was added the objective of striking at his industrial heart. Both these aims, for example, were present in the directive given to General Eisenhower for the European campaign. In Pacific strategy, it long remained a question whether invasion and defeat of the enemy forces on their home grounds would even be necessary. Strategic bombardment and sea power gave support for a strategy of attrition and destruction of industrial capacity. But the thesis that the sole aim of strategy was to support air power in order to annihilate the enemy's industrial power was only partially tested. It had been impossible to put all production and manpower into the air strategy. Nor were ground and sea enthusiasts on the basis of World War II experiences disposed to yield the field of strategy completely to the new arm. Allied military strategy, despite attention to "principles," had been a hybrid product, largely hammered out on the anvil of necessity. Final answers in military strategy to the problems of proper aims and emphases, and of best combinations and permutations, were still not clear.

It was clear, however, that modern war had grown more total than ever, and that science, industry, diplomacy and psychology all had to be harnessed to it. Campaign strategy could no longer be clearly differentiated from national strategy; military strategy —in the sense of disposing armed forces for battle—was overshadowed by grand strategy. Where all this would lead was still not clear at the end of World War II, for hanging over all society was the weapon that threatened to revolutionize all warfare, the atom bomb.

7. Cold War in the Atomic Age.—A great debate rose over strategy in the postwar years. Underlying it were a number of crucial factors. World War II had further upset the balance of power in the world, earlier disturbed in World War I. Germany and Japan had been defeated, Britain and France further weakened, and the United States and the Soviet Union had emerged as the two strongest powers in the world. The growing rift between the east and west led some critics in the western world to point to mistakes in Allied political and military strategy of World War II —especially to an overconcentration on military objectives at the expense of political aims, on gaining military victories rather than winning the peace. While part of this criticism seemed overly harsh and not always founded on well-established fact, it was clear from the growing flood of literature on World War II that political and military strategy in World War II had not been systematically meshed. At the same time the war had loosed a wave of nationalistic and revolutionary movements in the world with strategies of their own. The struggle by the committed powers to win the non-committed was intense. Through co-ordinated political, economic, military and psychological pressures, the Soviet Union sought to incline the new nationalistic and revolutionary movements toward its orbit. The growth of the cordon of Soviet satellite states in central and eastern Europe was one result of its postwar strategy; Soviet partnership with Red China was another. Meanwhile the west, led by the United States, moved to shore up the weakened European powers. The Truman doctrine, the Marshall plan and the emergence of the North Atlantic Treaty organization (NATO) were diverse strategic weapons to counter the many-sided Soviet threat. The "cold war" moved strategy to all fronts—political, diplomatic, economic, propagandistic, as well as military—in both the east and west. New phrases entered the vocabulary of western thought—"strategy of containment," "strategy of total war," "strategy of massive retaliation," "strategy of preventive war," "strategy of deterrence," "strategy of graduated deterrence."

Though the great concern in the cold war era was the possible outbreak of World War III, strategy also had to consider more limited warfare. The Korean war, beginning in 1950, brought the North Korean and Chinese Communist forces into conflict with the South Korean and United Nations forces. It demonstrated that warfare in the cold war could also be limited and "hot," and that strategy would have to be tailored to local as well as world-wide needs.

Meanwhile, on the military plane, the impact of science and technology on warfare was promising revolutionary changes in strategy. To the growing arsenal of powerful new weapons begun with the invention of the atom bomb was added the hydrogen bomb —a bomb that could wipe out a whole city. New developments such as atomic-powered air and naval craft, atomic artillery, guided missiles that could cross oceans and continents, chemical, biological and radiological weapons, supersonic speeds and other revolutionary adaptations in communication and transportation pointed to many changes in the art of war and in strategy. A great debate was joined between those who placed primary reliance on the new methods and means of war, and those who championed more orthodox and conventional approaches supplemented by the new weapons. The debate was particularly sharp between the advocates of a national military strategy based upon air power alone and the defenders of traditional strategy. Some air enthusiasts felt that the defeat of an enemy's armed forces on the battlefield was no longer a primary aim of strategy; that a new atomic air strategy offered unparalleled opportunity to break an enemy's will to resist by paralyzing his industrial power. Some advocates believed that the new atomic air strategy could be superimposed on the old strategy based on land, sea and air forces; others that a choice had to be made between the two. Inevitably, the new developments in warfare also led authorities to take a fresh look at the "principles of war." Some began to argue that dispersion rather than concentration had become a basic goal; others that concentration merely required new adaptation.

Still other theorists in the cold war period were pointing up other new trends in strategy. Some went so far as to assert that not only was conventional military strategy outmoded but, along with it, national strategy. This belief was based on the idea that the 20th century was the age of great coalitions. This condition had resulted from the shrinkage of the world, the growing interdependence of nations and the nature of modern war. This point of view was perhaps most emphatically asserted by Sir John Slessor in his *Strategy for the West* (1954). He wrote, "Strategy today is world strategy and it is no more possible to shape it merely to suit the limited interest of one particular country than it is to confine a typhoon to a potato patch." (John C. Slessor, *Strategy for the West*, p. 47, William Morrow & Co., New York, 1954, and Cassell & Co., Ltd., London, 1954.)

As the cold war advanced, the belief was growing that this was primarily a war of ideas, of opposing creeds, that strategy must be offensive but waged primarily with political, economic and

propagandistic means. Preparedness on all fronts—military and nonmilitary—became a significant factor in that strategy. The belief seemed to be growing that in the holocaust of all-out modern war, all sides might lose. The ultimate aim of strategy in the cold war had become the achievement of national and international objectives without touching off World War III. Strategy had expanded into every phase of national endeavour, just as had war itself. The line of demarcation between war and peace had further broken down, and with it the broadest definitions of strategy yet offered were no longer adequate. From the stratagems of the commander in war, strategy had come to pre-empt more and more of the spheres of warfare and statecraft. The line dividing it from policy had become indistinct and its highest aim had become the defense and survival of civilization itself. (*See* also LOGISTICS.)

BIBLIOGRAPHY.—Hermann Foertsch, *The Art of Modern Warfare* (New York, 1940); Edward Mead Earle *et al.* (eds.), *Makers of Modern Strategy* (Princeton, 1948); Maurice Matloff and Edwin M. Snell, *Strategic Planning for Coalition Warfare, 1941–1942* (Washington, D.C., 1953); B. H. Liddell Hart, *Strategy, The Indirect Approach* (New York, 1954); John C. Slessor, *Strategy for the West* (New York, London, 1954); John Ehrman, *Grand Strategy*, vol. v and vi (London, 1956); Maurice Matloff, "The Soviet Union and the War in the West," *United States Naval Institute Proceedings*, vol. 82, no. 3 (March 1956); Edgar J. Kingston-McCloughry, *Global Strategy* (New York, London, 1957); Henry A. Kissinger, *Nuclear Weapons and Foreign Policy* (New York, 1957). (M. MR.)

STRATEMEYER, GEORGE E. (1890–), U.S. air commander in Asia during World War II and the Korean War, was born in Cincinnati, O., Nov. 24, 1890. He joined the U.S. army air arm shortly after graduation from West Point in 1915 and served in a number of school, staff and command positions for the next two decades. He was chief of the air staff of the U.S. army air forces, 1942–43, and from Aug. 1943 until the end of World War II he commanded U.S. and Allied air forces in the China-Burma-India area.

After the war, Stratemeyer served as commander of the air defense command in the United States and of the far east air forces in Japan and Korea. He directed the air war in Korea until physical disability forced his return to the United States in 1951 and his retirement in 1952. (A. GG.)

STRATFORD, JOHN DE (d. 1348), archbishop of Canterbury, was born at Stratford-on-Avon and educated at Merton college, Oxford, afterward entering the service of Edward II. He served as archdeacon of Lincoln, canon of York and dean of the court of arches before 1323, when he became bishop of Winchester, an appointment which was made during his visit to Pope John XXII at Avignon and which was very much disliked by Edward II. In 1327 the bishop joined Queen Isabella's partisans; he drew up the six articles against Edward II, and was one of those who visited the captive king at Kenilworth to urge him to abdicate in favour of his son. Under Edward III he became a member of the royal council, but his political importance dates from the fall of Roger Mortimer. In Nov. 1330 Stratford became chancellor, and for the next ten years he was the king's most prominent adviser and the "head of the Lancastrian or constitutional party" (Stubbs). In 1333 he was appointed archbishop of Canterbury, and he resigned the chancellorship in 1334. He was again chancellor from 1335 to 1337 and for about two months in 1340. In Nov. 1340 Edward III, humiliated, impecunious and angry, returned suddenly to England from Flanders and vented his wrath upon the archbishop's brother, the chancellor, Robert de Stratford. Fearing arrest, John de Stratford fled to Canterbury, but after a violent dispute the principle was established that peers were only to be tried in full parliament before their own order (*en pleyn parlement et devant les piers*). The archbishop acted as president of the council during Edward's absence from England in 1345 and 1346, although he never regained his former position of influence. He died at Mayfield in Sussex on Aug. 23, 1348.

STRATFORD, a city and port of entry of Ontario, Can., and capital of Perth county, 83 mi. W.S.W. of Toronto on the Avon river, pop. (1961) 20,467, is noted for its annual summer Shakespearean festival. It is served by the Canadian National railway. Repair and engineering shops of the railway, flour and woolen mills, furniture and agricultural implement works are the principal industries. It is in the heart of rich dairy country, and there is a large export trade in cheese and farm produce. The festival, first held in 1953, attracts thousands of visitors.

STRATFORD DE REDCLIFFE, STRATFORD CANNING, VISCOUNT (1786–1880), British diplomat notable for his influence in Turkish affairs, was born in London on Nov. 4, 1786, the fourth son of Stratford Canning, a merchant who had been disinherited following his marriage to Mehitabel Patrick. The elder Canning died in April 1787, endowing his family with little but the friendship of his nephew, George Canning (*q.v.*). The latter's influence enabled Stratford Canning to be educated at Eton and King's college, Cambridge. His studies there were interrupted when George Canning, then foreign minister, appointed him in 1807 as a précis writer at the foreign office and later to missions at Copenhagen and Constantinople. In 1810 he became chargé d'affaires at Constantinople and, without instructions from England, helped arrange the peace of Bucharest (1812) between Turkey and Russia, just as Napoleon was about to invade the latter. On returning home, poverty thwarted his ambition to enter political life, but Lord Castlereagh appointed him minister to Switzerland (1814–18) and used him at the Congress of Vienna as adviser on that country. He then went as minister to Washington (1819–23) and, although he sometimes clashed with John Quincy Adams, the U.S. secretary of state, on returning to London he negotiated a convention which was, however, refused ratification by the U.S. senate.

In 1824 George Canning appointed him ambassador at Constantinople, but sent him first to discuss North American frontier questions at St. Petersburg. He also raised, unofficially, the Greek issue. At Constantinople he joined the Russian and subsequently the French ambassadors, but the battle of Navarino (*q.v.*) on Oct. 20, 1827, forced all three to retire to Corfu. In 1828–29 Canning resumed negotiations concerning Greece at a conference at Poros and accepted an extension of its frontiers, which was however repudiated by the earl of Aberdeen, then foreign minister. Lord Palmerston sent him again to Constantinople in 1831, when he finally obtained these frontiers, chiefly because of the threat to the sultan from Mohammed Ali (*q.v.*), the pasha of Egypt. Palmerston, whom the cabinet prevented from accepting Canning's advice to protect the sultan, next appointed him ambassador to St. Petersburg. The tsar Nicholas I, for reasons still not completely explained, refused to receive him, so that henceforward there was a personal issue between the two.

During this controversy Stratford Canning was sent on a special mission to Spain (1832–33) which was not a great success, and then embarked on a political career. He had sat for the pocket borough of Old Sarum (1828) and Stockbridge (1831), both abolished by the Reform bill, and 1835–41 was member for King's Lynn, speaking in the house mainly on foreign affairs. He impressed little and although in 1851 Lord Stanley promised that he should be the next Tory foreign minister, Canning was never to hold office. Meanwhile he had accepted a fourth mission to Constantinople (1841–46), when he established a unique position of influence over the sultan and his ministers, and tried hard without much success to reform the Ottoman empire. Palmerston afterward dispatched him on missions to Switzerland and central Europe, and then began another period at Constantinople (1848–51), when he supported the sultan in retaining Polish and Hungarian refugees. Created viscount in 1852 he was reappointed (1853–58), and thus served throughout the Crimean War (*q.v.*). Contrary to the view often held by historians, researches have shown that, though he saw the danger of the demands of Nicholas I, he worked unceasingly to obtain a peaceful solution of the questions at issue, but was thwarted by the refusal of the sultan to accept his advice. During the war he strove to improve hospital and army administration. He was in continual controversy with the French, jealous of his influence which in fact declined over this period. Nevertheless he was given a magnificent last farewell from Constantinople in 1858.

In 1816 Canning married Harriet Raikes, (d. 1817) and in 1825,

very happily, Eliza Alexander. From 1858 he lived quietly, occasionally speaking in the lords and publishing a few poems and books of devotion. He became a knight grand cross of the Bath in 1829, and received the Garter in 1869. He died on Aug. 14, 1880, in Frant, Sussex.

See also EASTERN QUESTION, THE.

See S. Lane Poole, *Life of Lord Stratford de Redcliffe* (London, 1888); E. F. Malcolm-Smith, *The Life of Stratford Canning* (London, 1933); H. W. V. Temperley, *England and the Near East: The Crimea* (London, New York, 1936). (C. K. W.)

STRATFORD-UPON-AVON, a municipal borough in the Stratford parliamentary division of Warwickshire, Eng., 24 mi. S.S.E. of Birmingham and 91 mi. N.W. of London by road. Pop. (1951) 14,983. Area 10.8 sq.mi. It stands mainly on the left bank of the river Avon at the point where it is crossed by a 14-arch stone bridge presented by Sir Hugh Clopton in the reign of Henry VII; there is also a 19th-century bridge. As *Æt-stretfordae* (signifying "the ford by which a Roman road crossed the river"), Stratford was mentioned as early as 691. The manor was granted by Offa to the bishopric of Worcester; and it was under the protection of the bishops of Worcester that the inhabitants of the town early assumed burghal rights. The Gild of the Holy Cross, founded in the 13th century for the support of poor priests and others, exercised authority over the town for many years. Its dissolution was the cause of the incorporation charter of Edward VI in 1553. Another charter, altering the constitution of the corporation, was granted in 1611. Other charters were granted in 1664 and 1674. Stratford is a picturesque market town and road centre and has one or two industries, but its world fame arises from its connection with William Shakespeare, born there in 1564 and whose plays and sonnets contain many allusions to local scenes. Relatively little is known of Shakespeare's life, so that the few genuine local traditions have, as is understandable with so famous a man, become encrusted with picturesque legend. It was not until 1769, when David Garrick inaugurated the first of the annual birthday celebrations, that any attempt was made to preserve buildings and other memorials of Shakespeare's life in the town, but many valuable links had already vanished.

Shakespeare was born in a half-timbered house in Henley street, bought by public subscription in 1847 and later handed to the care of the Shakespeare Birthplace trust. Here are gathered a number of documents and relics and an early portrait of him. He attended the grammar school adjoining the chapel built in the 13th century by the Gild of the Holy Cross and rebuilt in the 15th century. The chapel has a well-restored medieval Doom painting and there is a reference to a schoolmaster there as early as 1295. The hall was used from time to time by traveling companies of actors and it may have been that in this room Shakespeare had his first introduction to the stage. In 1597 Shakespeare bought the house known as New Place, and there he died in 1616. The house was pulled down in 1759, and only the foundations remain. By it is Nash's house, a museum of both local and Shakespearean relics.

Shakespeare's grave is in Holy Trinity church, a fine cruciform building on the site of a monastery that existed before 691. The earliest parts of the church (transepts, tower and parts of the nave walls) date from 1189 to 1272; the aisles are Decorated, the chancel is Perpendicular. Here is the font in which it is said Shakespeare was baptized, and the chancel where he is buried, beside his wife and other relatives. On the adjoining wall is a monument with the well-known bust of Shakespeare by Gerard Johnson (Janssen).

The group of buildings by the river known as the Shakespeare Memorial comprise a library and art gallery (opened in 1881)—all that was left of the original memorial built in 1879—and a theatre (opened in 1932) designed by Elisabeth Scott to replace that destroyed by fire in 1926. The Memorial houses a valuable collection of Shakespeareana, including early folio editions of the plays and the famous "Droeshout print" of Shakespeare which was a frontispiece to the *Plays* of 1623. The library (10,000 vol.) is one of the most complete of its kind.

April 23—the date of Shakespeare's death and, approximately, also of his birth—is annually celebrated with a procession attended by distinguished representatives of many nations, and every year from March until October is the Shakespeare festival when his plays are acted at the Memorial theatre.

Hall's croft, a Tudor building with rare old furniture, was the home of John Hall, Shakespeare's son-in-law. The Quiney house was the home of Judith Shakespeare, the poet's youngest daughter. Harvard house (1596) was the girlhood home of the mother of John Harvard who endowed Harvard university in the United States. It is one of the few notable buildings (such as the 18th-century Town hall, Shrieve's house, rebuilt 1614, and Emm's court, dating from the 17th century) which are not connected with Shakespeare. Mason croft became in 1951 the Shakespeare institute of Birmingham university. At Shottery, 1 mi. W., is the thatched cottage which was the early home of Shakespeare's wife, Anne Hathaway. It was bought for the nation in 1892.

The poet's mother was Mary Arden, the name of an ancient county family that survives in the forest of Arden, which formerly covered a large area northwest of Stratford. Her birthplace, a Tudor farmhouse at Wilcote (about 5 mi. N.W.), is preserved as a museum, and it has a collection of local farming implements.

STRATHAVEN (local pronunciation "Straven"), an upland manufacturing and market town in Avondale civil parish, Lanarkshire, Scot., on the Avon, 8 mi. S. of Hamilton by road. Pop. (1951) 4,288. It is the centre of a dairy-farming and pastoral country; industries include textiles (silk and hosiery), grain milling, agricultural engineering, printing and motor-coach building. The ruins of Avondale castle, which commanded the route into central Ayrshire, stand on a crag above Powmillion burn. At Drumclog, 6 mi. S.W., a school and a granite obelisk commemorate a victory of the Covenanters over Graham of Claverhouse on June 1, 1679. (G. P. I.)

STRATHCLYDE, the name given in the 9th and 10th centuries to the British (Welsh) kingdom, which from the 7th century onward was probably confined to the basin of the Clyde, together with the adjacent coast districts, Ayrshire, etc., on the west of Scotland. Its capital was Dumbarton (fortress of the Britons), then known as Alcluith. On the south this kingdom bordered on the territories of the Niduari Picts of Galloway, including the modern counties of Wigtown and Kirkcudbright. Strathclyde is also sometimes called Cumbria, or Cumberland, and the survival of the latter name on the English side of the border preserves the memory of a period when the territories of the northern Welsh were of much greater extent.

After the withdrawal of the Romans in the 5th century the northern Britons seem to have shown greater determination in maintaining their independence than any of the southern kingdoms and, according to Welsh tradition, Cunedda, the ancestor of the kings of Gwynedd, had himself come from the north. In the *Historia Brittonum* we read of princes of the northern Britons.

Such notices as we have of the history of Strathclyde in the 7th and 8th centuries are preserved only in the chronicles of the surrounding nations and even these supply us with little more than an incomplete record of wars with the neighbouring Scots, Picts and Northumbrians. It is probable that the Britons were allied with the Scots when Aidan, the king of the latter, invaded Northumbria in A.D. 597. In 642, however, we find the two Celtic peoples at war with one another, for in that year the Britons under their king Owen defeated and slew the Scottish king Domnall Breac. In the same year they came into conflict with the Northumbrian king Oswio. In 649 there appears to have been a battle between the Britons and the Picts, but about this time the former must have become subject to the Northumbrian kingdom. They recovered their independence, however, after the defeat of Ecgfrith by the Picts in 685. In 711 and again in 717 we hear of further wars between the Britons and the Scots of Dalriada, the former being defeated in both years. Toward the middle of the 8th century Strathclyde was again threatened by an alliance between the Northumbrians and Picts, and in 750 the Northumbrian king Eadberht wrested from them a considerable part of their territories in the west including Kyle in Ayrshire. In 756 the North Britons are said to have been forced into submission and

from this time onward we hear very little of their history.

In 870 Dumbarton was attacked and destroyed after four months' siege by the Scandinavian king Ivarr, and for some time after this the country was exposed to ravages by the Norsemen. It is believed that the native dynasty came to an end early in the 10th century and that the subsequent kings belonged to a branch of the Scottish royal family. Toward the end of the reign of Edward the Elder the Britons of Strathclyde submitted to that king together with all the other princes of the north. In the reign of his successor Aethelstan, however, they joined with the Scots and Norwegians in attempts to overthrow the English supremacy, attempts which were ended by their defeat at the battle of Brunanburh in 937. In 945–946 Strathclyde was ravaged by King Edmund and given over to the Scottish king Malcolm I. The fall of the kingdom was only temporary, for we hear of a defeat of the Scottish king Cuilean by the Britons in 971. In the 11th century Strathclyde appears to have been finally incorporated in the Scottish kingdom, and the last time we hear of one of its kings is at the battle of Carham (*c.* 1016 or 1018?) when the British king Owen fought in alliance with Malcolm II.

BIBLIOGRAPHY.—*Chronicles of the Picts and Scots,* edit. by W. F. Skene (1867); W. F. Skene, *Celtic Scotland* (1876); and Sir John Rhys, *Celtic Britain* (1904). (F. G. M. B.)

STRATHCONA AND MOUNT ROYAL, DONALD ALEXANDER SMITH, BARON (1820–1914), Canadian statesman and financier, was born at Forres, Scotland, on Aug. 6, 1820. In 1838 he was appointed a junior clerk in the Hudson's Bay company, sent to Labrador and stationed at Hamilton Inlet. For 13 years he roughed it there, learning the fur trade. Then he was for ten years on Hudson bay, rising in the company's service to be a chief trader and then a chief factor.

In 1868 he was appointed to the post of resident governor, with headquarters at Montreal.

The construction of the Canadian Pacific railway had become a practical question in 1872. Smith, together with his cousin George Stephen, and other Canadian and U.S. financiers, bought the insolvent St. Paul & Pacific railway which by 1873 had been completed from St. Paul to Breckenridge. The new owners persuaded the government of Manitoba to build a line from Winnipeg to Pembina on the U.S. frontier. This done, in 1879 the partners formed the St. Paul, Minneapolis & Manitoba Railway company, and by continuing the line from Breckenridge to Pembina united Manitoba with the south and west. In 1880 the government made a contract for building the Canadian Pacific railway with a syndicate of which Stephen was the chief director, and in which Smith, from the first largely interested, came more and more to the front. On Nov. 7, 1885, at Craigellachie in the Rocky mountains, Donald Smith drove the last spike of the first Canadian transcontinental railway.

In 1882 he left parliament, to which he had been elected first in 1871, but returned to it in 1887, and represented Montreal West until 1896, when he was appointed high commissioner for Canada. In that year he was made G.C.M.G.; in 1897 he was raised to the peerage and in 1909 made G.C.V.O. In 1889 he became governor of the Hudson's Bay company.

He died in London on Jan. 21, 1914.

STRATHNAIRN, HUGH HENRY ROSE 1ST BARON (1801–1885), British field marshal, to whose military genius and personal bravery the final suppression of the Indian mutiny (*q.v.*) was largely due, was born on April 6, 1801, at Berlin, Ger., where he was also educated. He entered the army in 1820 and his promotion was rapid. In 1840, now a lieutenant-colonel, he was selected for special service in Syria against Mohammed Ali's Egyptian army, and after distinguishing himself in several engagements, succeeded to the command of the British detachment and in 1841 was appointed British consul general. For several years, amid political complications and intrigues, Rose, by his energy and force of character, did much to arrest the horrors of civil war, to prevent the feuds between the Maronites and Druses from coming to a head, and to administer justice impartially. In 1848 he returned to England but in Jan. 1851 Lord Palmerston appointed him secretary of the embassy at Constantinople. In the following year he was chargé d'affaires in the absence of Sir Stratford Canning during the crisis connected with the "holy places." In the Crimean War he distinguished himself before Sebastopol and at the battles of the Alma and Inkerman.

During the Indian Mutiny Rose was placed in command of the Central India force. Marching from Mhow in Jan. 1858, he captured Rathgarh and defeated the rajah of Banpur near Barodia in the same month. He then relieved Saugor, captured Garhakota and the fort of Barodia, and early in March defeated the rebels in the Madanpur pass, capturing Madanpur and Chanderi. He arrived before Jhansi on March 20, and during its investment defeated a relieving force under Tantia Topi. Jhansi was stormed on April 3 and the fort occupied on the 5th. Kunch was captured early in May. His crowning success was the capture of Kalpi on May 22. When news came of the defection of Sindhia's troops and the occupation of Gwalior by Tantia Topi, Rose moved on Gwalior by forced marches, and on June 16 won the battle of Morar. By June 20 both the city and the fortress of Gwalior had been successfully stormed.

Owing to official jealousy Rose's outstanding services during the mutiny were only tardily recognized. In June 1860 he succeeded Lord Clyde as commander in chief in India. He held the Irish command from 1865 until 1870 and played an important part in the suppression of the Fenian conspiracy. He was promoted general in 1867 and field marshal in 1877. He died at Paris on Oct. 16, 1885, and was buried in the graveyard of the Priory church, Christchurch, Hampshire.

See O. T. Burne, *Clyde and Strathnairn* (1892).

(R. H. V.; C. C. D.)

STRATHSPEY, a Scottish type of slow dance-tune in common (4/4) time, characterized by frequent use of the "Scotch snap"; *i.e.*, a dotted quaver preceded by a semiquaver. Strathspeys appear to have originated *c.* 1700 in the valley (or strath) of the river Spey, probably as tunes for Scottish reels. Tradition in 1790 named the Browns of Kincardine, a noted family of fiddlers, as the originators.

After *c.* 1745 strathspeys came into general use throughout Scotland as tunes for reels and country dances.

Another Scottish dance which was performed to strathspeys was the Strathspey minuet, or Strathspey, a dance for two people similar to a quick minuet. This dance was popular only for a brief period *c.* 1770 but survived as a competition piece until 1832, when it finally disappeared.

See J. F. and T. M. Flett, "Some Early Highland Dancing Competitions," *Aberdeen University Review,* vol. xxxvi (1956). (T. M. FL.)

STRATIFICATION, as a term applied to rocks, refers to their layering. Strata or layers are distinctive features of most sedimentary rocks and are common also in those igneous rocks formed at the earth's surface—lava flows and volcanic fragmental deposits. No connotation of thickness is implied by the term and individual layers or strata may range from a fraction of an inch to many feet. Likewise the shapes of individual strata vary greatly according to the type of rock and the conditions of deposition. They range from thin sheets that cover many square miles to thick, lenslike bodies extending only a few feet laterally.

In sedimentary rocks the term stratum is not synonymous with the terms bed or lamination, although some geologists have used it in that sense. Both bed and lamina carry definite thickness connotations, beds referring to rock layers greater than one centimetre in thickness and laminations to layers that are one centimetre or less in thickness. Thus, both beds and laminations are varieties of strata; likewise bedding and lamination are terms restricted to stratification having definite thickness limitations.

Planes of parting or separation between individual rock layers commonly are referred to as stratification planes. They tend to be even and to be parallel to the strata that they bound, for they are essentially horizontal where sediments are deposited as flat-lying layers and they slope correspondingly where the deposits are laid down with an initial dip. Notable exceptions to the evenness of stratification planes develop where deposition takes place on irregular surfaces. In such situations the bottom surface of a stratum normally conforms roughly to the irregu-

larities of the underlying surface, but the stratification plane above the stratum will tend to be nearly horizontal. In many places original, nearly horizontal surfaces of stratification planes have subsequently been destroyed by erosion during pauses or breaks in sedimentation.

Stratification in sedimentary rocks may result from changes in texture or in composition during deposition of the sediment; it also may result from pauses in deposition that allow the first deposits to undergo certain changes before additional sediments cover them. A sequence of strata, therefore, may appear as alterations of coarse and fine particles, as a series of colour changes resulting from differences in mineral composition or merely as layers of similar aspect, separated by distinct planes of parting. No direct relationship exists between the thickness and extent of strata and the rate of deposition or time represented; for example, a 1-in. stratum of limestone may take longer to form than a 10-ft. stratum of sandstone.

The most frequent cause of stratification is variation in velocity of the depositing agent. During transportation water and wind sort sediments according to size, weight and shape of particles and these sediments settle in layers of relative homogeneity. A second cause is differences in composition of sediments from different sources, illustrated by successive floods spreading over a flood-plain from several tributaries of a river. Still another cause is variation in sediment brought about by change in agents of deposition, as where muds of a delta plain are covered by wind-blown sands, or where limy ooze of the sea floor is buried beneath ash of a volcanic eruption. Thus, variations of any type that affect the uniformity of sediment being deposited result in stratification.

Stratification in sedimentary rocks varies greatly both in degree of prominence and in details of structure. In general, it is best developed in fine-grained sediments and is least apparent and least persistent in coarse-grained materials such as conglomerates. Two important and distinctive structural types are recognized as characteristic of particular environments. These types are: (1) current bedding or cross-stratification in which layers have been deposited at one or more angles to the original dip of the formation; and (2) graded bedding in which particles grade upward in a stratum from relatively coarse at the bottom to relatively fine at the top.

Cross-stratification is the product of deposition by wind or water currents as they lose velocity, whereas graded bedding results from simultaneous settling of coarse and fine sediments in waters of a comparatively quiet bottom.

Cross-stratification, which may consist of medium to thick layers (cross-bedding) or of very thin layers (cross-lamination), includes many variations, the most basic of which are the simple, planar and trough types. Simple cross-stratification involves sets of sloping layers formed by deposition only; planar cross-stratification results from alternating deposition and partial planation of the strata; trough cross-stratification, from deposition together with scouring or channeling. Cross-stratification is developed in a number of different depositional environments including the fronts of deltas, the lee sides of dunes, the channels of rivers and streams and various other places where initial dip is a characteristic feature. Commonly the nature of this environment can be recognized by analysis of the structures.

Graded bedding is typical especially of dark sandstones or graywackes and of conglomerates formed in areas of relatively rapid sedimentation under unstable crustal conditions that characterize the rapid sinking of a geosyncline or depositional trough. Under such conditions thick sequences of graded beds may accumulate. The mixing of coarse and fine particles has been attributed to submarine temblors (earthquakes), to movement by density currents down steep slopes and to other causes.

Seasonal deposits or varves of glacial lakes are another kind of graded bedding. In them light-coloured silt grades upward into dark-coloured mud forming a couplet a few millimetres thick. Each such couplet is the deposit of a single year and is separated from other, similar couplets by an abrupt change in colour.

Stratification in volcanic rocks differs in some respects from that in sedimentary rocks. Fragmental volcanic material, thrown into the air during violent eruptions, becomes stratified in flight under the influences of gravity and wind. Falling to the ground, it may form well-sorted layers. If it falls into lakes or the sea, it becomes layered like any other water-borne detrital matter; however, when a mixture of coarse, volcanic debris full of gas bubbles (pumice) and fine ash falls into quiet water, reverse graded bedding may result, with the finer but heavier grains at the bottom grading up into the coarser, lighter fragments. Stratification also may result from successive outflows of liquid lava or alternations between lava flows and ash falls. A type of false stratification sometimes develops in flows due to contraction on cooling.

Not all sedimentary deposits are stratified. Those transported by ice, either pushed ahead of a glacier or carried by it, form unlayered deposits called till. They are piled up or dumped as the ice melts, forming moraines and other features lacking stratiform structure. Likewise materials accumulated through the force of gravity, such as talus deposits at the bases of cliffs and landslides, may exhibit no sorting and no stratification. Still other deposits are structureless because original stratification has been destroyed through the work of plant roots, or the borings of worms and various other animals, or by recrystallization. Residual soils and reefs built by organisms likewise have no stratification.

Stratification is one of the geologist's most useful tools. Its structure provides a clue to the environment in which the sediment or igneous material accumulated. Furthermore, where once-horizontal layers have been deformed, the record of past earth movements is preserved in the stratification. Folding, faulting and other forms of disturbance commonly are recorded by tilted or broken strata and these make possible the interpretation of geologic events. From this record such practical results as the location of mineral deposits, petroleum fields and ground water are obtained.

See also SEDIMENTARY ROCKS; GEOLOGY. (E. D. MCK.)

STRATIGRAPHY is a branch of geology concerned with the sequence of strata formed by sedimentation from water, wind, ice and volcanic and human activity. From these layers a sequence of events may be deduced and, closely co-operating with petrology and paleontology (*qq.v.*), stratigraphy provides the basis for historical geology, or the history of the earth's surface.

The first to appreciate the concept of stratigraphy was probably Leonardo da Vinci, but it was only during the 19th century that the scientific foundations were laid in England by Sir Roderick Murchison, Adam Sedgwick and William Smith, in Germany by A. G. Werner and F. A. von Quenstedt, in Bohemia by J. Barrande and in North America by Charles Schuchert. By the "law of superposition," *i.e.*, of two layers the upper one must have been deposited on top of the lower and must therefore be younger, an age sequence can be established wherever no complications have occurred as the result of tectonic movements; *e.g.*, upheavals such as faulting, thrusting and folding which can occasionally turn the sequence upside down. In the analysis of such complicated cases the contained fossils are of paramount importance and it was not until it was fully realized that the earth was populated with different forms of life at different times that stratigraphy could progress.

Thus the stratigraphical sequence was developed in the first instance for the last 500,000,000 years, during which there was abundant life and many plants and animals were equipped with hard structures that are readily fossilized.

This sequence was divided into three major eras or groups: the Paleozoic, the Mesozoic and the Cenozoic, each subdivided into periods or systems. The Paleozoic era (divided into Cambrian, Ordovician, Silurian, Devonian, Carboniferous and Permian periods) is known from the evidence of radioactivity to have lasted approximately from 500,000,000 to 190,000,000 years ago; the Upper Carboniferous was the time of the "Coal Measures," when coal was formed in swamp forests. The Mesozoic era (divided into Triassic, Jurassic and Cretaceous periods) lasted from about 190,000,000 to 70,000,000 years ago and was the time of the dinosaurs and ammonites; in the Upper Cretaceous the chalk of Britain

was laid down as an organic calcareous mud in the sea. The Cenozoic era (divided into Tertiary and Quaternary or Pleistocene periods) covered the last 70,000,000 years (Pleistocene only 1,000,000); the Tertiary was the age of mammals, the Quaternary the age of man.

These periods have been further subdivided into epochs too numerous to mention. Many were characterized by zone fossils which lived for only a brief time in many parts of the world; these made long-distance, even intercontinental, correlation possible.

The extension into the Pre-Cambrian part of the earth's history aims at establishing a sequence back to the origin of the earth's crust, covering (known from radioactivity dating) the last 3,000,000,000 to 500,000,000 years. Three eras are distinguished: the Proterozoic (Pre-Cambrian or Algonkian), the Archaean and the Azoic. The Proterozoic has yielded fossils of organized forms of life; the Archaean indirect evidence of life; and the Azoic is the period before life had appeared. The subdivisions of the Pre-Cambrian eras are based on the breaks in the sequences caused by earth movements that were restricted in time.

The most recent development is the application of stratigraphy to archaeology. The earlier part of prehistory, covering Paleolithic man living during the Pleistocene, is connected with the stratigraphy of this epoch, and fossil remains of man and especially human stone artifacts replace fossil plants and animals. When a site is excavated, attention is paid to the details of stratification and climatic sequences are established into which the prehistoric industries can be fitted. In post-Pleistocene times, too, the stratigraphical approach has become important in prehistoric and early historic archaeological research. All sites where the products of early man are found are carefully excavated according to layers, starting from the top. As in geological stratigraphy, the layers are studied by petrological methods to ascertain their composition and the contained fossils are identified and interpreted. The tells or occupation mounds of eastern countries illustrate this method particularly well; *e.g.*, ancient Jericho. These mounds consist of man-made deposits, later generations building their houses on the debris of earlier habitations.

Archaeological stratigraphy has the advantage over the geological in that tectonic disturbance is rare, but higher strata are liable to cut into lower ones and the composition of the strata tends to be much the same throughout the sequence. Consequently, petrological methods are increasingly employed. *See* also GEOCHRONOLOGY; GEOLOGY.

BIBLIOGRAPHY.—C. O. Dunbar, *Historical Geology* (1949); K. M. Kenyon, *Beginning in Archaeology* (1953); J. F. Kirkaldy, *General Principles of Geology* (1954); P. Lake and R. H. Rastall, *Textbook of Geology* (1941); R. E. M. Wheeler, *Archaeology from the Earth* (1954); F. E. Zeuner, *Dating the Past* (1958). (F. E. Z.)

STRATO OF LAMPSACUS (d. *c.* 270 B.C.), Greek philosopher, became head of the Peripatetic school in succession to Theophrastus. As the main subject of his writings, apart from logic, was natural science, he was called Strato Physicus in antiquity.

Strato's theory was a combination of Aristotelian and atomistic elements. Thus he taught with Aristotle that there is only one world, instead of many; but he abandoned his master's teleological interpretation of nature and his hypothesis of a fifth element. The insistence on causality and the materialism that he shared with the atomists led him not only to remove every theological element but also to deny the immortality of the human soul.

Strato's writings as a whole are lost, but in antiquity he was highly esteemed and influential. His doctrine that all substances contain void and that differences in the weight of substances are caused by differences in the extension of void served as the theoretical base for the Hellenistic construction of air- and steam-engines as described in Hero of Alexandria's book.

See the fragments ed. by F. Wehrli, *Straton von Lampsakos* (1950), being part v of *Die Schule des Aristoteles: Texte und Kommentar;* G. Rodier, *La Physique de Straton* (1890). (F. WI.)

STRATOSPHERE, the second layer of the atmosphere, above the troposphere and below the mesosphere. Its lower boundary, the tropopause, is located where the tropospheric temperature decrease of about 6° C. per km. is replaced either by a much smaller decrease or, more often, by constant temperature or slight increase with elevation. Approximate heights and temperatures of the tropopause are shown in the table. These figures are averages along the meridian 80° W. but they are representative for other longitudes and for the southern hemisphere. The tropopause is lower at higher latitudes. It is not a continuous

Heights and Temperatures of the Tropopause

Latitude	90° N.	60° N.	30° N.	0°
January	8.5 km., −60° C.	9 km., −55° C.	12 km., −55° C. 17 km., −70° C.	18 km., −80° C.
July	9 km., −45° C.	10 km., −50° C.	16 km., −70° C.	16 km., −73° C.

surface between the poles and the equator. The polar tropopause extends to about 30° lat. in winter and 35° lat. in summer, while the tropical tropopause can be followed as far as 45° lat. There is thus a multiple tropopause at middle latitudes (indicated by the two values for 30° N. in the table). The polar tropopause is lowest in winter, the tropical tropopause in summer. The temperature in the stratosphere is generally highest in summer, except for one region at about 45° lat. between 14 and 18 km. where the temperature is about 2° C. higher in January than in July. The greatest seasonal warming in the stratosphere occurs at high latitudes because of the length of polar summer days.

The latitudinal temperature variation is smaller in the stratosphere than in the troposphere. At 20 km. elevation, for instance, the January temperature increases from about −70° C. at the equator to −55° C. or higher around 55° N. and then decreases to −75° C. at the north pole. In July, at the same height, the temperature rises continuously from −60° C. at the equator to −40° C. at the north pole. Above approximately 30 km. the temperature increases again because of the absorption of ultraviolet solar radiation by ozone and reaches a maximum of roughly 0° C. in the vicinity of 50 km. This is the level of the stratopause which separates the stratosphere from the next higher layer, the mesosphere, where the temperature decreases with elevation. Variations in space and time of the stratopause are not well known.

Outstanding features of the stratospheric winds in the northern hemisphere at 80° W. are: during January, a strong westerly jet with a maximum of 100 knots centred at the tropopause at 40° lat.; another strong west wind maximum, 80 knots, at 70° lat. and 25 km. altitude; and an easterly current with a maximum of 40 knots at 20° lat. and about 25 km. During the month of July the midlatitude jet at the tropopause has a velocity maximum of only 50 knots, and the higher stratosphere has easterly currents at all latitudes.

The air density at 10 km. is about 30% of the sea level value, at 30 km. only slightly more than 1%. But the composition of the air, especially the relative proportion of the two main constituents, molecular oxygen and nitrogen, is the same as in the troposphere, indicating considerable vertical mixing. Ozone (*q.v.*) is more abundant than below. The water-vapour density is quite low because of the low temperature, and saturation seems to occur only rarely. Between 20 and 30 km. one occasionally observes clouds which are called mother-of-pearl or nacreous clouds because of their characteristic colouring.

The stratosphere is under more radiational control than the troposphere, where horizontal and vertical motions are of greater importance. The characteristic stratospheric temperature distribution is due mainly to the interaction of solar and terrestrial radiation. But strong atmospheric disturbances extend well up into the stratosphere and produce day-to-day changes there just as in the troposphere.

See also ATMOSPHERE; GEOCHEMISTRY: *Geochemistry of the Atmosphere.* (B. Hz.)

STRAUBING, German town in Lower Bavaria, on the Danube, 25 mi. S.E. of Regensburg, by rail to Passau. Pop. (1950) 36,147. Straubing is a town of remote origin, believed to be identical with the Roman station of *Sorbiodurum.* In definite history, however, it is known only as a Bavarian town, and from 1353 to 1425 it was the seat of the ducal line of Bavaria-Straubing.

Its oldest building is the tall square tower dating from 1208. The industries of Straubing are iron and tin founding, tanning, brewing, manufacture of machinery and chemicals, trade in wine and cattle.

STRAUS, a family of U.S. merchants, industrialists and diplomats. Lazarus Straus emigrated to the United States from Rhenish Bavaria in 1852. Two years later he was joined in Georgia by his wife and four children. Lazarus and his son Isidor (1845–1912) formed the partnership of L. Straus and Son which took over the pottery and glassware department of the R. H. Macy and Company department store in New York city. Two sons, Nathan (1848–1931) and Oscar Solomon (1850–1926) joined the firm in 1881. In 1888 Isidor and Nathan became partners in Macy's and in 1896 became its sole owners. Under their direction Macy's developed into one of the world's largest department stores. Isidor and Nathan also established a Brooklyn department store, Abraham and Straus. Isidor served a short term in the U.S. congress as a result of a special election in 1893. Both Isidor and Nathan had many philanthropic interests, Nathan being very active in public welfare work with a particular interest in child health. Isidor and his wife died in the unfortunate sinking of the "Titanic" in 1912.

Oscar Solomon Straus (1850–1926), although he spent some years in the family business, was primarily a lawyer, diplomat and author. He graduated from Columbia law school in 1873. From 1887 to 1889 and again from 1898 to 1900 he was U.S. minister to Turkey. He served as secretary of commerce and labour under Pres. Theodore Roosevelt. In 1909 he returned to Turkey as the first U.S. ambassador to the Ottoman empire. He was appointed chairman of the New York public service commission in 1915. He published several books on U.S. history and in 1922 published an autobiographical work, *Under Four Administrations: From Taft to Cleveland.*

Jesse Isidor Straus (1872–1936), son of Isidor, became an executive of Abraham and Straus, and from 1919 to 1933 served as president of Macy's. During the last few years of his life he was U.S. ambassador to France. Percy Selden Straus (1876–1944), brother of Jesse Isidor, assumed the presidency of Macy's in 1933 and served in that capacity until 1940 when he became chairman of the board. In 1940 Jack Isidor Straus (1900–), the grandson of one of the two Straus family members who were first partners in the Macy firm, and son of Jesse Isidor, became president. Several other Straus family members served as vice-presidents of Macy's.

Roger Williams Straus (1891–), son of Oscar Solomon, departed from the family business just as his father had before him. After graduating from Princeton in 1913 he joined the American Smelting and Refining company. After holding various other positions with the company he became president in 1941 and chairman of the board in 1947. (H. J. Sg.)

STRAUSS, DAVID FRIEDRICH (1808–1874), German theologian and biographer, was born at Ludwigsburg near Stuttgart, on Jan. 27, 1808. After a short interval of teaching he went to Berlin (1831) where he was influenced by F. D. E. Schleiermacher and Hegel's followers. On his return to Tübingen he lectured for a time, but soon found it necessary to give his whole energies to the preparation of his book, the *Leben Jesu.* The work, published in 1835–36, produced an immense sensation since it applied the myth theory to the life of Jesus, denied the historicity of all supernatural elements in the Gospels and interpreted the growth of primitive Christianity in Hegelian terms. In 1840 and the following year he published *Die christliche Glaubenslehre* (2 vol.), the principle of which is that the history of Christian doctrines is their disintegration.

Strauss had been elected to a chair of theology in the university of Zürich but so great was the hostility aroused by his *Leben Jesu* that the authorities pensioned him off before he was installed. With his *Glaubenslehre* Strauss abandoned theology for 20 years, publishing during the period a series of biographical works that secured a permanent place for him in German literature. In 1864, however, he returned to theology with his *Leben Jesu für das deutsche Volk.* His *Christus des Glaubens und der Jesus der Geschichte* (1865) is a severe criticism of Schleiermacher's lectures on the life of Jesus, which were then first published. From 1865 to 1872 Strauss resided in Darmstadt, and in 1870 published his lectures on Voltaire. His last work, *Der alte und der neue Glaube* (1872; Eng. trans. by M. Blind, 1873), caused some consternation among his friends; essentially it rejects Christianity in favour of scientific materialism, and like all his critical works suffered from his lack of critical study of the texts. Strauss died on Feb. 8, 1874.

Strauss's works were published in a collected edition in 12 vol. by E. Zeller (1876–78) without his *Christliche Dogmatik.* His *Ausgewählte Briefe* appeared in 1895.

BIBLIOGRAPHY.—E. Zeller, *David Friedrich Strauss in seinem Leben und seinen Schriften* (1874; Eng. trans., 1874); A. Hausrath, *D. F. Strauss und die Theologie seiner Zeit,* 2 vol. (1876–78); S. Eck, *D. F. Strauss* (1899); K. Harraeus, *D. F. Strauss, sein Leben und seine Schriften* (1901); T. Ziegler, *D. F. Strauss,* 2 vol. (1908–09); A. Schweitzer, *Von Reimarus zu Wrede* (1906), Eng. trans., *The Quest of the Historical Jesus,* pp. 68–120 (1910).

STRAUSS, JOHANN (1804–1849), Austrian violinist, conductor and composer of dance music, was born at Vienna, March 14, 1804. In 1819 he obtained his first engagement as a violist in Joseph Lanner's band, then playing at the Sperl, in the Leopoldstadt, and after acting as deputy conductor in another orchestra, he organized in 1825 a little band of 14 performers on his own account. It was during the carnival of 1826 that Strauss inaugurated a long line of triumphs by introducing his band to the public of Vienna at the Schwan, in the Rossau suburb, where his famous *Täuberlwalzer* (op. 1) at once established his reputation as the best composer of dance music then living. Upon the strength of this success he was invited back to the Sperl, where he accepted an engagement with a larger orchestra for six years.

Later he was appointed bandmaster to the 1st Bürgerregiment and charged with the duty of providing the music for the court balls. In 1833 he began a long series of tours throughout northern Europe, eventually visiting England in 1838 for the coronation of Queen Victoria and playing there at 72 public concerts besides innumerable balls and other private entertainments. On May 5, 1840, he moved with his band from the Sperl to the Imperial Volksgarten, which became the scene of his most memorable successes. He died in Vienna on Sept. 25, 1849.

Strauss was survived by three sons—Johann (1825–99) (*q.v.*), Josef (1827–70) and Eduard (1835–1916), all distinguished as composers of dance music.

STRAUSS, JOHANN, THE YOUNGER (1825–1899), Austrian musician, was born at Vienna, Oct. 25, 1825. He was educated at the Schottengymnasium and the Polytechnic institute in Vienna and then became a bank clerk since his father, Johann Strauss (*q.v.*), wished him to follow some other profession than music. He studied the violin without his father's knowledge, however, and in 1844 openly declared himself a musician by conducting a dance band formed by himself. In 1849 the elder Strauss died and Johann combined his own orchestra with his father's. With this orchestra he traveled all over Europe, including Russia (1865–66) and England (1869), winning great popularity wherever he went. In 1862 he married Henrietta (Jetty) Treffz, a popular singer, and in the same year relinquished the leadership of his orchestra to his two brothers, Josef and Eduard, preferring to spend his time writing music. His first wife died, the second, Angelica Dietrich, was divorced by him, and the third, Adele Deutsch, long survived him and published his letters in 1926.

His most famous single composition is *An der schönen blauen Donau (The Blue Danube)* (1867), but he was the composer of so many other melodious and highly successful waltzes, including most notably *Morgenblätter* (1864), *Künstlerleben* (1867), *Geschichten aus dem Wienerwald* (1868), *Wein, Weib und Gesang* (1869), *Wiener Blut* (1871) and *Kaiserwalzer* (1888), that he won the title of "the Waltz King." His waltzes number well over 150 (each with a number of sections) and he wrote more than 300 other dances.

Of Strauss's operettas, *Die Fledermaus* (1874) is the best known, but he wrote 15 others, including *Der Karneval in Rom* (1873); *Cagliostro in Wien* (1875); *Das Spitzentuch der Königin*

(1880); *Der lustige Krieg* (1881); *Eine Nacht in Venedig* (1883); *Der Zigeunerbaron* (1885); and *Waldmeister* (1895). He died in Vienna, June 3, 1899.

STRAUSS, JOSEPH BAERMANN (1870–1938), U.S. bridge engineer, was born in Cincinnati, O., Jan. 9, 1870. He received the degree of civil engineer from the University of Cincinnati in 1892. After several years of experience as a draftsman and designer he was appointed principal assistant engineer in the office of Ralph Modjeski in Chicago (1899–1902). In 1902 he entered private practice and in 1904 organized the Strauss Engineering corporation. He made a specialty of movable and long-span bridges. He originated the Strauss trunnion bascule bridge and the Strauss lift bridge. He designed and supervised construction of more than 500 bridges in the United States, Canada, Europe and Asia. Strauss was the engineer for the Columbia river bridge, Longview, Wash. He was a consultant for the Arlington Memorial bridge over the Potomac river at Washington, D.C.; the arch bridge at Bayonne, N.J.; and the George Washington suspension bridge across the Hudson river at 179th St., New York city. His greatest work was as chief engineer of the Golden Gate bridge at San Francisco. Strauss died in Los Angeles, Calif., on May 16, 1938. (S. C. Hr.)

STRAUSS, RICHARD (1864–1949), German composer, was born at Munich, Ger., on June 11, 1864, the son of Franz Strauss, an eminent horn player. To some extent a prodigy, Strauss was something of a pianist at four, a composer at six, and at ten he was already seriously studying music under F. W. Meyer, Munich court conductor. Soon the result of this study began to make itself apparent. Singers sang Strauss's songs; the Benno Walter quartet played his String Quartet, A major (op. 2); Hermann Levi performed his Symphony in D minor; and Hans von Bülow took the composer under his tutelage and introduced his early *Serenade* for wind instruments to the Meiningen public. For obvious reasons Strauss had not yet found himself. He had passed through the gymnasium and the university, and his music studies had been thorough, but all this had made of him merely an excellent technical musician, who in his Eight Songs (op. 10) and in his Quartet, C minor (op. 13) showed how strongly he was influenced by predecessors, Liszt on the one hand, Mendelssohn and Brahms on the other. It was to Bülow, and even more to Alexander Ritter, that Strauss owed the awakening of his interest in the modern development of music.

In 1885 Strauss succeeded Bülow as conductor of the Meiningen orchestra, but the appointment was held only for a few months, since in April 1886, Strauss resigned his post in order to travel in Italy, and on his return in the early autumn he became third conductor of the Munich opera under Hermann Levi. Four years later he was installed at Weimar as court conductor, but once again he held his post for only a brief period, for in 1894, the year of his marriage to Pauline de Ahna, an accomplished singer, he was promoted to be first conductor at Munich. Between these various appointments and that of court conductor in Berlin (1898) Strauss travelled considerably in the near east and over Europe. He first visited London in 1897 and in 1910 conducted a performance of *Elektra* at Covent Garden.

It has often been said that signs of the real Strauss are not to be perceived before his *Don Juan* (op. 20) and *Macbeth* (op. 23). A year only divided *Macbeth* (1890) from *Don Juan* (1889). The former is comparatively conventional in a Lisztian manner, but *Tod und Verklärung* (1891) was thought a sensational work. Between it and *Till Eulenspiegels lustige Streiche* (1895), Strauss's first opera, *Guntram*, finds place (first performance, Weimar, 1894), a work in the Wagnerian manner which failed to hold a position upon the stage.

In *Till Eulenspiegel*, a brilliantly clever score, is to be found a great sense of fun, and it is perhaps worth recording that no more noteworthy example of the rondo form exists in modern music, while its approximate successor, *Don Quixote* (1898), is an equally outstanding example of the variation form. In *Don Quixote* (especially the incidents of the sheep and the windmills) Strauss's zenith as a musical realist was reached. In between came another symphonic poem on the most ambitious scale, inspired by and named after Friedrich Nietzsche's *Also sprach Zarathustra* (1896), which stirred up more temporary strife than any of its predecessors, if not so much perhaps as was aroused later by the production of *Ein Heldenleben* (1899) or by the comparatively ingenuous *Sinfonia domestica* (1904), both autobiographical works. The latter marked a decline in Strauss's orchestral music toward the *Alpensinfonie* (1915).

Up to 1910 Strauss had composed four operas. After *Guntram* came *Feuersnot* (1901), which despite the fact that it is largely in the nature of a satirical skit at the expense of Munich and its citizens, remained sufficiently alive to merit performance at His Majesty's theatre, London, in June 1910. At Covent Garden in March 1910 *Elektra* was played to crowded houses and aroused great discussion by what was then regarded as its unbridled violence and sensationalism. Later in December came the not less novel and startling setting of Oscar Wilde's *Salome*, which had been first produced at Dresden in 1905.

A still more important work than any of these was the bitter-sweet *Der Rosenkavalier*, which was first brought out at Dresden in 1911. Later came *Ariadne auf Naxos*, forming part originally of a setting of Molière's *Le Bourgeois gentilhomme* (1912) with incidental music but subsequently rearranged as an independent work (1916). Other important works included: the dramatic ballet *Josephslegende* (*The Legend of Joseph*) (1914); *Die Frau ohne Schatten*, first heard at Vienna in 1919; an autobiographical "comic play-opera" *Intermezzo* (with text by the composer, 1924); a ballet-pantomime, *Schlagobers*, produced at Vienna in 1924; and the operas *Die ägyptische Helena* (*Helen of Egypt*) (1928), *Arabella* (1933), *Die schweigsame Frau* (1935), written to Stefan Zweig's libretto of Ben Jonson's *Epicoene*, or *The Silent Woman*, *Friedenstag* (1936), *Daphne* (1937), *Die Liebe der Danae* (1940) and *Capriccio* (1941). His *Olympische Hymne* (*Olympic Hymn*) was a feature of the Berlin Olympiad (1936).

In addition to his operas and orchestral works (which in his last years tended toward the chamber orchestra) Strauss wrote more than 100 songs, not a few of which rank him a true descendant of the line of German song writers.

During World War II Strauss was absent from public view and was thereafter a controversial figure, having taken no stand on the political issues of naziism. In 1947 he resumed the baton at the Strauss festival arranged in his honour in London, where he was warmly received. In 1948 he was officially cleared of all charges by a denazification court. He died at Garmisch-Partenkirchen, Bavaria, Sept. 8, 1949.

Bibliography.—W. Mann, *Strauss* in "Master Musicians" series (1954); E. M. von Asow, *Richard Strauss: thematisches Verzeichnis* (1948); K. Pfister, *Richard Strauss: Weg, Gestalt, Denkmal* (1949); W. Schuh, *Über Opern von Strauss* (1947); R. Strauss and H. von Hofmannsthal, *Briefwechsel* (1952). (E. W. Bm.; X.)

STRAVINSKY, IGOR FEDOROVICH (1882–), Russian composer, who later took French and then U.S. nationality, by general consent one of the great masters of the 20th century, was born at Oranienbaum, near St. Petersburg (Leningrad), June 5 (old style), 1882. His father, Feodor Stravinsky, was a distinguished operatic bass singer. Stravinsky first studied law, turning to music only upon the advice of Rimsky-Korsakov, to whom he was introduced by a fellow law student.

Rimsky-Korsakov, who became his teacher, bequeathed to Stravinsky the legacy of a brilliant technique but his immediate influence was temporary. His technique is evident in the gaudy orchestral panoply of *The Firebird* (1910), in *Petrushka* (1911) and in *The Rite of Spring* (1913), three ballets commissioned by S. Diaghilev (*q.v.*). But that these masterly works, still probably the composer's most popular, were a beginning and not an end is shown by the case of the opera *The Nightingale* (1914). The first act was written before *The Firebird*, in Stravinsky's lushest Rimsky-Korsakov manner; the second came after the three ballets, and in its clearer, leaner lines indicated that Stravinsky was more vividly aware than his admirers that the stunning success of *The Rite* could not be repeated. The score of *The Rite*, excelling even *Petrushka* in its novelty of idiom and far outstripping anything ever written in its violence and complexity, was the cause of a famous riot at its *première* on May 29, 1913, in Paris—a backhand compliment to what was later regarded as one of the greatest peaks of 20th-century music.

When World War I broke out, Stravinsky settled in Switzerland, and there produced three works whose effect has been hardly less than that of the preceding three. *The Soldier's Tale* (1918) is a haunting setting of a grim little story for only a handful of instruments with a narrator and dancers, and shows the marvelous economy of style that is a truer characteristic of Stravinsky than the blinding exuberances of *The Rite*. "One must be stingy with music," he told a young composer; and it is the precision of his seemingly limitless capacity for invention that is the hallmark of his music. *Renard* (1917) is a return to Russian themes in this new manner, as is *The Wedding* (*Les Noces*) (1923), a ballet-cantata of astonishing and reverberating originality.

Between World Wars I and II, Stravinsky established himself as a kind of *elegantiae arbiter*, the brilliance of his musical personality and the assurance of his novel art acting as a focus and

stimulus for those in the Parisian circles he inhabited. On the other hand, he himself was more than ever dependent on outside stimulus to provide him with the materials for composition; *e.g.*, the music of G. B. Pergolesi for his ballet, *Pulcinella,* (1919); and from this derives the succession of different manners (not "styles": Stravinsky is always unmistakably Stravinsky) that has provided his opponents with ammunition. The essential point is not that he has himself been influenced so much as that he has taken what he needs and used it for works of such originality and power that they then become the influence on his contemporaries.

The variety of manners has been fascinating. *Mavra* (1922) is a Russo-Italian comic opera of minor interest; but with the opera-oratorio *Oedipus Rex* (1927, revised 1948) there came a classical masterpiece, transcending fashionable trappings by the monumental power and aloof dignity of the musical invention. The ballet *Apollo Musagetes* (1928) is still more classical of mien, its chaste beauty earning the composer not for the first time the charge of coldness. That he is an anti-romantic is indeed true, as the many concert works in a manner dubbed "neoclassical" testify; but objectivity need not imply chill of heart, and the enduring spell that Stravinsky's music of all periods holds over music-lovers is the best witness to his inventive genius of mind. One may point to passages in many of his works where a tender, if unsentimental, heart is the mind's motivation. One example is the ballet-mime *Persephone* (1934), a work of deep poetry and compassion.

The beginning of the 1930s brought to a close Stravinsky's "neoclassical" period, of which the finest concert works are perhaps the *Symphonies of Wind Instruments* (1920), the rather difficult piano concerto (1924), the piano serenade (1925) and the graceful, gay *Capriccio* for piano and orchestra (1929). The dawn of a new phase of creativity was signaled with one of his masterpieces, the *Symphony of Psalms* (1930, revised 1948), a choral work of elevated, humbling inspiration that points to Stravinsky's increasing fascination with pre-classical art.

With the outbreak of World War II, Stravinsky settled—and found new stimulus—in the United States. Two magnificent symphonies ensued, one in C major (1940), which goes a good way to countering the persistent charge of melodic poverty, and the *Symphony in Three Movements* (1945), which concerns itself more with massive blocks of orchestral sound. The beautiful ballet *Orpheus* (1947) seemed to hark back to earlier glories; but the arresting, deeply devotional *Mass* (1948) showed that Stravinsky's questing mind was moving still forward, not backward. His witty, tender opera, *The Rake's Progress* (1951), with libretto by W. H. Auden, certainly quarries the past (including Stravinsky's own); but the postwar rediscovery of the small, remote, impeccably fashioned art of Anton Webern (*q.v.*) found Stravinsky once more the undisputed leader of a new trend in music. The septet (1953) already shows signs of an interest in serial technique; and Stravinsky's fascination for Webern subsequently found expression in a *Canticum Sacrum* (1955) for St. Mark's, Venice; a bright, entertaining ballet, *Agon* (1957); in *Threni* (1958), a religious meditation from Jeremiah as devout as the *Mass;* and the enigmatic *Movements* for piano and orchestra (1960). The variety of these four works testifies yet again to the range of Stravinsky's invention within whatever idiom he adopts, while their unfailing mastery of the sheer stuff of music and, especially in *Threni,* their concentration of feeling make irrelevant any charge of parasitism. The cantata *A Sermon, a Narrative and a Prayer* (1962) and the television opera *The Flood* (1962) mark an advance in melodic power in the new phase from *Movements* and a return to the dramatic and religious intensity which was always the deepest force in his art. Stravinsky has been accused of opportunism, of having declined from his early talent, of inventive poverty (especially as regards melody) and coldness, but no one, however hostile, denies him skill, an incredible rhythmic sense, the capacity to arouse a vigorous response to whatever he does. This unique figure is one of the great masters of music in the long European tradition, straddling past and present with his creative interest in the music of bygone ages, a hope and inspiration for the future.

BIBLIOGRAPHY.—Stravinsky's own writings include an autobiography, *Chronicle of My Life* (Eng. trans., 1936); *The Poetics of Music* (1947); and three books in the form of "conversations" with Robert Craft (1953, 1958 and 1959); E. W. White, *Stravinsky's Sacrifice to Apollo* (1930) and *Stravinsky* (1947); R. Vlad, *Stravinsky* (1960).

(J. WK.)

STRAW. Straw is the generic name applied to the dried stems or the stalks of certain cereals, such as wheat, barley, oats and rye. The first use of straw was probably as bedding or for floor coverings for mankind or as fodder and litter for cattle. The value of straw for cattle was early recognized, for in making conditions of land tenure, it was, and still is, customary to prohibit the removal of the straw residue which, as litter, became the only fertilizing agent. Thatching for primitive shelters and body coverings for man were some of the succeeding utilitarian developments of straw.

Straw is now the raw material of many important industries, viz., mat-making; the stuffing of bedding; the weaving of hats, plaits and baskets; twisting into ropelike bands for tying up sheaves and trusses of corn and hay and for rolling into "skeps" or beehives; for various methods of ornamentation; for the making of pulp for paper and the species of cardboard known as strawboard and the making of "envelopes" for the protection during transit of glass and earthenware.

STRAWBERRY, the well-known fruit plant of about eight species of the genus *Fragaria.* They are native to the temperate regions of the old and new world, but one species is also found in southern Chile. The common cultivated varieties are derived from the meadow strawberry (*Fragaria virginiana*) of eastern North America crossed with the beach strawberry (*F. chiloensis*), native from Alaska to central California, the coast and Andes mountains of southern Chile and the mountains of the Hawaiian Islands. Several varieties, hybrids of cultivated varieties with *F. ovalis* native to western North America from New Mexico to the Pacific, north to Alaska, have been introduced. *F. vesca,* a fourth species, is circumpolar and in the mountains of the West Indies, Mexico and South America. It is a fragile woodland species, lacking the vigour of garden strawberries. The "alpine" varieties with small aromatic fruit are everbearing selections of *F. vesca,* originally from the mountains of northern Italy. In several varieties it is fairly common in European gardens but rare in America. The alpine varieties come true to seed in contrast to the great diversity of the seedlings of ordinary garden varieties. A few varieties of *F. moschata,* a fifth species native to central Europe, are also raised in European gardens. Its berries are small, vinous in flavour and the most aromatic of all. Allied to *F. vesca* are three other rather distinct species, *F. nilgerrensis* of southern Asia, the similar *F. hayatai* of Formosa, and *F. viridis* of central Europe, all having small and unpromising fruit. The so-called Indian strawberry naturalized throughout eastern North America, having yellow flowers and succulent insipid berries, is not a true strawberry but belongs in an allied genus *Duchesnea.* The strawberry-raspberry is not a hybrid but a raspberry (*Rubus illicebrosus*) of eastern Asia.

JOHN MARKHAM

WILD STRAWBERRY (FRAGARIA VESCA) IN FLOWER

Before the white man reached America the Indians of Chile cultivated strawberries, bearing fruit of exceptional size—"as large as a walnut." The berries were pale red with firm, meaty, almost white flesh and a delicate aroma. In 1714 M. Frezier took to Europe five Chilean strawberry plants. Plants of the North American meadow strawberry were already growing in European gardens, and from crosses of these two the modern strawberry evolved.

In the United States about half the crop is ordinarily processed as frozen food. The plant is grown in all the states and in most provinces of Canada on a wide range of soil types, from sand to

clay. In all except sections near the Gulf and Pacific coasts the plants are usually set in early spring two to three feet apart in rows three and one-half feet distant and fruit the following year. Near the Gulf coast plants set in the fall bear their fruit during the following winter or in early spring. On the Pacific coast plants may be set from early to late winter.

During the summer runners grow from the original plants and form new plants at their tips. By fall a matted bed of plants will have formed. Flower buds form in the crowns of the plants in the fall. These develop into flowers and berries the following spring. Because of the warmer winter and the short days from eastern North Carolina west to Louisiana and southern Texas, flower buds form all winter and into the spring, resulting in a second crop of fruit following the first crop. Plants formed late in the fall or crowded in dense rows so that they have only two or three leaves may bear five to ten berries in the spring. Well-spaced plants formed early in the summer with 20 or more leaves may bear as much as a quart the following spring. To obtain such large plants the original mother plants set 2 ft. apart in the row may form 5 runner plants spaced about 12 in. apart or 8 runner plants spaced 8 in. apart. All runners, after these have rooted, are cut off. The larger the plant in the fall the greater the crop the next spring, and runner restriction, use of fertilizers, irrigation or any other cultural practice to increase the size of the plant in the fall will increase the next year's crop.

The strawberry is only a half-hardy plant and may be badly injured by temperatures of 10° to 15° F. in the fall and winter. For this reason the plants are covered with one to six inches of wheat straw or other mulching material before temperatures as low as 15° F. occur. One inch of straw is sufficient from Maryland to Kansas, two inches from New England to Illinois, and three to six inches in Wisconsin, Iowa, Minnesota and northward. Sprinkler irrigation for frost protection is becoming important at flowering time.

Less than 20 varieties constitute 95% of the acreage in the U.S. Important varieties are Missionary and Florida Ninety in Florida; Klonmore in Louisiana; Blakemore from Maryland to Oklahoma south to 100 mi. from the Gulf of Mexico; Armore and Tennessee Beauty in Missouri, Kentucky, southern Illinois and Indiana; Fairfax, Dixieland, and Pocahontas from the eastern shore of Virginia to southern New Jersey; Sparkle, Catskill, Robinson, and Howard 17 (Premier) from northern New Jersey to Canada and west to Minnesota; Dunlap and Beacon in Wisconsin, Iowa and Minnesota; Marshall and Northwest in Oregon and Washington; and Shasta and Lassen in California.

J. HORACE MCFARLAND CO.

CULTIVATED VARIETY OF STRAWBERRY (FRAGARIA)

Everbearing varieties differ from others in forming flower buds in the long days of summer at relatively high temperatures. They are adapted to northern sections with cool summers, as from New England to Iowa and Minnesota and northward. Red Rich, Rockhill and Gem are good everbearing varieties for the U.S. and Canada.

Kent county is the largest producer in England, with nearly one-third of the entire acreage. Hampshire, Cambridge and Norfolk counties also have large acreages. Plants are set in the spring 12 to 15 in. apart in the rows with the rows 30 in. distant. Talisman is best on heavy soil, Royal Sovereign on lighter soil. The fields may be kept until the fourth year but not usually more than that long.

Serious diseases are the leaf spots, red stele root disease, spring and summer dwarf (due to nemas), fruit rots and virus diseases. Serious insects are the bud weevil, root weevils, crown borer, cyclamen mite, leaf rollers, white grubs and the red-spider mites.

BIBLIOGRAPHY.—U.S. Department of Agriculture, "Strawberry Improvement" in U.S. Department of Agriculture *Yearbook* (1937), "Strawberry Varieties in the U.S.," *Farmers' Bulletin 1043* (1958), *Farmers' Bulletin 1026,* strawberry culture in the southern states (1958), *1027,* western states (1947) and *1028,* eastern states (1944).

(G. M. D.)

STRAWBERRY TREE (*Arbutus unedo*), a beautiful ornamental tree of the heath family (Ericaceae, *q.v.*), native to southern Europe and also to southwestern Ireland (Kerry), and widely planted in warm countries. It grows from 10 to 30 ft. high and has smooth red bark, which peels in thin plates, and oblong, glossy, dark-green leaves. The white or rose-coloured flowers, arranged in short, drooping clusters, are borne in autumn intermingled with the scarlet granular berries, about three quarters of an inch across, which ripen only the second season after flowering. *See* MADROÑA.

STRAWBOARD: *see* CARDBOARD.

STREATHAM, a residential and shopping district in Wandsworth (*q.v.*), London, Eng. Streatham common (67.75 ac.) rises to about 250 ft.; there are several sports grounds in the area. Streatham place, a house on the south side of Tooting Bec common, pulled down in 1863, belonged to the brewer Henry Thrale (d. 1781). For many years Dr. Johnson (*see* JOHNSON, SAMUEL) spent long periods there; other visitors included Boswell, Fanny Burney, David Garrick, Sir Joshua Reynolds and Edmund Burke.

STREET, GEORGE EDMUND (1824–1881), English architect noted for his many Gothic churches, was born at Woodford in Essex on June 20, 1824, and died on Dec. 18, 1881. He worked for five years with the architect Sir George Gilbert Scott in London and in 1849 began to practise on his own account. He was a draftsman of a very high order. In 1855 he published *The Brick and Marble Architecture of Northern Italy,* and in 1865 *The Gothic Architecture of Spain,* with beautiful drawings of his own.

Street's personal taste led him in most cases to select for his design the 13th-century Gothic of England or France, his knowledge of which was very great. The majority of his buildings were for ecclesiastical uses, the chief being the convent of East Grinstead, the theological college at Cuddesden and a large number of churches, such as St. Philip and St. James's at Oxford, St. John's at Torquay, All Saints' at Clifton, St. Saviour's at Eastbourne, St. Margaret's at Liverpool and St. Mary Magdalene, Paddington. His largest works were the nave of Bristol cathedral, the choir of the cathedral of Christ Church in Dublin and the new courts of justice in London. The competition for the latter was prolonged, and the building was not completed at the time of his death. Street was professor of architecture at the Royal Academy, where he lectured on medieval architecture. He was buried in the nave of Westminster abbey.

See A. E. Street, *Memoir of G. E. Street, R.A.* (1888).

STREETCAR: *see* ELECTRIC TRACTION.

STREET CRIES. The itinerant vendor has always adopted a distinctive cry to draw attention to his wares. The earliest record possessed in English is that given in the *Oxford English Dictionary* of a cry of 1393: "Kokes and here knaues crieden hote pyes, hote!"—a cry which with but slight variation was preserved to the close of the 18th century. The earliest collection of street cries is to be found in the 14th-century poem *London Lackpenny.* The author of this poem is, on John Stow's authority, John Lydgate (1370–1450), a Benedictine monk of the abbey of Bury St. Edmunds. Among the many cries included are those of "Hot peascods," "Strawberries ripe and cherries in the rise." *Pammelia: Musicke's Miscellanie* was published in 1600–18 and contains a number of cries, but the most important collection of the time is P. Tempest's *The Cryes of Old London* (1668). In Smollett's *Humphrey Clinker* the author protests against "noisy rustics bellowing 'green peas,' " and Herrick long before had introduced stray street cries in his *Hesperides.* The cries of Paris are of very remote origin and *Le Livre des Mestiers* contains examples of the time of Saint Louis; Guillaume de la Villeneuve compiled a famous collection known as *Les Crieries de Paris.*

BIBLIOGRAPHY.—Charles Hindley, *A History of the Cries of London* (1881); Sir Frederick Bridge, *The Old Cryes of London* (1921);

Victor Fournel, "Cries of London," in Grove's *Dictionary of Music; Les Cris de Paris* (1887). (H. L. Mo.)

STREETER, BURNETT HILLMAN (1874–1937), English theologian and biblical scholar, best remembered for his original contributions to knowledge of Gospel origins, was born Nov. 17, 1874, at Croydon. Most of his life was spent at Queen's college, Oxford. He wrote or contributed to a dozen volumes in the fields of philosophy of religion, comparative religion and New Testament studies.

While Streeter regarded his work as primarily philosophical—the contemporary interpretation of religion—he is more widely known as a student of the New Testament. He contributed an important essay to the *Oxford Studies in the Synoptic Problem* (1911), but his most important work was *The Four Gospels: A Study of Origins* (1924). In this he originated a "four document hypothesis" (including a Proto-Luke) as a solution to the synoptic problem, and developed the theory of "local texts" in the manuscript transmission of the New Testament. His section on the Caesarean text was very influential in subsequent studies in this field. This work was followed in 1929 by the *Primitive Church*, in which he argued that there were three (not one) systems of church government in the earliest Christian churches.

Streeter followed "the hue and cry after new discovery" into fields as diverse as *Foundations: A Statement of Christian Belief in Terms of Modern Thought, by Seven Oxford Men* (1912), *The Chained Library* (1931), *Reality: A New Correlation of Science and Religion* (1926) and *The Buddha and the Christ* (Bampton Lectures, 1932). He also was an enthusiastic follower of Frank Buchman and the Oxford Group. Streeter and his wife were killed in an airplane crash in Switzerland on Sept. 10, 1937.

(E. C. Co.)

STRELTSY (plural, from singular *strelets* "musketeer"), a Russian military corps that came into being in the middle of the 16th century and for about 100 years made up the bulk of the Russian Army. By the middle of the 17th century, the *streltsy* had developed into a hereditary military caste. In peacetime the Moscow *streltsy* provided the tsar's bodyguard and performed police and security duties. The municipal *streltsy* had a similar function in the border towns that they garrisoned. In the second half of the 17th century the government adopted a policy of substituting landholdings for the *streltsy's* payment in money and grain, and this was one of the causes of their discontent and consequent unreliability. The poorer *streltsy* more than once sided with the rioters or rebels whom they were supposed to quell. In spite of this, their number grew from about 40,000 in 1632 to 55,000 in 1681, and 22,500 out of these were stationed in Moscow. The *streltsy* lived in their own settlements (*slobody*) and were organized in regiments administered by the *streletski prikaz* ("office") in Moscow.

In 1682 the revolt of the *streltsy* against the family of Naryshkin (*q.v.*), which had been fomented by the rival family of Miloslavski, was skilfully exploited by the regent Sophia (*q.v.*) Alekseevna and by F. L. Shaklovity, whom Sophia had appointed chief of the *streletski prikaz* after the execution of the unreliable I. A. Khovanski. When in June 1698 the *streltsy* attempted to restore Sophia to power, they were defeated by Tsar Peter I's foreign mercenaries under Gen. Patrick Gordon (*q.v.*), while Peter himself was abroad (*see* PETER I the Great). Returning to Moscow early in September, Peter dealt ruthlessly with the *streltsy:* in the following six weeks 799 were either broken on the wheel or impaled, while 222 were deported. Though in 1702, in the first period of the Northern War, Peter saw fit to revive the *streltsy*, the Moscow *streltsy* were finally disbanded in 1713.

See R. Brix, *Geschichte der alten Russischen Heeres-Einrichtungen* . . . (1867); A. V. Chernov, *Vooruzhennye sily russkogo gosudarstva v XV–XVII vv.* (1954). (L. R. LR.)

STREPTOMYCIN is an antibiotic substance produced in cultures of *Streptomyces griseus* (*Actinomyces griseus*). It was first described by A. Schatz, E. Bugie and S. A. Waksman in 1944. Of greatest use in the treatment of tuberculosis, streptomycin inhibits the multiplication of susceptible bacteria or kills those bacteria, depending upon the concentration of the drug achieved in the blood stream of the patient. The drug has only a local reaction in the gastrointestinal tract if ingested. For systemic action it is given by injection into muscle tissue, from where it is rapidly absorbed into the blood stream. Tubercle bacilli can become resistant to streptomycin during a single genetic mutation. Other drugs may be given concurrently to retard this phenomenon.

Streptomycin may cause vertigo (dizziness) or loss of hearing or both after prolonged treatment.

Although the chemical formulas are known, the drug is not synthesized. It is produced by growing the mold in suitable media and isolating the antibiotic by physicochemical methods.

A closely related antibiotic, dihydrostreptomycin, is produced by the chemical addition of hydrogen to streptomycin. Its antibiotic action and toxic effects are very similar to those of streptomycin. *See* also ANTIBIOTICS.

See monograph in American Medical Association, *New and Nonofficial Remedies* (1952). (WR. H. B.)

STRESA, a village of Piemonte, Italy, in the province of Novara, situated on the west side of Lago Maggiore, on the Simplon railway, 10 mi. N. of Arona, 673 ft. above sea level. Pop. (1961) 3,169 (commune). It is noted for the beauty of its scenery and its fine villas and is a health resort.

STRESEMANN, GUSTAV (1878–1929), German statesman, was born in Berlin on May 10, 1878. He was educated at the Berlin gymnasium and at the universities of Berlin and Leipzig where he studied philosophy and political economy. A liberal and an ardent patriot he drew his political inspiration from Baron H. F. K. vom und zum Stein, from Prince K. A. von Hardenberg and Rudolf von Bennigsen. He was a lover of art and literature. These early interests and ideals he maintained throughout his life. At the age of 28 he entered the Reichstag. On the death of Ernst Bassermann in 1917 he became leader of the National Liberal party.

During World War I he urged a more vigorous mobilization of the national effort, and he helped to bring about the fall of T. von Bethmann-Hollweg whom he accused of weakness and indecision. After the revolution he tried to form a united Liberal party. When this failed he became leader of the new German People's party which represented the right wing of the old National Liberals. Although he was himself a monarchist he fully accepted the republican constitution of Weimar. He had always been in favour of responsible parliamentary government. He hoped that his party would form a bridge between the old and the new Germany. He condemned every kind of intransigent opposition and afterward made repeated efforts to induce the German National party to take part in the responsibility of government. In 1919 he voted against ratification of the treaty of Versailles which he believed to be unjust and unworkable. But once the decision was taken he thought that Germany must accept the consequences; it was powerless and disarmed; recovery could be achieved only if the treaty were revised by agreement. The Franco-Belgian occupation of the Ruhr in 1923 was, he held, a crass infringement of the treaty intended to bring about the economic ruin and dismemberment of Germany. He was at first a strong supporter of passive resistance.

In Aug. 1923 Stresemann became chancellor and foreign minister at the head of a "grand coalition" of all parties from the Social Democrats to the German People's party. The separatists, who were encouraged by the French in the Rhineland, the Communists in Saxony and Thuringia and the right-wing extremists in Bavaria threatened the unity of the reich. The collapse of the mark made normal economic life impossible. Stresemann therefore decided to end the passive resistance. The reform of the currency was achieved by Hans Luther, and by Hjalmar Schacht, who was Stresemann's personal choice. The reichswehr, under emergency decree, took strong measures against the Communist-controlled government of Saxony and, after the Hitler-Ludendorff *Putsch*, they restored order in Bavaria. The Social Democrats resigned because they considered the action against Saxony unconstitutional. Stresemann tried to start negotiations with France but his overtures were rejected by Raymond Poincaré. The cabinet fell in December.

Stresemann remained foreign minister. He served in a series of

coalitions under Wilhelm Marx (centre), Luther (nonparty) and Hermann Müller (Social Democrat). At the London conference in 1924 he accepted the report of the Dawes committee on reparations. Germany received an international loan, its finances were reformed and a provisional scale of payments fixed according to its capacity to pay. There followed a period of rapid though rather spurious economic recovery. A year later the French and Belgian troops left the Ruhr. The Dawes plan was carried with the support of a section of the German National party, but as time went on Stresemann's policy of fulfillment won more support from the left than from the right.

In 1925 Stresemann, encouraged by the British ambassador Lord D'Abernon, took the initiative in seeking a political settlement between Germany, France and Great Britain. These negotiations culminated in the Locarno pact (*q.v.*). Stresemann had become convinced that there was no real conflict of interests between France and Germany and that, once the French demand for security had been satisfied, fruitful co-operation in political and economic matters would be possible. In face of great difficulties he maintained his belief that Aristide Briand would loyally support this policy. In 1926, after Germany's entry into the League of Nations, he held private conversations with Briand at Thoiry. They discussed the possibility of the evacuation of the Rhineland and the return of the Saar and of German financial assistance to France.

Stresemann's hopes for revision of the territorial clauses of the treaty of Versailles remained unfulfilled. In 1929 he accepted the Young plan which was intended to be a final settlement of reparations. The Allies agreed to evacuate the Rhineland in 1930. Stresemann carried his policy in the Reichstag, in spite of strong opposition from the Nationalists and Communists. Schacht, who had been the chief German expert on the Young committee, turned against him on the grounds that he had won meagre concessions in return for financial obligations which could not be fulfilled.

In 1926 Stresemann signed a treaty of neutrality and friendship with the U.S.S.R. He was anxious that Germany's entry into the League of Nations should not lead to a break with Soviet Russia. In the published sources there is no indication that he favoured an eastern in preference to a western policy. Nor is there any evidence that he ever envisaged a departure from the methods of peaceful negotiation. He came to identify the interests of Germany with the interests of Europe as a whole. He was awarded the Nobel Peace prize in 1926. He died in Berlin on Oct. 3, 1929.

See *Gustav Stresemann, His Diaries, Letters, and Papers,* ed. and tr. by Eric Sutton, 3 vol. (London, 1935–40); Rochus von Rheinbaben, *Stresemann: the Man and the Statesman* (1929). (A. H.-M.)

STRETFORD, a municipal (1933) and parliamentary borough of Lancashire, Eng., adjoining Salford and Manchester to the northeast. Pop. (1961) 60,364. Area 5.5 sq.mi. It contains a considerable part of the Trafford Park industrial estate, between the Bridgewater canal and the Manchester ship canal (*see* MANCHESTER). Old Trafford, the Lancashire County Cricket club's ground, is in the town. Longford hall houses the town's art collection and Longford park is the biggest of Stretford's five parks. There is a civic theatre. The two manors of Stretford and Trafford have, since about 1260, descended in the family of de Trafford. The name of the town refers to the crossing of a ford by the Roman road (*strata*) from Chester to Manchester.

STREUVELS, STYN (1871–), pen name of Frank Lateur, Flemish writer, and the nephew of the priest-poet Guido Gezelle. Essentially a man of the soil, he nevertheless possessed a delicate artistic sensibility and acquired considerable literary culture. He was the most popular writer of his time in Flanders and enjoyed a great reputation in Holland, although he frequently introduced the Western Flemish dialect into his writings. Many of his novels were translated. *De Vlaschaard* (*The Flax Field,* 1907) is considered one of his best works.

STRICKLAND, AGNES (1796–1874), English historical writer, the author of a number of biographies of British royalty, was born in 1796 in London. Her early literary works were historical romances in verse in the style of Sir Walter Scott, followed by prose histories. She then turned to the work for which she is best known—*Lives of the Queens of England,* from Matilda of Flanders to Queen Anne (12 vol., 1840–48). During this time she also edited *Letters of Mary, Queen of Scots.* Miss Strickland was a warm partisan on the side of royalty and the church, but she made industrious study of "official records and other public documents," gave copious extracts from them, and drew interesting pictures of manners and customs; her style was readable, and the volumes enjoyed considerable popularity.

Her later books included *Lives of the Queens of Scotland and English Princesses* (8 vol., completed in 1859), *Lives of the Bachelor Kings of England* (1861), *Lives of the Seven Bishops* (1866), and *Lives of the Last Four Princesses of the House of Stuart* (1872). She died in July 1874.

STRICKLAND, WILLIAM (1787–1854), U.S. architect and engineer, was born in Philadelphia, Pa., in 1787. He first became known as an artist, although he studied architecture under Benjamin Latrobe. In 1810, however, he designed the Masonic hall in Philadelphia and was thus well launched as an architect. Although the hall was Gothic in design, Strickland later turned to the classic Greek style, and his next important work, the United States bank in Philadelphia, took the Parthenon for its inspiration. Strickland also designed the Merchants' Exchange building, the United States Naval asylum, the United States mint and the United States custom house, all in his native city and Washington's marble sarcophagus at Mount Vernon.

His engineering projects were nearly as well known as his architectural. In 1825 he was sent to Europe to study internal improvements and on his return did much to encourage the construction of the original line of the Pennsylvania Railroad company, one of the first designed to carry passengers in the United States. He also constructed the Delaware breakwater, a commission he received from the U.S. government. At the time of his death he was in Nashville, Tenn., superintending the construction of the state house, which he designed and which is regarded by many as his best work. He died in that city April 7, 1854, and by a special act of the state legislature was buried in the building.

He was the author of several technical publications relating to engineering and architectural projects which he headed; author of *Reports on Canals, Railways, Roads and Other Subjects* (1826), the report he made on his return from Europe; and, with two others, of *Reports, Specifications and Estimates of Public Works in the United States* (1841).

STRIGIDAE: *see* OWL.

STRIKES AND LOCKOUTS. Strikes (often called turnouts or sticks in the early 19th century) are concerted refusals to work under the conditions required by employers. Lockouts, strictly speaking, are occasions on which employers close their establishments against their employees until they accept the terms offered them; but the word lockout is often used, especially by workers, to cover occasions on which the employer attempts to worsen the terms of employment and the workers abstain from work in protest. Lockouts in the strict sense became uncommon in the first half of the 20th century; formerly they occurred especially when an employer or group of employers was attempting to ban union membership and refused employment to workers who would not sign a pledge not to belong to a trade union. Later they arose most often when a body of employers retaliated against a strike at a particular works by closing all factories until the strikers returned to work. When the word lockout is used in its wider sense the distinction between it and a strike is often hard to draw, and workers will describe as a lockout what the employers call a strike. Statistically, no separation can be made, and strikes and lockouts are lumped together under the heading of stoppages, or trade disputes.

EUROPE AND BRITISH COMMONWEALTH

Types of Strikes and Lockouts.—Strikes (and lockouts) can be classified into a number of types. The simplest is the normal type of trade dispute turning on a difference about wages or conditions of employment, or about some special issue of discipline or trade union practice, when the strikers (or lockers-out) are the parties directly involved. Sympathetic strikes or lockouts are those in

which a part is taken by persons or associations not directly involved in the original dispute, but striking or closing their establishments in sympathy with one of the original parties. What are called general strikes may be either very extensive sympathetic strikes or concerted stoppages over a wide range of industries for a common object. The political general strike belongs to the latter type: it can be either a mere protest strike, called for a single day or for a limited period, or a strike designed to last until it has achieved its purpose or failed, as in the German general strike against the Kapp *Putsch* of 1920. The British general strike of 1926 was simply an extended sympathetic strike in support of the Miners' federation, whereas the threatened strike of 1920 against British participation in the Soviet-Polish war belonged to the same type as the strike against the Kapp *Putsch*. A special variety of strike, which may belong to any of the foregoing categories, is the sit-down or stay-in strike (stay-down, in mining) in which the strikers occupy their places of employment and either do no work or attempt to carry on production on their own account (as in the Italian stay-in strikes after World War I). The stay-in strike was extensively used in France in the 1930s at the time of the popular front and is also a familiar technique among native workers in colonial areas. The general strike as actually practised has to be distinguished from the general strike as aspiration—sometimes called the social general strike—which is a mass strike of workers designed to bring about a social revolution and the establishment of a workers' society. Its earliest advocates were the anarchist-communists (*see* A. Roller, *The Social General Strike*, 1905); and the idea was taken up by the French syndicalists. It was Georges Sorel, author of *Reflections on Violence* and other books of syndicalist theory, who introduced the idea of the general strike as a "social myth," inspiring working-class revolutionary feeling and giving a heightened meaning to ordinary strike action, even if in practice it never occurred. The French syndicalists also introduced the conception of *la grève perlée* (sometimes anglicized as "the strike with knobs on") by way of contrast with *la grève aux bras croisés* ("the strike with folded arms"), the former being accompanied by acts of sabotage or violence.

Official and Unofficial Strikes.—Strikes may be either spontaneous or organized. A great many strikes have begun as spontaneous reactions by a group of workers to a particular grievance, such as the dismissal of a fellow-worker whose "victimization" for trade union activity was suspected, or the employment of unskilled workers on jobs normally reserved for skilled workers, or the cutting of a piecework price, or any of a hundred other things. Such sudden strikes may become organized subsequently, if a trade union takes them up, or sometimes if the strikers throw up leaders of their own. Strikes may also be either official or unofficial: they are official when they are called or approved by a trade union, which takes control of the proceedings and of any negotiations designed to settle them; unofficial strikes are either strikes apart from trade unions, or strikes called by trade union branches or districts without the support of the union as a whole in accordance with its rules. Since, in many countries, organized national collective bargaining has been introduced into most industries and collective agreements have laid down procedures to be gone through before an official strike can be called, an increasing proportion of all strikes tends to be unofficial—for example, a group of workers, tired of waiting to have its grievance dealt with by often dilatory negotiating procedures, strikes contrary to the provisions of a national or other collective agreement. The proportion of unofficial strikes is also increased when laws have been passed either forbidding strikes or imposing a compulsory period of delay for investigation before they can be lawfully called (the Lemieux system, used in Canada, and called after its proposer, Rodolphe Lemieux). Groups of workers will often dare to strike "against the law" when trade unions, fearful of having their funds attacked or their recognition by the state suspended or removed, hesitate to endorse a strike and so make it official.

The Law Relating to Strikes and Lockouts.—The law relating to strikes and lockouts has varied greatly from country to country and from time to time. When there is no industrial conscription or compulsory labour, a man is free individually to withdraw his services without penalty, provided that he works out the period of notice stated or implied in his contract. When long contracts are allowed, this right may be greatly limited; and so it may be where there are laws, as there used to be in many countries (in Great Britain up to 1867), against "leaving work unfinished." Moreover, the laws relating to "breach of contract" differ widely.

In Great Britain it used to be a criminal offense for a workman to break a contract of service, and imprisonment or fine was the penalty; whereas a master breaking his contract could only be sued civilly for damages. Only in 1875 were employers and workers in Great Britain in normal cases made equal parties to a civil contract; and even in the middle of the 20th century exceptions remained so that a strike in essential public services or in circumstances which endanger life or valuable property could, if it involved breach of contract, render a striker liable to criminal prosecution. Usually, however, breach of contract gives rise only to a civil action. Breach of contract is, of course, an individual matter between an employer and a single workman. Trade union collective agreements in Great Britain are not contracts enforceable at law, though they can be invoked in establishing implied terms of individual contract (the "custom of the trade"). In Great Britain, except under wartime regulations, there have never been any penalties for breach of collective agreements, even where agreed or even compulsory reference of disputes to arbitration has been in force. In other countries which applied compulsory arbitration, there have often been penalties against trade unions which have called strikes contrary to law—the most usual being the withdrawal of recognition from the offending union so as to prevent it from bringing cases before the arbitration court, with the result either that a rival union can supplant it or that there remains no lawful way of pressing the case of the workers concerned.

In Great Britain, the Trade Disputes and Trades Unions act of 1927 (passed after the general strike, but repealed by the Labour government in 1946), created a special category of illegal strikes and lockouts (in which it was unlawful for trade unions or employers' associations to engage) and laid down a procedure of injunction for determining the illegality in any particular case; but this provision was never in fact used. In totalitarian countries strikes have usually been forbidden altogether, and severe penalties laid down against their inciters. In countries where personal freedom is recognized as a basic principle the question of legality is more difficult. If an individual is free to withdraw his labour, subject to not breaking his contract, what can make it unlawful for a number to do the same? The law courts used to answer this question by invoking the common law notion of conspiracy in restraint of trade, under which concerted action could render criminal what it would be fully lawful for one man to do. This was the legal doctrine under which strikes were held criminal, or usually held so when a case came before the courts, up to 1799, when a general act (amended in 1800) was passed forbidding all forms of working-class combination. Thereafter, until 1824, all strikes could be treated as criminal either under the Combination acts or as conspiracies at common law to violate the Combination acts. These acts were repealed in 1824, and under a statute of 1825 strikes for a narrow range of purposes (wages, hours, etc.) were barely legalized, provided they involved nothing that the courts regarded as violence or intimidation. As this latter term was not defined, legality remained highly precarious up to 1871, when the Trade Union act laid down that acts done by a number of persons in concert in pursuance of trade disputes were not to be regarded as criminal conspiracies—that is, that the mere fact of concerted action by a number could not, in such cases, constitute an otherwise lawful act a crime. Thereafter, in Great Britain, the legal right to strike seemed to be securely established, until it was challenged by the Taff Vale railway judgment of 1901, which decided that trade unions could be sued for damages arising out of their agents' actions. (*See* Labour [Trade] Union.) This decision was reversed by the Trade Disputes act of 1906, which gave trade unions and their agents a wide immunity, making trade unions virtually unsuable in their industrial activities and giving a wide freedom from prosecutions for incitement to breach of contract in connection with trade disputes.

Picketing.—Strikes are often accompanied by picketing—that is, mainly, stationing representatives at the entrance to a works to warn off potential "blacklegs," but also, sometimes, picketing the homes of such persons. A blackleg (or scab, or knobstick) is a worker who is prepared to work while other workers of the same group are on strike or locked out; he may be either a regular worker who refuses to join the strike, or a substitute brought in from elsewhere to help in breaking it. Picketing is normally done by one or a few men, stationed at each point; but there is also mass picketing, which means assembling a large body of strikers to obstruct entry to the place of work. Much trouble arose over the legal aspects of picketing. In Great Britain peaceful picketing was authorized in 1859, withdrawn in 1871, replaced only in a restricted form in 1875; fully restored in 1906 by the Trades Disputes act, drastically limited in 1927 by the Trade Disputes and Trade Unions act, which wholly forbade picketing of a worker's home, and restored in the 1906 form in 1946.

Political Strikes.—Strikes have taken place from time immemorial. They go back much further than trade unions and still occur frequently among unorganized workers in the less developed countries without any trade union support. They often lead to the formation of permanent trade unions, as in the West Indies, in British Africa and in many other places. In such areas it is often almost impossible to distinguish between strikes having industrial and strikes having political objects, for the two are often inextricably mixed up. A historical parallel can be seen in the extensive British strikes of 1842, known as the Plug plot. These, beginning as protests against wage reductions, partly turned into political strikes for the enactment of the People's Charter. There was a similar mingling of political and economic objects in the popular front strikes in France in 1936–37, and in the strike waves that accompanied the Russian revolutions of 1905 and 1917.

The Financing of Strikes.—Industrial strikes constitute the immense majority. When trade unions are behind them, there is usually strike pay (dispute benefit) unless the unions' funds run out. When unions have only scanty funds, relief in kind may be given instead of strike pay. Unions involved in costly disputes often receive financial help from other unions, particularly where national trade unions have not developed, or have not accumulated large funds. One of the principal purposes of local trades councils has been to authorize appeals for help of this kind. The London Trades council (1860) arose in this way, out of a temporary grouping formed to assist the London builders in their lockout of 1859. The London dock strike of 1889 and the northeast coast nine-hour strike of 1872 were both won largely because of help from other unions—in the dock strike from the Australian trade unions. The British miners have repeatedly sent large donations to help striking miners in France and elsewhere. In some cases there are central strike insurance funds on which federated unions can draw. This, for example, is one of the main functions of the British General Federation of Trade Unions founded in 1898.

TABLE I.—*Great Britain: Trade Disputes, 1893–1956*

Year	No. of stoppages	Workers involved (in 000)	Working days lost (in 000)	Year	No. of stoppages	Workers involved (in 000)	Working days lost (in 000)
1893*	599	634	30,440	1925	603	441	7,950
1894	903	322	9,510	1926*‖	323	2,734	162,230
1895	728	259	5,700	1927	308	108	1,170
1896	906	192	3,570	1928	302	124	1,390
1897†	848	229	10,430	1929	431	533	8,287
1898†	695	252	15,200	1930	422	307	4,399
1899	710	179	2,500	1931	420	490	6,983
1900	633	185	3,090	1932	389	379	6,488
1901	631	179	4,130	1933	357	136	1,072
1902	432	255	3,440	1934	471	134	959
1903	380	116	2,320	1935	553	271	1,955
1904	346	87	1,160	1936	818	316	1,829
1905	349	92	2,370	1937	1,129	610	3,413
1906	479	218	3,020	1938	875	274	1,334
1907	585	146	2,150	1939	940	337	1,356
1908‡	389	293	10,790	1940	922	299	940
1909	422	297	2,690	1941	1,251	360	1,079
1910	521	514	9,870	1942	1,303	456	1,527
1911§	872	952	10,160	1943	1,785	557	1,808
1912*§	834	1,462	40,890	1944	2,194	821	3,714
1913	1,459	664	9,800	1945	2,293	531	2,835
1914	972	447	9,880	1946	2,205	526	2,158
1915	672	448	2,950	1947	1,721	620	2,433
1916	532	276	2,450	1948	1,759	424	1,944
1917	730	872	5,650	1949	1,426	433	1,807
1918	1,165	1,116	5,880	1950	1,339	302	1,389
1919§	1,352	2,591	34,970	1951	1,719	379	1,694
1920*	1,607	1,932	26,570	1952	1,714	416	1,792
1921*	763	1,801	85,870	1953	1,746	1,374	2,184
1922†	576	552	19,850	1954	1,989	450	2,457
1923	628	405	10,670	1955	2,424	671	3,704
1924	710	613	8,420	1956	2,648	508	2,083

*Large coal-mining stoppage. †Large engineering stoppage. ‡Cotton and shipyard stoppages. §Transport stoppages. ‖General strike.

Statistics.—Statistics of strikes and lockouts are usually calculated in terms of numbers of disputes involving stoppages of work, numbers involved and numbers of working days lost. No single measure is satisfactory. One big national strike involves more workers and days lost than a great number of small disputes, and even a short national stoppage in a big industry outweighs many smaller disputes in terms of numbers involved. Still less satisfactory are the attempts to assess the results of strikes on a statistical basis by distinguishing between strikes won, lost and compromised. So high a proportion of all disputes has ended in compromise that the figures show almost nothing; for compromise solutions may range from almost complete successes to barely qualified defeats. All that is effectively possible is a very broad measurement of fluctuations in strike activity from year to year. After 1926 there was no year of considerable strike action in Great Britain, and the years after 1945 were remarkably free from disturbance in comparison with those after 1918. But only the broadest conclusions can be drawn from the available statistics. Table I shows the number of disputes, workers involved and working days lost in Great Britain on account of trade disputes from the inception of the returns in 1893.

It will be observed that there was a much greater fall in the number of strikers and of working days lost than in the number of strikes. This is because many disputes during World War II and after were small and short unofficial stoppages in a single works or colliery or dock. Attempts are sometimes made to classify stoppages by causes—for example, wages, hours, recognition, work rules, etc.—but the resulting statistics are of scant value, as more than one factor is often involved. Nor does the attempt to analyze methods of settlement yield valuable results; it would be much more important to know the numbers of disputes settled without stoppages of work by the various methods—for example, direct negotiation, conciliation, arbitration, etc.—but no statistics on this matter were available by the middle of the 20th century.

BIBLIOGRAPHY.—Pre-1939 data are summarized in the *I.L.O. Year Book of Labour Statistics* for 1939; but they are defective. References to some of the principal strikes are given in the annual reports of the ministries of labour of the various countries. *See* also S. and B. Webb, *History of Trade Unionism*, rev. ed. (London, New York, 1920); Lord Amulree, *Industrial Arbitration in Great Britain* (Oxford, New York, 1929); J. H. Richardson, *Industrial Relations in Great Britain*, I.L.O., 2nd ed. (Geneva, 1938); H. A. Marquand (ed.), *Organized Labour in Four Continents* (London, New York, 1939); G. D. H. Cole, *Short History of the British Working-Class Movement, 1789–1947*, rev. ed. (London, 1948; New York, 1949), *Introduction to Trade Unionism* (London, 1953; New York, 1954); K. G. J. C. Knowles, *Strikes: a Study in Industrial Conflict; With Special Reference to British Experience Between 1911 and 1947* (Oxford, 1952; New York, 1953).

(G. D. H. C.)

UNITED STATES

A strike of journeymen printers at New York, N.Y., in 1776 for an increase in wages is the earliest recorded strike in U.S. history. Prior to 1827 disputes were on a small scale for higher wages or the ten-hour day. The years 1835–37 saw "wildcat" prosperity accompanied by a rapid rise in the cost of living. To maintain their standards of living, workers organized trade unions and trades' unions (city centrals) to gain financial and moral support in their struggles for higher wages. While this "false" prosperity continued, the workers' demands tended to be granted, but the financial panic of 1837, with its consequent unemployment, meant defeat for the trade unions on the economic front.

After the Civil War, unionism took on a national character. The first strike of nation-wide extent was on the Pennsylvania and Baltimore and Ohio railroads in 1877, at the depth of a prolonged depression. The Pittsburgh freight yards of the Pennsylvania railroad were burned and looted. State troops brought in from Philadelphia, because of the inaction of local authorities and the unreliability of local state troops were forced to surrender. The strike, which many feared might be the U.S. counterpart of the

commune of Paris, was quelled by the dispatch of units of the U.S. army.

On May 1, 1886, approximately 340,000 workers went on strike in response to a call by the national trade union federation, which became the American Federation of Labor in 1886, to secure the eight-hour day. Explosion of a bomb in Haymarket square in Chicago, Ill., on May 4, 1886, supposedly by anarchists, and the consequent public clamour against radicalism and unionism was blamed by Samuel Gompers for arresting this promising move mid-career. A revival of the demand for an eight-hour day occurred in 1890, when carpenters struck throughout the nation and were generally successful in achieving their objective.

Skilled organized workers at the Homestead, Pa., plant of the Carnegie Steel Co. in July 1892 refused to accept a wage reduction, and in the process gained the support of the unorganized workers. Thereupon Henry C. Frick, general manager of the company, shut down the entire plant and refused to negotiate any further with the Amalgamated Association of Iron, Steel and Tin Workers, until then the strongest union in the world. As the strike issue settled down to one of continued union recognition, Frick imported strike-breakers and armed guards. In the clash which ensued between the guards and strikers, several were killed on both sides. The strike ended in U.S. unionism's most decisive defeat. Organized labour's defeat in the American Railway union strike, or the so-called Pullman company boycott of 1894 was of another order. The union, young and weak, instituted a sympathetic strike to assist the Pullman local's resistance to a 25% wage cut. Violence and mass destruction of property occurred, centring in the Chicago yards. The railroad asked for federal intervention, and 2,000 federal troops were sent to Chicago. However, the strike was really broken by the issuance of injunctions in the federal courts secured at the insistence of the attorney general of the United States on the basis of the Sherman Anti-trust law as one of the grounds.

The successful strike of the bituminous miners in 1897 resulted in establishment of an agreement for the so-called central competitive field, with Illinois as the union's stronghold for decades thereafter. The collective agreement, after an apparently hopeless strike begun with an empty union treasury, founded an "industrial government" based on reciprocal rights and obligations of both parties and pointed the way in all of the coal industry and other industries as well. In 1902, 150,000 anthracite miners struck to achieve a reduction of hours, an increase in wages and recognition of their union. Violence occurred in areas where strikebreakers were hired to mine coal. To preserve order the national guard was sent to certain cities. As winter came, the situation grew worse, and Pres. Theodore Roosevelt intervened and arranged for arbitration. The miners won a 10% increase in wages and a shorter working day, but their union was not recognized. The outstanding figure of this struggle was the miners' leader, John Mitchell, who sustained the hope of his polyglot following and won public sympathy by his moderation.

The next big strike was that of the coat and suit makers (women's) in 1910. It produced a second major pattern of industrial government, added to that of the soft-coal miners. Following a two months' strike in New York, a "protocol of peace" established a tripartite board to settle differences between the manufacturers and the union under the agreement. Louis D. Brandeis, later on the United States supreme court, mediated in the strike and devised a compromise in the "preferential union shop" in lieu of the closed shop vainly demanded by the union. This proved a major invention. A four-months' strike against Hart, Schaffner and Marx in Chicago established a similar pattern in the men's division. In this strike, as in the working of the agreement, Sidney Hillman, later a major leader in U.S. unionism, made his debut.

Stable industrial government in the bituminous coal industry was constantly menaced by nonunion competition. Continuous efforts were made by the United Mine Workers to organize the West Virginia mines; but the operators, with state support, succeeded in repelling organization. Union members were evicted from company houses and forced to live in tent colonies, and in Feb. 1913, an armed railway train machine-gunned a miners' tent colony at night. A similar incident of industrial autocracy occurred in Colorado. In 1913–14 the national guard machine-gunned miners' tent colonies, and the miners retaliated by killing mine guards and destroying mine property. Federal troops were called to bring order in April 1914, but the strike was not called off until December.

An attempt made to organize the iron and steel industry during and immediately after World War I was a major effort to recoup labour's loss in the Homestead strike in 1892. About 350,000 responded to a strike started in Sept. 1919, but strike-smashing was employed extensively. The men held out for six weeks at the cost of 18 lives, and then mass surrenders began.

The New Deal.—With the New Deal, strikes began to evolve to a new pattern of diminishing violence because of a more considerate treatment of strikers by government and employers. Yet this may be said to date only from the "little steel" strike of 1937 and the La Follette committee investigation of denial of rights to wage earners.

A longshoreman's strike in San Francisco, Calif., in May 1934 developed into a general strike. Union recognition and control of the hiring halls, in essence the closed shop, and better wages and hours were the issues. Serious rioting began when the employers attempted to operate the docks by force. The California national guard was called out. Labour became steadily more embittered, and in July sympathetic strikes started, and the principal highways into the city were closed. The general strike called on July 16 lasted four days, but the strike committee permitted the entrance of food trucks. The longshoremen's leader in this strike was Harry Bridges, subsequently known in Pacific coast and Hawaiian industrial relations.

To force "little steel" recognition of the Steel Workers' Organizing committee (C.I.O.), a strike was called in May 1937. The struggle was attended by considerable violence and, although "big steel's" recognition of the Steel Workers' Organizing committee had occurred early in 1937, "little steel" (the independent producers) continued to resist unionization for a few more years.

World War II and After.—Prior to the attack on Pearl Harbor two allegedly Communist-led strikes threatened to hold up the national defense program. In June 1941 leaders of the United Automobile Workers' local at the North American Aviation Co. in Inglewood, Calif., called a strike in violation of an interim agreement negotiated with the aid of the National Defense Mediation board. This unauthorized strike was broken through plant seizure by the government and a dispatch of troops. A strike at the Allis-Chalmers plant in Milwaukee, Wis. (1941), crippled vital defense production.

Production losses caused by wartime strikes were relatively small except for soft-coal mining. An outgrowth of two strikes during 1943 was the Smith-Connally or War Labor Disputes act.

With the termination of the no-strike pledge at the end of the war came renewed industrial warfare. The restriction of wage increases during the war, fear of unemployment and some reduction in income because of "cutbacks," and the conviction that organized labour was to be subjected to a concerted employer attack created a favourable climate for widespread strikes. It was expected that the events of World War I and its aftermath would be repeated. In that war the wartime truce had been followed by a triumphant employer counterattack which not only "contained" unionism but regained most of the ground lost to management. Just as 1919 had seen a new record in the number of strikers, it was expected that 1946 might approach that record and possibly bring about a comparable defeat for labour. What happened was something entirely different: prosperity held and the unions continued to gain in membership, notably in the newer of New Deal-sponsored unions.

A number of the postwar strikes were industry-wide, and most of them were of considerable duration. During the 12-month period following V-J day, work stoppages resulted in almost 12,000,000 man-days of idleness. Strikes were called in the oil, lumber, auto, steel, electrical manufacturing, meat-packing, coal, railroad and maritime industries. A strike by the United Automobile Workers against General Motors began in Nov. 1945 and

TABLE II.—*Number of Workers Involved in U.S. Work Stoppages, 1916–1955*
(in 000s)

Year	Workers	Year	Workers	Year	Workers	Year	Workers
1916	1,600	1926	330	1936	789	1946	4,600
1917	1,230	1927	330	1937	1,860	1947	2,170
1918	1,240	1928	314	1938	688	1948	1,960
1919	4,160	1929	289	1939	1,170	1949	3,030
1920	1,460	1930	183	1940	577	1950	2,410
1921	1,100	1931	342	1941	2,360	1951	2,220
1922	1,610	1932	324	1942	840	1952	3,500
1923	757	1933	1,170	1943	1,980	1953	2,400
1924	655	1934	1,470	1944	2,120	1954	1,530
1925	428	1935	1,120	1945	3,470	1955	2,560

continued until March 1946. The leader of this strike, Walter Reuther, championed the consumers' cause as well as that of the strikers. He spurned any wage increase that might result in a rise of the prices of cars, but he was certain that the demanded increase could be met out of the corporation's accumulated profits. This "public-minded" attitude proved somewhat embarrassing to Philip Murray of the Steel Workers union, whose members were equally pressing for a wage raise from an industry that did not have such profit reserves. The bituminous coal miners were on strike from April 1 to May 29, 1946, interrupted by a truce from May 13 to May 27. The railway trainmen and engineers struck on a nation-wide scale for two days in May 1946 but were defeated by Pres. Harry S. Truman's radio appeal. Smaller but equally crucial strikes were those of the tugboat and building employees in New York and the public utility strike in Pittsburgh.

On the all-important wages question, the matter arranged itself in three "rounds" over 1946–48. Peculiar to this period was the importance of "pattern settlements." The first settlement reached in a mass-production industry tended to set a pattern which all other industries accepted. As the cost of living became stabilized, unions turned their attention to nonwage demands, such as employer-financed pension and social insurance programs. This demand was stimulated by the inability of unions to break the "little steel" wage formula during the war and gained added momentum from an agreement signed in 1946 between the United Mine Workers and the government in which a health and welfare fund was set up to be financed by a contribution of five cents on each ton of coal mined. From coal, "social security by contract" spread to the other basic industries.

The postwar years intensified the periods of industrial strife in the bituminous coal industry. Work stoppages were variously termed walkouts, memorial holidays, stabilization periods and just plain strikes. Since the industry was a "sick industry"—the coal demanded by the nation could be produced if the miners worked only three days a week—the union assumed responsibility for stabilizing production. Industrial government in the industry at times seemed to be dominated by a labour autocracy under John L. Lewis. But Lewis was careful to promote the welfare of his men. Action for contempt of court under the emergency provision of the Taft-Hartley act failed in the 1949–50 strike, as the union and its leaders gained immunity in complying outwardly with a court order to resume work, but without actually doing so.

The steel strike of 1949 typified the new day in industrial conflict. While the cost in dollars was great, the jails were empty. Not a cent was spent on strikebreakers, tear gas, riot guns or arms of any kind. There was no loss of human life. This absence of violence, formerly so prevalent in U.S. labour disputes, stemmed from the Wagner act, the rise of a new kind of employer and a well-disciplined trade union membership.

Following the outbreak of the Korean war, the number of work stoppages increased sharply. In 1950 there were 4,843 stoppages, as opposed to 1948 with 3,419 stoppages. There had been 4,985 in 1946, however. The demand for wage increases came to the forefront as unions proposed and obtained increases in anticipation of federal wage controls. Following the establishment of federal wage controls in 1951, the unions once again shifted their emphasis to demands for fringe adjustments (*e.g.*, vacation and holiday pay, shift differentials and overtime). This policy continued into 1952–53 as the cost of living became stabilized and started to decline. In the decade following World War II the peak in work stoppages occurred in 1953, with 5,091. The number dropped to 3,468 in 1954 and then rose to 4,320 in 1955.

Among the outstanding events of the period following the 1949 recession was the contract signed by United Automobile Workers, C.I.O., with the General Motors corporation. The five-year contract provided for cost-of-living adjustments in hourly wage rates, provided for a pension fund and established a union shop. These negotiations were harmoniously conducted and gave evidence of the effect that expanding business activity and sustained near capacity production levels had on labour-management relations. This settlement influenced the peaceful conclusion of wage agreements made by the Chrysler and Ford companies with the U.A.W. in 1950.

Another important feature of labour-management relations after outbreak of the Korean war was the active participation of the federal government in the settlement of labour disputes.

The Wage Stabilization board was authorized in April 1951 to investigate and recommend settlement in any dispute which was not resolved by collective bargaining or mediation, which threatened national defense, where the dispute was referred to the board by the parties involved or by the president of the United States. In 1951 the president referred five disputes to the board for settlement, in the copper refining and transport equipment industries. Also in 1951 the president referred the important steel dispute to the board. The steel union demanded an hourly wage increase and fringe benefits including shift premiums and holiday pay. The board recommended a wage increase and acceptance of some form of union shop. This recommendation was criticized by industry members because it threatened the anti-inflationary program of the federal government and it imposed a union shop on the industry. The union issued a strike notice in April 1952 when negotiations broke down. In an effort to avert the strike, the steel plants were seized by the federal government in April 1952. This seizure was declared unconstitutional by the U.S. supreme court in June 1952, and the president directed return of the mills to the owners. Immediately following the return of the mills the steel workers went out on strike.

This steel strike lasted 54 days and was responsible for one-half of the 55,000,000 man-days lost in 1952 through work stoppages.

Steel production was not resumed until a new contract was signed granting an hourly wage increase and some of the fringe demands. Following the wage increase the acting defense mobilizer ordered the Office of Price Stabilization to authorize increases in the price of basic steel.

The government also participated directly in the settlement of two major railway disputes. Following an unsuccessful attempt to settle an 18-month dispute between the rail carriers and the major railway unions the government prevented a walkout by seizing the railways in Aug. 1950. The dispute centred about the unions' demands for a 40-hour week for yard workers and rules changes for road service employees. A settlement of dispute between the railways and the switchmen was reached in Sept. 1950. The dispute with the trainmen ended in May 1951. However, an agreement between the railways and the other operating unions was not reached for another year. Consequently federal operation of the railroads ended in May 1952.

Strike activity in 1953 measured in total time lost by workers involved was relatively low but, when measured in the number of strikes, was high. Most of these strikes involved demands for wage increase. One notable exception was the strike of the United

TABLE III.—*Man Days of Idleness in U.S. Work Stoppages, 1927–56*
(in 000s)

Year	Days of idleness	Year	Days of idleness	Year	Days of idleness
1927	26,200	1937	28,400	1947	34,600
1928	12,600	1938	9,150	1948	34,100
1929	5,350	1939	17,800	1949	50,500
1930	3,320	1940	6,700	1950	38,800
1931	6,890	1941	23,000	1951	22,900
1932	10,500	1942	4,180	1952	55,000
1933	16,900	1943	13,500	1953	28,300
1934	19,600	1944	8,720	1954	22,600
1935	15,500	1945	38,000	1955	28,200
1936	13,900	1946	116,000	1956	33,000

Hatters, Cap and Millinery Workers against the Hat Corporation of America to prevent the transfer of work from Norwalk, Conn., to another plant. It lasted more than a year, and a formula was finally found which guaranteed the right of the company to determine its production policy and gave the workers some protection against sudden shifts in operation.

In 1954 a sharp drop in the number of workers involved in labour disputes took place. Fewer workers were involved than had been in any other year after World War II, and the number of strikes was the lowest of any year in this period.

The number of workers involved and the number of disputes rose markedly in 1955. Several disputes involved large numbers of workers and were bitterly contested. The strike against the Southern Bell Telegraph and Telephone company which lasted 72 days as well as the strike of nonoperating workers on the Louisville and Nashville railroad which lasted 58 days were attended by violence and bitter recriminations. The latter dispute hinged around the refusal of the carrier to accept a health and welfare plan, and the former was over wages and arbitration of certain kinds of disputes. On the other hand, the strike of New England textile workers against a number of mills in that area which lasted 90 days at some mills was free from violence. One of the more bitter strikes in the 1945–55 period was that of 70,000 workers in the International Union of United Electric Workers against the Westinghouse Electric company. This strike lasted into 1956, almost six months, and was bitterly fought on the picket line.

The first part of 1956 was generally peaceful. The most important walkout was the shutdown of the basic steel industry by the failure of the union and the companies to agree on terms of a new contract. After a shutdown of five weeks an agreement was reached.

See also Labour Law; Hours of Labour; Labour (Trade) Union.

Bibliography.—Bureau of Labor Statistics, *Monthly Labor Review* (May issue, 1947–52); Ethelbert Stewart, "A Documentary History of the Early Organizations of Printers," U.S. Bureau of Labor *Bulletin No. 61* (1905); John R. Commons *et al.*, *History of Labour in the United States* (1921); Selig Perlman, *History of Trade Unionism in the United States* (1922); Selig Perlman and Philip Taft, *Labor Movements,* vol. iv in "History of Labor in the United States Series" (1935); Philip Taft, *Economics and Problems of Labor,* 2nd ed. (1948).

(P. Tt.)

STRINDBERG, (JOHAN) AUGUST (1849–1912), the greatest Swedish playwright, who exercised a profound influence on the drama, especially in Europe and the United States, novelist, short-story writer and poet, was born in Stockholm, Jan. 22, 1849. He was the son of a steamship agent and a former waitress, and had an unhappy childhood, marred by emotional insecurity, as he relates in his remarkable autobiography *Tjänstekvinnans son* (1886; Eng. trans., *The Son of a Servant,* 1913). He studied intermittently at Uppsala university, never taking a degree. He had to earn his living by working as a freelance journalist in Stockholm, among other occupations. Meanwhile he struggled to complete his first important work, the historical drama *Mäster Olof* (1880; Eng. trans., 1931), on the theme of the Swedish reformation, influenced by Shakespeare and by Ibsen's *Brand.* The Royal theatre's rejection of *Mäster Olof* deepened his pessimism and sharpened his contempt for official institutions and traditions. He continued to revise the play—later recognized as the first living piece of modern Swedish drama—for several years, and this delayed his development as a dramatist of contemporary problems.

In 1874 he became a librarian at the Royal library, and in 1875 met the Finno-Swedish Siri von Essen, then the wife of a guards officer; two years later they married. Their intense but ultimately disastrous relationship ended with divorce in 1891, when Strindberg, to his great grief, lost the custody of the children. At first, however, marriage stimulated his writing, and in 1879 he published his first novel *Röda rummet* (Eng. trans., *The Red Room,* 1913), a satirical account of abuses and frauds in Stockholm society: this was something new in Swedish fiction, and made its author famous.

He also wrote more plays, of which *Lycko-Pers resa* (1881; Eng. trans., *Lucky Peter's Travels,* 1931) contains the most biting social criticism. In 1883, after publishing *Det nya riket* ("The New Kingdom"), a withering satire on contemporary Sweden, Strindberg left Stockholm with his family and for six years moved restlessly about the continent. The first volume of his collection of stories *Giftas* (1884; Eng. trans., *Married,* 1913) led to a prosecution for blasphemy; he was acquitted, but the case affected his mind, and he imagined himself persecuted, even by Siri.

He returned to the drama with new intensity, and the conflict between the sexes inspired some of the outstanding works written at this period, such as *Fadren* (1887; Eng. trans., *The Father,* 1930), *Fröken Julie* (1888; Eng. trans., *Lady Julie,* 1930) and *Fordringsägare* (1888; Eng. trans., *The Creditor,* 1914). In these bold and concentrated works he combined the technique of dramatic naturalism with his own conception of psychology and thereby inaugurated a new movement in European drama. *Hemsöborna* (1887; Eng. trans., *The People of Hemsö,* 1959), a vigorous novel about the Stockholm skerries, always one of Strindberg's happiest sources of inspiration, was also produced during this intensively creative phase.

The years after his return to Sweden in 1889 were lonely and unhappy; in 1892 he went abroad again, to Berlin. His second marriage, to a young Austrian journalist, Frida Uhl, followed in 1893; they finally parted in Paris in 1895. A period of literary sterility, emotional and physical stress, and considerable mental instability, culminated in a kind of religious conversion, the crisis which he described in *Inferno* (1897; Eng. trans., 1912). Strindberg had earlier moved from Christianity to atheism by way of deism; scientific studies (1894–96) led to experiments with alchemy and to theosophy.

His new faith, coloured by Swedenborgian mysticism, recreated him as a writer. The immediate result was the drama in three parts, *Till Damaskus* (1898; Eng. trans., *The Road to Damascus,* 1939), in which he depicts himself as "the Stranger," a wanderer seeking and finally finding spiritual peace; another character, "the Lady" resembles both Frida Uhl and Siri von Essen.

By this time Strindberg had again returned to Sweden, settling first in Lund, then, in 1899, in Stockholm, where he lived (after 1908 in his apartment *Blå Tornet*—"the Blue Tower") until his death. The summers he often spent among his beloved skerries. His view that life is ruled by the "Powers," punitive but righteous, was reflected in a series of historical plays which he began in 1889. Of these *Gustav Vasa* (1899; Eng. trans., 1931) is the best, masterly in its firmness of construction and characterization and its vigorous dialogue. In 1901 he married the young Norwegian actress Harriet Bosse; in 1904 they parted, and again Strindberg lost the child, his fifth.

Yet out of this "spring in winter," as he called it, came *inter alia* the plays *Dödsdansen* (1901; Eng. trans., *The Dance of Death,* 1929), *Kronbruden* (1902; Eng. trans., *The Bridal Crown,* 1916) and *Drömspelet* (Eng. trans., *A Dream Play,* 1929), as well as the charming autobiography *Ensam* ("Alone," 1903) and some lyrical poems. Renewed bitterness after his parting from his wife provoked the grotesquely satirical novel *Svarta Fanor* ("Black Banners" pub. 1907), which attacked the vices and follies of Stockholm's literary coteries, as Strindberg saw them. *Kammarspel* ("Chamber Plays," 1907), written for the little Intima theatre, which Strindberg ran for a time with a young producer, August Falck, embody further developments of his dramatic technique: of these, *Spöksonaten* (Eng. trans., *The Ghost Sonata,* 1929) is the most fantastic, anticipating much in later European drama. His last play, *Stora Landsvägen* (Eng. trans., *The Great Highway,* 1954), a symbolic presentation of his own life, appeared in 1909.

Up to the last Strindberg debated current social and political ideas (returning to the radical views of his youth) in polemical press articles, while his philosophy was expounded in the aphoristic *Blå Böcker* (1907–12; Eng. trans., *Zones of the Spirit,* 1913). He died of cancer on May 14, 1912, in Stockholm, ignored in death, as in life, by the Swedish Academy, but mourned by his countrymen as their greatest writer. On Swedish life and letters he has exercised a lasting influence and is admired for his origi-

nality, his extraordinary vitality and his powerful imagination, which enabled him to transform autobiographic material into dramatic dialogue of characteristic brilliance.

The pregnant, colloquial style of Strindberg's early novels and, especially, of his short stories, brought about a long overdue regeneration of Swedish prose style, and *Tjänstekvinnans son* gave perhaps the strongest impulse since Rousseau's *Confessions* to the publication of discreditable self-revelations. His greatest influence however was exerted in the theatre, through his critical writings (such as the introduction to *Fröken Julie*), his plays and the devices of production which their staging enjoined. The continuous, brutal action and phonographic dialogue of *Fröken Julie* and other plays written between 1887 and 1893 reached the *ne plus ultra* of naturalistic drama, as presented simultaneously in André Antoine's Théâtre Libre in Paris. With the later phantasmagoric plays, such as *Till Damaskus*, *Drömspelet* and *Spöksonaten*, Strindberg then led that section of the revolt against stage realism issuing in the expressionist drama, which was developed mainly in Germany after 1912 by Fritz von Unruh, Franz Werfel, Ernst Toller and Georg Kaiser, but left notable traces elsewhere, as in the drama of Sean O'Casey, Elmer Rice, Eugene O'Neill, Luigi Pirandello and Pär Lagerkvist.

BIBLIOGRAPHY.—*Editions: Samlade skrifter*, ed. by J. Landquist, 55 vol. (1911–21); *Samlade otryckta skrifter*, ed. by G. Carlheim-Gyllenskjöld, 2 vol. (1918–19).

Letters: A. Strindbergs brev, ed. by T. Eklund (1948 *et seq.*); *Från Fjärdingen till Blå Tornet*, ed. by T. Eklund, a selection (1946); *Strindbergs brev till H. Bosse* (1932).

Bibliography: R. Zetterlund, *Bibliografiska anteckningar om A. Strindberg* (1913); K. Lundmark and K. Kärnell—a bibliography in preparation.

Criticism and biography: E. Hedén, *Strindberg, en ledtråd vid studiet av hans verk* (1921); M. Lamm, *Strindbergs dramer*, 2 vol. (1924–26) and *Strindberg*, 2 vol. (1940–42); G. Ollén, *Strindbergs dramatik* (1948); G. Brandell, *Strindbergs Infernokris* (1950); see *Svensk Litteraturhistoria*, vol. iv, ed. by E. N. Tigerstedt (1956).

Non-Swedish works: K. Jaspers, *Strindberg und Van Gogh* (1922); M. Gravier, *Strindberg et le théâtre moderne* (1949); J. Bulman, *Strindberg and Shakespeare* (1933); E. Sprigge, *The Strange Life of August Strindberg* (1949); B. M. E. Mortensen and B. W. Downs, *Strindberg: an Introduction to His Life and Work* (1949).

(B. M. E. M.; B. W. D.)

STRINGED INSTRUMENTS owe their existence to the musical properties inherent in an elastic filament when it is made to vibrate regularly by being suitably energized while stretched between two fixed points. The tension, length, thickness and mass of the filament, or string, all affect the rate of vibration and consequently the note it will produce. The shorter or tenser the string, the higher the note; a heavy string produces a lower note than a lighter one of the same length and tension.

Resonators.—For musical purposes, the stretched string is not enough; it must, if the result is to be musically acceptable, be in contact with some sort of resonator that is capable of radiating vibrations. All stringed instruments therefore possess a resonant body, usually of wood, the essential feature of which is always the soundboard or belly, which lies directly under the strings and receives their vibrations through the bridge over which they are stretched. The form of the belly varies with the instrument, but it is in effect a thin diaphragm capable of responding to the vibrations transmitted by the bridge and of propagating them freely and instantaneously all over its area, which is made as large as practical considerations allow. Straight-grained pine is the favoured material for the bellies of the majority of European instruments on account of its excellent resonant properties and the rapid and even distribution of musical vibrations both along and across the grain. A stretched skin membrane has, however, been used, notably in the European banjo, the Indian *sarinda* and *sarangi* and the Arab *kamanga* and *rabab*. The wooden bellies of all developed instruments are reinforced in a variety of ways by crossbars glued to the underside; these bars serve as girder stiffeners and prevent the belly from sinking under pressure of the strings, while preserving its general flexibility and response. Such bars also assist in the propagation of transverse vibrations across the pine belly. The bowed stringed instruments, whose string tensions and bridge load are relatively high, employ a distinctive vertical strut, the soundpost, which not only gives greater support to the belly but also is a vital agent in the tonal adjustment of the instrument (*see* VIOLIN FAMILY).

With only a few primitive exceptions, the belly forms the "lid" of a box that is in fact the body of the instrument. If the air space thus enclosed is to resonate freely it must be in contact with the surrounding atmosphere, and it is usual to cut soundholes through the belly to ensure this. The form of the soundholes varies enormously, but to some extent their shape, position and area are functional, serving to regulate both the stresses to which the belly is subject and the effect of the resonant air space. In the violin family especially they have a marked influence on the over-all tone quality.

String Supports.—There must of course be two limiting supports between which a string is made to vibrate. The bridge has already been described. In practice the second "bridge," called the nut, is usually fixed rigidly to the fabric of the instrument beyond the area of the belly. Exceptions to this are the virginals and instruments of the dulcimer family, where both bridges stand on the belly and share the function of transmitting vibrations thereto. The upper extremity of the string is carried over and beyond the nut and there wound upon a peg for tuning. At the other end it is hitched to a string holder or a fixed pin attached to the instrument. Both these terminals are arranged below the level of the bridge and nut to secure a firm down-beating upon them when the string is tightened to its working tension. The bridge, and through it the belly, are then subject to a vertical pressure from the string that ensures sensitivity of response to its vibrations. This pressure bridge is the same type as used on the violin and other bowed instruments, on all keyboard instruments and certain plucked instruments (mandolin, banjo, cittern, etc.).

In keyboard instruments, with their many, carefully graded strings, the bridge is necessarily fixed to the belly, usually by glue. But bridges of the violin type are held in position only by the pull of the strings, thus allowing some latitude for adjusting their position to get the best result.

A contrast to the pressure bridge is the tension bridge. In this type the strings are attached direct to the bridge unit, which is glued to the belly. The pull of the strings is sustained by the bridge alone instead of by a separate anchorage beyond the bridge. Such bridges are found on instruments of the guitar and lute families, which are characterized by low-lying, rather slack strings and very thin, light bellies. In the harp the strings pull upward, away from the plane of the belly, the bridge being a central rail to which they are pinned.

Production of Scales.—The scale of notes that a stringed instrument will give is produced in one of two different ways: either by providing separate strings for every note or by "stopping" a given string somewhere along its working length, thus temporarily shortening the portion left free to vibrate and raising the pitch of the resultant note. Keyboard instruments, dulcimers and zithers belong to the first class, as also do harps, although for some centuries European harps have employed a limited form of stopping device for the chromatic alteration of some or all of the notes of the scale (*see* HARP).

In the majority of cases, stopping is accomplished by the player's fingers; he presses the string against a board (the fingerboard) that is either raised above the belly as in some forms of zither or, more characteristically, carried on a handle or neck extending beyond the belly to give the hand easier access to the strings. These necked instruments, which always carry the nut and tuning pegs in a head at the upper extremity of the fingerboard, exhibit a greater variety of form and size than any other species. In many instances the fingerboard is fretted; that is, set with a number of raised markers, usually spaced at semitone intervals along the string so that each marker acts as an auxiliary nut when the string is pressed against it. Fretting is not, as is sometimes supposed, merely a guide to correct intonation in the early stages of learning to play the instrument; rather it is an integral part of the tone and technique of those instruments that use it. Among the common western instruments only the members of the violin family are not fretted; in that family the pressure of the fleshy part of the fingertip on a plain fingerboard seems to provide

exactly the right kind of stop for a bowed string.

In instruments of the harp kind the many strings have to be graded in length and thickness from the lowest to the highest. None of the developed necked instruments derives the whole of its scale merely from stopping a single string all the way up. All employ several strings of the same length, laid close to each other on the fingerboard so that the same stopping technique may be used on all; but they are graded in thickness and tension in such a way that they may be tuned to widely spaced intervals. The number of strings and their tuning varies with particular instruments, but all have a common purpose: to produce the scale, or large parts of it, by a combination of finger-stopping and crossing from one string to the next above or below, the fingered notes bridging the gap or interval between a string and its neighbours. It is also possible, by stopping and sounding selected notes on different strings simultaneously, to produce chordal effects on such an instrument. Normally, the lowest-tuned string lies to the left of the neck, viewed from the front, and the highest to the right, those between being arranged in order of pitch. It is customary to refer to the highest-pitched string as no. 1 and to number the others in descending order.

Methods of Inducing Vibrations.—Musical strings may be set in vibration by applying energy in a variety of ways, and it is of course the precise method employed, more than anything else, that determines the character of the instrument and the kind of sound it produces. Energy may be applied either by plucking or by striking or by subjecting the string to continuous friction or wind pressure.

Plucking.—The first method of inducing vibrations is at once the most primitive and universal. If the string is pulled aside and suddenly released, it will continue vibrating until the initial energy stored by its elasticity is dissipated and it comes to rest. Such a note, consisting of a sharp *ictus* followed by a more or less rapid decay of loudness, does not really "sing," although the ear readily accepts the convention that it does, and both the performer's skill and the maker's craft are directed to sustain the illusion. Moreover, a wide variety of tone colour can be obtained even from a single string by altering the plucking point, or the nature of the material that does the plucking, or the string itself. The player's own fingertips pluck the strings directly in the lutes, guitars and harps and their derivatives, where the strings are mainly of highly elastic catgut and the playing of chords is an everyday occurrence, involving as it does the plucking of several strings at once with different fingers. For many wire-strung instruments, such as the mandolin, a plectrum is used. This is a small lozenge-shaped spatula of hard, flexible material, such as tortoise-shell, held in the player's hand and struck across the strings. (For a description of the keyboard plucking mechanism of the harpsichord *see* the article HARPSICHORD.)

Striking with some form of hammer presupposes that the string shall offer rather more resistance to the blow than strings intended for plucking. Metal strings at comparatively high tension behave best under these conditions and afford the least latitude in the point of striking. What has been said above regarding the *ictus* and decay of the note applies equally here. This system of tone production is not adaptable to instruments with fingerboards, although for completeness it should be said that it has been so applied in one solitary instance, the short-lived English keyed guitar of Christian Clauss (1783). The basic type is rather the dulcimer, a struck version of the older plucked psaltery (*q.v.*). The dulcimer has separate strings (usually duplicated) for every note; the strings are stretched over a flat body or soundboard and are played with hand beaters. This ancient instrument survives in central and east Europe, the Hungarian *cimbalom* and Swiss *hackbrett* being best known. The classic example, however, is of course the pianoforte (*q.v.*) with its elaborate keyboard hammer-mechanism.

Friction or Wind.—The application of continuous friction to the string in order to obtain sustained sound is the principle of the violin bow. The bow, which also is used with such other instruments as the oriental rebab and the European rebec, consists of a flat ribbon of resined horsehair that is kept under tension by a light, specially curved stick. The bow is drawn across the string at right angles to the string and rather near the bridge. The bow will naturally be designed to suit the weight and size of the instrument, and it is of course used in both directions of travel: downstroke when the hand is moving away from the strings and upstroke when moving towards them. Because of the nature of this technique the player has absolute control over the pressure of the bow on the string, its velocity in traveling across it, and of the exact point at which it is applied to the string. All of these factors are employed to vary not only the attack or articulation by which the note is started but also the loudness or softness and actual quality of the tone produced, and these varieties can, if necessary, be introduced within the duration of a single stroke. In its higher flights the bowed technique is probably capable of a wider range of musical effect than any other form of tone production.

A more mechanical application of the friction principle is that of the medieval *organistrum* or *symphony*, later to become known as the hurdy-gurdy (*q.v.*), in which a resined wheel turned by a handle passes over the strings. Nearly all such instruments, which have never died out at the folk level, have a system of touch-pieces for stopping the strings, as opposed to the direct contact with the player's fingers, and are equipped with "drones" (one-note strings tuned to the prime tones of the piece being played) that form a simple and continuous accompaniment to the tune in bagpipe fashion. The hurdy-gurdy principle has been applied to keyboard instruments many times during the past four centuries but has met no general acceptance.

The final method of eliciting tone from strings is by wind pressure. This is the principle of the Aeolian harp (*q.v.*) and of one or two abortive keyboard instruments where a wind jet directed at the string sets it vibrating. Attractive though the resulting sound is, it is too uncontrollable in attack and duration to have any practical application.

See also SOUND: *Vibrations of Strings; the Helmholtz Resonator.* (E. HA.)

STROBL DE KISFALUD, SIGISMUND (1884–), Hungarian sculptor, eminent among 20th-century portraitists, was born at Alsorajk. He studied at the Royal Academy in Budapest, at the Academie Julian in Paris, at Vienna and in Italy. His early work, including the "Bather" at Amsterdam and "Woman With Lizard" at Budapest, was in the strictly classic style and was all in marble.

Later he did much work in contemporary portraits; a distinguishing characteristic of these is that men are in bronze, whereas women and children are in marble. He spent much time in England, where he did busts of George Bernard Shaw, Sir Austen Chamberlain, Lord Allenby, Lady Astor, Princess Elizabeth and other notables.

Among his best-known busts of Hungarian statesmen is that of Count Albert Apponyi. A huge liberation memorial completed in 1948 in Budapest is one of 30 monuments he designed. He had several exhibitions at the Royal Academy in London.

In 1924 he was appointed professor in the Royal Academy of Art, Budapest.

STROBOSCOPE, an instrument that utilizes periodic, intermittent light for the observation of vibrating or rotating mechanical objects. The instrument is usually calibrated to indicate the frequency of vibration or rotation directly, when the rate of light interruption is so adjusted that the observed motion appears to stop. As an instrument for the measurement of speed of rotation, it can be considered a special type of tachometer. Originally a laboratory device, it is now a widely used industrial instrument.

A single flash of light, if short enough, will produce a motionless image of any moving object. With repetitive light flashes, if the number of flashes per minute is adjusted to correspond exactly to the number of revolutions per minute (r.p.m.) in the motion, the subject will be in the same phase of its rotation whenever it is illuminated and will appear to be stationary. Apparent ultra-slow motion will result when the flashing rate and the rate of rotation are slightly different by a small amount. The apparent motion will be forward if the light flash rate is slightly less than

the rotational rate of the object, backward if it is slightly greater.

Stroboscopic observation depends upon persistence of vision in the eye of the observer. Below about 50 cycles per second the observer is conscious of the intermittent nature of the light, but, in spite of the flicker, he can still follow the action down to about 10 cycles per second. Below this he can "see" the object illuminated by each flash if the intensity of the flash is sufficient to outshine the ambient continuous light (and if his eyes are focused upon the desired spot when the flash occurs).

History.—The basic principle of the modern stroboscope is attributed to Joseph Plateau (1836). It had been used earlier in a less precise manner by Michael Faraday, and August Töpler (1866) was among the first to reduce the method to an exact science. In the early researches, particularly those concerned with observing acoustical vibrations, two types of procedures were employed. The first made use of intermittent vision; the second employed interrupted light. In both cases a spinning or vibrating disk equipped with a narrow slit allowed a view of the object at periodic intervals of time, or it permitted light to illuminate the body at successive instants. Thus, the object was exposed to view, or was illuminated, at precisely those times when it passed a given point in its motion.

Research reports and technical literature from the mid-19th century described many types of stroboscopic devices. Plateau's disk with radial slits was used to determine the rotational speed of wheels and shafts. The prongs of the stroboscopic tuning fork were equipped with slotted diaphragms to create periodic interruptions in a beam of light. Rotating and vibrating mirrors were used to produce periodic pulsing of light for stroboscopic observation.

Mechanical shutter types of stroboscopes are particularly well adapted for outdoor use in sunlight and are still used. Most stroboscopes for indoor use now employ the method of successive light flashes from an electronic flash lamp, as explained below.

A direct outcome of the early stroboscopic experiments is the modern motion picture. In fact, the first motion pictures were a series of hand drawings of a subject. These were viewed intermittently through slots that often were a part of the image rotating system. Moving wagon wheels viewed on a motion picture screen provide a familiar example of stroboscopic effects. Here the camera shutter serves to produce intermittent vision. If the shutter frequency is greater than that of the successive passages of spokes, the wheel appears to turn slowly backward, a phenomenon very puzzling to the audience.

Stroboscopic Effects.—Disks bearing symmetrical patterns are often used to illustrate stroboscopic effects. Two such disks, which show interesting and useful illusions, are shown in fig. 1. Fig 1A will show different rows of squares going apparently in different directions at certain ratios of rotation speed and flashing frequency. Any one row will be stationary when the following mathematical ratio holds: $F = kR/n$, where R is the speed of rotation in r.p.m.; F is the frequency of the stroboscope in flashes per minute; k is the number of repetitive patterns such as spokes of a wheel; and n is any integer such as 1, 2, 3, 4, etc.

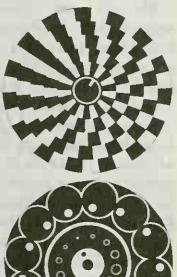


FIG. 1.—DISKS, WHEN ROTATED, SHOW USES OF STROBOSCOPIC LIGHT

Confusion can and often does result, since the stroboscope can give several readings where a rotating shaft appears to be stationary. For example, a shaft will seem stationary for one revolution per flash, for two revolutions per flash, etc. How can the operator know he has $F = R$?

One method is to increase the frequency of the stroboscope slowly, noting the frequencies that show a stationary fundamental pattern. The highest frequency found will correspond to the speed of rotation, since this is one flash per revolution. Any higher frequency will leave a multiple pattern on the shaft. For example, if $F = 2R$, then there will be a double pattern on the shaft, since the stroboscope flashes twice per revolution.

Measurement of speeds above the highest frequency of the stroboscope can be calculated by using the multiple image relationships mentioned above. Let us assume that the highest frequency of the stroboscope that shows a stationary single pattern is F_1 cycles per second. The next lowest stationary-pattern-producing frequency is F_2. Since these have a ratio of $F_1/F_2 = n/n - 1$ where $F_1 = R/n$ then the speed of rotation is $R = F_1F_2/F_1 - F_2$ r.p.m. This method has been of great assistance in determining the speeds of very high frequency repetitive motion without the necessity of using a stroboscope of a similar high speed.

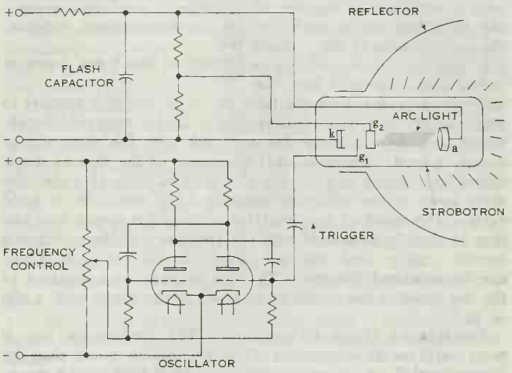


FIG. 2.—SCHEMATIC WIRING DIAGRAM OF A STROBOSCOPIC TACHOMETER. (*see* TEXT)

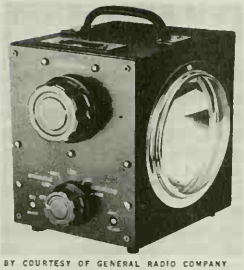
Stroboscope Circuits.—The modern electronic stroboscope employs a low voltage gas-filled discharge tube as a lamp, since this type of lamp can produce a very short flash of brilliant peak light. The lamp is caused to light up intermittently by the discharge of a condenser. Between flashes the condenser voltage builds up its stored energy slowly. The instant at which the flash occurs is controlled by means of a grid in the flash tube. An adjustable electronic oscillator is used to furnish the pulses of grid voltage which control the recurrent discharges in the gas.

A special neon lamp (strobotron) is one type of commonly used gas-filled arc-discharge stroboscope tube (*see* fig. 2). A potential of about 350 v. D.C. is applied to the discharge condenser and electrodes a and k, of the lamp. Except when pulsed, a suppressor grid, g_2, prevents the discharge of the condenser through the gas. Then an oscillator supplies voltage impulses to a starting grid, g_1, at a rate determined by the adjustment of the oscillator frequency. At each impulse the lamp produces a pulse of light that has a duration of about 40 microseconds with a peak light of about 50,000 candle power.

Applications of the Stroboscope.—The stroboscope has a great variety of applications for precise frequency determinations and for observations of moving parts. A familiar demonstration in the physics laboratory or in the study of musical instruments is "freezing" of the motion of a vibrating string to show the number and position of nodes and loops. Thousands of these instruments (fig. 3) are in use in engineering and science laboratories and in industry. One stroboscope has a 2-microsecond flash duration and a speed range of 60–25,000 r.p.m.

Slip Speed Determinations.—A common application of the stroboscope is the determination of the "slip" speed in electrically driven induction motors. The electrical currents in the stationary part of the induction motor causes magnetic fields that rotate in the air gap and link the windings in the rotating part. This magnetic field rotates at a speed of two poles per cycle of the current. For example, a two-pole motor excited from a 60-cycle line will have a field that rotates 60 times per second or 3,600 r.p.m. The speed of the two-pole induction motor rotor, on the

other hand, is less by an amount that depends upon the load. With no load, the rotor speed is very nearly the same as the speed at which the field rotates. As load is added to the motor, the speed decreases. The difference in speed is the slip speed, and it is here that the stroboscope can be used, since it measures this speed directly as described below.



BY COURTESY OF GENERAL RADIO COMPANY

FIG. 3.—STROBOSCOPIC TACHOMETER

Frequency of light is read directly on dial on top

A black circular cardboard disk with a single white spot near the edge is attached to the end of the induction motor shaft for the measurement of slip. When the disk is illuminated by a pulsed source operated from the same 60-cycle line, with one flash per cycle, the disk appears to rotate slowly in a backward direction. If, for example, the observer sees one white spot for each pole pair, the spots will be equally spaced. The rotational frequency of the disk as determined by counting and the use of a stop watch is 36 r.p.m., the shaft speed of the induction motor is 3,600 minus 36 or 3,564 r.p.m. The speed of the shaft for a particular motor load can thus be directly compared with the frequency of the alternating current supply, and the slip speed, in this case 36 r.p.m. can be measured directly. The slip is given as a fraction of the slip speed to the synchronous speed in the above case, a slip of .01.

Synchronous Motor Observations.—The stroboscope also is most useful for the observation of the synchronous motor, since the "power angle" can be seen and photographed both in the steady state and in transient conditions. The rotating part of a synchronous motor rotates at the same speed (synchronously) as the rotating magnetic field set up by the currents from the line. There is a magnetic field that is stationary with respect to the rotating parts which is caused either by a direct current in a winding or by magnetic induction (salient poles) or by permanent magnets. Without load these two fields, one caused by stator currents and the other by rotor currents, line up as a result of forces acting between them. A stroboscope that flashes in phase with the line voltage will show the rotor in a zero phase position in space in alignment with one of the fields. As load is applied, this phase angle (power angle) will increase until the "pullout" torque is reached at which time the motor speed will drop below synchronous speed and the induction effects supply the torque. A disk on the shaft marked in degrees enables the experimenter to read directly the power angle. Photographs taken with a stroboscope on a moving film permit the plotting of data deriving transients where the human observer could not take down the information because of the rapid variations.

Use as Tachometer for Small Motors.—The speeds of fractional horsepower motors cannot be measured by ordinary means, because the load of a conventional tachometer would alter operating conditions. The stroboscopic tachometer, because it requires no mechanical or electrical connection to the motor, is the ideal tachometer for this use. Speed can be measured under any driving or load conditions.

Miscellaneous Uses.—The stroboscope has a wide range of industrial applications. The following are typical examples.

1. Electric motor operation can be studied, particularly as to brush action and vibration.

2. The stroboscope is used in the design of electric fans, particularly to discover sources of vibration. Chemical "smokes" introduced into the air stream permit study of air currents around blades.

3. Vibration patterns produced when a vacuum cleaner vibrates a rug can be studied.

4. In industrial maintenance work, the operation of gears, cams and other moving parts can be studied and misadjustments, misalignment, wear, sources of noise etc., can be detected before damage occurs.

5. The various reciprocating parts of internal combustion engines can be studied when a stroboscope is flashed from a contactor on the crankshaft. To study the action of pistons and other enclosed parts a window in the side of an experimental engine is ordinarily used.

6. The stroboscope is used in the textile industry to adjust a variety of rotating and reciprocating machines for more efficient operation.

7. In radio manufacturing the stroboscope is used to adjust power-supply vibrators. Both frequency and amplitude of vibration can be studied.

8. The effect of vibration on components intended for aircraft use can be observed under simulated service conditions.

Slow Speed Stroboscopy.—Very slow speeds such as those below 600 r.p.m. are quite difficult to measure and observe, since the human eye lacks the ability to retain the image, during the interval between flashes.

For technical studies at slow speeds, which require a high light intensity, a more powerful electronic flash can be operated from the stroboscope. This device has greatly expedited the use of stroboscopic methods at low speeds, as, for instance, in the observation of textile looms where the shuttle may make only one or two picks per second. An observer can see the shuttle with a single flash of light, but he must be in a dark environment and must have his eyes directed at the subject when the flash occurs.

Photography With a Stroboscope.—The brilliant, short-duration flash produced by the gas-discharge lamp developed for stroboscopic use is admirably suited for the photographing of rapidly moving objects. Single flashes of one microsecond or less in scientific photography, and, in usual photography, flash durations of one millisecond, are common. High-speed, single-flash photography is popularly called "strobe" photography and the lamps "strobe" lights, although no stroboscopic effect is involved other than stopping motion with a single flash.

With the camera shutter open, a series of exposures, resulting from successive flashes of a strobe light, can be made. Thus a number of superimposed images on a single sheet of film show various stages of a continuous motion such as a golfer's swing. Here again, no stroboscopic effect is involved. *See* also PHOTOGRAPHY: *Stroboscopic Photography: Electronic Flash.*

(H. E. ED.)

STRODE, RALPH (*fl.* 1350–1400), English schoolman, was probably a native of the West Midlands. He was a fellow of Merton college, Oxford, before 1360, and famous as a teacher of logic and philosophy and a writer on educational subjects. He belonged, like Thomas Aquinas and Bonaventura, to that "School of the Middle" which mediated between realists and nominalists. Besides his *Logica,* which has not survived, he wrote *Consequentiae,* a treatise on the syllogism, and *Obligationes* or *Scholastica militia,* a series of "formal exercises in scholastic dialectics."

He had some controversy with his colleague John Wycliffe, against whom he defended the possession of wealth by the clergy, and whose doctrine of predestination he attacked. Strode is also associated with John Gower in Chaucer's dedication of *Troylus and Cryseyde,* and Strode himself, according to the 15th-century *Vetus catalogus* of fellows of Merton, was a "poeta nobilis."

John Leland and John Bale confirmed this testimony, and I. Gollancz suggested the identification of the *Phantasma Rodulphi* attributed to Strode in the *Vetus catalogus* with the beautiful 14th-century elegiac poem *The Pearl.* If this be true, Strode wrote also *Cleanness, Patience,* and *Sir Gawayne and the Green Knight.* From 1375 to 1385 this Strode or another of the same name was common sergeant of the city of London and died in 1387.

STRODE, WILLIAM (1598–1645), English parliamentarian, second son of Sir William Strode, of Newnham, Devonshire (a member of an ancient family long established in that county, which became extinct in 1897). He was admitted as a student of the Inner Temple in 1614, matriculated at Exeter college, Oxford, in 1617 and took the degree of bachelor of arts in 1619. He was returned to parliament in 1624 for Beeralston, and represented the borough in all succeeding parliaments until his death. He from the first threw himself into opposition to Charles I and took a leading part in the scene of March 2, 1629, when the Speaker, Sir John Finch, was held down in the chair. (*See* HOLLES, DENZIL HOLLES.)

Prosecuted before the Star Chamber, he refused "to answer anything done in the House of Parliament but in that House." On

May 7, a fresh warrant was issued, and a month later, to prevent his release on bail, he was sent by Charles with two of his fellow members to the Tower. Refusing to give a bond for his good behaviour, he was sentenced to imprisonment during the king's pleasure, and was kept in various prisons for 11 years.

In Jan. 1640, in accordance with the king's new policy of moderation, he was liberated; and on April 13, took his seat in the Short Parliament, with a mind embittered by the sense of his wrongs. In the Long Parliament, which met on Nov. 3, 1640, he was the first to propose the control by parliament over ministerial appointments, the militia and its own duration; supported the Grand Remonstrance of Nov. 7, 1641; and displayed a violent zeal in pursuing the prosecution of Strafford, actually proposing that all who appeared as the prisoner's counsel should be "charged as conspirators in the same treason." He was included among the five members impeached by Charles of high treason on Jan. 3, 1642. He opposed all suggestions of compromise with Charles, urged on the preparations for war, and on Oct. 23, was present at the battle of Edgehill. In the prosecution of Laud he showed the same relentless zeal as he had in that of Strafford, and it was he who, on Nov. 28, 1644, carried up the message from the commons to the lords, desiring them to hasten on the ordinance for the archbishop's execution. Strode did not long survive his victim. He is mentioned as having been elected a member of the assembly of divines on Jan. 31, 1645. He died Sept. 9, 1645, and by order of parliament was accorded a public funeral in Westminster abbey. The body was exhumed after the Restoration.

Clarendon speaks of Strode as a man "of low account and esteem" who gained his reputation by his accidental association with those greater than himself; but to his own party his "insuperable constancie" gave him a title to rank with those who had deserved best of their country at a time when the liberties of England hung in the balance.

STROKE (Apoplexy) is the term used to describe the combination of signs and symptoms produced by damage to a portion of the brain as result of or in conjunction with at least one of the following four events:

(1) A blood clot forms within a blood vessel of the brain (thrombosis). This is the commonest cause of strokes.

(2) A blood clot lodges in an artery supplying brain tissue after originating in another portion of the body (usually a damaged heart) and traveling to the brain. This is known as an embolism. Both types of clotting reduce or stop the flow of blood to brain cells. During the early stages of apoplexy from these two causes, the illness may be further complicated by the oozing of blood and fluid into the surrounding areas (edema).

(3) Intermittent insufficiency of specific arteries results temporarily due to spasm of the arteries or sludging of the blood as it passes through segments of the vessels narrowed by atherosclerosis (arteriosclerosis).

(4) Hemorrhage occurs after an artery ruptures, usually as result of a weakening of the arterial wall due to atherosclerosis or due to a thinning of the wall with bulging (an aneurysm), which may be congenital or develop later in life.

Consequences.—The initial onset may be massive in its effects, producing widespread paralysis, inability to speak, coma or death within a short time, usually within several hours or days. On the other hand the onset may be manifested by a series of transient "little strokes" during which the patient may experience weakness and numbness of an arm, leg or the side of his face. There may be temporary difficulty in speech, confusion of thought or visual disturbances. These may recur many times, but eventually they are usually followed by more widespread and permanent paralysis.

The groups of muscles involved are a direct reflection of the artery and brain tissues involved. Most commonly the left side of the brain (dominant side for most persons) is affected, producing paralysis of the right side of the body because most of the nerves cross to the opposite side of the body from their origin in the brain. However, the combinations of signs and symptoms are innumerable and cannot be described in detail here.

Minute strokes may develop as result of closure of many tiny arterioles and produce gradual mental and physical deterioration.

Hemorrhagic strokes are commonly associated with high blood pressure. They frequently produce massive neurological signs, and death is commoner with them than with thrombo-embolic strokes. However, hemorrhagic strokes may be minute and recurrent, gradually producing widespread damage.

Treatment.—Precise history and physical examination, especially for neurological changes, are essential to differentiate apoplexy from tumours and brain injury from other causes. It is also important to determine whether the stroke is due to a thrombosis or embolism on the one hand, or hemorrhage on the other. Examination of the spinal fluid for evidence of blood is essential. X-ray studies following the injection of radiopaque substances may clarify the diagnosis. Establishing a differential diagnosis has become essential because anticoagulant drugs are now widely used in the treatment of strokes due to thromboses or emboli but are contraindicated when due to hemorrhage. A stroke may have both a clotting and a significant hemorrhagic factor present, and this presents difficulties.

Many strokes are due to closure of one of the two carotid arteries that supply the brain after passing up the sides of the neck from the aorta. If the closure involves only a small segment, surgery may be attempted to remove the obstruction or to place a graft or bypass using one of the synthetic materials such as Dacron.

Early and persistent efforts for rehabilitation are essential. These should begin within a day or two after the stroke so that muscle tendon contractures will not occur.

For most patients apoplexy or a stroke should not be considered as an isolated disease but as a chapter in the natural history of the over-all condition of atherosclerosis that tends to involve other organs. However, the outlook has been improved by modern treatment.

See Arteries, Diseases of; Brain: *Nutrition of the Brain;* Paralysis: *Diseases That Produce Paralysis;* Rehabilitation, Medical and Vocational; Thrombosis and Embolism.

Bibliography.—E. Hugh Luckey and Irving S. Wright, *Cerebral Vascular Diseases,* Transactions of First Conference (1956); Clark H. Millikan and Irving S. Wright, *Cerebral Vascular Diseases,* Transactions of Second Conference (1958); Clark H. Millikan, Robert G. Siekert and J. Whisnant, "Anticoagulant Drugs in Cerebral Vascular Disease: Current Status," *J.A.M.A.,* 166:587 (Feb. 8, 1958); Irving S. Wright, "The Pathogenesis, Diagnosis, and Treatment of Strokes: a Progress Report," *Ann. Intern. Med.,* 49:1004 (Nov. 1958). (I. S. W.)

STROMNESS, a small burgh and fishing port on the island of Mainland (Pomona), county of Orkney, Scot. Pop. (1961) 1,477. It is situated on the west side of a well-sheltered bay opening off Hoy sound. Many of the houses are within tidal limits and are furnished with quays and jetties. The harbour admits vessels of all sizes. There is an egg-packing station, and eggs are exported to the Scottish mainland. The neighbourhood suffered much damage from the great storm of Jan. 1953.

Stromness is a centre for visiting the magnificent scenery of the west coast of Mainland, including Black craig (400 ft.) and the Hole of Row, a natural arch carved out by the ocean. At the southern edge of the bay of Skaill lies the prehistoric village of Skara Brae (*q.v.*).

STRONG, EUGENIE (nee Sellers) (1860–1943), English archaeologist, married in 1897 the late S. A. Strong, librarian to the House of Lords. She was educated in France and at Girton college, Cambridge, of which she became the first life fellow in 1910. She lectured on Greek art and archaeology in London, Rome and Dublin, and in 1913 delivered the Norton lectures for the Archaeological Institute of America. In 1904 she became librarian to the duke of Devonshire, in succession to her husband, and in 1909 was appointed assistant director of the British School of Archaeology in Rome, a post which she held until 1925. She was elected fellow of the Society of Antiquaries, 1920; a member of the Accademia Ponteficia of Rome, 1921; of the Reale Accademia dei Lincei in 1928 and the Reale Accademia di San Luca in 1932. She was also elected to various other German, Italian and British societies.

In 1938 she was awarded the Serena gold medal for Italian

studies by the British academy, and in the same year she received the gold medal of the city of Rome.

Her publications include *Roman Sculpture from Augustus and Constantine* (1907; new ed. in Italian, 1923); *Apotheosis and After Life* (1915); *Art in Rome* (1927). She was also the English editor of G. Schuchardt's *Schliemann's Excavations*, Furtwängler's *Masterpieces of Greek Sculpture* and F. Wickhoff's *Roman Art*.

Eugenie Strong died at Rome in Sept. 1943.

STRONG, LEONARD ALFRED GEORGE (1896–1958), British author and journalist, was born March 8, 1896, at Plympton, Devon, and received his education at Brighton college and at Wadham college, Oxford, where he was awarded his bachelor's degree in 1920. His first work was in verse (*Dublin Days*, *The Lowery Road*, *Difficult Love*) and showed the influence of William Butler Yeats.

Hist first novel was *Dewer Rides* (1929), a story laid in Dartmoor; later books included *The Jealous Ghost* (1930), *The Garden* (1931), *The Brothers* (1932), *Sea Wall* (1933), *The Open Sky* (1939), *The Bay* (1941), *The Director* (1944) and *Trevannion* (1948). His last novel was *Deliverance* (1955).

Perhaps best known as a writer of short stories, many of which he successfully adapted for broadcasting, Strong published several collections, including *Doyle's Rock* (1925), *Don Juan and the Wheelbarrow* (1932), *Travellers* (1945) and *Darling Tom* (1952). For *Travellers* he received the James Tait Black memorial prize. His critical works include *Sacred River: an Approach to James Joyce* (1951) and as a supporter and adviser to other writers. *English for Pleasure* (1942) and *The Writer's Trade* (1953). Other non-fiction works include *Flying Angel* (1956) and *The Rolling Road* (1956).

For 12 years he taught English and classics at Summer Fields, Oxford, and in 1938 he became a director of Methuen Ltd., a publishing house.

Strong died at Guildford on Aug. 17, 1958.

STRONG ROOMS: *see* SAFES, STRONG ROOMS AND VAULTS.

STRONGYLION, a Greek sculptor. His bronze figure of a horse, representing the wooden horse of Troy with the Greek heroes inside and looking forth, was set up on the Acropolis of Athens late in the 5th century B.C. Other works of Strongylion were a figure of Artemis at Megara, a group of the Muses and an Amazon which was greatly admired by the emperor Nero.

STRONTIANITE, a mineral consisting of strontium carbonate and a principal source of strontium. It takes its name from Strontian in Argyllshire, where it appears to have been known as far back as 1764, but it was not recognized as a distinct mineral until later, when the examination of it led to the discovery of the element strontium. The formula is $SrCO_3$. It crystallizes in the orthorhombic system and is isomorphous with aragonite and witherite. Distinctly developed crystals are, however, of rare occurrence; they are usually acicular with acute pyramid planes and are repeatedly twinned on the prism. Radiating, fibrous or granular aggregates are more common. The colour is white, pale green or yellowish brown. The hardness is 3.5 and the specific gravity 3.7. The mineral occurs in metalliferous veins in the lead mines of Strontian in Argyllshire, abundantly in veins in calcareous marl near Münster and Hamm in Westphalia, in limestone near Barstow, San Bernardino county, California, and in geodes at Schoharie, N.Y. It is also found in Washington, Arizona, Texas and other states.

It is used for producing red fire in highway and railway fusees, marine distress signals and pyrotechny and for refining sugar.

(CL. F.; X.)

STRONTIUM, a metallic element belonging to the alkaline earth family of elements in group two of the periodic table. Strontium (symbol Sr, atomic weight 87.63, atomic number 38) composes about .00019% of the igneous rocks and about .02% of the entire crust of the earth. It is distributed in small quantities in many different rocks and soils. Some is present in bones, in which it replaces small amounts of calcium. Its principal sources are strontianite, $SrCO_3$, and celestite, $SrSO_4$.

William Cruikshank, in 1787, first detected the existence of the element in the strontianite found at Strontian in Argyllshire. The metal was isolated in 1808 by Sir Humphry Davy, who electrolyzed a mixture of the moist hydroxide or chloride with mercuric oxide, using a mercury cathode. It was obtained in a state of purity by A. Guntz and Roederer by heating the hydride in a vacuum to 1,000° C. It may be obtained in the form of sticks by the "contact cathode" method, in which a cooled iron rod, acting as a cathode, just touches the surface of a fused mixture of potassium and strontium chlorides and is raised as the strontium collects on it. Because calcium and barium, which it resembles closely, occur in much greater abundance, strontium is not produced in commercially important quantities.

There are four stable isotopes of the element; their mass numbers, in the order of abundance, are 88, 86, 87 and 84. Several radioactive isotopes are produced by nuclear reactions. Of these, the isotope having a mass number of 90 has received most attention because of its presence in the radioactive fall-out from atomic explosions. The distribution of electrons in the levels and sublevels of the strontium atom is: $1s^2$, $2s^2$, $2p^6$, $3s^2$, $3p^6$, $3d^{10}$, $4s^2$, $4p^6$, $5s^2$.

Properties.—Strontium has a silvery white colour. It is malleable and ductile and acts as a good conductor of electricity. Other physical properties are listed below:

Atomic radius	2.13 Å
Ionic radius	1.13 Å
Ionization potential (gaseous element) 1st electron.	5.69 v.
2nd electron.	10.98 v.
Single electrode potential (between metal and molal solution of Sr^{++} ion)	2.89 v.
Melting point	757 ± 1° C.
Boiling point	1,366° C.
Density	2.60 g. per cu.cm.

The chemical properties of strontium are similar to those of calcium and barium, in keeping with its position in the periodic table, where it is above barium and below calcium. Generally speaking, its base-forming characteristics are somewhat less pronounced than those of barium but more pronounced than those of calcium. In its compounds, the element has a valence number of +2. As indicated by its atomic radius, ionization potential and single electrode potential, the atom of strontium easily loses the two electrons in the 5s level when it reacts with nonmetallic elements, thus forming the Sr^{++} ion. The element is an active reducing agent. It reacts readily with water to form hydrogen and the hydroxide, $Sr(OH)_2$. It oxidizes rapidly when exposed to the air and burns brilliantly when heated in air, oxygen, chlorine, the vapour of bromine or the vapour of sulfur. With oxygen, the element forms the monoxide, SrO, and the peroxide, SrO_2.

Compounds.—Compounds of strontium are not so extensively used as those of calcium or barium. The solubilities of the salts of the element, generally speaking, are intermediate with respect to the solubilities of the corresponding salts of calcium and barium.

The hydride, SrH_2, obtained by Guntz on heating strontium amalgam in a current of hydrogen, is a white solid, which readily decomposes water in the cold and behaves as a strong reducing agent. The monoxide or strontia, SrO, is formed by strongly heating the nitrate, or commercially from the hydroxide which is produced by heating the sulfide or carbonate in superheated steam (at about 500°–600° C.). It is a white powder which resembles lime in its general character. By heating the powder in the electric furnace, H. Moissan succeeded in obtaining crystals of the oxide. The compound readily slakes with water, and the aqueous solution yields a crystalline hydrated hydroxide. It is used in the extraction of sugar from molasses, since it combines with the sugar to form a soluble saccharate, which is removed and then decomposed by carbon dioxide. A hydrated peroxide, approximating in composition to $SrO_2.8H_2O$, is formed as a crystalline precipitate when alkali is added to an aqueous solution of a strontium salt containing hydrogen peroxide. If all the solutions are above 50° C. and concentrated, the anhydrous peroxide results.

Strontium fluoride, SrF_2, is obtained by the action of hydrofluoric acid on the carbonate, or by the addition of potassium fluoride to strontium chloride solution. Strontium chloride,

$SrCl_2.6H_2O$, is obtained by dissolving the carbonate in hydrochloric acid or, commercially, by fusing the carbonate with calcium chloride and extracting the melt with water. It crystallizes in small colourless needles and is easily soluble in water; the concentrated aqueous solution dissolves bromine and iodine readily. By concentrating the aqueous solution between 90° and 130° C. or by passing hydrogen chloride into a saturated aqueous solution, a second hydrated form of composition, $SrCl_2.2H_2O$, corresponding with dihydrated barium and radium chlorides, is obtained. The anhydrous chloride is formed by heating the hydrated chloride in a current of hydrogen chloride.

Strontium sulfide, SrS, is formed when the carbonate is heated to redness in a stream of hydrogen sulfide. It phosphoresces slightly when pure. Strontium sulfate, $SrSO_4$, found as the mineral celestite, is formed when sulfuric acid or a soluble sulfate is added to a solution of a strontium salt. It is a colourless solid, which is almost insoluble in water, its solubility diminishing with increasing temperature; it is appreciably soluble in concentrated sulfuric acid. When boiled with alkaline carbonates, it is converted into strontium carbonate.

Strontium nitride, Sr_3N_2, is formed when strontium amalgam is heated to redness in a stream of nitrogen or by igniting the oxide with magnesium. It is readily decomposed by water, with liberation of ammonia. Strontium nitrate, $Sr(NO_3)_2$, is obtained by dissolving the carbonate in dilute nitric acid. It crystallizes from water (in which it is soluble) in monoclinic prisms which approximate in composition to $Sr(NO_3)_2.4H_2O$. It is used in pyrotechny for the manufacture of red fire.

Strontium carbide, SrC_2, is obtained by heating strontium carbonate with carbon in the electric furnace. It resembles calcium carbide, decomposing rapidly with water, giving acetylene. Strontium carbonate, $SrCO_3$, found in the mineral kingdom as strontianite, is formed when a solution of a carbonate is added to one of a strontium salt. It loses carbon dioxide when heated above 1,250° C.

Analysis.—Strontium ion is colourless, almost tasteless and feebly toxic. Volatile strontium salts impart a carmine colour to the flame. The spectrum of the element is characterized by the two prominent lines corresponding to wave lengths of 422 and 461 $\mu\mu$ (10^{-6} mm.).

In the usual schemes of analysis, strontium is separated, along with calcium, barium, magnesium and the alkali metals, by making use of the fact that none of these ions is precipitated by sulfide ion in either an acidic or an alkaline solution.

After precipitation of all other positive ions by the sulfide ion, calcium, strontium and barium are precipitated as carbonates in the presence of ammonium ion. The precipitate is then dissolved in acetic acid; and, in the usual procedure, barium is then separated by precipitation as the chromate, leaving strontium and calcium ions in solution. Strontium ion can then be separated from calcium by the addition of a dilute solution of sulfate ion which, because of the considerably greater solubility of calcium sulfate, precipitates strontium but not calcium. *See* also Index references under "Strontium" in the Index volume.

BIBLIOGRAPHY.—J. W. Mellor, *Comprehensive Treatise on Inorganic Chemistry*, 16 vol. (1922–37); H. Remy, *Treatise on Inorganic Chemistry*, 2 vol. (1956). W. M. Latimer and J. H. Hildebrand, *Reference Book of Inorganic Chemistry*, 3rd ed. (1951); F. Ephraim, *Inorganic Chemistry*, P. C. L. Thorne and E. R. Roberts (eds.), 4th rev. ed. (1943). (H. T. BE.)

STROPHANTHUS, a genus of poisonous plants of the dogbane family (Apocynaceae), derives its name from the long, twisted threadlike segments of the corolla, which in one species attain a length of 12 to 14 inches. The genus comprises about 28 species, mainly of tropical Africa but extending into South Africa, with a few species in Asia, from eastern India to the Philippines and China. The seeds of the plant contain the biologically active materials. Some African tribes made a poison for arrows by pounding the seeds into an oily reddish mass, which was then smeared on the arrow immediately behind the barb.

The chief active principles in these plants are cardiac glycosides called strophanthins. The strophanthins obtained from the various species differ somewhat in chemical structure, for this reason they are named according to their source. Thus K-strophanthin and G-strophanthin are the glycosides obtained from *S. kombe* and *S. gratus* respectively. G-strophanthin, also called ouabain, is an official drug in the British and U.S. pharmacoepeias. Ouabain was originally applied to a poison derived from the bark of another plant. It was later found that this poison was identical with G-strophanthin. It is a pure, crystalline substance, whereas K-strophanthin is an amorphous preparation. The biological activity of both substances is similar, but in K-strophanthin the activity varies according to the composition.

Strophanthus sarmentosus has been found to be a source of cortisone.

The strophanthins have a digitalislike action on the heart (*see* DIGITALIS). By a direct action on the heart muscle fibres they increase the force of contraction and it is this property that accounts for their value in the treatment of heart failure. Other muscles of the body are not affected. Secondary to the increased force of contraction of the heart muscle the strophanthins cause a decrease in venous pressure and a slower and more regular heart beat in patients with cardiac failure.

G-strophanthin is used for the treatment of heart failure. Intravenous injection is preferred in order to produce a rapid and brief stimulation or in cases of heart failure where vomiting prevents taking digitalis by mouth. When given by mouth the drug is apt to be destroyed in the intestine; its action by this route is therefore erratic.

Toxic doses of the strophanthins cause nausea, vomiting and gastrointestinal distress; headache and nervous irritability may precede the gastrointestinal effects. Greenish-yellow vision is a frequent and almost certain symptom of poisoning. The heart becomes hyperirritable and ventricular fibrillation and death may result. (K. P. DU.)

STROPHE, a term in versification which properly means a turn, as from one foot to another, or from one side of a chorus to the other (Gr. *strophē*, "a turning"). Hence, the song of a chorus during one turn or figure of a dance, and thus a collection of various prosodical periods combined into a structural unit. When the strophe is answered by another unit of identical scansion (called an antistrophe) and accompanied by a third of somewhat different structure (epode), the three form a period.

STROSSMAYER, JOSEPH GEORGE (JOSIP JURAJ STROSSMAJER) (1815–1905), Croatian Roman Catholic bishop and leader of the Pan-Slav movement, was born at Esseg in Croatia-Slavonia on Feb. 4, 1815, of German parentage. He became lecturer on theology at Vienna, chaplain to the Austrian emperor and director of the Augustinian body at Rome. In 1850 he was consecrated bishop of Djakovo, with the official title bishop of Bosnia and Sirmium. Strossmayer fostered the growth of Slavonic nationalism in Croatia-Slavonia, in Dalmatia and among the Slovenes of south Austria, aiding the Ban Josef Jellacic (*q.v.*) in his campaigns against Hungary (1848–49) and subsequently becoming a recognized leader of the opposition to Hungarian predominance. (*See* CROATIA-SLAVONIA.) Besides being foremost among the founders of the South Slavonic academy in 1867, and of Agram university in 1874, he helped to reorganize the whole educational system of Dalmatia and Croatia-Slavonia. He built a palace and cathedral at Djakovo, founded a seminary for the Bosnian Croats, presented the South Slavonic academy with a gallery of valuable pictures and published collections of national songs and tales.

Strossmayer also aided Augustin Theiner, then the librarian at the Vatican, to compile his *Vetera monumenta Slavorum meridionalium historiam illustrantia* (1863). At the Vatican council of 1869–70 he was firmly opposed to the definition of papal infallibility. Strossmayer died on April 8, 1905.

STROZZI, the name of an ancient and noble Florentine family, which was already famous in the 14th century. Palla Strozzi (1372–1462) founded the first public library in Florence in the monastery of Santa Trinita. Filippo Strozzi il Vecchio (1426–1491), son of Matteo and of Alessandra Macinghi, began to build the beautiful Strozzi palace in Florence. Another Filippo Strozzi (1488–1538), who, although married to a Medici, opposed

the hegemony of that house and was one of the leaders of the rising of 1527. On the final overthrow of the republic in 1530 Alessandro de' Medici attempted to win over Filippo Strozzi, but Strozzi had no faith in the tyrant and retired to Venice. After the murder of Alessandro he undertook the leadership of a band of republican exiles with the object of re-entering the city (1537), but, having been defeated and captured and put to the torture, he committed suicide. His son Leone (1515–1554) was a distinguished admiral in the service of France and fought against the Medici; he died of a wound received while attacking Sarlino. The Strozzi palace, which belonged to the family until 1907, was bequeathed by will to the Italian nation.

BIBLIOGRAPHY.—A. Bardi, *Filippo Strozzi* (1894); B. Niccolini, *Filippo Strozzi;* C. Guasti, *Le Carte Strozziane* (1884–91).

STRUCTURAL ENGINEERING: *see* ARCHITECTURAL ENGINEERING.

STRUENSEE, JOHAN FREDERICK, COUNT (1737–1772), Danish statesman, was born at Halle on Aug. 5, 1737. His father, subsequently superintendent general of Schleswig-Holstein, was a rigid pietist; but young Struensee, who settled down in the 1760s as a doctor at Altona, revolted against his father's narrow creed and became a fanatical propagandist of the doctrines of the encyclopaedists. In 1768 he became court physician to Christian VII and rapidly gained authority in affairs.

For a time Struensee kept himself discreetly in the background, but he soon grew impatient of his puppets. He dismissed the heads of departments, and abolished the Norwegian stadholderships. Henceforth the cabinet, with himself as its motive power, was to be the one supreme authority in the state. He then began to reform the administration and the law without regard to national customs and predilections. Between March 29, 1771, and Jan. 16, 1772—the ten months during which he held absolute sway—he issued no fewer than 1,069 cabinet orders, or more than three a day. In order to be sure of obedience he dismissed wholesale without pension or compensation the staffs of all the public departments, substituting nominees of his own.

The general ill will against Struensee, which had been smouldering all through the autumn of 1771, led to a conspiracy against him, headed by Rantzau-Ascheburg and others, in the name of the queen-dowager Juliana Maria. On Jan. 17, 1772, Struensee, E. Brandt, the unhappy king's keeper, and Queen Caroline were arrested, and "the liberation of the king," who was driven round Copenhagen by his deliverers in a gold carriage, was received with universal rejoicing.

The interrogation of Struensee began on Feb. 20, and on Feb. 22 he confessed that he had been Queen Caroline's lover. On April 25 he and Brandt were condemned first to lose their right hands and then to be beheaded; their bodies were afterward to be drawn and quartered. The sentences were carried out on April 28, Brandt suffering first.

BIBLIOGRAPHY.—Elie Salomon François Reverdil, *Struensee et la cour de Copenhague 1760–1772* (1858); Karl Wittich, *Struensee* (1879); Peter Edward Holm, *Danmark-Norges Historie,* vol. iv (Copenhagen, 1897–1905); Gustave Bascle de Lagrèze, *La Reine Caroline-Mathilde et le Comte Struensee* (1887); Robert Nisbet Bain, *Scandinavia,* cap. xv (1905); William Henry Wilkins, *A Queen of Tears* (1904); Georg Friedrich von Jenssen-Tusch, *Die Verschwörung gegen die Königin Karoline Mathilde und die Grafen Struensee und Brandt* (1864).

STRUTHIONIDAE: *see* OSTRICH.

STRUTT, the name of two British manufacturers and inventors.

JEDEDIAH STRUTT (1726–1797), British manufacturer and inventor of the Derby hosiery frame, was born at South Normanton, Derbyshire, on July 25, 1726. He was educated at a country school with a view to becoming a farmer, but he showed mechanical aptitude and was apprenticed in 1740 to a wheelwright at Findern, Derbyshire. During his apprenticeship he inherited a farm and it was while occupying this that his attention was brought, by his brother-in-law, William Woollatt, a hosier, to the unsuccessful efforts to produce ribbed as well as plain goods on the stocking frame. Strutt invented a ribbing machine which could be attached to the stocking frame and used whenever ribbed work was required. Patents were taken out by Strutt and Woollatt in 1758 and 1759, a factory was started at Derby and, in 1762, Samuel Need was taken in as a partner. In 1768 Richard Arkwright (*q.v.*) consulted Need as to the possibilities of his cotton-spinning frame. Strutt at once realized its value and solved one or two minor difficulties which had interrupted the smooth working of the new mechanism. The firm of Arkwright, Strutt and Need started its first cotton mill at Nottingham and later works were erected at Cromford and Belper. After dissolving the partnership with Arkwright in 1781, Strutt retained the Belper works. He died at Derby on May 6, 1797.

WILLIAM STRUTT (1756–1830), Jedediah's eldest son, born in 1756, was one of the pioneers of the iron-framed building. In 1792 he designed a calico factory at Derby which was the first multistory fireproof building with an integrated frame structure. He also studied the house-heating problem, inventing the Belper stove in 1806. A fellow of the Royal society, he died at Derby on Dec. 29, 1830.

See R. S. Fitton and A. P. Wadsworth, *The Strutts and the Arkwrights 1758–1830* (1959). (K. R. G.)

STRUVE, FRIEDRICH GEORG WILHELM VON (1793–1864), German astronomer, who made a systematic survey of the whole available sky for the discovery and measurement of visual double stars, was born at Altona on April 15, 1793. He attended the University of Dorpat, and was appointed to a professorship of astronomy and mathematics in 1813. He was summoned by Emperor Nicholas I of Russia in 1835 to superintend the building of the Pulkowa observatory, near St. Petersburg; he became its director in 1839. He died on Nov. 23, 1864.

Struve was a highly skillful observer. The catalogue in which his results were published, *Mensurae Micrometricae,* is one of the classics of double-star astronomy. (W. W. M.)

STRUVE, OTTO (1897–1963), Russian-born U.S. astronomer who made some of the major astronomical discoveries of the 20th century, was born at Kharkov, on Aug. 12, 1897. A fourth-generation member of a distinguished astronomical family (his great-grandfather was F. G. W. von Struve [*q.v.*]), he graduated from the University of Kharkov in 1919 and settled in the United States in 1921. He received the Ph.D. degree from The University of Chicago in 1923 and continued as a member of the staff of Yerkes observatory, where he was director (1932–47). He became a U.S. citizen in 1927. He was director of the Leuschner observatory of the University of California (1950–59), after which he became director of the National Radio Astronomy observatory in Green Bank, W.Va. He died in Berkeley, Calif., on April 6, 1963.

Struve occupied a unique place in modern astronomy with respect both to his own discoveries and to the effect he had on other investigators. He was the first to demonstrate that a considerable percentage of stars of high temperature rotate rapidly on their axes. Struve showed that some blue stars have rotation periods as short as a day or less. Perhaps his greatest effect on astronomy was due to work on the nature and distribution of the gas located between the stars. Some of his earlier investigations showed that the interstellar calcium spectral lines strengthen in intensity with increasing distance. Somewhat later he discovered the large-scale distribution of hydrogen between the stars, a discovery that was basic to mid-20th-century advances in radio astronomy and in galactic structure. (W. W. M.)

STRYCHNINE, an alkaloid discovered in 1818 in St. Ignatius' beans (*Strychnos ignatii*); it also occurs in other species of *Strychnos—e.g., S. nux-vomica, S. colubrina, S. tieute*—and is generally accompanied by brucine (dimethoxy strychnine). Strychnine crystallizes from alcohol in colourless prisms, which are practically insoluble in water and soluble with difficulty in the common organic solvents. It has an exceptionally bitter taste. Strychnine is a tertiary monoacid base giving an alkaline reaction.

In Medicine.—The 1955 *British Pharmacopoeia* included strychnine hydrochloride and several preparations of nux vomica. The latter is prepared from seeds of that species. No preparations of strychnine were listed in the 1955 *United States Pharmacopoeia,* although several were included in the 1955 *National Formulary.*

Strychnine rapidly enters the blood, whether given orally or

hypodermically. In small doses it acts as a bitter, increasing the secretion of gastric juice. The specific effects of the drug, however, are upon the central nervous system. It excites the motor areas of the spinal cord and increases their reflex irritability. Small doses increase the sensibility of touch, sight and hearing; large doses cause twitching of the muscles and difficulty in swallowing. In overdoses violent convulsions are produced. All portions of the central nervous system are stimulated but the predominant effects are localized in the spinal cord. Not only is the respiratory centre stimulated, but the vasomotor centre is acted upon both directly by the drug and indirectly by the convulsion and asphyxia. Ordinary doses have no effect upon the temperature, but in overdose the temperature rises during a convulsion. Strychnine is partly detoxified in the liver and partly excreted unchanged by the kidneys.

Strychnine has no consistently beneficial effect on the cardiovascular system or on the gastrointestinal tract.

Toxicology.—The symptoms of strychnine poisoning usually appear within 20 minutes of the ingestion of a poisonous dose, starting with stiffness at the back of the neck, twitching of the muscles and a feeling of impending suffocation. The patient is then seized with violent tetanic convulsions characterized by opisthotonos, in which the body is arched and the head bent backward. After a minute the muscles relax, and the patient sinks back exhausted, heightened perceptiveness being preserved throughout due to sensory cortex stimulation.

Strychnine has been used in rodent poisons and in cathartic preparations, and these uses have led to considerable accidental poisoning in children. Suicidal use of the drug has been wide, although other agents have been held in more favour. In treating strychnine poisoning, antidotes such as charcoal or egg white should be administered promptly. Emetics should be avoided as they tend to precipitate convulsions, as does morphine. The patient should be kept quiet and protected from external stimuli. The convulsions can be controlled best by intravenous barbiturates or general anesthesia. Artificial respiration may be necessary. *See* also Alkaloids; Brucine; Nux Vomica; Poison.

Bibliography.—Frank E. Hamerslag, *The Technology and Chemistry of Alkaloids* (1950); Louis S. Goodman and Alfred Gilman, *The Pharmacological Basis of Therapeutics* (1956); Torald A. Sollman, *A Manual of Pharmacology and Its Application to Therapeutics and Toxicology* (1949); H. L. Holmes, "The Strychnos Alkaloids," in R. H. F. Manske and H. L. Holmes, *The Alkaloids* (1950). (W. J. V.)

STRYPE, JOHN (1643–1737), English historian and biographer, was born in Houndsditch, London, on Nov. 1, 1643. He was the son of John Strype, or van Stryp, a member of a Flemish family settled in Strype's Yard in Petticoat Lane, as a merchant and silk thrower. The younger John was educated at St. Paul's school, and at Jesus college, and Catherine Hall, Cambridge. In 1670 he became perpetual curate of Theydon Bois, Essex, and subsequently received the curacy of Leyton and a sinecure living in Sussex. He was lecturer at Hackney from 1689 till 1724. He died at Hackney on Dec. 11, 1737. He was buried in the church at Leyton.

The most important of Strype's works are the *Memorials of Thomas Cranmer, Archbishop of Canterbury, 1694*, ed. for the Eccl. Hist. Soc., 3 vol. (Oxford, 1848–1854), and in 2 vol. with notes by P. E. Barnes (London, 1853); *Life of the Learned Sir Thomas Smith* (1698); *Life and Acts of John Aylmer, Lord Bishop of London* (1701); *Life of the Learned Sir John Cheke, with his Treatise on Superstition* (1705); *Annals of the Reformation in England*, 4 vol. (vol. i, 1709 [reprinted 1725]; vol. ii, 1725; vol. iii, 1728; vol. iv, 1731); 2nd ed., 4 vol. (1735); 3rd ed., 4 vol. (1736–1738); *Life and Acts of Edmund Grindal, Archbishop of Canterbury* (1710), *of Matthew Parker, Archbishop of Canterbury* (1711), and *of John Whitgift, Archbishop of Canterbury* (1718); *An Accurate Edition of Stow's Survey of London* (1720), a valuable edition of Stow, although its interference with the original text is a method of editing which can scarcely be reckoned fair to the original author; and *Ecclesiastical Memorials*, 3 vol. (1721); 3 vol. (1733). His *Historical and Biographical Works* were reprinted in 19 vol. at the Clarendon Press, Oxford, between 1812 (*Cranmer*) and 1824 (*Annals*).

STUART, ARABELLA (1575–1615), daughter of Charles Stuart, earl of Lennox, younger brother of Lord Darnley and of Elizabeth, daughter of Sir William Cavendish and "Bess of Hardwick," was (by strict pedigree) next in succession to James VI. of Scotland to the thrones of England and Scotland, after Queen Elizabeth I. She became the centre of the intrigues of those who refused to accept James as Elizabeth's successor. Suitors for her hand included Henry IV. of France, the earl of Northumberland, and Esmé Stuart, duke of Lennox. In 1590 a scheme was formed of marrying her to Ranuccio, eldest son of the duke of Parma, who was descended from John of Gaunt, and of raising her with Spanish support to the throne. She was regarded with suspicion by Elizabeth and closely guarded at Hardwick by the dowager countess of Shrewsbury. In 1602 the queen's suspicions were increased by the discovery of a plot to marry Arabella to Edward, eldest son of Lord Beauchamp, who as grandson of Edward Seymour, earl of Hertford, and of Lady Catherine Grey, was heir to the throne after Elizabeth according to Henry VIII.'s will. According to other accounts the intended husband was Thomas Seymour, a younger son of the earl of Hertford. Arabella planned an escape from Hardwick with the aid of her chaplain Starkey, who after its failure committed suicide. In December she wrote secretly to Lord Hertford proposing her marriage with his grandson, but the latter immediately informed the council. In February 1603 another attempt at escape failed, and she was then transferred to the care of the earl of Kent at Wrest House. Arabella was received at the court of James I. and treated with favour, and she showed her fidelity to James by revealing a communication made to her by the conspirators in the Main and Bye Plots, in which her name had been used without her sanction. Every effort, however, was made to prevent her marriage. In December 1609 her plan to escape with Sir George Douglas to Scotland, apparently with a view of arranging a marriage with Stephen Bogdan, pretender to Moldavia, was discovered, and she was arrested. She was, however, granted a pension of £1,600 a year by James. In 1610 she was married secretly in despite of the King's prohibition to William Seymour, younger brother of Edward, and grandson of Lord Hertford. They were imprisoned, Arabella at Lambeth and her husband in the Tower. In 1611 she was placed in charge of the bishop of Durham. She escaped on June 3, 1611, and succeeded in boarding a ship bound for Calais. Her husband had also effected his escape and was sailing towards the French coast. Arabella was captured and brought back to the Tower, where she spent the rest of her unhappy career. She sank into melancholy, and, according to some accounts, insanity, and died on or about Sept. 25, 1615. She was buried in the tomb of Mary Queen of Scots in Henry VII.'s chapel in Westminster Abbey.

See also *The Life and Letters of Arabella Stuart*, by E. T. Bradley (1889), which supersedes the *Life* by E. Cooper (1866); and *Lives* by M. Lefuse (1913), and B. C. Bradley (1913).

STUART, GILBERT CHARLES (1755–1828), U.S. painter, one of the great portraitists of his era and the creator of the most popular image of George Washington, was born on Dec. 3, 1755, in his father's snuff mill near Narragansett, R.I. He grew up in nearby Newport and there learned the rudiments of painting from Samuel King and Cosmo Alexander, a visiting Scotsman. He accompanied Alexander to Edinburgh about 1771, but returned home a year later. In 1775 he went to London and entered the studio of Benjamin West, with whom he worked for about six years. His mature style owed more, however, to Thomas Gainsborough and Sir Joshua Reynolds than to West. In 1782 he opened his own London studio and for five years enjoyed great success, but in 1787 he fled to Dublin to escape his creditors. After six years in Ireland he returned to the U.S. in

BY COURTESY OF BOSTON ATHENAEUM. ON LOAN AT MUSEUM OF FINE ARTS, BOSTON, MASS.

"ATHENAEUM" HEAD OF GEORGE WASHINGTON ON AN UNFINISHED CANVAS BY GILBERT STUART

1793, worked for a year in New York, then settled in Philadelphia, then the political and cultural capital of the nation. He quickly established himself as the nation's leading portrait painter, a position he held until his death on July 9, 1828, in Boston, where he had settled in 1805.

Stuart's work was hailed by his contemporaries and subsequent critics have confirmed this judgment, especially praising its painterly brushwork, luminous colour and psychological penetration. Of his nearly 1,000 portraits, undoubtedly the most famous is the unfinished head of George Washington (1796) at the Boston Athenaeum. Other fine portraits are those of Mrs. Richard Yates and "The Skater," both at the National gallery, Washington; Mrs. Perez Morton, Worcester (Mass.) art museum; and John Adams, privately owned.

Although Stuart had no formal pupils, many young artists, including John Vanderlyn, Thomas Sully and John Neagle, availed themselves of the advice he freely gave. Less talented artists, including his own daughter Jane, simply reduced his style to a formula which gave a Stuartesque appearance to much U.S. portraiture of the succeeding generation.

BIBLIOGRAPHY.—William T. Whitley, *Gilbert Stuart* (1932); Lawrence Park, *Gilbert Stuart: an Illustrated Descriptive List of His Works* (1926); William Sawitsky, "Some Unrecorded Portraits by Gilbert Stuart," *Art in America* (1932); James Thomas Flexner, *Gilbert Stuart* (1955). (D. H. W.)

STUART, JAMES EWELL BROWN (1833–1864), American soldier, was born in Virginia on Feb. 6, 1833 and entered West Point military academy in 1850. In 1859 Stuart was sent to assist Col. R. E. Lee in suppressing John Brown at Harpers Ferry. Two years later when Virginia seceded from the Union, Stuart resigned his commission in the United States army to share in the defense of his state. He had resigned as a lieutenant but he was at once made a colonel. With the scantiest of formal training his regiment was mustered into the Confederate army, and assigned to Joseph Johnston's force in the Shenandoah valley. At the first battle of Bull Run, Stuart distinguished himself by his personal bravery. Later in the year 1861 he was promoted brigadier general and placed in command of the cavalry brigade of the Army of Northern Virginia. Just before the Seven Days' battle (*q.v.*) he was sent out by Lee to locate the right flank of McClellan's army, and not only successfully achieved his mission, but rode right round McClellan's rear to deliver his report to Lee at Richmond. In the next campaign he had the good fortune, in his raid against General Pope's communications, not only to burn a great quantity of stores, but also to bring off the headquarters' staff document of the enemy, from which Lee was able to discover the strength and positions of his opponents in detail. Stuart, now a major general and commander of the cavalry corps, was present at the second battle of Bull Run, and during the Maryland campaign he brilliantly defended one of the passes of South mountain (Crampton's gap), thus enabling Lee to concentrate his disseminated army in time to meet McClellan's attack. At Fredericksburg Stuart's cavalry were as usual in the flank of the army, and his horse artillery rendered valuable service in checking Franklin's attack on "Stonewall" Jackson's corps. At Chancellorsville Stuart was specially appointed by Lee to take over command of the 2nd army corps after Jackson had been wounded. The next campaign, Gettysburg, was preceded by the cavalry battle of Brandy station, in which for the first time the Federal cavalry showed themselves worthy opponents for Stuart and his men. The march to the Potomac was screened by the cavalry corps, which held the various approaches on the right flank of the army, but at the crisis of the campaign Stuart was absent on a raid. He attempted to rejoin Lee during the battle, but was checked some miles from the field. Shortly after the opening of the campaign of 1864 Stuart's corps was drawn away from Lee's army by the Union cavalry under Sheridan; part of it was defeated at Yellow Tavern on May 10, and Stuart himself was killed.

See *Life* by H. B. McClellan (1885).

STUBBS, WILLIAM (1825–1901), English historian and bishop of Oxford, son of William Morley Stubbs, solicitor, of Knaresborough, Yorkshire, was born on June 21, 1825, and was educated at the Ripon Grammar school and at Christ Church, Oxford, where he graduated in 1848, with a first class in classics and a third in mathematics. He was elected a fellow of Trinity college in the same year, was ordained priest in 1850, and held the college living of Navestock, Essex, from 1850 to 1866. In 1862 he was appointed librarian at Lambeth, and in 1866 regius professor of modern history at Oxford. He held this chair till 1884. Many of his lectures were published in book form, including *Seventeen Lectures on the Study of Mediaeval and Modern History, etc.* (1886, 3rd ed. 1900); *Lectures on European History* (1906); *Germany in the Early Middle Ages* (1908); and *Germany in the later Middle Ages* (1908). Stubbs aimed at the organization of a school of history in Oxford after the German model, but his lectures were thinly attended and he gave up the idea.

As a historian Stubbs was eminent alike in ecclesiastical history, as an editor of texts and as the historian of the English Constitution. In 1858 he published his *Registrum sacrum anglicanum*, which sets forth episcopal succession in England, and he edited with A. W. Haddan, vol. iii. of *Councils and Ecclesiastical Documents covering the History of the Anglo-Saxon Church* (1878). He edited for the Rolls series 19 volumes of editions of the chronicles, among the most notable of which are the *Gesta Regum* of William of Malmesbury (1867); the *Gesta regis Henrici II.* (1867); Roger Hoveden's *Chronica* (4 vols., 1868–71); the *Memorials of St. Dunstan* (1874); *The Historical Works of Ralph Diceto* (1878) and *The Historical Works of Gervase of Canterbury* (1879–90). The prefaces to these volumes contain some brilliant sketches of character.

His most famous work is his *Constitutional History of England* (3 vols., 1873, 75, 78, French trans. 1907), preceded by the *Select Charters, and other Illustrations of English Constitutional History from the Earliest Times to the Reign of Edward I.*, in 1870. The appearance of the *Constitutional History*, which traces the development of the English Constitution from the Teutonic invasions of Britain till 1485, is a landmark in the study of mediaeval English history, and has not been superseded.

In character Bishop Stubbs was modest, kind and sympathetic, ever ready to help and encourage serious students, generous in his judgment of the works of others, a most cheery companion, full of wit and humour. His wit was often used as a weapon of defence, for he did not suffer fools gladly. He died on April 22, 1901. In 1859 he had married Catherine, daughter of John Dollar, of Navestock, and had a numerous family.

See *Letters of William Stubbs, Bishop of Oxford*, edit. W. H. Hutton (1904).

STUCLEY (or STUKELY), **THOMAS** (*c.* 1525–1578), English adventurer, who became a popular hero as a result of his exploits against Spanish ships on the high seas. The son of Sir Hugh Stucley, of Affleton, near Ilfracombe, a knight of the body to King Henry VIII, he was supposed to have been an illegitimate son of the king. He was a standard-bearer at Boulogne from 1547 to 1550, and entered the service of the duke of Somerset. After his master's arrest in 1551 a warrant was issued against him, but he escaped to France, and served in the French army. He was sent in Aug. 1552 with a letter of recommendation from Henry II of France to Edward VI. On his arrival he proceeded on Sept. 16, 1552, to reveal the French plans for the capture of Calais and for a descent upon England, which had, according to his account, been the object of his mission to England. Stucley was imprisoned in the Tower for some months. A prosecution for debt on his release in Aug. 1553 compelled him to become a soldier of fortune once more, but he returned to England in Dec. 1554 in the train of Philibert, duke of Savoy. He married an heiress, Anne Curtis, but in a few months had to return to the duke of Savoy's service. From 1558 onward he seems to have been engaged mostly in buccaneering, and although Elizabeth I was compelled to disavow Stucley, who surrendered in 1565, his prosecution was merely formal.

Stucley was next engaged by Sir Henry Sidney in Ireland, then by Bertrand de Salignac Fénelon, the French ambassador in London, and finally by Philip II of Spain, but he fell into disgrace at Madrid. He commanded three galleys under Don John of Austria

at the battle of Lepanto, where his exploits restored him to favour at Madrid. On March 2, 1572, he was at Seville, offering to hold the narrow seas against the English with a fleet of 20 ships. In four years (1570–74) he is said to have received more than 27,000 ducats from Philip II. Wearied by the Spanish king's delays, he sought assistance from Gregory XIII, who aspired to make his illegitimate son, Giacomo Buoncompagno, king of Ireland. He set sail from Civitavecchia in March 1578, but put into Lisbon, where he was to meet his confederate, James Fitzmaurice Fitzgerald, and to secure better ships before sailing for Ireland. There he was turned from his purpose by King Sebastian, with whom he sailed for Morocco. He commanded the centre in the battle of Alcazar on Aug. 4, 1578, and was killed.

Several plays and ballads based on his adventures are printed, with a biography, in R. Simpson, *School of Shakespeare,* vol. 1 (1878).

STUDEBAKER, the family name of five brothers associated in the manufacture of wagons, carriages and automobiles in the United States. CLEMENT STUDEBAKER (1831–1901) was born in Pinetown, Pa. March 12, 1831, and worked in his father's blacksmith shop in Ashland, O. In 1852, with his older brother HENRY STUDEBAKER (1826–95), he founded the firm of H. & C. Studebaker, blacksmiths and wagonmakers, at South Bend, Ind. In 1858 another brother, JOHN MOHLER STUDEBAKER (1833–1917), bought out Henry. The American Civil War provided the firm with a large government contract, and in 1868 the company was incorporated as the Studebaker Brothers Manufacturing Co., at which time another brother, PETER EVERST STUDEBAKER (1836–97), joined the firm. A fifth brother, JACOB FRANKLIN STUDEBAKER (1844–87), joined the firm in 1870. The company opened its first branch office at St. Joseph, Mo. to outfit settlers moving west. At that time the firm was the largest manufacturer of horse-drawn vehicles in the world. With the advent of self-propelled vehicles, the firm began experiments in their manufacture in 1897. Soon after Clement's death in 1901, the firm started building electric and gasoline automobiles. Clement was succeeded as president by John. In 1911 the firm was absorbed by the Studebaker Corp., and in 1954 the Studebaker Corp. merged with the Packard Motor Car Co. (W. H. D.)

STUDENT AID is an award, loan or opportunity designed to help young people with the expenses of education in a college or university. In most countries this (apart from prizes and fellowships) takes the form of scholarships, exhibitions or bursaries. The technical distinctions between the three terms are now blurred. In European usage at any rate it is not quite correct to take exhibition as merely an early word for scholarship: the distinction is rather that an exhibition is a small scholarship, a scholarship a more valuable exhibition. In England especially, the word bursary came to be applied in the 20th century to a conditional award, having about it some connotation of charity.

National Traditions.—In many countries student aid may also take the form of a long-term loan upon specially favourable terms, or of special opportunities for the paid employment of the student. In this matter national traditions and historical factors have caused a marked divergence in aims and methods between the educational systems of the United States and Great Britain and it is convenient to consider these separately. Student aid in most other countries approximates either the type in Britain or that in the U.S. In both countries colleges and universities charge substantial fees and these, plus vacation, maintenance and incidental costs would often impose a prohibitive burden upon students and their parents without some form of student aid. The difference is that the English educational system, with its medieval traditions, never sought to provide grammar school facilities—and at that time there was no other kind of secondary education—for more than a small proportion of the population, with university education for a still smaller proportion. Until the 20th century only very small quotas of these proportions were selected by real or supposed merit. So the characteristic feature of the English system has been the scholarship, often quite generous in amount, but available only to very few students and this (especially in the 19th and first half of the 20th centuries) after a fiercely competitive examination. Although the U.S. system bears in it many marks of its English origin, it was established to meet quite different needs and was molded by thinkers having quite different concepts. Despite the enormous increase in university and college provision in England in the 1940s and 1950s, the proportion of U.S. citizens who may look forward to full-time higher education after the age of 18 is very much higher than the comparable figure in England. In 1957 England had in its universities an annual enrollment of 27,000, with a total student body of about 95,000. To these may be added about 77,000 full-time students in other colleges. In the U.S. in the same year there was a fall enrollment of full-time students amounting to 2,700,000. It is not unreasonable, then, that in the U.S. there should seem to be no hardship, and a great many advantages, in a substantial proportion of these being expected to earn or to borrow some proportion of the cost incurred in their education.

The U.S. System.—In the 1930s the normal level of attainment of students leaving school was that represented by eighth grade; in the 1950s a majority of young people completed four years in high school. About 25% of the high school graduates entered college or university as freshmen; only half of these completed their course and graduated. A major reason for dropping out of school is the heavy cost involved in a four-year college or university course. The methods adopted to meet this situation may be classified as scholarships, loans, self-help and, for mature students, fellowships.

Scholarships.—Though U.S. scholarships *in toto* involve the expenditure of vast sums, they are usually small in individual amount, quite often designed to cover only tuition costs. Most colleges and universities offer some institutional scholarships, some on evidence of need, some for high grades or special promise and some on open competitive examination. Most states also offer state scholarships, as well as certain grants for postgraduate and professional training, the largest funds having been provided in this way in New York and California. American business and industry also provide a great many scholarships.

Loans.—Many colleges also administer loan funds with long repayment periods and favourable interest rates. The largest modern loan fund, however, is one established by congress in 1959, with a total appropriation, over a period of years, of $295,000,000. This is available for loans up to $1,000 per annum, for terms up to five years, with no interest until the sixth year and repayments over the next ten years. There are many other institutions which also administer loan funds, *e.g.,* lodges, sororities, benevolent foundations and churches, and there are also some funds in the hands of individuals. Such nationally available funds include the Knights Templar Masonic Loan fund and the P.E.O. National sorority's scholarships for women.

Self-Help.—In the U.S. it is quite usual for high school students, while still at school, to earn money toward their further education; similarly college students can usually earn part (rarely the whole) of the money needed to meet educational costs, by full-time employment during the vacations and by part-time employment in term. Often it is possible for this to be done on or near the university campus and many colleges operate a student employment service. The appointments found through such a service vary widely, ranging from housework, through typing, to tutoring or helping faculty members in classrooms or laboratories. Usually the opportunities for employment of this kind are much better for men than for women students.

Fellowships.—In the U.S., as elsewhere, this term is usually applied to grants or stipends available to encourage advanced or postgraduate work. The college or university itself is of course the best source of information on institutional fellowships. From 1902 the Rhodes scholarships (about 100 in all, of which a substantial proportion are reserved for American graduates) have been available at Oxford. They are without any restriction as to specific fields of study. American inquiries on them may be addressed to the president, Swarthmore college, Swarthmore, Pa. The Fulbright awards (from 1946) have been instrumental in financing exchange visits for teaching, lecturing, study or research by American graduates in about 50 foreign countries. Particulars of them may be had from the Board of Foreign Scholarships, De-

partment of State, Washington, D.C.

The English System.—*The Schools.*—From very early years both law and canon in England have provided that, at any rate for poor scholars, education shall be quite free. There is a mass of documentary proof of this, dating from the 10th century onward. But changing costs and price levels early caused the introduction of fee-payers and by the 18th century in many instances these had come to monopolize, or almost monopolize, the foundations. It was the work of more than a century to reform this, not by expelling the fee-payers, but in providing for the admission of a due proportion of scholarship holders and ensuring that the entrance examinations (from 1840 onward) should be a genuine test of intellectual ability and of capacity to profit by the school. In most independent and direct-grant schools it is a major condition for the admission of the majority of the boys that their parents shall be able to pay the substantial fees charged. But, from April 1, 1945, every place in a maintained (by state and local authority) grammar or technical school has been free. There is now selection for the various types of secondary school, which are sometimes housed together in a comprehensive school, but no differentiation between the types from the point of view of fees—all alike are free. Direct-grant grammar schools receive part of their costs directly from the state in return for accepting a high proportion of children from the local authority primary schools. Independent schools (including the so-called public schools) often have a few boys whose fees are paid by the local authorities, in addition to a fair number of foundation scholars. Often these have won their scholarship awards as pupils of the (independent) preparatory schools. In both direct-grant and the independent schools, fees have often been graded according to family income.

The Universities.—The story of scholarship provision in the universities closely parallels that in the grammar schools. The English universities have always welcomed the able, poor student and college statutes are studded with poverty clauses. To this day the junior members of a college foundation are the scholars; the commoners (Oxford) and pensioners (Cambridge) are but, as it were, paying guests of the society. As well as scholarships created at the foundation of colleges, many others were endowed by individual benefactors in medieval, Tudor, Stuart and even Georgian times. Nevertheless, like the grammar schools with which they are closely connected (and probably for much the same reasons), by the 18th century the universities were entering upon a period of stagnation, on occasion indeed almost of corruption. The poverty clauses for scholars were generally evaded or ignored and a clever poor boy "going up" might have to do so in the menial capacity of sizar (Oxford) or servitor (Cambridge), as did Richard Bentley, Samuel Johnson, Isaac Newton and George Whitefield.

Scholarship reform followed the introduction of the modern examination system (*see* EXAMINATIONS). After the reform of Oxford and Cambridge universities (*qq.v.*), many scandalous anomalies and anachronisms were swept away, including most of the old restrictions as to close, or restricted or "founder's kin" scholarships. There were further developments after 1902, when the newly established local education authorities were empowered to offer scholarships available to sons and daughters of residents in their areas; these extensively supplemented the scholarships offered by the ancient foundations. The system was crowned by the introduction of state scholarships in 1920, generous in amount, having enormous prestige value and given only after a rigorous examination. Following the further reform of the universities by statute in 1923, the proportion of students receiving aid of one kind or another at Oxford and Cambridge is only a shade lower than the national average (roughly 80%).

It should be noted that under the English system there are no recognized loan facilities for university students and few opportunities for students to work their way through college. Some students work in the vacations, generally for brief periods, but the English student is expected to devote much of his vacations to reading and other academic work.

See also EDUCATION, HISTORY OF: *National Systems;* EDUCATION, INTERNATIONAL.

BIBLIOGRAPHY.—S. N. Feingold, *Scholarships, Fellowships and Loans,* vol. 1–3 (1949–55), with bibliography; university calendars, or catalogues, announcements, bulletins, etc., usually issued annually; *Commonwealth Universities Yearbook* (annual); *Higher Education in the United Kingdom* (biennial). (W. J. GR.; W. E. T.)

STUDENT VOLUNTEER MOVEMENT, a religious movement among college students of the United States and Canada that began in the 1880s as a fresh expression of the evangelical missionary interest that had first appeared in the colleges 75 years earlier. During 1883–84 Robert G. Wilder organized the Princeton Foreign Missionary society on that campus; the members, "willing and desirous, God willing, to become foreign missionaries," took as their watchword "The Evangelization of the World in This Generation," by which Wilder meant that "the lifetime of one generation is all the time available for evangelizing that generation." Missions were then being stressed in the college Y.M.C.A's by their first secretary, Luther D. Wishard, who persuaded the evangelist Dwight L. Moody to invite a student group to meet at Mount Hermon school for the month of July 1886. The 235 men who came from 96 colleges both pioneered the student summer conference and crystallized the widespread missionary concern when, due largely to Wilder's influence, 100 signed the Princeton pledge. During 1886–87 Wilder visited 162 colleges, chiefly in the midwest, obtaining 2,106 volunteers; the next year 1,000 more were added. The Student Volunteer movement (S.V.M.) was organized in 1888. Adopting the Princeton watchword and pledge, it channeled for more than half a century the missionary concern of the North American Student Christian movement, notably through quadrennial conferences. John R. Mott was its chairman until 1920; its pioneer secretaries were Robert E. Speer and Wilder, who estimated in 1936 that 13,000 Volunteers had become foreign missionaries.

With the rise of denominational student organizations and basic changes in the Protestant missionary enterprise after 1930, the S.V.M. widened its areas of concern, affiliated with the National Council of Churches of Christ, and in 1959 became the Commission on World Mission of the National Student Christian federation.

See J. R. Mott, *Addresses and Papers,* vol. i (1946); C. P. Shedd, *Two Centuries of Student Christian Movements* (1934). (C. H. Hs.)

STUKELEY, WILLIAM (1687–1765), English antiquary, was born at Holbeach, Lincolnshire, on Nov. 7, 1687. His principal works, *Itinerarium Curiosum* (1725) and finely illustrated accounts of Stonehenge (1740) and Avebury (1743), did much to popularize the supposed druidical associations of these monuments. He died in London on March 3, 1765.

See S. Piggott, *William Stukeley* (1950). (R. J. C. A.)

STUMPF, CARL (1848–1936), German philosopher and theoretical psychologist, known principally for his work on the psychology of music, was born in Wiesentheid, Bavaria, April 21, 1848. A precocious child and a musical prodigy, he received his education at home, in the *Gymnasia* of neighbouring towns, at Würzburg with Franz Brentano, and at Göttingen with Hermann Lotze, from whom he received his Ph.D. in 1868. After two years of further study with Brentano at Würzburg he habilitated as a *dozent* at Göttingen in 1870. In the spring of 1873 he published his first book, *Über den psychologischen Ursprung der Raumvorstellung,* which won him the chair at Würzburg that autumn. During the next two decades, Stumpf changed his affiliation many times, moving to Prague (1879), Halle (1884), Munich (1889) and finally to Berlin (1894), where he held the chair in philosophy and the directorship of the laboratory until he retired in 1921. Stumpf was joint president of the International Congress of Psychology in Munich (1896); founded the *Beiträge zur Akustik und Musikwissenschaft* (1898); established the *Phonographische Archiv* of primitive music (1900); and served as rector of the university (1907–08). He died at Berlin on Dec. 25, 1936.

Stumpf's chief books are *Tonpsychologie,* 2 vol. (1883–90); *Über Leib und Seele* (1897); *Philosophische Reden und Vorträge* (1910); *Die Anfänge der Musik* (1911); and *Die Sprachlaute* (1926).

See H. S. Langfeld in the *Amer. J. Psychol.,* 49:316–320 (1937). (K. M. D.)

STUMPF, JOHANN (1500–1576), writer on Swiss history and topography, was born at Bruchsal (near Carlsruhe) in 1500 and died at Zürich in 1576. He studied theology at Strasbourg and Heidelberg, and in 1522 was ordained priest at Basel and assigned to the parish of Bubikon (in Canton Zürich). He soon, however, adopted the Protestant faith, and he remained at Bubikon (most of his parishioners following him) as Protestant pastor till 1543. Thereafter he removed to Stammheim, but retired in 1561 and spent the rest of his days at Zürich. The chief result of his historical studies and travels was his *Gemeiner loblicher Eydgnossenschaft Stetten, Landen, und Volckeren Chronikwirdiger Thaaten Beschreybung* (1548). He also wrote on the Council of Constance (1541) and on the emperor Henry IV. (1556).

STUPA, in architecture, a specific type of Buddhist religious building, consisting of a solid mass of masonry, built above a receptacle containing a sacred relic. In India it is commonly called a tope (*q.v.*). Although the origin of the stupa was probably the more primitive tumulus, or mound of earth over a grave, its development in historic times was toward greater and greater height and richness of treatment. Outside of Tibet and China, a profile generally conical and of great height was developed, as in the famous gilded example in the centre of the Shwe Dagon, in Rangoon. The Tibetan form, developed to a great degree of beauty and richness in China, has a bulbous silhouette, wider near the top than at the bottom. *See* TEMPLE; INDIAN ARCHITECTURE.

STURDEE, SIR (FREDERICK CHARLES) DOVETON (1859–1925). British naval commander, whose squadron won the battle of the Falkland Islands (*q.v.*), was born at Charlton, Kent, on June 9, 1859. He entered the navy in 1871 and was promoted lieutenant (1880), commander (1893), captain (1899), rear admiral (1908), vice-admiral (1913), admiral (1917) and admiral of the fleet (1921). He saw service in Egypt (1882) and in Samoa (1899), when he was in command of the Anglo-American force. He was assistant director of naval intelligence to the admiralty (1900–02) and chief of staff, Mediterranean fleet (1905–07) and Channel fleet (1907). In 1910 he became rear admiral of the first battle squadron and commanded the 2nd cruiser squadron (1912–13). During World War I Sturdee was made chief of the war staff in Nov. 1914. As commander in chief in the south Atlantic and Pacific, he led the squadron which won the battle of the Falkland Islands (Dec. 8, 1914), and he was in command of the 4th battle squadron when it took part in the battle of Jutland (*q.v.*). He was created knight commander of the Bath in 1913, knight commander of St. Michael and St. George in 1916, and a baronetcy was conferred on him in 1916, with the title "of the Falkland Islands." From 1918–21 he was commander in chief at the Nore. He died on May 7, 1925.

STURDZA or STURZA, the name of an ancient Rumanian family, of unknown origin, which probably came from Trebizond and settled in Moldavia. The Sturdza family has been long and intimately associated with the Government first of Moldavia and afterwards of Rumania. Its members belong to two main divisions, which trace their descent respectively from John (Ioan) or from Alexander (Sandu), the sons of Kirak Sturdza, who lived in the 17th century, the founder of the family.

1. To the first division belongs MICHAEL [Michail] STURDZA (1795–1884), prince of Moldavia from 1834 to 1849. A man of liberal education, he established the first high school, a kind of university, in Jassy. He brought scholars from foreign countries to act as teachers, and gave a very powerful stimulus to the educational development of the country. In 1844 he decreed the emancipation of the gypsies. Until then the gypsies had been treated as slaves and owned by the Church or by private landowners; they had been bought and sold in the open market. Michael Sturdza also attempted the secularization of monastic establishments, which was carried out by Prince Cuza in 1864, and the utilization of their endowments for national purposes. Under his rule the internal development of Moldavia made immense progress; roads were built, industry developed, and Michael is still gratefully remembered by the people.

See *Michel Stourdza et son administration* (Brussels, 1834); *Michel Stourdza, ancien prince regnant de Moldavie* (1874); A. A. C. Sturdza, *Règne de Michel Sturdza, prince de Moldavie 1834–1849* (1907)

2. GREGORY [Grigorie] STURDZA (1821–1901), son of the above, was educated in France and Germany, became a general in the Ottoman army under the name of Muklis Pasha, and afterwards attained the same rank in the Moldavian army. He was a candidate for the Moldavian throne in 1859, and subsequently a prominent member of the Russophil party in the Rumanian parliament. He wrote *Lois fondamentales de l'univers* (Paris 1891).

3. JOHN [Ioan] STURDZA, prince of Moldavia (1822–1828), was the most famous descendant of Alexander Sturdza. Immediately after the Greek revolution, Prince John Sturdza took an active part in subduing the roving bands of Greek Hetairists in Moldavia; he transformed the Greek elementary schools into Rumanian schools and laid the foundation for that scientific national development which Prince Michael Sturdza continued after 1834. In 1828 the Russians entered the country and took Prince John prisoner. He died in exile.

4. DEMETRIUS [Dimitrie] STURDZA (1833–1914), Rumanian statesman, was born in 1833 at Jassy, and educated there at the Academia Michaileana. He continued his studies in Germany, took part in the political movements of the time, and was private secretary to Prince Cuza. Demetrius afterwards turned against Cuza, joined John Bratianu, and became a member of the so-called Liberal Government. In 1899 he was elected leader of the party in succession to Bratianu and was four times prime minister. He died at Bucharest on Oct. 21, 1914. (M. G.)

STURE, STEN, commonly called the Younger (1492–1520), succeeded his father Svante (d. 1512) as regent of Sweden. His regency was wrecked by the feud between the Sture and Trolle families, and his own personal feud with Gustaf Trolle, who invoked the aid of Christian II. of Denmark. The war with Denmark was begun in 1516 and resumed in 1518, and in 1520 culminated in the battle near Börgerund on Lake Aarunden (Jan. 19). At the very beginning Sture was hit by a bullet and his peasant levies fled to the wild mountainous regions of Tiveden where they made a last desperate but unsuccessful stand. The mortally-wounded regent took to his sledge and posted towards Stockholm, but expired on the ice of Lake Mälar two days later.

STURGEON (*Acipenser*), a small group of fishes, of which some 20 species are known, from European, Asiatic and North American rivers. The distinguishing characters of this group are dealt with in the article FISHES. Most of them pass a great part of the year in the sea, but periodically ascend large rivers, some in spring to deposit their spawn, others later in the season for some purpose unknown; a few are confined to fresh water. None occur in the tropics or the southern hemisphere. By a decree of Edward II, English sturgeons are the property of the king.

Sturgeons are found in the greatest abundance in the rivers of southern Russia, more than 10,000 fish being sometimes caught at a single fishing station in the fortnight of the upstream migration, and in the fresh waters of North America. In Russia the fisheries are of immense value. Early in summer the fish migrate into the rivers or towards the shores of fresh-water lakes in large shoals for breeding purposes. The ova are small and numerous. The growth of the young is rapid. After the sturgeon attains maturity, growth continues for some years, but slowly. Some attain great age, and von Baer claimed that observations made in Russia indicate that the Hausen (*A. huso*) may attain an age of between 200 and 300 years. Sturgeons ranging from 8 to 11 ft. in length are by no means scarce, and some species grow to a much larger size. Sturgeons are ground-feeders. The more important species are:

1. The common sturgeon of Europe (*A. sturio*) occurs on all the coasts of Europe, but is absent from the Black sea. It is not rare on the coasts of North America. It reaches a length of 12 ft., and is always caught singly, or in pairs. The form of its snout varies with age (as in the other species), being much more blunt and abbreviated in old than in young examples. There are 11–13 bony shields along the back and 29–31 along the body.

2. *A. güldenstädtii* is one of the most valuable species of the

rivers of Russia; it inhabits the Siberian rivers also, eastward as far as Lake Baikal. It attains the same size as the common sturgeon, and abounds in the rivers of the Black and Caspian seas.

3. *A. stellatus* occurs in abundance in the rivers of the Black sea and of the Sea of Azov. It has a long and pointed snout, like

JOHN H. GERARD

SHOVEL-NOSED STURGEON (SCAPHIRHYNCHUS PLATORYNCHUS)

the sterlet, but simple barbels without fringes. Though growing only to about half the size of the preceding species, it is of no less value, its flesh being more highly esteemed and its caviar and isinglass fetching a higher price.

4. The sturgeon of the Great Lakes of North America (*A. rubicundus*) has been made the object of a large industry on Lakes Michigan and Erie. The sturgeons of the lakes are unable to migrate to the sea.

5. *A. huso* has bony plates on the snout and flattened, tapelike barbels. It is one of the largest species, reaching the length of 24 ft. and a weight of 2,000 lb. It inhabits the Caspian and Black seas, and the Sea of Azov. Its flesh, caviar and air bladder are of less value than those of the smaller kinds.

6. The sterlet (*A. ruthenus*) is one of the smaller species, which inhabits both the Black and Caspian seas.

The family Acipenseridae includes one other genus, *Scaphirhynchus*, the shovelhead or shovel-nosed sturgeon, distinguished by the long, broad and flat snout, the suppression of the spiracles, and the union of the longitudinal rows of scales posteriorly. All the species are confined to fresh water. One of them is common in the Mississippi and other rivers of North America, the other three occur in the larger rivers of eastern Asia.

STURGEON BAY, the seat of Door county, in northeastern Wisconsin, U.S., 145 mi. N. of Milwaukee is a lake port. The city is on an inlet of the same name on the western or Green bay side of the Door peninsula in Lake Michigan. The federal government constructed a ship canal across the peninsula there in 1878 and has since maintained it. Activities of earlier days were sawmilling, which began in the 1850s, and the quarrying of limestone. In the mid-20th century the city derived income from the processing of cherries, for which the surrounding region was famous; manufacture of shoes and leatherwork; shipbuilding and the handling of water freight. The scenic beauty of the area makes the city the headquarters of an important tourist trade. It has one of the largest yacht storage facilities on the Great Lakes. For comparative population figures *see* table in WISCONSIN: *Population.*

(W. F. RY.)

STURGIS, RUSSELL (1836–1909), U.S. architect and author of a number of books on art, was born in Baltimore county, Md., on Oct. 16, 1836. He graduated from the Free academy in New York (now the College of the City of New York) in 1856, studied architecture under Leopold Eidlitz and then for two years in Munich, returning to the United States in 1862. He designed the Yale university chapel and the Farnham and Durfee dormitories at Yale, the Flower hospital, New York, the Farmers' and Mechanics' bank, Albany, N.Y., and many other buildings. An active member in many professional societies, he was president of the Architectural League of New York in 1889–93, first president of the Fine Arts federation in 1895–97, and a member of the National Society of Mural Painters, the National Sculpture society, the National Academy of Design and the New York chapter of the American Institute of Architects. He lectured on art at Columbia university, the Metropolitan Museum of Art in New York, the Peabody Institute of Baltimore and the Art Institute of Chicago. His works include *A Dictionary of Architecture and Building,* 3 vol. (1901–02) and the English version of Wilhelm Lüebke's *Outlines of the History of Art,* 2 vol. (1904), both of which he edited; and *European Architecture* (1896), *How to Judge Architecture* (1903), *The Appreciation of Sculpture* (1904), *The Appreciation of Pictures* (1905), *A Study of the Artist's Way of Working in the Various Handicrafts and Arts of Design,* 2 vol. (1905), and an unfinished *History of Architecture* (1906 *et seq.*). He died in New York city on Feb. 11, 1909.

STÜRGKH, CARL, COUNT (1859–1916), Austrian politician, was born at Graz on Oct. 30, 1859. In 1891 he entered the *Reichsrat* as representative of the constitutional landed proprietors. From Feb. 10, 1909, to Nov. 3, 1911, he was minister of education, and a zealous advocate of the humanistic education traditional in the *gymnasia.* On Nov. 3 he became prime minister and formed a cabinet without regard to nationality. Because of the incessant parliamentary obstruction, however, Stürgkh prorogued parliament indefinitely on March 16, 1914, with an announcement that "Parliaments were only means to an end, and where they failed, other means must be employed." So on the outbreak of World War I, Austria was without a parliament, and the decisions lay with its few rulers alone. Stürgkh was one of the committee of five ministers who decided on the ultimatum to Serbia. He refused to convoke parliament, and as a protest against this the Social Democrat, Friedrich Adler, assassinated him in a Viennese restaurant on Oct. 21, 1916 (*see* AUSTRIA, EMPIRE OF).

STURM, (JACQUES) CHARLES FRANÇOIS (1803–1855), Swiss mathematician of Alsatian parentage, who was the originator of the Sturm theorem of equations, was born at Geneva on Sept. 29, 1803, and entered the university there in 1818. The death of his father meant that while there he had to support his family by private coaching. He visited Paris in 1823, as tutor to the de Broglie family, through whom he met many of the leading French men of science. He finally settled there in 1825. With Colladon, in 1826, he made the first accurate determination of the velocity of sound in water, and a year later wrote a prize essay on compressible fluids. The theorem which bears his name, regarding the number of real roots of a numerical equation included between given limits, was discovered in 1829 (*see* EQUATIONS, THEORY OF). His work on what is now called the Sturm-Liouville theory of differential equations of the second order was published in 1834 and won for him the Grand Prix des Sciences Mathématiques (*see* also LIOUVILLE, JOSEPH). He was elected to the Académie des Sciences in 1836, and was appointed to the staff of the École Polytechnique in 1838. Finally he succeeded Siméon Denis Poisson in the chair of mechanics in the Faculté des Sciences in Paris in 1840. He died in Paris, on Dec. 18, 1855. His *Cours d'analyse de l'école polytechnique* (1857–63) and *Cours de mécanique de l'école polytechnique* (1861) were published posthumously. Although primarily an analyst, Sturm made significant contributions to projective geometry and to the differential geometry of curves and surfaces. (G. A. B.)

STÜRMER, BORIS VLADIMIROVICH (1849–1917), Russian politician of German origin, started his career in the chamberlain's department of the imperial court. When in 1892 the government rejected the candidate nominated to the presidency of the executive board of the Tver zemstvo, Stürmer, whose name was on the list of the Tver gentry, was appointed to this office, the first case of a president being appointed instead of elected. In 1894 Stürmer was appointed governor of the Novgorod, and later of the Yaroslavl province. In Jan. 1916, he was made prime minister. He was opposed in liberal and patriotic circles. Accusations of connections with Germany were brought to the Duma by Milyukov and resulted in Stürmer's resignation in Nov. 1916. After Sazonov's dismissal Stürmer took the portfolio of foreign affairs. He was arrested after the revolution, and died in prison in Sept. 1917.

STURNIDAE, a family of perching birds called starlings (*q.v.*), stocky birds with long pointed bills and short tails.

ŠTURSA, JAN (1880–1925), one of the leaders of the modern school of Czech sculpture, was born at Nové Město in Moravia on May 15, 1880. In 1899 he entered the Academy of Fine Arts at Prague as a pupil of the influential teacher J. V. Myslbek. His early works manifest the character of the symbolic literature of the period, and reflect also the plastic qualities of Rodin's sculpture. His individuality began to proclaim itself in such lyric works as "Puberty" (1905), "The Melancholy Girl"

(1906), and "Primavera" (1907). A journey to Italy in 1907 was the opening of a new epoch in his development. "Eve" (1908, State gallery, Munich), "Hetaira" (1909), and the later "Gift of Heaven and Earth" demonstrate his sensitivity to feminine beauty as well as, stylistically, a tendency toward simplification of sculptural form and increasing expressive strength. Occasional excursions into extremes of sensuality, as in his monumental representation of the dancer "Sulamith Rahu" (1911, Venice), were not sustained. His decorative commissions, notably the groups for the pylons of the Hlavka bridge at Prague, are marked by intricate compositional massing restrained by strong, almost geometric, tensions. Štursa's war experiences gave to his mature style a tragic and more profound character, as in "Wounded Man" (1916), expressed in bold and vigorous sculptural terms. In 1916 he became professor in the Academy of Fine Arts at Prague, and in 1923–24 its director. He produced a number of portrait busts, the most noteworthy being "President Masaryk." Štursa died in Prague on May 2, 1925.

See A. Matějček and Z. Wirth, *Modern and Contemporary Czech Art* (1924). (N. L. R.)

STURT, CHARLES (1795–1869), British explorer of Australia, was born in Bengal on April 28, 1795. He was educated in England, joined the army in 1813, and in 1826 was sent to Australia. In 1828 he explored the Macquarie river and discovered the Darling; in 1829 he followed the Murrumbidgee river and discovered the Murray and Lake Alexandrina; and in 1844–46, while attempting to reach the centre of Australia, discovered Cooper's creek and the stony desert which bears his name. From 1839 Sturt held government offices in South Australia but retired in 1851. He died at Cheltenham, Eng., June 16, 1869. (T. M. Py.)

STURTEVANT, EDGAR HOWARD (1875–1952), U.S. linguist, was the originator of the controversial Indo-Hittite hypothesis, according to which the Anatolian languages, among them Hittite, and the Indo-European languages are separate offshoots of a common ancestral language, called Indo-Hittite. Sturtevant was born at Jacksonville, Ill., on March 7, 1875. After studying at Indiana university and The University of Chicago, he taught at Indiana, at Columbia and from 1923 to 1943 at Yale. Sturtevant's hypothesis is summarized in *The Indo-Hittite Laryngeals;* his *The Pronunciation of Greek and Latin* is one of the standard works on the subject. He died on July 1, 1952, at Branford, Conn.

See for complete bibliography, *Language,* vol. xxviii, no. 4, pt. 1, (1952). (My. F.)

STURZO, LUIGI (1871–1959), Italian priest and political organizer, promoter of the Christian Democratic movement in Italy. Born at Caltagirone, Sic., on Nov. 26, 1871, he studied at the Gregorian university, Rome, and became professor of philosophy and sociology in his native town. On the outbreak of World War I he went to Rome as secretary of Azione Cattolica (Catholic Action). In 1919 he founded the Partito Popolare and became its first political secretary. Sturzo and his party based themselves on papal encyclicals concerning the social order and advocated reforms of a radical nature, especially in agriculture.

At the elections of Nov. 1919 the newly founded Partito Popolare secured 101 seats and became a dominant force in politics, if not with the Giolitti cabinet certainly during the subsequent Bonomi cabinet. After the Fascist "march on Rome" several Popolari entered the first Mussolini cabinet, but soon he and his party joined the opposition. In 1923, however, the Vatican, subjected to pressure from Mussolini, withdrew its support from Sturzo in an attempt to reach some understanding with the Fascists. Sturzo's party had by now much shrunk as an outcome of Fascist electoral methods. He resigned the leadership in 1923 and left Italy in Oct. 1924. Sturzo spent his 20 years of exile first in London and after the outbreak of World War II in the U.S. His party in Italy was driven underground, but after the overthrow of Fascism many of his ideas re-emerged in the Christian Democratic party, headed by Alcide de Gasperi and other followers of Sturzo's who revered him as the father of their ideas. He returned to Italy after the war and was made a senator for life in 1952. He died in Rome on Aug. 8, 1959. His books include *Italy and Fascism* (1927), *Church and State* (1939), *Italy and the Coming World* (1945) and other works not translated into English. (B. Wl.)

STUTTGART, the capital of the *Land* Baden-Württemberg, Germany, lies on both banks of the Neckar, a tributary of the Rhine, approximately 120 mi. N.W. of Munich and 70 mi. E.N.E. of Strasbourg in France. Stuttgart lies in the midst of woods, orchards and vineyards, but it is also an important industrial town and notable for its modern architecture. Some of the earliest work on the invention of the motorcar was done in Stuttgart and the oldest motorcar factory in the world is there. Pop. (1950) 497,677.

In the centre of the town is the Schillerplatz with a statue of Schiller by B. Thorwaldsen (1839), the old castle which dates from the 13th century (completed 1553–78) and now houses the state museum with its collection of prehistoric and historic objects, and the Stiftskirche (collegiate church), a 12th-century Romanesque basilica completed in the late Gothic style (1436–95). Immediately to the northeast is the Schlossplatz with its Jubilee column (1841–46), one of the most beautiful squares in Germany. In it is the new palace (1746–1807), built on a typical 18th-century plan; it was burned in 1944 and under reconstruction in the 1950s. Farther to the northeast are the Theater-Anlagen, with rose and other gardens, and the opera house of the Württemberg State opera along its eastern side. These gardens are continued northeast by the Schlossgarten Anlagen, with the main railway station on its west, to the Neckar, which there takes a right-angle turn to the northeast. On its left bank are the Wilhelma botanical and zoological gardens and the Rosenstein palace (1825–29), now the natural history museum. To the north and over the river is Bad-Cannstatt which, with Stuttgart-Berg on the other bank, has many mineral springs; the number of baths and swimming pools, mineral and otherwise, is a notable feature of the town. The city church in Bad-Cannstatt was built in 1471–1506. Along the east bank of the river lies the Cannstatter Wasen, a large sports ground with a sports stadium holding 100,000 spectators; every autumn the big Cannstatter folk festival is held there. The rococo Solitude palace (1763–67) stands in the wooded hills to the west of the town; Hohenheim palace (1768–85), housing the college of agriculture, is to the south. In the north of the city is the Höhenpark Killesberg, a large park with exhibition halls, an observation tower, chair lift and miniature railway, and adjoining it to the southeast is the Weissenhof estate, built in 1927 by 16 German and other architects. Other modern buildings include the town hall (1954–56) in the market place, south of the Schillerplatz; the television tower (1955), to the south, which is 633 ft. high with an observation platform and restaurant; the Stuttgarter Liederhalle (concert and congress hall) built in 1954–56. Not far from the market place is the late Gothic St. Leonard's church (1463–74), of special interest as being of the hall type with nave and aisles of the same height.

Stuttgart has a technical college and institutions for agriculture, music, art and architecture. The Staatsgalerie, in Neckarstrasse, has German paintings up to the 20th century and 15th- to 17th-century Dutch and Italian pictures. The Württemberg Land library in the Urbanstrasse contains the second largest collection of Bibles in the world. One of the most interesting of the museums is the Daimler-Benz automobile museum in Untertürkheim on the east bank of the Neckar. Gottlieb Daimler's tomb is in the Uff cemetery in Bad-Cannstatt and a memorial to him stands in the park there.

Stuttgart is an important transport junction at the centre of railway connections stretching all over Europe. It lies on two *Autobahnen* (Frankfurt-Munich-Salzburg and Stuttgart-Heilbronn) and has an airport 8 mi. from the city. The port, which is at the end of the Neckar, was opened in 1958 and will take shipping up to 1,200 tons. The city is the largest industrial centre in southwest Germany. Electrical engineering is of primary importance, followed by vehicle and machine construction. Textiles, clothing, precision tools, foodstuffs, luxury wooden and leather goods, shoes, musical instruments, chemicals and paper are manufactured. There are also large breweries. Stuttgart has a special reputation as a book centre, with numerous printing works and over 200

publishing houses.

Settlements existed in the area of what is today Bad-Cannstatt in prehistoric times, and the Romans established a fort there. Stuttgart itself developed from a *Stuotgarten*, a *Gestüt* or stud farm, which was set up in about 950. It developed into a town about 1254 under Count Ulrich "the Founder." In 1495 it became the capital of the new dukedom of Württemberg, which in 1806 became a kingdom. With the industrial revolution, the court and administrative town quickly expanded, its population rising from 48.000 in 1850 to 177.000 in 1900 and 460,000 in 1939. (U. S.)

STUYVESANT, PETER (1592–1672), Dutch colonial governor, was born in Scherpenzeel, in Southern Friesland, in 1592. He studied at Franeker, entered the military service in the West Indies about 1625, and was director of the West India Company's colony of Curaçao 1634–44. In April, 1644 he attacked the Portuguese island of Saint Martin and was wounded; he had to return to Holland, and there one of his legs was amputated. Thereafter he wore a wooden leg ornamented with silver bands. In May, 1645, he was selected by the West India Company to supersede William Kieft as director of New Netherland. He arrived in New Amsterdam (later New York) on May 11, 1647, and was received with great enthusiasm. In response to the demand for self-government, in Sept., 1647, he and the council appointed—after the manner then followed in Holland—from 18 representatives chosen by the people a board of 9 to confer with him and the council whenever he thought it expedient to ask their advice.

The leading burghers were, however, soon alienated by his violent and despotic methods, by his defence of Kieft, and by his devotion to the interests of the company; the nine men became the centre of municipal discontent, and a bitter quarrel ensued. In 1650 the states-general suggested a representative government to go into effect in 1653, but the company opposed it; in 1653, however, there was established the first municipal government for the city of New Amsterdam modelled after that of the cities of Holland. Stuyvesant also aroused opposition through his efforts to increase the revenues of the company, to improve the system of defence, and to prevent the sale of liquor and firearms to the Indians, and through his persecution of Lutherans and Quakers, to which the company finally put an end. In 1650, he came to an agreement with the commissioners of the united colonies of New England at Hartford upon the boundary between New Netherland and Connecticut, involving the sacrifice of a large amount of territory. On Long Island, during Stuyvesant's rule, Dutch influence was gradually undermined by John Underhill. Stuyvesant's dealings with the Swedes were more successful. With a force of 700 men he sailed into the Delaware in 1655, captured Ft. Casimir (Newcastle)—which Stuyvesant had built in 1651 and which the Swedes had taken in 1654—and overthrew the Swedish authority in that region. He also vigorously suppressed Indian uprisings in 1655, 1658 and 1663.

In March, 1664, Charles II. granted to his brother, the duke of York, the territory between the Connecticut river and Delaware bay, and Col. Richard Nicolls with a fleet of 4 ships and about 300 or 400 men was sent out to take possession. Misled by instructions from Holland that the expedition was directed wholly against New England, Stuyvesant made no preparation for defence until just before the fleet arrived. As the burghers refused to support him, Stuyvesant was compelled to surrender the town and fort on Sept. 8. He returned to Holland in 1665 and was made a scapegoat by the West India Company for all its failings in New Amsterdam; he went back to New York again after the treaty of Breda in 1667, having secured the right of free trade between Holland and New York. He spent the remainder of his life on his farm called the Bouwerie, from which the present "Bowery" in New York city takes its name. He died in Feb. 1672, and was buried in a chapel, on the site of which in 1799 was erected St. Mark's church.

See Bayard Tuckerman, *Peter Stuyvesant* (1893), in the "Makers of America" Series; and Mrs. Schuyler Van Rensselaer, *History of the City of New York in the Seventeenth Century* (1909).

STYLE. It is desirable to insist at the outset on the dangers of a heresy which found audacious expression towards the close of the 19th century, namely, that style is superior to thought and independent of it. Against this may be set at once one of the splendid apophthegms of Buffon, "Les idées seules forment le fond du style." Before there can be style there must be thought, clearness of knowledge, precise experience, sanity of reasoning power. A confusion between form and matter has often confused this branch of our theme. Even Flaubert, than whom no man ever gave closer attention to the question of style, seems to dislocate them. For him the *form* was the work itself: "As in living creatures, the blood, nourishing the body, determines its very contour and external aspect, just so the *matter*, the basis, of a work of art imposes, necessarily, the unique, the just expression, the measure, the rhythm, the *form* in all its characteristics." This ingenious definition seems to strain language beyond its natural limits. If the adventures of an ordinary young man in Paris be the *matter* of *L'Education sentimentale* it is not easy to admit that they "imposed, necessarily," such a "unique" treatment of them as Flaubert so superlatively gave. They might have been recounted with feebler rhythm by an inferior novelist, with bad rhythm by a bad novelist and with no rhythm at all by a police-news reporter. What makes that book a masterpiece is not the basis of adventure, but the superstructure of expression. The expression, however, could not have been built up on no basis at all, and would have fallen short of Flaubert's aim if it had risen on an inadequate basis. The perfect union is that between adequate matter and an adequate form. We will borrow from the history of English literature an example which may serve to illuminate this point. Locke has no appreciable style; he has only thoughts. Berkeley has thoughts which are as valuable as those of Locke, and he has an exquisite style as well. From the artist's point of view, therefore, we are justified in giving the higher place to Berkeley, but in doing this we must not deny the importance of Locke. If we compare him with some pseudo-philosopher, whose style is highly ornamental but whose thoughts are valueless, we see that Locke greatly prevails. Yet we need not pretend that he rises to an equal height with Berkeley, in whom the basis is no less solid, and where the superstructure of style adds an emotional and aesthetic importance to which Locke's plain speech is a stranger. At the same time, an abstract style, such as that of Pascal, may often give extreme pleasure, in spite of its absence of ornament, by its precise and pure definition of ideas and by the just mental impression it supplies of its writer's power and placidity of mind. But whether in the abstract or concrete style, what Rossetti called "fundamental brain-work" must always have a leading place.

When full justice has been done to the necessity of thought as the basis of style, it remains true that what is visible, so to speak, to the naked eye, what can be analysed and described, is an artistic arrangement of words. Language is so used as to awaken impressions, and these are roused in a way peculiar to the genius of the individual who brings them forth. The personal aspect of style is therefore indispensable, and is not to be ignored even by those who are most rigid in their objection to mere ornament. Ornament in itself is no more style than facts, as such, are thought. In an excellent style there is an effect upon our senses of the mental force of the man who employs it. D'Alembert said of Fontenelle that he had the style of his thought, like all good authors. In the words of Schopenhauer, style is the physiognomy of the soul: in the Renaissance phrase, it is *mentis character*. All these attempts at epigrammatic definition tend to show the sense that language ought to be, and even unconsciously is, the mental picture of the man who writes.

To attain this, however, the writer must be sincere, original and highly trained. He must be highly trained, because, without the exercise of clearness of knowledge, precise experience and the habit of expression, he will not be able to produce his soul in language. Nor can anyone who desires to write consistently and well, afford to neglect the laborious discipline which excellence entails. He must never rest until he has attained a consummate adaptation of his language to his subject, of his words to his emotion. This is the most difficult aim which the writer can put before him. Perfection is impossible, and yet he must never desist from pursuing perfection.

"If all the pens that ever poets held
Had fed the feeling of their masters' thoughts,
And every sweetness that inspired their hearts,
Their minds, and muses, on admirèd themes—
If all the heavenly quintessence they 'still
From their immortal flowers of poesy,
Wherein, as in a mirror, we perceive
The highest reaches of a human wit—
If these had made one poem's period,
And all combined in beauty's worthiness,
Yet should there hover in their restless heads
One thought, one grace, one wonder, at the least.
Which into words no virtue can digest."
—Marlowe, *Tamburlaine the Great.*

Flaubert believed that every thought or grace or wonder had one word or phrase exactly adapted to express it, and could be "digested" by no other without loss of clearness and beauty. It was the passion of his life, and the despair of it, to search for this unique phrase in each individual case. Perhaps in this research after style he went too far, losing something of that simplicity and inevitability which is the charm of natural writing. The greatest writing is that which in its magnificent spontaneity carries the reader with it in its flight; that which detains him to admire itself can never rise above the second place. Forgetfulness of self, absence of conceit and affectation, simplicity in the sense not of thinness or poorness but of genuineness—these are elements essential to the cultivation of a noble style. Here again, thought must be the basis, not vanity or the desire to astonish. We do not escape by our ingenuities from the firm principle of Horace, "scribendi recte sapere est et principium et fons."

Of the errors of style which are the consequences of bad taste, it is difficult to speak except in an entirely empirical spirit, because of the absence of any absolute standard of beauty by which artistic products can be judged. That kind of writing which in its own age is extravagantly cultivated and admired may, in the next age, be as violently repudiated; this does not preclude the possibility of its recovering critical if not popular favour. Perhaps the most remarkable instance of this is the revolt against Ciceronian prose which occurred almost simultaneously in several nations toward the middle of the 16th century and in England is best represented by Lyly in his celebrated *Euphues.* Montaigne in France and Castiglione in Italy, by their easiness and brightness, their use of vivid imagery and their graceful illumination, marked the universal revulsion against the Ciceronian stiffness. Each of these new manners of writing fell almost immediately into desuetude, and the precise and classic mode of writing in another form came into vogue (Joseph Addison, Jacques Bossuet, Giovanni Vico, Samuel Johnson). In the 19th century admiration of the ornamental writers of the 16th and 17th centuries revived. A facility in bringing up before the memory incessant analogous metaphors is the property, not merely of certain men, but of certain ages; it flourished in the age of Giovanni Marini and was welcomed again in that of George Meredith. A vivid, concrete style, full of colour and images, is not to be condemned because it is not an abstract style, scholastic and systematic. It is to be judged on its own merits and by its own laws. It may be good or bad; it is not bad merely because it is metaphorical and ornate. The amazing errors which lie strewn along the shore of criticism bear witness to the lack of sympathy which has not perceived this axiom and has wrecked the credit of dogmatists. Yet that particular species of affectation which encourages untruth, affectation, parade for the mere purpose of producing an effect, must be wrong, even though Cicero be guilty of it.

See Walter Pater, *An Essay on Style* (1889); Walter Raleigh, *Style* (1897); Antoine Albalat, *La Formation du style par l'assimilation des auteurs* (1901) and *Le Travail du Style* (1904); Remy de Gourmont, *Le Problème du style* (1902); J. Middleton Murry, *The Problem of Style* (1922); H. W. Fowler, *A Dictionary of Modern English Usage* (1926). (E. G.; X.)

STYLOLITE. A common feature in limestone, marbles and similar rock is a series of relatively small, alternating, interlocked, toothlike columns of stone called stylolites. The term is taken from the Greek *stylos* meaning "pillar" or "column." The individual columns never appear singly but occur as a succession of interpenetrations which in cross section make a zigzag suture across the face of the stone. There has been much controversy over their origin since they were first described in 1751. However, geologists generally submit to the preponderance of field evidence supporting the theory of secondary origin resulting from differential chemical solution by ground waters circulating along some sort of a parting in hardened rock. (P. B. Se.)

STYLOPS, the name of a genus of insects belonging to the order Strepsiptera allied to the beetles. The order is distributed over most parts of the world, but numbers less than 200 species. The males are minute black or brown insects with branched antennae and vestigial mouth parts; the forewings are reduced to clublike scales but the hind wings are large and fanlike. They are seldom seen and lead a very brief life. The females are mostly degenerate saclike parasites that live partly protruding from the bodies of bees, wasps or leafhoppers and other insects. The young larvae are minute and known as triungulins, which hatch within the parent and probably escape onto flowers, etc., where they await the presence of a host. On finding the latter they bore their way into the body, change into legless maggots and pupate there; in most species, the males alone subsequently emerge. The presence of these parasites frequently induces structural and other changes in their hosts, including the acquisition of characters pertaining to the opposite sex (parasitic castration).

STYRENE is a colourless, liquid, aromatic hydrocarbon, important chiefly for its marked ability to undergo polymerization (*q.v.*). It was first produced commercially in 1935. Production thereafter increased steadily and by 1955 had passed 1,000,000,000 lb. per year in the U.S. Nearly all of this large production is converted to synthetic rubber (*see* RUBBER), molded plastic materials (*see* PLASTICS) or synthetic resins (*see* RESINS). Styrene polymerizes slowly on standing, rapidly on heating, to form polystyrene, an inexpensive, clear, rigid plastic with excellent electrical and optical properties and with good strength and chemical resistance. Polystyrene is somewhat brittle; some of the copolymers of styrene are tougher.

Styrene was first found in storax balsam (oriental sweet gum). It was isolated, analyzed and named in 1839. It forms in the coking of coal and appears in small amounts in the crude xylene fraction. Styrene is produced by combining benzene (*q.v.*) from coal or petroleum with ethylene (*q.v.*) from the cracking of petroleum. The resulting ethylbenzene is then dehydrogenated to styrene. Styrene has the structure $C_6H_5—CH=CH_2$; a boiling point of 145° C., a freezing point of −30.6° C., a specific gravity at 25° C. of 0.904 and a refractive index of 1.544. Styrene has a burning taste and a sweetish, penetrating, aromatic odour like manufactured (coke-oven) gas, in which it is present in very small amount. The reactive double bond in styrene participates readily in chemical reactions with halogens, halogen acids, oxygen and other reagents. *See* CHEMISTRY: *Organic Chemistry.*

See R. H. Boundy and R. F. Boyer, *Styrene: Its Polymers, Copolymers and Derivatives* (1952). (Fk. R. M.)

STYRIA (Ger. STEIERMARK), the second *Bundesland* (federal state) of Austria, covering 6,326 sq.mi., is distinguished for its scenery, forests and mineral wealth. It is divided into Upper Styria, the mountain and basin lands of the upper courses of the Enns and the Mur with the Mürz; Middle Styria, the region of the middle Mur and the headwaters of the Raab; and Lower Styria, of which only the borderlands toward Yugoslavia remained Austrian after 1918. The northernmost part comprises mountains of the Limestone Alps; the other mountains belong to the Slate and Central Alps, with most peaks between 6,000 and almost 10,000 ft.

History.—The area now known as Styria was inhabited from prehistoric times, despite its mountainous nature, partly because of its mineral wealth. In Roman times it was peopled by the Taurisci and the Norici. From the 6th to the 9th centuries it was occupied by the Slavs from the northeast and the Bavarians from the northwest and Charlemagne incorporated the area into Carinthia. In 976 Carinthia became a separate duchy and Styria was separated from it as the Mark of Carinthia. In 1035 it was granted to Arnold of Wels-Lambach and in 1050 became an independent margravate under Otakar von Steyr, from whom it took its name (Steiermark). For an account of its incorporation into the Baben-

berg and Habsburg lands, *see* AUSTRIA, EMPIRE OF. During the 13th and 14th centuries, it often passed to junior branches of the line, but was reunited with the other Austrian lands under Frederick III (*q.v.*).

During his reign, Styria was troubled by internal disputes, by Turkish invasions (especially in 1480) and by wars with Hungary; in 1485, the Hungarian king drove Frederick out of Vienna and occupied most of Styria for a few years. It was reintegrated with Austria under Maximilian I, but in 1564 again fell to a younger son, Duke Charles. Protestantism had taken firm hold early in the 16th century, and religious liberty was granted, but in 1572 Charles called in the Jesuits, and Ferdinand II, who succeeded in 1590, ruthlessly enforced Roman Catholicism. Under Ferdinand, Styria was again reunited with the Habsburg lands (1619), of which it remained a part until 1918, when it became a province of the Austrian republic. In 1919 most of the German-mixed and Slovene-speaking area of Lower Styria (the Drava valley and the territory south of it) was given to Yugoslavia.

Agriculture and Industry.—Despite its mountainous character only 9% of the area is classed as barren land. Forests cover 49%, cultivated land 15% and the remainder (26%) is good grassland. In Upper Styria cultivation is mainly on the fertile glacial moraines and alluvial soils of the stream terraces and basins, rye and oats being the most common cereals. Potatoes, beets, flax and hemp are also grown. Middle and Lower Styria produce wine, fruit, wheat, maize, millet and hops. Cattle breeding flourishes, as does the breeding of draft horses in the marshy basins of the north and pigs and poultry in the south. Sheep are on the decrease everywhere. Large numbers of game are to be found on the highlands. Forestry and the timber industry are important.

The Erzberg iron mines have been worked from early days and yield nearly the whole of the Austrian production (3,258,000 tons). Styria supplies more than 60% of Austria's lignite. There are important deposits of magnesite (at Veitsch, Breitenau and Trieben). Other mineral resources include salt, graphite, bauxite, talc, gypsum, marble and building stone. The basin of Leoben and Donawitz and the adjoining Mur-Mürz valley, near coal and iron supplies, is the main district of the iron and steel industry. Graz (*q.v.*), the administrative, religious and intellectual capital, is also an industrial and commercial centre.

Population.—The population in 1961 was 1,126,587 (178.1 persons per square mile), but it is very unevenly distributed. In the remote highlands scattered farmsteads and hamlets predominate. Settlements in the valleys are usually small and either strung like beads along the river courses, lower slopes or terraces or clustered on fans. (L. G. S.)

STYX, a stream near Nonacris in Arcadia, Paus., viii, 17, 6 ff., where see Frazer's commentary; the modern Mavro Nero (Black Water). It was thought to be virulently poisonous and capable of dissolving any vessel it was put in, save one made from the hoof of a horse. From the time of Homer, Styx (The Hateful) is one of the rivers of the underworld, generally its boundary, and the gods, if they swear by it, dare not break their oath. In Hesiod (*Theogonia*, 383 ff.) the nymph Styx, daughter of Ocean, with her children (Power, Might, Victory and other abstractions) helps Zeus against the Titans, and is therefore honoured by him. If a god does break the oath, he is insensible for a year and then banished for nine years; mortals on occasion might take this oath, and in either case it would seem to have involved drinking the water (Hesiod, *Theogonia,* 783 ff.; Herodotus, vi, 74). We may suppose that the oath was really an ordeal; the water was a magical poison, perhaps originally fatal only to perjurers.

SUAKIN (SAWAKIN), on the Red sea coast in the province of Kassala, once the major port of the Sudan, was abandoned in 1922, being replaced by Port Sudan, 36 mi. to the north. Pop. (1956) 4,228. Suakin was built on a coral island and connected by a causeway with El Geif, a suburb on the mainland. The natural harbour can take large vessels, but the approach is so difficult that they cannot use it.

Suakin seems to have started about the 12th century as a rival port to Aidhab, still known as Old Suakin (Suakin el-kadim), farther north, since merchants from the Yemen and India could trade there with the Beja without paying the Egyptian dues levied at Aidhab. Toward the end of the 13th century Suakin came under the rule of Egypt, but Egyptian control was not permanent. Suakin increased in importance after the destruction of Aidhab *c.* 1428. The pilgrim route has crossed the Red sea between Suakin and Jedda ever since. In the 16th century, Suakin was occupied by the Turks, and by the 18th century its prosperity had seriously declined, although pilgrims continued to come from Sennar to Mecca through it, and there was still an export of slaves, gold and ivory. In 1821 Mohammed Ali, after conquering the Sudan, leased Suakin from Turkey. This lease lapsed at his death, but in 1865 Ismail Pasha reacquired the port for Egypt; and it was the only part of the Sudan that did not fall into dervish hands during the mahdia. (A. J. AL.)

SUARDI, BARTOLOMMEO: *see* BRAMANTINO.

SUÁREZ, FRANCISCO (1548–1617), Spanish theologian and philosopher, most important figure (*doctor eximius et pius*) of the second flowering of Scholasticism, and generally considered the greatest theologian of the Society of Jesus, was born in Granada on Jan. 5, 1548, to a family of jurists. He studied law in Salamanca and in 1564 joined the Jesuit order. After completion of his theological studies, he became (1571) instructor and professor of philosophy in Segovia, Valladolid and for a short time in Rome. From 1585 he taught theology in Alcalá and Salamanca. In 1597 he was appointed by Philip II *professor primarius* at Coimbra, where he died after a studious, yet humble and simple life, on Sept. 25, 1617.

Suárez' principal work in philosophy is the *Disputationes Metaphysicae*, in which he follows mostly Aristotle and Aquinas, yet taking in consideration the criticism of Duns Scotus and of late Scholasticism. He works out in some important points (*e.g.*, principle of individuation, potency and act) his own solution, and the points in which he departs from Thomism are considered by some scholars to be significant enough to warrant description as a separate system, called Suarezianism. The *Disputationes* was used for more than a century as the textbook in philosophy in most European universities, Catholic and Protestant alike, and Leibniz speaks of Suárez' philosophy as the *Philosophia recepta*. In theology Suárez intended to write a full commentary on the *Summa* of Thomas Aquinas in the form of treatises, but he did not succeed, since, upon the request of his general and of Pope Paul V, he had to write apologetic works on the religious state (*De Virtute et Statu Religionis*, 1608–09) and a *Defensio Fidei* (1613) against Anglican theologians, the oath of supremacy and the divine right theory of James I. In the grand dispute on the efficacy of divine grace, free will and predestination he took a stand, in the system called Congruism (*De vera intelligentia auxilii efficacis*, 1605, published 1655), more with Molina against Domingo Báñez.

Suárez' treatises *De Legibus* (1612), *De Bello et de Indis* and the "Defense of the Faith" contain his political theory and legal philosophy, in which he is both most original and personal. He refutes the divine right theory and explains that the people itself is the original holder of political authority: thus all constitutions can only be of positive law, neither of divine nor of natural law; the state must be thought to come into existence (ideally) by the consent (social compact) of those forming it. He argues strongly for the natural rights of the human person such as life, liberty and property, and repudiates the Aristotelian thesis of "natural slavery," affirming that all forms of servitude are only of human positive law. He criticizes most of the practices of Spanish colonization in the newly discovered Indies. These lands are for him sovereign states, members of the community of nations legally equal to Spain. Neither pope nor emperor as supposedly lord of the world (an absurd theory to him) nor Christian princes have the right to order an invasion or to conquer these states by reason of the idolatry or supposed backwardness of their citizens or for the propagation of Christianity. This would be against what Suárez calls for the first time the *jus inter gentes*, which, resting on natural law, contains customary and treaty law and, as the public law of the international community, is to be distinguished from the Roman *jus gentium*, which contained also private or civil law in addition to public law.

Suárez' collected works, were published in 23 volumes at Venice (1740–51) and in 28 volumes at Paris (1856–78). Neither is complete; a complete and critical edition is in preparation.

BIBLIOGRAPHY.—R. de Scorraille, *François Suárez . . . d'après ses lettres, etc.*, 2 vol. (1912–13); J. H. Fichter, *Man of Spain* (1940); a full bibliography of and on Suárez is in *Atti del IV nascita di S.* (1949–50); Recasens Siches, *Legal Philosophy of Suárez*, in Spanish; H. Rommen, *Staatslehre des Suárez* (1927); J. Brown Scott, *Catholic Conception of International Law* (1934); G. Ambrosetti, *Philosophia delle Leggi di Suárez* (1950); Thomas U. Mullaney, *Suárez on Human Freedom* (1950). (H. A. Ro.)

SUBCONSCIOUS, in psychology, means anything that is neither in the focus nor in the margin of consciousness (that is, does not receive attention, and so cannot be recognized in the actual experience of the moment), but which nevertheless must be assumed to be influencing us in some way, as a disposition, etc.

See PSYCHOLOGY; PSYCHOLOGY, ABNORMAL; UNCONSCIOUS, with bibliographies.

SUB-DOMINANT, in music, the 4th degree of the diatonic scale, as F in the scale of C. *See* HARMONY.

SUBIACO (anc. *Sublaqueum*), a town of Italy, Lazio region, situated on the Anio, 1,339 ft. above sea level. Pop. (1957 est.) 9,241 (commune). It was so called from its position under the three artificial lakes constructed in the gorge of the Anio in connection with the aqueduct of the Anio Novus, which had its intake at the lower end of the lowest of them (the *Simbruina stagna* of Tacitus). On the banks of this lake Nero constructed a villa, in the remains of which was found the beautiful headless statue of a youth kneeling, now in the Museo delle Terme at Rome. The lakes ceased to exist, the last dam being washed away in 1305. In 494 St. Benedict retired as a hermit to a cave (Sacro Speco) above the lakes of the Anio. In 505, probably, he founded the first of his 12 monasteries. The church dedicated to Ste. Scholastica, St. Benedict's sister, was erected in 981. In 1053 it was restored and a campanile built, which still exists; and in the middle of the 13th century the church was rebuilt in the Gothic style. Other buildings grew up round it; the cloister on the right is a fine·Romanesque arcaded court with twisted columns and Cosmatesque (13th cent.) mosaics, the south side by Lorenzo Cosmati, the other three sides by his sons (*c.* 1227–1243). *See* BENEDICTINES.

Arnold Pannartz and Conrad Schweinheim, two German ecclesiastics, set up here the first printing press in Italy, issuing an edition of Donatus (1464), followed by one of Cicero (1465) and of Lactantius (1465). Copies of the Lactantius, of the Augustine of 1467, printed not here but in Rome, and of other rare *incunabula* are preserved there.

Still more interesting is the monastery of the Sacro Speco, higher up the hill. The Grotta dei Pastori has frescoes of the 9th century, while the Sacro Speco, or cave of St. Benedict, contains frescoes of the 13th century, and so does the lower church, the latter having been in part repainted in the latter half of that century by an unknown master Conxolus. The upper church contains scenes from the life of Christ by a Sienese master of the end of the 14th century, to whom is also attributable a remarkable fresco of the triumph of death and some 15th-century work, and in the chapel of St. Gregory a portrait of St. Francis of Assisi (who was perhaps here in 1218), probably painted before 1228, as it lacks the halo and the stigmata. The whole group of buildings is constructed against the rocky sides of the gorge, part of it on massive substructions.

The town contains various buildings constructed by Pius VI, who as cardinal was commendatory abbot of Subiaco. It is crowned by a medieval castle constructed originally by Gregory VII. (T. A.)

SUBJECT: *see* TERM; LOGIC.

SUBJECTIVISM, a philosophical term, applied in general to all theories which lay stress on the purely mental sides of experience, opposed to objectivism.

In the narrowest sense subjectivism goes to the logical extreme of denying that mind can know objects at all (*cf.* SOLIPSISM; IDEALISM).

SUBLEYRAS, PIERRE (1699–1749), French painter and engraver whose best-known works are on religious subjects, was born at Uzès, Gard, in 1699. He gained the *Grand Prix* and went to Italy in 1728. His first important work was "Christ's Visit to the House of Simon the Pharisee," which made his reputation. His "St. Basil and the Emperor Valens" was executed in mosaic for St. Peter's. Benedict XIV and all the princes of Rome sat to him, and the pope commanded two great paintings—the "Marriage of St. Catherine" and the "Ecstasy of St. Camilla"—which he placed in his private apartments.

Subleyras shows greater individuality in his curious genre pictures, which he produced in considerable numbers. In his illustrations of La Fontaine and Boccaccio his true relation to the modern era comes out; and his drawings from nature are often admirable. He died in Rome·on May 28, 1749.

His wife, Maria Felice Thibaldi, was a celebrated miniature painter.

SUBMARINE. The submarine first became a major factor in naval warfare during World War I, when Germany demonstrated its full potentialities. However, its advent at that time, marked by wholesale sinkings of Allied shipping, was in reality the culmination of a long process of development.

For the use of submarines in naval warfare, *see* CONVOY; LAWS OF WAR: *Conduct of Hostilities: Submarines;* WASHINGTON CONFERENCE; WASHINGTON, TREATIES OF; WORLD WAR I: *Introduction* and *The Penultimate Year: The Mastering of the Submarine;* WORLD WAR II: *The War at Sea.*

History.—It might well be said that the invention of the submarine goes back to the third century B.C., when Archimedes discovered the laws of floating bodies. From then on, various men realized that by increasing the weight, or by decreasing the displacement of a floating body, it could be made to submerge, and by reversing the process, it could be surfaced again.

However, significant progress beyond this point had to await two developments of the latter part of the 19th century. These were the internal combustion engine for surface propulsion, and the electric storage battery and motor for submerged propulsion. The next significant date is Jan. 17, 1955, when the signal "Underway on nuclear power," from U.S.S. "Nautilus" marked the advent of the true submarine, capable of submergence for unlimited periods.

Not until 1578 does any record appear of a craft designed to be navigated under water. In that year William Bourne, a British mathematician and gunner, published *Inventions and Devises,* in which the "18th Devise" was the design of a completely enclosed boat which could be submerged and rowed under the surface. It consisted of a wooden framework covered with waterproofed leather. It was to be submerged by reducing its volume by contracting the sides through the use of hand vises. Although Bourne never built this boat, a similar construction, sponsored by Magnus Pegelius, was launched in 1605. It is to Cornelis Drebbel, a Dutch inventor, that credit is usually given for building the first submarine. To him is conceded the honour of successfully maneuvering his craft during repeated trials in the Thames, at depths of 12 to 15 ft. beneath the surface.

Drebbel's craft resembled those of Bourne and Pegelius in that its outer hull consisted of greased leather over a wooden framework. Oars, extending through the sides and sealed with tight-fitting leather flaps, provided propulsion either on the surface or when submerged. Drebbel built his first boat in 1620 and followed it later with two others, both larger but embodying the same principles.

Submarine boats seem to have been numerous in the early years of the 18th century. By 1727 no fewer than 14 types had been patented in England alone. An unidentified inventor, whose work is described in the *Gentleman's Magazine* of 1747, introduced an ingenious method of submerging and surfacing his submarine. His craft was to have had a number of goatskins built into the hull, each of which was to be connected to an aperture in the bottom. He planned to submerge the vessel by filling the skins with water and to bring it to the surface again by forcing the water out of the skins with a twisting rod. This was the first approach to the modern ballast tank.

In the American Revolutionary War, a submarine was first used as an offensive weapon in naval warfare. The "Turtle," a one-man submersible invented by David Bushnell and hand-operated by a screw propeller, attempted to sink a British man-of-war in New York harbour. The plan was to attach a charge of gunpowder to the ship's bottom with a screw and explode it with a time fuse. After repeated failures to force the screws through the copper sheathing on the hull of H.M.S. "Eagle," the submarine gave up, released the charge and withdrew. The powder exploded without result, except that the "Eagle" at once decided to shift to a berth farther out to sea.

Although his name is most often associated with the invention of the steamboat, Robert Fulton experimented with submarines at least a decade before he sailed the "Clermont" up the Hudson. Fulton's "Nautilus" was built of steel, in the shape of an elon-

gated oval. It was somewhat similar in structure to the modern submarine. A sail was employed for surface propulsion and a hand-driven screw propeller drove the boat when submerged. A modified form of conning tower was equipped with a porthole for underwater observation since the periscope had not yet been invented. In 1801 Fulton tried to interest France, Britain and the U.S. in his idea, but no nation ventured to sponsor the development of the craft, even though his model was superior to any submarine designed up to that time.

Development of the submarine boat was held back during all of this period by lack of any adequate means of propulsion. Nevertheless, inventors continued resolutely with experiments upon small, hand-propelled submersibles, carrying a crew of not more than six or eight men. On Feb. 17, 1864, a Confederate vessel of this type, the "Hunley," sank the "Housatonic," a Federal corvette that was blockading Charleston harbour. This is the first recorded instance of a submarine sinking a warship. The "Hunley" accomplished the feat by using a torpedo suspended ahead of her bow as she rammed the corvette. Interest in improvement of the submarine was active during the period of the American Civil War, but the problem of a suitable means of propulsion continued to limit progress. Steam was tried and finally in 1880 an English clergyman, the Reverend George William Garrett, successfully operated a submarine with steam from a coal-fired boiler which featured a retractable smokestack. During the same period, a Swedish gun designer, Thorsten Nordenfelt, also constructed a submarine using steam and driven by twin screws. His craft, which could submerge to a depth of 50 ft., was fitted with one of the first practical torpedo tubes.

Meanwhile, electric propulsion machinery had proved its utility in many fields. An all-electric submarine invented in 1884 by two Englishmen, Campbell and Ash, was built in 1886 by Wolseley and Lyon. Their boat was propelled at a surface speed of six knots by two 50-h.p. electric motors operated from a 100-cell storage battery. However, this craft suffered one major handicap —its batteries had to be recharged and overhauled at such short intervals that its effective range never exceeded 80 mi.

Antedating the efforts of Nordenfelt were the experiments of J. P. Holland of New Jersey, who launched his first boat in 1875. Although his early models embodied features that were discontinued as development progressed, many of his initial ideas, perfected in practice, came into later use. Outstanding in importance was the principle of submergence by water ballast and the use of horizontal rudders to dive the boat.

Not until 1895, however, did Holland, in competition with Nordenfelt, finally receive an order for a submarine from the United States government. The vessel, named the "Plunger," was propelled by steam on the surface and by electricity when submerged. The original craft was redesigned frequently during construction and was finally abandoned altogether in favour of a newer model then being built at Elizabeth, N.J. This was Holland's ninth submersible, but it was the first to be delivered to the U.S. government. It was delivered in 1900 and was the basic design of all submarines to follow.

Simon Lake, who began building submarines in 1894, designed them primarily with peacetime uses in mind. His vessels could travel about the sea bottom and had an air lock which permitted a passenger in a diving helmet to emerge from the hull to walk about and explore. In fact, Lake used his extensively in commercial salvaging operations. His first model, the "Argonaut, Jr.," was solely an experimental one. It was built of two layers of yellow pine with a sheet of canvas between them and was operated by hand.

It was followed in 1897 by the "Argonaut," a cigar-shaped hull 36 ft. long and powered by a 30-h.p. gasoline engine. This craft could submerge to the bottom of a lake or river and roll along the bottom on three wheels. For navigating, the wheels could be raised and carried in packets in the keel. In 1898 the "Argonaut" traveled under its own power through the heavy November storms from Norfolk to New York and was thus the first submarine to navigate extensively in the open sea. In 1906 Lake built the "Protector" and sold it to Russia. After it had successfully passed various severe tests there, Lake built a number of submersibles on contract for the Russian government. Thus, the fundamental principles of construction and operation of submarine boats had been determined and demonstrated before the outbreak of World War I. By that time, too, internal-combustion engines, both gasoline and diesel, were available for use as practical power plants. The invention of the periscope had materially increased the feasibility of underwater navigation (*see* PERISCOPE: *Submarine Periscopes*). And the primary weapon of the submarine, the torpedo, had been perfected for use (*see* TORPEDOES). Thus, the preliminary development of the submarine was finished and the vessel was ready to take its place as a major factor in naval strategy. In place of the tiny, one-man contraptions that first dared to venture beneath the surface submarines became effective weapons.

Construction.—While details of modern submarines vary, the general principles of their design are quite similar. The World War II U.S. fleet submarine typifies these general principles and it is this type of submarine that for the most part is described at this point. Nuclear-powered submarines are described below under *Developments After World War II*.

When the submarine rests on the surface, so little of it is seen above the water that it has the appearance of being longer and more slender than it really is. Actually the World War II U.S. submarine was 312 ft. long with a superstructure deck tapering almost to a point, both fore and aft, from its greatest width of approximately 16 ft. amidship.

The personnel aboard the World War II U.S. fleet submarine numbered from 66 to 78 of which from 6 to 8 were officers.

Beneath the superstructure deck is the all-welded hull (actually two hulls). To understand the construction of a submarine, the conditions under which the vessel operates below the surface must first be appreciated. The submarine must at all times be watertight; otherwise, self-destruction would result. Construction is therefore on the basis of the fabrication of a series of watertight containers into one large watertight cylinder by means of watertight joints. The containers must be pressure vessels; that is, watertight containers or cylinders capable of withstanding great pressure. The pressure vessels, being subject to mechanical action (leverage), must be secured to each other by one common strength member (the keel) as well as by watertight connections (bulkheads).

In the double-hull type of submarine, the pressure hull is inside the outer hull; between the two hulls are the water and the fuel oil tanks. On U.S. submarines the double-hull construction extends from the after bulkhead of the forward torpedo room to the forward bulkhead of the after torpedo room. The pressure or inner hull extends from the forward bulkhead of the forward trim tank to the after bulkhead of the after trim tank. Above the hull is built a nonwatertight superstructure which forms the main deck, for use when surfaced.

A topside nonwatertight superstructure deck extends the length of the ship. The space between this deck and the pressure hull is used as a locker for stowing anchor gear and other equipment undamaged by immersion. Ready ammunition in watertight boxes easily accessible to gun crews was kept in the superstructure of World War II U.S. fleet submarines. However, after 1945 the guns of U.S. submarines were removed and ready ammunition stowage deleted. The deck is perforated on each side to prevent air pockets from forming in the superstructure when submerged. A watertight tower (the conning tower), extending up through the superstructure amidships was found during World War II, but is now obsolete for U.S. submarines. When surfaced a submarine is operated from a bridge atop the conning tower. Submerged navigation is accomplished within the conning tower or in a compartment (the control room) directly below. In submerged operation a periscope contained in the conning tower is used for making observations.

The modern fleet-type submarine is built to withstand the pressure of a head of sea water, consistent with requirements as shown by battle experience and with naval specifications. The pressure is measured in actual submergence tests from the surface of the

water to the axis of the vessel through its pressure hull.

Individual compartments are air tested for tightness only, to a pressure of 15 lb. per square inch. Watertight bulkheads are designed structurally and strengthened through reinforcements to withstand the pressure at test depths.

Armaments.—Torpedo tubes are the main offensive and defensive armament of the submarine. A maximum of ten 21-in. tubes is carried by a modern U.S. fleet submarine. The tubes can be fired electrically or by hand when surfaced or submerged. The condition of the tube is indicated by the torpedo ready lights.

During World War II, U.S. fleet submarines carried at least one, and sometimes two, 5-in. 25 calibre deck guns which were capable of use against both air and surface targets. Usually, however, these were used only in an emergency or against small, lightly armed craft not considered worthy of a torpedo. Also light antiaircraft guns (40 mm. and 20 mm.) were mounted forward and aft of the conning tower. Although after the war, deck and antiaircraft guns were removed from U.S. fleet submarines, most non-American submarines carry similar armaments. In addition, submarines carry other small arms and pyrotechnics. Chief among these are machine guns and submachine guns. During World War II, some submarines also carried mines which could be discharged from the torpedo tubes.

The submarine is essentially a torpedo vessel, launching its torpedoes while submerged or surfaced. The size of the torpedo and the number of tubes gradually increased, until by mid-20th century the 21-in. torpedo was being launched from six torpedo tubes forward and four aft, the submarine carrying 24 torpedoes. The tubes are horizontal fore and aft and can be loaded from within the submarine. The tube is fitted with doors at either end so that the inner door can be opened when the outer door is closed and vice versa. The torpedoes are discharged by air pressure either by remote control from the conning tower or at the tubes by operating personnel.

Subdivision.—A modern submarine contains, in addition to the mechanisms required to operate it on the surface, a multitude of operating machinery and tanks which enable it to dive, surface and proceed submerged. Although it is one of the most compact vessels afloat, it is designed and arranged along simple and logical lines, and in spite of the seeming confusion of valves, lines and apparatus, everything in the submarine is situated to ensure the maximum of speed and efficiency.

World War II U.S. fleet submarines consist of a superstructure and a hull surrounded for the most part by various fuel and water ballast tanks. The pressure hull, designed to withstand the sea pressure, houses most of the ship's machinery and provides the living quarters for the officers and the crew. It is divided into eight watertight compartments, separated by pressure bulkheads provided with watertight pressure-resisting doors. The ninth compartment, the conning tower, in the shape of a cylinder placed on its side, is located above the control room and connects with the control room through the access hatch.

Exterior.—The main deck is attached to the exterior hull by means of the framing and rounded sides. Limber holes in the sides allow sea water to enter all hollow spaces in the superstructure and the deck when diving, and drain off when the submarine is surfacing. The midship section of the main deck is occupied by the conning tower, which is surmounted by the bridge deck, with periscope shears, periscope, radio compass loop and radar gear.

The after section of the bridge deck contains the ship's pelorus and one antiaircraft gun (where installed) with adjacent ammunition ready lockers. The forward part of the bridge may be similarly equipped with antiaircraft guns and ammunition lockers.

The forward section of the main deck houses the capstan, marker buoy, forward escape and rescue hatch, cleats, hydrophone, forward torpedo loading hatch and radio antenna. The after section of the deck contains a deck gun, when it is installed, and the galley access hatch, after engine room access hatch, after torpedo loading hatch, marker buoy and capstan.

The bow is equipped with six torpedo tube shutters, three on the port and three on the starboard side, and also the bow diving planes. The underside of the hull contains ballast tanks flooding ports and underwater sound heads.

Tanks.—In a submarine, the principal ballast is water. The main ballast tanks make possible the controlled diving and surfacing of the submarine. In the surface condition they are empty, and in the submerged condition they are completely full. To accomplish the rapid flooding of these tanks, there are large, power-operated vents at the tops of the tanks, and the tank bottoms are either open to sea or have large power-operated flood valves. For more effective control, the tanks are arranged in groups. The power for operating the vents (and flood valves, if installed), is usually hydraulic, although pneumatic equipment has been used. Some of the ballast tanks are designated fuel ballast tanks. While fuel is being carried in these tanks they are kept flooded at all times. When emptied of fuel they are kept empty in the surface condition, and flooded on diving, as are the main ballast tanks.

To dive a submarine, all vents (and flood valves, if fitted) are opened. After leveling off at the desired depth, and ascertaining that the ballast tanks are completely flooded, the vent valves are usually closed. The submarine is then ready to be surfaced at any time by admitting high pressure air to the tanks and expelling the water through the floods in the bottom of the tanks.

To enable the submarine to be kept in neutral buoyancy, trimming tanks are provided to compensate for changes in weights due to expenditure of food, ammunition, fuel, etc. These consist of one or more auxiliary tanks located approximately amidships, a forward trimming tank and an after trimming tank. The former tanks are for control of over-all weight, and the latter for correcting any alteration in the longitudinal distribution of weight. These tanks are equipped with remote reading gauges, and the trimming system is designed for the greatest flexibility in shifting water ballast.

In addition to the above-named water ballast tanks, there are the normal fuel oil tanks, collecting tank, expansion tank, clean fuel oil tank, main sump tanks, reduction gear sump tanks, fresh-water tanks, emergency fresh-water tanks, battery fresh-water tanks and sanitary tanks.

"Water around torpedo" tanks are fitted below the centre of length of the torpedo tubes and contain sufficient water to fill the spaces around the torpedoes when in the tubes. By their use the torpedoes can be left dry in the tubes and withdrawn for examination. If it is desired to fire a torpedo the tube is closed, the water from around the torpedo tanks is blown into the tube and when the door of the tube is opened no water enters and the weight of the submarine is unaltered. It is necessary to keep a constant check of variations in weight on the submarine to maintain a good "diving trim" at all times.

The fuel oil is the main item for which compensation is needed. The fuel is stored in the normal fuel oil tanks or in the fuel ballast tanks. The fuel tanks are completely filled with oil and are symmetrically positioned both transversely and longitudinally. When fuel from a tank is required it is forced to the clean fuel oil tanks and thence to the engines by pumping sea water into the bottom of the tank. The oil passes out through a line at the top of the tank.

Air Systems.—The importance of the air systems to a submarine cannot be overemphasized, for virtually every function in the diving and surfacing procedure stems initially from air provided by one or more of the air systems. The main hydraulic system, for example, operates because of the air pressure maintained in the air-accumulator flask; torpedoes are discharged by air; and the main propulsion engines are started by air. Air or, more specifically, compressed air, is necessary to surface, submerge, attack and cruise. Compressed air together with oxygen is used to revitalize the air in the ship after long periods of submergence. Pressure in the boat, a test for tightness, utilizes air.

Compressed air is stored in banks consisting of seven or more flasks with a total capacity of 300–700 cu.ft. at 2,000–3,000 lb. per square inch. It is supplied by two high-pressure air compressors located in the pump room, a cellarlike compartment located directly beneath the control room. Compressed air can be ob-

tained also from a tender ship through outside connections.

Trim and Drain Systems.—A submarine is designed to dive or surface rapidly under complete control; it must be able to proceed on the surface and to submerge at the desired rate of speed to the depths required. To do so quickly and efficiently, it must maintain fore-and-aft balance and athwartship stability. The chief function of the trim and drain systems is to maintain this fore-and-aft balance by controlling the amount and distribution of water in the various tanks used for this purpose. Motor-driven pumps are provided for pumping out the submarine's tanks or bilges; they are capable of pumping against any depth pressure at which the vessel is designed to operate. These pumps deliver or discharge overboard or from one tank to another, as desired. They are also capable of pumping certain ballast tanks in case of compressed-air failure.

The rudder of the submarine is similar to that of any surface vessel. Under normal operation the steering system has its own source of electrohydraulic power. The principal control units are assembled in the steering stand, located in the control room. However, since there is a steering wheel in the conning tower, to allow for every contingency the steering system is so planned that three different methods of steering are available, based on three different sources of hydraulic power—normally used power, hand and emergency.

The rudder itself is moved by hydraulic power in all three cases; the only difference between these methods is in the manner in which the power is developed.

Bow and Stern Planes.—The hydroplanes or bow and stern planes, which are actually horizontal rudders, are used for depth and angle control of the submarine when operating submerged. They are rectangular in shape and can be turned through 35° either way. They also are controlled electrohydraulically. The after pair when in a neutral position give the submarine a certain stability. The forward pair can be folded into the side of the superstructure which is called the housing. This decreases water resistance and the planes are not subject to additional strain when on the surface. The bow planes also assist in submerging the vessel as they overcome the slight amount of buoyancy carried by the submarine, present even in a submerged condition. Hydraulic power is used to tilt the bow and stern planes. Each system (bow and stern planes) has its own power supply system. The control units for diving and rising are assembled in a diving control stand located in the control room. There is a set of controls for stern plane tilting, a set for bow plane tilting and control valves for bow plane rigging. The control panel has diving indicator gauges and motor switches. The planes are operated by crew members.

Periscopes.—The periscope is the eye of the submarine. It was invented and developed solely for the purpose of providing a means to view the surface without detection by surface craft. While it is primarily simple in principle, actually it is a complicated piece of apparatus. It is probable that all navies of the world have similar instruments with only minor variations.

The earliest submarines were built without provision for periscopes and when submerged were forced to grope their way blindly. In 1854 E. M. Marié-Davy, a Frenchman, designed a submarine sight tube containing two mirrors, one above the other, held at a 45° angle and facing in opposite directions. These, while providing some degree of sight to the submerged vessel, were faulty at best, and in 1872 prisms were substituted for mirrors.

A U.S. naval man, Thomas H. Doughty, was the inventor of the original modern periscope, during the American Civil War.

The essential function of a periscope is to give the person conning for the submarine a view of the surrounding horizon while his vessel remains submerged. The periscope must be long enough to extend beyond the surface and must deflect the horizontal rays of light first in a downward direction and then horizontally to the eye of the observer. In addition, the part above water must be as inconspicuous and streamlined as possible, in the form of a long narrow tube. The instrument used on U.S. submarines is of 40-ft. nominal length and $7\frac{1}{2}$-in. outer diameter. It is equipped with a tilting head prism capable of elevating the line of sight 74.5° above the horizon and of correcting for the roll or pitch of the vessel.

Propulsion Engines.—The first U.S. submarines utilizing internal-combustion engines for propulsion were powered by 45-h.p., two-cylinder, four-stroke cycle gasoline engines produced by the Otto company of Philadelphia, Pa. Meanwhile, early British submarines made use of 12- and 16-cylinder gasoline engines.

The inherent hazards accompanying the use of such a highly volatile fuel as gasoline were quickly realized. Stowage was a constant problem and handling of the fuel was extremely dangerous. Internal explosions were frequent; many of the engines gave off considerable carbon monoxide fumes creating a menace to personnel.

In the meantime, M.A.N. (Maschinenfabrik Augsberg-Nürnberg A.-G.) of Germany built and experimented with two-stroke cycle diesel engines for submarine propulsion. However, insufficient progress had been made in metallurgy to provide metals capable of withstanding the greater heat and stress inherent in diesels. M.A.N. then turned its efforts toward production of a four-stroke cycle diesel engine capable of developing 1,000 h.p. While fairly successful, these engines eventually developed structural weaknesses at the crankcase. Prior to 1930 the engines used in most submarines of all the larger naval powers with the exception of Great Britain were four-stroke cycle diesels. The U.S. navy, however, experimented with a two-stroke cycle Busch-Sulzer engine and equipped a number of boats with this type. About this time also the U.S. navy turned to the use of solid-injection engines, with their advantage in weight reduction.

The need for more powerful engines paralleled development of the U.S. fleet-type submarine in the 1930s. Eventually, a 16-cylinder single-acting engine was built (General Motors) as well as 9-cylinder opposed piston and 8-cylinder double-acting engines (Fairbanks-Morse and Hooven-Owens-Rentschler respectively).

Along with the development of these engines there was experimentation in the method of transmitting the power of an engine to a submarine's propellers. Originally, engines were coupled to the propellers through friction clutches. With the advent of high-speed engines came reduction gears to keep propeller speed in an efficient range, and the mechanical clutch was replaced by electric drive. The power developed by the engine-driven generators is used in the two main motors for surface propulsion, or for charging batteries.

The modern fleet-type submarine is usually driven by four main propulsion diesel engines, each capable of producing 1,600 h.p. The four main generators each produce 1,100 kw. There are four slow-speed motors, driven by the generators or batteries, each producing about 1,375 h.p. The auxiliary engine is rated at 450 h.p. and drives a 300-kw. generator.

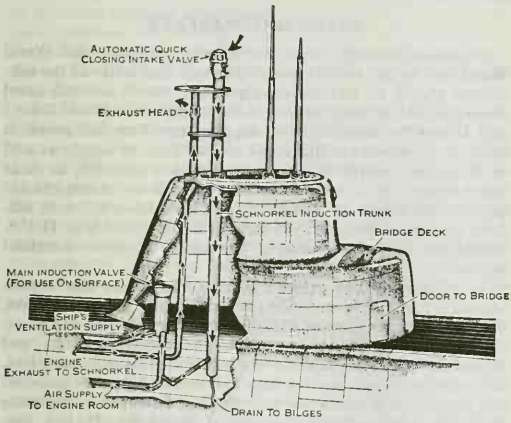
Batteries.—When submerged the diesels and generators are not used, and power for the motors is supplied by storage batteries which are charged by the auxiliary and main generators during surface operations.

Most U.S. submarines have two main storage batteries of 126 cells each. Usually, the forward battery is installed below decks in the area of the wardroom while the after battery is located in the crew's space. Each cell weighs approximately 1,650 lb. Each battery is fitted with an exhaust ventilating system to remove battery gases. The air required to operate this system is supplied through inlets located at opposite ends of each battery well. The free air in the compartment is drawn through the filling vent connection of each cell. This is necessary in order to eliminate battery gases and hydrogen which develop when recharging the batteries.

Escape Apparatus.—For many years there were three vital pieces of gear used in escaping from a submerged submarine. The first of these, the marker buoy, is released by a disabled submarine to assist in guiding rescue parties to the area where the boat rests. If a properly equipped rescue party is able to locate the submarine, the second piece of gear, a rescue chamber, is lowered by cable over an escape hatch and the crew members of the submarine are removed in relays to the surface. However, if rescue parties

are unable to locate the submarine individual ascent is necessary. For many years this ascent was made with the assistance of an artificial lung of which the two main types were the Davis submerged escape apparatus, used by the British, and the Momsen lung, used in the United States. By the mid-1950s, however, the use of these lungs was abandoned by both British and American navies in favour of a buoyant ascent, in which the escapee is assisted by no equipment at all other than a life jacket.

The Schnorkel and Guppy.—In the latter part of World War II the Germans adopted a radical change in submarine design



SCHEMATIC DRAWING OF THE SCHNORKEL WITH RETRACTABLE PERISCOPES AND MAST SHOWN HERE IN UP POSITION

known as the schnorkel—a Dutch invention. The spelling was reduced by the Americans to snorkel and further changed by the British to snort. Development of improved sonar (underwater sound-detecting and ranging equipment) and radar (*see* NAVIGATION: *Electronic Aids to Navigation;* SOUND: *Applications of Sound; Military Acoustics: Underwater Sound*), used by both aircraft and surface craft against axis submarines, forced the Germans to develop the new device.

The schnorkel is a breathing tube which is raised while the submarine is at periscope depth. With the schnorkel in raised position, air for the diesels can be obtained from the surface. This considerably reduces the visible portion of the submarine and also electrical consumption, since the submarine can cruise almost totally submerged on its engines and conserve its battery power for attack and evasive maneuvers.

At the end of World War II the United States navy developed an improved schnorkel and also the guppy submarine. The guppy ("greater underwater propulsion power") had the same type hull as that used for the fleet-type submarine of World War II fame. It was 306 ft. long and displaced about 1,800 tons. The only change in the hull was in the superstructure, which was radically changed by reducing the surface area and streamlining every protruding object. The life lines and all guns were removed, the bitts made retractable and the periscope shears (supports) enclosed in a streamlined metal fairing. All topside equipment and armament were either removed, made retractable or streamlined. The speed of the new guppy was considerably greater than that of the fleet-type boat, and nearly twice that of the old-style submarines.

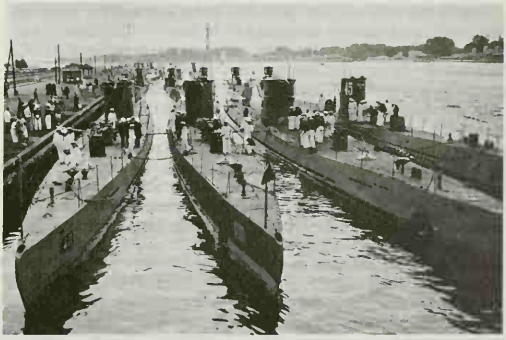
Developments After World War II.—The lessons of World War II wrought great changes in submarine construction. Torpedoes came to be fired by hydraulic (water) pressure instead of compressed air as previously, eliminating any chance of tell-tale air bubbles escaping to the surface and betraying the submarine's presence.

Hulls were given greater strength for deeper diving that enabled the submarine to pierce depths far below those previously possible. Hydraulic mechanisms (oil) were employed extensively throughout the ship for quiet, efficient operation. Radar and sonar equipment increased in importance as it increased in efficiency. Torpedoes no longer had to be fired with sharpshooting accuracy; fire-control equipment and homing torpedoes (the torpedo that seeks its target) eliminated much of the guesswork and chance in firing torpedoes. Engines, electrical motors and generators were made more compact.

On Jan. 17, 1955, the U.S.S. "Nautilus" made history by getting underway on nuclear power. In this propelling plant, the heat of a nuclear reactor is transmitted to water, generating steam which is used to drive the main propulsion turbines, and turbo-generators. The same power plant is used surface and submerged, thus freeing the ship from the necessity of coming to the surface to charge batteries. Here at last was the true submarine. "Nautilus" can maintain submerged speeds in excess of 20 knots for several weeks, during which time she could, if called on to do it, more than circumnavigate the globe. Her capabilities were shown in Aug. 1958 when she made the first sea voyage to the North Pole, passing beneath the polar ice cap on her way from Hawaii to England. The ship's air-conditioned interior is designed for the greatest possible comfort of the crew. Air is filtered and recirculated through carbon dioxide scrubbers, and oxygen is added as required.

Although "Nautilus" is capable of previously unattainable speeds submerged, her speed is not the ultimate, as her hull is designed along more or less conventional designs. The U.S.S. "Albacore" was designed and built for the sole purpose of carrying out experimentation in hydrodynamics. The greatly increased submerged speeds raised problems in this field somewhat similar to those encountered in passing the sonic barrier, in a much smaller degree. Her function was to prove the hull design, as that of "Nautilus" was to develop the power plant. Fat and short, with only one propeller, she was the fastest submarine afloat.

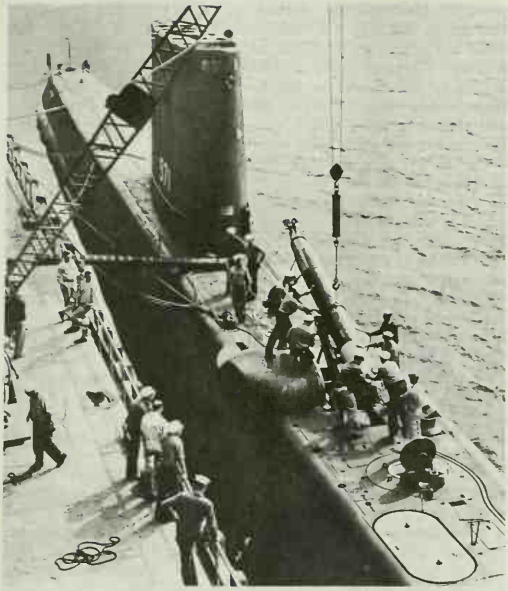
Inasmuch as in the late 1950s, the United States, Great Britain, France, and presumably the Soviet Union, all were either building or planning to build nuclear-powered submarines, it seemed that this new era would be characterized by submarine fleets which would become more and more completely nuclear powered. The new era also would be one of more specialized uses for submarines.



EUROPEAN

GERMAN U-BOATS USED IN THE EARLY PART OF WORLD WAR II

For example, radar picket submarines, such as the U.S.S. "Triton" (laid down 1956), would maintain radar lookouts far at sea. "Triton," said to be the largest naval submarine ever attempted, displacing about 5,500 tons, was to have as its main "armament" elaborate radar equipment, supported by an extensive communications system to transmit the radar intelligence into appropriate defense systems. As navies came to depend more and more upon missiles to augment the traditional armament of gun and torpedo, submarines such as the U.S.S. "Halibut" (laid down 1956), whose main armament would consist of missiles, were being built. Many



BY COURTESY OF U.S. NAVY

LOADING A TORPEDO ABOARD THE U.S.S. "NAUTILUS," THE FIRST NUCLEAR SUBMARINE

of these submarines would carry rockets similar to "Polaris," a development of the U.S. navy, which could be fired up to ranges of 1,500 mi. from a submerged position. As a result of such specialization, submarines would become of great value for exceedingly long-range shore bombardment and for the collection of strategic intelligence, as well as acting as the highly efficient commerce destroyers, antisubmarine weapons, and scouting craft which they had always been.

But in this radically new period of underseas warfare more conventional developments continued to be made. The British navy had great success, for example, with the experimental H.M.S. "Explorer" (accepted into service 1956), in using hydrogen peroxide or closed circuit engines for submerged propulsion. As with the use of nuclear power, the benefit of this system is that a submarine can run submerged on the same engines it uses on the surface, with long endurance and a high rate of speed. Basically, such performance is made possible by using the oxygen in liquid hydrogen peroxide, rather than that of the atmosphere, to burn fuel oil. As a result of World War II experiences in which midget submarines were shown to have value in harbour and riverine warfare, as well as in scouting, several navies continued the development of this type. Probably the most extensive work was done by the British. Typically these submarines are about 50 ft. in length and carry a crew of five officers and men. Some carry as armament in place of or in addition to torpedoes, magnetic limpet mines, which, like the "torpedo" carried by the "Turtle" in her unsuccessful attack upon H.M.S. "Eagle" during the American Revolution, are intended to be attached to the hulls of enemy vessels until detonated by a timing device.

By the mid-1960s the size of submarine fleets and the increasing attention being paid to their development clearly indicated their importance. The largest force, that of the U.S.S.R., was composed of over 440 submarines. The United States had over 130 boats, and Great Britain had about 44. A number of those in each fleet were nuclear powered and carried ballistic missiles.

BIBLIOGRAPHY.—F. W. Lipscomb, *The British Submarine* (1954); Robert H. Barnes, *United States Submarines*, 3rd ed. (1946); for further detailed bibliographical information, *see* National Research Council, Committee on Undersea Warfare, *Bibliography of the Submarine, 1557–1953* (1954). (H. S. Dy.; F. K. L.)

SUBMARINE WARFARE

Submarine warfare did not gain major importance until World War I, but in that conflict and in the years that followed the submarine played an increasingly significant role in over-all naval strategy. The primary mission of the submarine in World Wars I and II was the destruction of shipping, merchant and naval, in order to disorganize military and civilian lines of supply as well as to contest control of the sea in areas not accessible to other naval forces. In both wars Germany failed in her attempt to cut supply lines across the Atlantic though her submarines sank millions of tons of shipping. In the Pacific in World War II U.S. submarines virtually destroyed the entire Japanese merchant marine.

In addition to their primary mission, submarines also demonstrated a remarkable versatility at a number of subsidiary tasks. They landed raiders, spies and saboteurs; carried arms and provisions to advanced or isolated elements and to guerrilla forces; and reconnoitered landing beaches and bombarded shore installations. U.S. submarines on scouting duty first reported the movement of the Japanese fleet which was subsequently destroyed in the battle of Surigao strait. Toward the end of World War II, U.S. submarines were largely employed on lifeguard duty to rescue crews of bombers forced down off the coast of Japan. After the war the United States modified a few submarines specifically as radar pickets, troop transports and tankers. In short, submarines are capable of carrying out in a limited way almost all of the tasks assigned to naval forces.

World War II brought about sweeping changes in the character of submarine warfare, chiefly through the introduction of radar. Radar-equipped U.S. submarines were able to conduct devastating night attacks on Japanese convoys. Radar-equipped Allied aircraft denied the German submarines the use of the surface of the sea, forcing them to resort to the schnorkel breathing tube in order to operate their diesel engines. Attempts to streamline submarines and provide greater storage-battery capacity came too late to be factors in deciding the battle of the Atlantic.

Although the schnorkel reduced the toll taken by radar-equipped aircraft, it still tied the submarine to the atmosphere and to the surface of the ocean where it was noisy, detectable by radar and sonar and vulnerable to attack by air or surface forces. To overcome these limitations, efforts were made to propel submarines with closed-cycle engines that carried their own oxidants

U.S.S. "TRITON," TWIN-REACTOR NUCLEAR SUBMARINE WHICH IN 1960 BECAME THE FIRST SUBMARINE TO ENCIRCLE THE GLOBE WITHOUT COMPLETELY SURFACING

BY COURTESY OF U.S. NAVY

within the submarine. The possibility of using primary batteries containing active elements that could be replenished by materials carried aboard was also studied. Most importantly, nuclear power pointed the way to a "true" submarine—one that would be independent of the atmosphere and the surface of the ocean.

The significance of nuclear-powered submarines is tremendous. Heretofore antisubmarine forces knew that a submarine would eventually have to come to the surface or near it to get air for its diesel engines. The faster it ran or the more violently it maneuvered under water, the sooner it would have to come up. The nuclear submarine can remain under water indefinitely and travel at maximum speed as long as necessary. It does not have to expose itself to observation by aircraft radar. It is independent of the weather and has an endurance, in a single cruise, of tens of thousands of miles. The problems posed by such a ship to the antisubmarine forces is staggering.

But the nuclear submarine has its problems too. While deeply submerged it cannot use its periscope, radar or radio, for they require operation close to the surface at slow speed. In order fully to exploit its capability of remaining indefinitely in the depths the submarine must navigate, search, attack and escape almost entirely by underwater sound, since underwater acoustic energy is the best known agency for penetrating the opacity of the ocean. Similarly, antisubmarine forces must search for the submarine by underwater sound, for vision, radar or magnetic devices are of limited effectiveness against the deep-running submarine.

The picture given by underwater sound is at best a distorted one. Acoustically the ocean is like a giant layer cake within whose layers sound is channeled, refracted, reflected or absorbed, depending upon the physical characteristics of the water at the moment. Because the submarine can make itself very quiet and can select the best depth for listening purposes, it is particularly suited for fighting in the depths against other submarines. Killer submarines lying quietly in ambush can detect noisy submarines at great distances. Submarines proceeding to or from patrol stations are thus exposed to attack from the time they leave port.

The relative invulnerability of the submarine when it chooses to operate quietly makes it a logical choice for a missile-launching platform. Submarines may carry either the cruise-type missile, such as the U.S. navy's Regulus, which requires that the submarine surface before firing, or they may carry ballistic missiles such as the 1,500-mi. Polaris that can be launched while the submarine is submerged. Missile-launching submarines make it virtually impossible for the enemy to locate the launching sites. Massive retaliation through submarine-launched missiles is almost impossible to blunt. The location of the launch sites at sea draws counterattack away from the continental limits of the country. For these reasons the U.S. navy proceeded with the Polaris program at the highest priority.

The advent of the nuclear submarine opened to naval operations vast areas of the Arctic ocean, hitherto inaccessible to ships. Operations of the nuclear submarines Nautilus and Skate under the polar ice pack in 1958 and 1959 proved that these ships could operate effectively under the ice in both winter and summer. The military significance of this area was expected to increase as further experience was gained.

The potential of the nuclear submarine for the future seems almost limitless. The true submarine can be expected to generate air, water and perhaps even food from the sea surrounding it. It may become the most self-sufficient war vessel of all time. Since nuclear submarines do not operate on the surface for long periods of time, their hull forms can be modified to obtain greater hydrodynamic efficiency with increased underwater speed and maneuverability. These qualities, coupled with the submarine's capability to hunt and attack while completely hidden in the depths of the ocean, promise that the submarine will play an increasingly important role in naval strategy with the passing years. *See* WORLD WAR I: *Naval;* WORLD WAR II: *The War at Sea; see* also references under "Submarine" in the Index.

See U.S. Navy, Bureau of Naval Personnel, *U.S. Submarine Operations in World War II* (1949); H. F. Cope and Walter Karig, *Battle Submerged* (1951). (F. D. WR.)

SUBMEDIANT, in music, the sixth degree of the diatonic scale, as A in the scale of C. *See* HARMONY.

SUBOTICA, a town of Serbia, Yugoslavia. Pop. (1961) 74,832. It is the centre of an immense agricultural district, and in the adjoining territory of Puszta Telecska large herds of cattle are reared. The chief industries are the making of iron articles of furniture, railway trucks and boots. The town was occupied by Hungary during World War II.

SUBPOENA, the common procedure for requiring the attendance of a witness at a judicial or investigative proceeding. It is a writ or official command to a named person to appear at a particular time and place to testify as a witness. The *subpoena duces tecum* additionally requires the prospective witness to bring with him books, records, documents or other things in his possession relevant to the subject matter under inquiry. The mandatory writ or process of a court of chancery requiring a person to appear and answer matters charged against him is also termed a subpoena. (C. E. CL.)

SUBSIDY, a term denoting the promotion of private business enterprise by direct or indirect government action, frequently involving the granting of special privileges. Government subsidies to shipping, aviation, mining, agriculture, and other industries are granted in many countries throughout the world. *See* AGRICULTURAL ECONOMICS: *Agricultural Prices and Incomes;* AVIATION, CIVIL; SHIPPING, HISTORY OF: *International Rivalry;* SHIPBUILDING (MERCHANT AND NAVAL): *Subsidies.*

Though some subsidies may be criticized as obvious favours to special interest groups at the expense of the economy as a whole, others may serve to encourage activities that are generally regarded as socially desirable. Subsidies to aid the settlement of new lands, to support bona fide infant industries, to provide government statistical services, and to carry on basic scientific research (with results made available to all) are rarely questioned. On the other hand, paying above-market prices for certain metals, subsidizing irrigation when excess acreage already exists, making payments to a select group of growers of agricultural products, and giving favoured tax treatment to certain industries may be of dubious merit.

It is necessary to distinguish direct or visible subsidies, such as airmail subsidy grants, direct payments for ship construction, and payments to cover operating deficits, from indirect or concealed subsidies, which are more widespread but are often not recognized as subsidies. The latter include tariffs, quotas, and other trade restrictions that exclude foreign competitors; the provision of government services for which no charge is made; the sale of goods and services for which no, or an insufficient, charge is made; the purchase of goods and services for more than they are worth; government loans at lower than the market rate of interest; and the exemption of some enterprises from various forms of taxes that others in a similar position must pay.

It is obvious, of course, that the very existence of a government, which includes a legal system, a monetary system, and the establishment of standards and grades of quality, facilitates the conduct of business. A subsidy goes beyond such activities, however; its purpose usually is to maintain or improve the position of certain businesses or private interest groups, without commensurate public gains. In this sense, subsidies are an outgrowth of the protectionism of the mercantile system (*q.v.*), under which governments used special forms of regulation or restraint, particularly import duties, quotas, and fiscal expedients, in order to maintain, encourage, and promote "essential" industries that were endangered by foreign competition. Tariffs, import quotas, and other trade restrictions serve to exclude competitors from the home market or limit their access to it, thus enabling protected sellers to charge higher prices. The result is a transfer of income (or subsidy) from consumers of the protected goods to the favoured domestic producer.

In the case of industries that do not need protection to remain in business, tariffs and other restrictive devices merely enhance profits. Industry protection through trade restrictions causes a shift in the nation's resources from more productive to less

productive uses. The objective underlying a policy of protection, aside from questions of national security and power, is to provide a balance of economic activities and power among various economic groups, both nationally and internationally. Use of such a doctrine to protect infant industries of a country until they expand sufficiently to attain an efficient level of operation often is extended to justify a variety of protectionist measures, without demonstrable public gains to either the domestic or international community. Once protectionist measures are justified on the ground that they benefit infant industries, they are often extended to cover lost markets (both real and imagined), as well as new markets. Thus, protectionist devices and subsidies are in reality practical expedients, not unlike other forms of controls designed to deal with specific problems or to accommodate special interest groups.

Nations have retreated in varying degrees from mercantilist protective measures toward freer trade, but other methods of protection, support, and promotion of private industry have become important in the functioning of economic systems of today. In the United States, where businessmen lay great stress on preserving the free-enterprise system, subsidy has become so pervasive as a means of effectuating public policy that it can properly be said the economy has entered an age of government by subsidy. In the 1960s, for example, up to $15,000,000,000 per year, or roughly 15% to 20% of the total U.S. budget, was allocated for benefits to interest groups, such as farmers, small and big businesses, veterans, and others, not including hidden subsidies covered under such items as interest, national security, federal highway program, and international affairs.

Direct Subsidies.—Railroads, transoceanic shipping companies, shipbuilders, and airlines have been given various forms of subsidy by almost all developed countries at one time or another. Direct gifts of land and construction funds were given to the railroads in connection with the settlement and development of the United States. In more recent years, railroad subsidies have consisted of authorization for railroads to amortize new investment more rapidly than usual, thus reducing their tax liability, and government guarantee of repayment of private loans to the railroads for the purchase of new equipment.

The most important subsidies in this field are for ship construction, payments for operation of transoceanic merchant vessels, establishment of cargo preferences for domestic lines, and exemption of profits from corporate income taxes under certain conditions. The U.S. government ship construction subsidy amounts to the differential between cost of construction in U.S. and in foreign shipyards, not to exceed one-half the cost (except where defense features are included). The liner "United States," for example, cost approximately $77,000,000, of which taxpayers provided $44,000,000. The subsidy to operators of transoceanic vessels, hidden for many years in excessive payments for carrying the U.S. mail, was brought out in the open during the 1930s as an "operating differential" between the cost assumed to be involved in operating under the U.S. flag rather than under a foreign flag.

Operating subsidies, or airmail subsidies, are fairly common as a means of promoting airlines. As early as 1924, for example, the British government provided a subsidy for Imperial Airways, Ltd., to develop air routes to the Commonwealth countries. During the 1950s the United States paid commercial airlines $500,000,000 in such subsidies.

Agriculture has received various direct subsidies, including gifts of land, export subsidies, and provision of a variety of services, such as market information, agricultural research and education, and protection from crop pests and animal diseases. Federal support for agricultural education and research alone have averaged several hundred million dollars per year in the United States.

Indirect Subsidies.—Indirect subsidies are the most widespread form of assistance to private enterprise, particularly in the United States. One type, the payment in excess of actual cost or market value, includes the purchase of silver for monetary purposes, the purchase of certain metals for stockpiling, and the purchase of agricultural commodities. Under domestic producer preference laws, government procurement officers must purchase goods made at home unless their price exceeds the price of goods made abroad by more than a specific percentage. The U.S. mining industry benefits from a federal law requiring the government to buy supplies at above market prices and in excess of its needs. For many years the entire domestic output of silver has been sold to the U.S. government at prices in excess of the international price, and most of it has been reburied; meanwhile the silver used by industry for commercial purposes was purchased from foreign sources at lower prices. Federal, state, and local governments all have provided subsidies to shipping by dredging rivers and building canals. Aviation has benefited from the construction of airports, provision of airways, traffic control equipment, weather-reporting facilities, radio beams, and instrument landing systems.

One of the most important subsidies in the United States is the price support program for 12 basic crops, with special emphasis on wheat, corn, and cotton. The subsidy operates through government "nonrecourse" loans: if market prices go above an arbitrary support level (minimum price), the farmer can pay off the loan and repossess his crop pledged as collateral. If prices fall below the support level, the farmer defaults on his loan, keeps the money loaned him, and the government takes the loss. These subsidies amounted to about $1,000,000,000 per year in the early 1960s, with principal benefits accruing to the large, corporate-sized farms.

Indirect subsidies to farm operators also arise in connection with the building of farm-to-market roads, which in part are paid for with city gasoline tax revenues and federal support.

Other indirect subsidies consist of free or partially requited services, ostensibly available to all persons and businesses but actually of benefit to only a few enterprises. The trucking industry, for example, is subsidized indirectly to the extent that taxes imposed on trucks to cover the cost of highway construction and maintenance do not equal the benefits received. In addition, the tax burden falls heaviest on light, gasoline-powered vehicles (including passenger cars) as against heavy, diesel-powered vehicles.

Owners of irrigated farmlands are heavily subsidized; settlers are given long-term, low-interest loans that work out so that they pay only about one-half of the real cost to the government; in multipurpose water projects, the direct beneficiaries of federal irrigation pay only a minor part of their cost, the remainder being paid by taxpayers and other water users. Power companies receive subsidies in connection with the generation of electricity from atomic fuels; *i.e.*, in government-supported research, the price charged for the nuclear material, and the price paid by the government for nuclear wastes and residues.

Two other important types of indirect subsidies arise out of government guarantees and various kinds of special tax reduction to private industry. In the first, the government assumes a contingent liability on private investments, debts, and contracts, without requiring an equivalent public benefit, such as an equivalent reduction in price. Included here are guarantees on investments abroad, on loans to feeder or local-service airlines, on mortgage lenders' loans to home buyers, and on loans to small businesses, farmers, and railroads. Tax reduction as a subsidy includes the exemption of new businesses from local property taxes and of income from certain securities, accelerated amortization of plants (which as a minimum provides an interest-free loan), and the special "depletion" allowance for the mining industry. In the case of mining, not only are the costs of exploration and development deductible as current costs for tax purposes but mining companies also have the option of making an arbitrary deduction for the "depletion" of assets (in the United States at rates of from 5% of gross income for coal, sand, and gravel up to 27½% for oil and gas), instead of deducting the *actual* depletion through the estimated life of the assets. In the United States the depletion allowance is limited to a maximum of 50% of net income, but this can reduce corporate tax liability from 52% down to 26%.

Other indirect subsidies include the sale of war plants and merchant vessels to private industry at a fraction of their original costs and disposal of surplus commodities at less than buying prices plus storage and interest costs.

BIBLIOGRAPHY.—Walter Adams and Horace M. Gray, *Monopoly in America: the Government as Promoter* (1955); James C. Nelson, *Railroad Transportation and Public Policy* (1959); Murray R. Benedict, *Can We Solve the Farm Problem?* (1955); J. C. Lessard, *Transportation in Canada* (1956); Otto Eckstein, *Water-Resource Development* (1958); Commission on the Organization of the Executive Branch of the Government, *Task Force Report on Lending Agencies*, (1949); Clair Wilcox, *Public Policies Toward Business* (1960); Howard S. Piquet, *Aid, Trade and the Tariff* (1953); Edmund Blair Bolles, *How to Get Rich in Washington* (1952); Eli Clemens, *Economics and Public Utilities* (1950). (R. F. LA.)

SUBSTANCE. In traditional philosophy a substance is a thing that exists independently and persists through change; but the notion has been developed in different ways by different philosophers, and it is impossible to judge of its value without a summary of its history.

Origins of the Term.—The Greek word sometimes translated as "substance" is *ousia*. Its earliest and most widespread use in Greek literature is for property or estate, but it is similar in origin to "being" and it was in fact used by Plato and Aristotle in a non-technical way to signify being in the sense of "essence," being in the sense of "existence" and being as opposed to becoming. The beginning of the technical usage now rendered by "substance" is to be found in Aristotle's *Categories*, 5:

"Substance (*ousia*) in the . . . primary . . . sense of the word is that which is neither predicable of a subject nor present in a subject. . . . But in a secondary sense those things are called substances within which, as species, the primary substances are included. . . . Thus everything except primary substances is either predicated of primary substances or is present in them. . . . The fact that parts of substances appear to be present in the whole, as in a subject, should not make us apprehensive lest we have to admit that such parts are not substances: for in explaining the phrase 'present in a subject,' we stated that we meant 'otherwise than as parts in a whole.' . . . The most distinctive mark of substance appears to be that, while remaining numerically one and the same, it is capable of admitting contrary qualities. . . . The same individual . . . is at one time warm, at another cold. . . . This capacity is found nowhere else."

Use of "being" as a translation for *ousia* in this passage might perhaps give a clearer impression of Aristotle's intention. For he does not wish to fix the sense of a term that would otherwise be unintelligible. On the contrary, assuming that his reader knows the meaning of *ousia*, he asserts the existence of certain *protai ousiai* ("prime beings,"; *e.g.*, Socrates), presupposed by all talk of other things. Similarly, in his *Metaphysics* he announces that *ousia* is the principal subject of his study, as though everyone knew what he meant by that term, and then sets out to determine where it may be found. His conclusion in this work is similar to that in the *Categories*, but in his argument he attacks the doctrines of other philosophers (*e.g.*, Ionian theories about a primal stuff and Plato's assertion that "forms" exist independently), and he also makes a rather bewildering shift in his use of *ousia*. If we here translate the Greek by "substance," we must allow for a distinction between *a* substance (*i.e.*, an individual thing) and *the* substance of a thing (*i.e.*, what is fundamental in its being). According to *Metaphysics*, vii, the substance of a thing is what makes it what it is, and this is its form or essence (*i.e.*, the defining character of the species which in the *Categories* he described as a substance in a secondary sense). In addition to their form all ordinary substances have matter, and it is this which distinguishes them as individuals, though matter cannot exist without form. There are, however, some substances (God and the Intelligences in control of the planetary spheres) which have no matter; Aristotle does not explain how they are individuated.

The Latin word *substantia* appears first in the younger Seneca's *Epistulae*, 58, with the sense of "reality"; and slightly later occurrences show that at the end of the 1st century A.D. it had the whole range of uses of the Greek *ousia*. This exact correspondence is curious, since the etymology of *substantia* (literally "standing under" or "near") is different from that of *ousia*, but the circumstances in which *substantia* occurs suggest that it was not a deliberate neologism introduced for the purpose of translating the Greek. In later Latin *substantia* retains its wide range of usages: thus *consubstantialis* occurs in the Latin version of the Nicene creed as a translation for the Greek *homoousios*, meaning "of the same essence." In English "substance" has acquired yet other meanings (*e.g.*, "kind of stuff"), but is now generally reserved by philosophers for Aristotle's "prime being," while other terms such as "essence" and "existence" are used for the other senses of *ousia*.

Medieval and Modern Theories.—Medieval Aristotelians, such as St. Albertus Magnus and St. Thomas Aquinas, retained the master's doctrine of substance for the most part unchanged; but they felt difficulties about individuation that had not distressed him. According to St. Thomas, angels are without matter (as Aristotle said of his Intelligences) and must therefore be distinguished one from another in respect of their forms; *i.e.*, must each be of a different species (*Summa Theologica*, I. x. 6). The problem presented by human souls is even more serious. According to Aristotle's *De Anima*, which St. Thomas followed, a soul is the form of the living body that it animates. This view may perhaps be reconciled with Christian doctrine, but what of the interval between death and the resurrection of the body? During that time, it seems, souls must continue to exist as separated forms, though this condition is unnatural for them (*Summa Theologica*, I. lxxvi). These were among the difficulties that produced a reaction against Aristotelianism during the Renaissance; but to the founders of modern physics the most irritating feature of the system was that it deluded men into thinking that they had given an explanation of a fact when they had merely formulated it in a technical jargon. The statement that a substance originated through the reception of a form in matter might perhaps be defensible by an array of definitions, but it was useless for science.

This is not to say, however, that physicists of the 17th century had no use for the concept of a thing that existed independently of all else and persisted through change. On the contrary, many of them adopted the atomic theory suggested in antiquity by Democritus, and their reasons were *a priori*, like those of its ancient champions; for it was not until the time of John Dalton (1766–1844) that the atomic hypothesis was used in a scientific fashion to explain empirical generalizations. Since ordinary observable things come into being by the aggregation of parts and are destroyed by the separation of parts, it is plausible to suppose that there are very small things which never come into being or pass away but constitute all observable things by entering into close spatial relations with each other. This conclusion seems, indeed, to be required by the principle that nothing is made of nothing or resolved into nothing (Lucretius, *De Rerum Natura*, i. 147 ff). Aristotle did not accept the atomic theory of his day and did not define substances as things that are sempiternal; but there is a connection of thought between the atomist's belief in simple bodies and his doctrine of ultimate subjects. In a surviving fragment of his work *On Democritus*, he himself speaks of atoms as little substances, and it is not surprising that some philosophers of the 17th century came to think of atoms as the only genuine substances.

The new orthodoxy established at the end of the 17th century by Newton and Locke allowed for two kinds of basic things, atoms and souls; but there were some curious developments among philosophers who tried to take seriously the notion of a substance as something which requires nothing else for its own existence. Descartes (who gave this definition in his *Principia Philosophiae*, i. 51) agreed with the demand for mechanical explanation of natural phenomena, but rejected atomism as self-contradictory (*ibid.*, ii. 20). According to his philosophy there is only one *res extensa*, though there are many *res cogitantes*, or minds. Spinoza went farther and argued that there could be only one substance, Nature or God, which comprised in its infinity the attributes both of extension and of thought. Whereas Descartes had allowed the title of substance by courtesy to things which were independent of all else except God, Spinoza denied the existence of created substances (*Ethica*). Frightened by this pantheism, Leibniz reverted

to the view that there were many substances, but maintained that nothing extended could be a substance, since its existence presupposed the existence of its parts. According to his view, all substances must be conceived by analogy with minds, because these alone have the simplicity necessary for independence; and it must further be supposed that each has its experiences without being affected by any other, because causal interaction would be inconsistent with independence (*Discours de métaphysique; Monadologie*).

These surprising results were all derived from attempts to take the notion of independent existence in a much stricter way than Aristotle ever contemplated. So too the criticism which the doctrine of substance received a little later from Locke, Berkeley and Hume was based on a concept foreign to Aristotle. These philosophers all maintained that ideas must be derived from experience; but Locke held also that the ideas obtained in sense-perception were representative of the qualities of physical things and found himself driven to admit that he had no idea of a substance except that of a something-he-knew-not-what serving as support to qualities (*Essay Concerning Human Understanding*, I, iv, 18). It is not surprising that Berkeley declared talk of material substance to be meaningless, or that Hume went on to declare talk of a spiritual substance equally meaningless. Throughout the argument it was assumed that "substance" meant something standing under qualities. But Aristotle never taught that a substance was a featureless substratum of qualities; and apart from that, it is surely a mistake to speak of qualities as though they were drapings. (*See* QUALITIES.)

Hume, who held every impression or idea to be a distinct existence, says in one place that if anything may be called a substance which may exist by itself, then his perceptions are substances (*Treatise of Human Nature*, I, iv, 5). The remark is interesting because it had always been assumed without question that anything which satisfied Aristotle's first requirement must also persist through change. When he tried to rehabilitate the doctrine of substance after the attacks of the empiricists, Kant made this a matter of definition (*Critique of Pure Reason*, First Analogy of Experience). But it is difficult to be certain about the interpretation of Kant's doctrine; for he tried to justify it by saying that it was presupposed in our perception of change, suggesting also that it was the same as the principle of the conservation of matter. Clearly the continuants (to adopt a term from W. E. Johnson's *Logic*) which undergo perceived changes are themselves perceived, whereas the particles which scientists have supposed to be conserved through all change are not perceived.

In the 20th century Bertrand Russell ("The Philosophy of Logical Atomism," in *The Monist*, 1918–19) and A. N. Whitehead (*An Enquiry Concerning the Principles of Natural Knowledge*, 1919) made once more the distinction that Hume had indicated between Aristotle's two notions. According to their arguments, there must be simple and ultimate particulars which are the subjects of basic facts: Russell called them logical atoms. These satisfy Aristotle's first requirement for a substance, though it may be undesirable to call them by that name, since it suggests permanence, or at least persistence for a long time, whereas they are minimal events. To say that such events happen to a continuant can only mean that they form a temporal series in which each has a close resemblance to its neighbours and is causally connected with them.

The Value of the Notion.—There are undoubtedly substances such as Aristotle describes in his *Categories*. For according to his explanations a substance is just a thing, not in the sense in which things are contrasted with persons, but in the larger sense in which things are contrasted with attributes, such as manhood, with abstractions, such as grammar, and with the events that occur to them. It is not quite correct to say, as Aristotle does, that these things are the only subjects of change (for even Greek grammar has changed since his day); but they can undoubtedly change in respect of inessential characters. Are there also substances corresponding to the narrower definitions of later philosophers? In order to answer this question we must distinguish between the requirement of independence and that of permanence.

If we insist that nothing with parts shall be called a substance, we must agree with Leibniz that there can be no extended substances. This consideration is fatal to the *a priori* atomism of Democritus and also to that of Whitehead and of Russell. For in each case that which is offered as a substance is complex in the same way as entities disqualified for the title. There may, of course, be good scientific reasons for believing in physically indivisible particles or in events that are, so to say, natural units; but an *a priori* argument for atomism of either kind is fallacious in the same way as an argument from the principle of universal causation to the necessity of a first cause. It would be an equally serious mistake, however, to suppose that the extended world is composed of unextended substances. For although anything punctiform or instantaneous is not dependent for its existence on other things in the way in which a whole is dependent on its parts, it is dependent in another way, namely, in the way in which a boundary is dependent on that of which it is a boundary. There is, indeed, no absurdity in the suggestion that everything is dependent on something else in some way. Why then should we suppose that some thing or things must be independent of everything in every way?

On the other hand, the search for something permanent through all change has been of the greatest importance in the development of science. For it has led not only to the primitive atomic theory of Democritus, but also to the principle of the conservation of energy (*q.v.*). Unlike the Democritean theory, the latter principle was formulated in connection with experimental inquiries, but it is a postulate that guides research rather than a generalization that might be refuted by a counterexample. Nature is intelligible only insofar as it is conservative, but it is not necessary for its intelligibility that what it conserves should be stuff.

On Aristotle's doctrine *see* W. D. Ross, *Aristotle*, 5th ed. (1949). Some aspects of the later history are treated in E. Cassirer, *Substance and Function*, Eng. trans. (1923). (W. C. K.)

SUBWAY (UNDERGROUND RAILWAY), an underground electric railway system used to transport large numbers of passengers. Subways are usually built under city streets, although they may cross under rivers and may emerge from underground in outlying sections of cities and become elevated railways. Trains made up of several cars are usually operated in a subway as distinguished from single-car operation typical of surface systems of transportation (*see* ELECTRIC TRACTION).

History.—The world's first subway system was proposed for London by Charles Pearson, the city solicitor, shortly after the opening of the Thames river pedestrian tunnel in 1843. This idea was advanced to cope with the great traffic congestion hampering life in London. After years of discussion, Parliament in 1853 gave approval for construction of a 3¾-mi. underground roadway, called the Metropolitan railway, to run between Farringdon street and Bishop's road. Work was started in 1860 by digging an open ditch, constructing a single-arched structure of brick 17 ft. high and 27½ ft. wide in the trench and covering it with dirt, a technique now called "cut-and-cover." The line was opened Jan. 9, 1863, using coal-burning steam locomotives for motive power, and, in spite of the sulfurous fumes, it was a success from its inception, carrying 9,500,000 passengers in the first year of its existence.

In 1886 the City and South London Railway Co. started excavations for a tube railway, using a shield developed by James Henry Greathead (*see* TUNNEL). This shield was a great improvement over the Sir Marc Isambard Brunel's sectional rectangular shield, developed for the Thames river tunnel, and the circular shield invented by Greathead and Peter W. Barlow for the Tower tunnel, a 7-ft.-diameter tube used for a pedestrian crossing under the Thames. The 12-ft.-diameter Greathead shield, prototype of modern tunneling shields, used hydraulic jacks with a thrust of 200 tons to force it through the ground. The shield provided protection for the workmen excavating for and erecting the permanent cast-iron lining of the tunnel. Compressed air was used to tunnel through water-bearing sand but was not required for clay. The twin tunnels were driven deep in the excellent London clay and consequently neither underpinning nor support of buildings was required, nor was there any disruption of

traffic and utilities. The original plan called for the use of steel cables to haul the trains, but before the tunnel was completed electricity was substituted; in 1890 this 53-mi. section of the London underground was opened, becoming the first underground electric railway. A fare of twopence was charged for the ride. Charles Tyson Yerkes, an American, went to London in 1900 and was responsible for the final electrification of the London underground, the steam era ending in 1905. Yerkes also was instrumental in consolidating the various underground railways into one system called the London Transport. Brunel's Thames river tunnel, the first underwater tunnel, was incorporated into this system.

The London Transport System.—London Transport trains in the 1960s served 277 stations and covered 253 route miles, of which 180 mi. were underground; 90 route miles were in tube tunnels; 67 mi. in deep-level tubes; and 23 mi. in subsurface (or cut-and-cover) double-track tunnels. The single-track tube tunnels were lined with 12-ft.-diameter cast iron, except for a short section of the Central line where reinforced concrete segments were used. The track gauge was standard (4 ft., 8½ in.); 600-v. D.C. power for the multiple-unit trains was supplied mainly from London Transport's own generating stations. Automatic block signaling, with two-aspect coloured lights and automatic train stops, permitted running a maximum of 40 trains an hour each way; the average speed, including stops, was 20½ m.p.h. Speed-control signaling was employed on inner area tracks bearing heavy traffic.

Rolling stock on the London Transport was of two sizes, the subsurface cars of almost main-line size and tube cars small enough to fit 12-ft.-diameter tunnels. Where tube and subsurface lines connected, safety devices prevented the larger rolling stock from entering the tube tunnels. Both types of stock had 40 seats and could accommodate about 200 passengers. Trains were made up of five to eight cars and shortened to three to six in slack periods. The system in the 1960s carried about 670,000,000 passengers annually, the average journey per passenger being about five miles.

Because of the depth of the tunnels, hydraulic lifts for passengers were first installed at stations, but in 1911 the first escalator supplanted them. In the 1960s more than 180 escalators were in use in the London system with one running as deep as 80 feet in Leicester square; there were approximately 100 electric elevators in service also. The deep London stations were of great value during World War II as bomb shelters for thousands of people.

Other European Systems.—Many years passed before any other city started a subway system. In 1891 the city of Glasgow, stimulated by the success of the London underground, began a 6½-mi. double-track subway loop; however, a steam suburban line, put in operation in 1886, already had two miles of underground tunnel in the heart of the city.

The Glasgow tube was built by shield and open-cut methods. Considerable difficulty was encountered in driving the shields under the river Clyde. Work was abandoned after a disastrous series of blowouts under the Clyde in 1894 but George Talbot, a contractor, resumed work on the 11-ft.-diameter tunnel and was able to complete the 410-ft. section tunnel under the river in three and one-half months; his predecessors had taken five months to build 85 ft. The 4-ft. gauge Glasgow trains were pulled through the tunnels by cables and it was not until 1935 that 570-v. D.C. propulsion (three-rail) was instituted.

A 2½-mi. subway section in Budapest, opened in 1896, used electric trolleys and was the continent's first system. The Budapest subway was constructed by the cut-and-cover method and is of historical importance as the first to be built with a flat roof, with steel beams spanning the masonry walls. The flat roof has the advantage of requiring a shallower trench than that needed for an arched roof.

Paris started its system, called "the Metro" (Métropolitain), in 1898 and in two years completed an 8¾-mi. section. Rapid progress was attributed to the wide streets of Paris and the modified cut-and-cover method devised by chief engineer F. Bienvenue and contractor L. Chagnaud. In this system, the street is excavated to the level planned for the underside of the tunnel roof. The roof is constructed, the street is replaced on top of it, and the subway then is constructed under the protection of the roof. Disturbance to street traffic thus is greatly reduced. This method is still being used in Paris, although other cities have never copied the scheme to any great extent.

The Metro in the early 1960s consisted of 15 lines and 274 stations designed in a comprehensive layout with many transfer points to give excellent coverage to the large city. It was the third largest system in the world, with 120 route miles, and was second only to New York's in the number of passengers carried, over 1,100,000,000 per year.

The second largest system in Europe, constructed by the cut-and-cover method, was the Berlin subway, called the U-Bahn. The initial section, running from Potsdamer platz to the zoological gardens, was opened in 1902 and by the 1960s comprised 58 route miles with 94 stations. It consisted of five radial lines extending from the center of the city and carried about 128,000,000 passengers per year. Another large German system was the Hamburg subway, third largest in Europe, with 41½ route miles of line and 62 stations; it carried 156,000,000 passengers per year.

The last large system to be constructed in Europe was the Moscow subway, distinguished by its lavishly decorated, spacious stations; no two stations were alike and they were finished in polished marble and granite with mosaic decoration and statuary. A 7-mile subway was started in 1931 under the direction of Lazar Kaganovich, but during the first two and a half years little progress was made, only 10% of the line being completed. Acceleration of the work was ordered in 1934 and a labour force of 75,000 men and women was used, supplemented by volunteers on Sundays. In May 1934 the subway was opened for regular operation, remarkable progress in view of the difficult ground conditions in Moscow. Various methods were used—50% was constructed using a tunnel without a shield, 5% by shield, 20% by the Parisian trench method and 25% by open cut. Freezing was used to consolidate the soil before excavating for shafts and escalator inclines.

The Moscow subway was the fourth largest in Europe in the 1960s, having four lines, totaling about 37 miles in length, and 62 stations, and using a 5-ft. track gauge similar to that of the Russian railways. It carried 800,000,000 passengers per year, a very heavy concentration of traffic for its length as compared with other lines.

Underground railways existed in 17 continental cities of Europe by 1960. In order of length of subway these were Paris, Berlin, Hamburg, Moscow, Stockholm, Vienna, Madrid, Athens, Leningrad, Rome, Barcelona, Naples, Budapest, Oslo, Brussels, Warsaw and Copenhagen. Some were extensive self-contained urban and suburban networks, such as that in Paris, while others were short sections of main lines carried in tunnels under city centres, as in Copenhagen and Warsaw. Brussels had a six-track subsurface tunnel, 112 ft. wide, with three intermediate stations connecting the north and south city railway terminals. Leningrad's 6¼-mi. subway was opened in November 1955, and in 1957 Milan started work on a subway system.

In very few instances was an entire system in tunnel; as in London, the tracks rose to the surface in outer districts. In the case of Hamburg, part of the system was underground and part elevated. Most lines were of standard gauge, except that in Moscow and Barcelona's Metropolitano Transversal, with a 5-ft., 6-in. gauge. Nearly all had automatic block signaling with colour-light signals. Stockholm, however, utilized cab-signaling—the first in Europe—with line-side signals only at junctions. Oslo's two-car trains operated "at sight" on a tramway basis, without signals.

Subways in the U.S. and Other Countries.—The first subway line was constructed in the United States in Boston from 1895 to 1898. It was 1½ mi. long, built by the cut-and-cover method, and was used to take streetcars off the crowded, narrow Boston streets. Later regular subway lines were built from Boston to Cambridge, Mass.

New York city tried to solve its transportation problem by constructing elevated railroads—a move that was later much regretted—and these elevated railroads have now been demolished. It was not until 1900 that the New York subway program was

started, but work progressed thereafter at a rapid rate. The first subway to be built in New York was a 300-ft. experimental line built in 1870 on lower Broadway by Alfred E. Beach, then editor of the *Scientific American*. The cars were propelled in the tube by compressed air, using the cars as a plunger and the tube as a cylinder. The system was financially unsuccessful, and it operated for only two years; it finally was destroyed by subsequent subway construction.

The first New York subway was built from city hall to Dyckman street under the direction of the Rapid Transit commission, with William Barclay Parsons as chief engineer and John B. MacDonald as contractor. The first section was opened from Brooklyn bridge to 145th street and Broadway on October 27, 1904. Other lines were opened in rapid succession and were operated by the Interborough Rapid Transit company (IRT) and the Brooklyn-Manhattan Transit company (BMT). In 1925 the Independent System was opened; owned and operated by the city, it consisted of the Eighth avenue subway. These lines later were consolidated into the New York Transit authority, controlled by the state of New York with capital funds supplied by the city. The authority runs the largest system in the world, consisting of 240 mi. of subway and costing over $1,700,000,000 to construct. By 1960 it was collecting over 3,700,000 fares a day.

Other subways in the United States are in Philadelphia, which opened its first section in 1928, and in Chicago, opened in 1943. In Canada, Toronto opened a subway in 1954. In South America, the Buenos Aires system was opened in 1928. Japan has two subways—one in Tokyo opened in 1927 and the other in Osaka opened in 1933.

After World War II the pace of subway construction was much slower, with only Leningrad, Toronto and Milan starting systems. New York did not expand its subway lines, but modernized its existing system and rolling stock. The slackening in new construction was undoubtedly caused by the great expense of subway construction (from $6,000,000 to $8,000,000 per mile in New York city) and the greater use and reliance on automobiles and superhighways.

Construction Problems in New York.—The varying ground conditions, river crossings, tall buildings and narrow streets in New York presented construction problems never encountered in other cities. Most of the New York subways were constructed by cut-and-cover methods, with shield tunnels used for the river crossings and rock tunnels under hills. Some elevated lines were used in outlying areas at the early stages of development but construction of these was discontinued in later years because of their blighting effect on adjacent property and the hazards their supports presented to street traffic.

Subway construction in business districts or crowded residential areas requires temporary decking over the trenches to maintain traffic. This was not done in the early days, and great disruption to the life of the city and bankruptcy of many commercial establishments resulted. Utilities, including sewers, water mains, telephone and electric lines, steam mains and mail tubes, are kept in service under the decking. Gas mains are by-passed on top of the decking to prevent the leakage of gas into the excavation, a cause of dangerous explosions that were common in the early days of subway construction. The sides of the earth trenches are supported by driving steel beams, called soldier beams, vertically along the sides of the cut. The beams are driven 7½ to 10 feet apart and horizontal wood lagging is placed between the flanges of the beams to support the earth. The steel beams supporting the timber decking rest on these soldier beams and also serve as bracing against lateral earth pressure.

One of the great problems of subway construction in New York city was the support or underpinning of large buildings. This hazardous operation did not exist in London because the tunneling was done in the excellent London clay far below the foundations of buildings. In Paris the wide boulevards enabled the engineers to build subways far enough from buildings to prevent damage from subway construction. In 1916 an unsuccessful court action was introduced to prevent the construction of the subway on William street, a vital link in the New York system, because it was believed that the subway could not be constructed without seriously damaging the large buildings on this narrow street. It was on this section of subway that the pretest pile for underpinning in wet ground was developed by Lazarus White and Edmund A. Prentis. The pretest pile is installed by jacking steel cylinders under a building to support the building at a level below the subgrade of the new subway. The piles are preloaded to carry the weight of the building and wedged up to prevent the elastic rebound of the earth.

Thousands of buildings have been underpinned successfully in New York and in many cases subways have been constructed without difficulty underneath heavy structures, with the building finally resting on the roof of the subway.

Traffic Control.—The signal system is a vital part of subway operation. Its purpose is twofold: to insure safety with speed, since signals prevent conflicting movements of trains; and to secure proper spacing of trains in order to preserve schedules and insure maximum track usage.

Train movements are controlled by coloured light indications placed either at the roadside or in the motorman's cab. These lights, in turn, are controlled automatically through the track circuits. In the subway block-signal system the track length is divided into blocks so that a train in any block automatically sets the light signals which regulate the movement of following trains. At switches, train movements are controlled by an interlocking device which can be set up by the switchman to direct the train on the desired route.

Special features of modern signal systems include a tripping device which sets the train brakes automatically if a red stop signal is passed and a centralized traffic control, called CTC. In the CTC system, each switch and signal is controlled by a push-button switch on a miniature model of the track system. This model shows the operator, by means of lights, the position of each train in the controlled area. (E. E. WH.)

SUCCESSION: *see* INTESTATE SUCCESSION.

SUCCESSION, STATE: *see* SUCCESSOR STATE.

SUCCESSOR STATE, a term applied to a state acquiring all or part of the territory of another state. "Total succession" occurs when a sovereign state ceases to exist because all its territory is annexed by another state (*e.g.*, the U.S. annexation of Texas in 1845 and Hawaii in 1898), or by several states (*e.g.*, partition of Poland among Austria, Prussia and Russia in 1795). The total successor succeeds to all vested rights of the defunct state and all its debts, appraised at market rather than face value, except those which arose from the war ending its existence, but not to its obligations under executory treaties on unliquidated tort claims. If there are several successors, the general debt is usually distributed among them in proportion to the tax-producing capacity of the territory each acquires.

"Partial succession" occurs when a state or several states acquire a portion of the territory of another state. The United States and Cuba, for example, were partial successors to Spain in 1898, when Cuba won its independence and the U.S. acquired Puerto Rico, Guam and the Philippines. In this case the successor state or states succeed only to the vested rights of the ceding state within the acquired territory and only to those debts especially associated with, and beneficial to, that territory. The ceding state usually has been considered bound only by its general debt. This principle, however, was not observed in the treaties ending World War I.

The new Germany, Austria, Hungary and Turkey were respectively deemed to be the same states as the old German empire, the old Austrian empire, the old Hungarian kingdom and the old Ottoman empire, but, because the substantial losses of territory by these states reduced their capacity to pay their national debts, portions of the debts were distributed to the successor states. This principle did not, however, apply to the French acquisition of Alsace-Lorraine because Germany had not accepted any portion of the French debt when it annexed that territory in 1871. Nor did it apply to the colonial mandates because, in principle, the mandatories did not acquire sovereignty or gain any material benefits. In the treaties after World War II, however, the older

principle was reverted to. The successors to the substantial territories renounced by Italy, Japan and other states of the Axis alliance were not burdened with any portion of the general debt of those countries, though they accepted debts incurred for the benefit of an acquired territory.

A successor state should not be confused with a successor government. The status of the latter is a problem primarily of constitutional law though it may be influenced by the policies of other states concerning recognition. The rights and obligations of the state under international law are not, in principle, affected by changes in its government, whether effected by constitutional or revolutionary means, so long as its territory remains the same (Tinoco Arbitration, *Costa Rica* v. *Great Britain*, 18 *Am. Journ. Int. Law*, 1924, p. 147).

Apart from succession to all or a portion of the vested rights and public debts, a successor state is generally acknowledged to have obligations to the rights of individuals or corporations within the annexed territory but the extent of these obligations has been controversial. Courts have said that after annexation the people change their allegiance but their relations to each other and their rights of property remain undisturbed (*U.S.* v. *Percheman*, 7 Pet. 51) and that property includes rights which lie in contract (*Soulard* v. *U.S.*, 4 Pet. 54). The same has generally been held true of concessions of a mixed public and private character in that their "nationalization" involves a duty of fair compensation. On the annexation of Hanover, Prussia proclaimed: "We will protect everyone in the possession and enjoyment of his duly acquired rights."

The problem arises because of the difficulty of determining what are vested or acquired rights. A concession may properly be canceled when "(a) the grant of the concession was not within the legal powers of the late government, or (b) was in breach of a treaty with the annexing state, or (c) when the person seeking to maintain the concession acquired it unlawfully or by fraud, or (d) has failed to fulfil its essential conditions without lawful excuse" (Report Transvaal Concession Commission, U.K., Parl. Pap., South Africa, 1901, Cd. 623). Furthermore, a concession, not protected by treaty, may be canceled on payment of fair compensation if its maintenance would be injurious to the public interest.

Clearly the successor is not bound to respect titles to public office or other claims under the public law of the ceding state. In general, while the private-law system continues, except insofar as modified by explicit legislation of the successor, and rights acquired under it must be respected, the public-law system ends with the transfer, and claims under it need not be respected. Furthermore, while international tribunals apply international law and will protect vested private rights against the successor (German Settlers in Poland, Per. Ct. Int. Jus. Ser. B, no. 6, p. 6), national courts, bound by municipal law, cannot do so if the successor government has rejected them (*Cook* v. *Sprigg*, 1899, A. C. 572).

Bibliography.—Max Huber, *Die Statensuccession* (1898); A. B. Keith, *Theory of State Succession* (1907); E. H. Feilchenfeld, *Public Debts and State Succession* (1931); Wesley L. Gould, *An Introduction to International Law*, pp. 411–437 (1957). (Q. W.)

SUCCINIC ACID is an organic compound occurring in amber (3% to 8%), from which fossilized gum it is obtained by distillation. It is also found in other resins, in lignites, in fossilized woods and in many plants.

It is present at low concentration in a great variety of animal tissues, where it has an important function in intermediary metabolism (*see* Biochemistry).

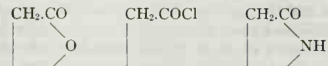
Chemically, succinic acid, or ethylenesuccinic acid, $CO_2H.CH_2.CH_2.CO_2H$, is a saturated dibasic acid which crystallizes in colourless prisms or plates melting at 186° C. and boiling at 235° C. Its vapour readily loses water to form succinic anhydride (formula I), which crystallizes in plates melting at 120° C.

The acid itself is soluble in water, and its salts with the alkali and alkaline-earth metals are also soluble in water. Barium succinate is precipitated from aqueous solution by alcohol; ferric succinate is insoluble in water and is sometimes employed in the analytical separation of iron from other metals.

When heated with phosphorus trisulfide, sodium succinate yields thiophene.

Succinic acid is produced during the bacterial fermentation of ammonium tartrate or calcium malate. It may be prepared synthetically from ethylene, C_2H_4, through ethylene dichloride, $C_2H_4Cl_2$, ethylene dicyanide or succinonitrile, $C_2H_4(CN)_2$. The last compound yields the acid on hydrolysis.

Maleic acid, now produced industrially by the catalytic oxidation of benzene, yields succinic acid on being reduced with sodium amalgam.



I II III

Succinyl chloride (formula II) is obtained by the interaction of phosphorus pentachloride and succinic acid. Succinic anhydride (formula I), also produced by heating the acid or its salts with acetic anhydride or by heating succinyl chloride with anhydrous oxalic acid, is employed as an intermediate in dye making. It is employed also in the manufacture of alkyd resins (*see* Resins) and is an important reagent in organic chemical synthesis.

Succinimide (formula III), produced by heating succinic anhydride in ammonia, crystallizes in colourless octahedra melting at 125°–126° C. and readily soluble in water. This imide when distilled with zinc dust furnishes pyrrole (*q.v.*).

Succinonitrile, $CN.CH_2.CH_2.CN$, obtained as above and also by electrolysis of potassium cyanoacetate, is a colourless solid melting at 54°–55° C. On reduction with sodium and alcohol it is converted into putrescine (tetramethylenediamine), $NH_2.[CH_2]_4.NH_2$, and into pyrollidine (tetrahydropyrrole; *see* Pyrrole).

Methylsuccinic acid or pyrotartaric acid,

$$CO_2H.CH_2.CH(CH_3).CO_2H$$

formed during dry distillation of tartaric acid, by heating pyruvic acid with concentrated hydrochloric acid, or by the reduction with sodium amalgam of citraconic and mesaconic acids, is obtained in colourless prismatic crystals soluble in water and melting at 112° C. It forms an anhydride and its sodium salt heated with phosphorus trisulfide yields β-methylthiophen. Isosuccinic acid (ethylidenesuccinic acid), $CH_3.CH(CO_2H)_2$, an isomeride of ordinary succinic acid, is obtained on hydrolyzing α-cyanopropionic acid, $CH_3.CH(CN).CO_2H$, or by the action of methyl iodide on ethyl sodiomalonate, $CHNa(CO_2.C_2H_5)_2$. It crystallizes in prisms melting at 120° C. and is soluble in water. Unlike ordinary succinic acid, it does not yield an anhydride on heating, but loses carbon dioxide and passes into propionic acid.

See T. E. Thorpe, *Dictionary of Applied Chemistry*, vol. xi, 4th rev. ed. (1954). (G. T. M.; X.)

SUCCULENT. A succulent plant, strictly speaking, is one that is full of sap, but in horticultural usage it is any thick, fleshy and juicy plant. Custom determines what plants are included, although for the most part they are native in regions that are arid or semiarid for at least a part of the year, and succulence is related to the storage of water in them. Many genera in a number of families are involved and the individual plants may assume odd, even grotesque, forms, as in the cacti (*see* Cactus). Most of them need a full sun exposure and a soil which is gritty and porous for good drainage but which will retain sufficient water for the needs of the plants. A few of the genera containing well-known succulents are *Agave*, *Aloe*, *Echeveria*, *Kalanchoe*, *Sedum* and *Sempervivum*. (J. M. Bl.)

SUCEAVA (Suczawa), a town of Rumania, in Suceava region on the Suceava river. Pop. (1956) 20,828. It was from 1401 the seat of the metropolitan of Moldavia and until 1565 the capital; its 14th-century church contains the grave of the patron saint of the Bukovina. It was many times besieged by Poles, Hungarians, Tatars and Turks and also suffered in World War I. Nearby is the early 17th-century monastery of Dragomirna in the Byzantine style.

SUCHET, LOUIS GABRIEL, Duc D'albufera Da Va-

LENCIA (1770–1826), marshal of France, one of the most brilliant of Napoleon's generals, was born at Lyons on March 2, 1770. As *chef de bataillon* he was present at the siege of Toulon in 1793, where he took General O'Hara prisoner. After serving under Joubert in Tirol in 1797 and also in Switzerland under Brune in 1797–98, he was made chief of staff to Brune. In July 1799 he was made general of division and chief of staff to Joubert in Italy, and was in 1800 named by Masséna his second in command. His action contributed to the success of Napoleon's crossing the Alps, which culminated in the battle of Marengo on June 14. In the campaigns of 1805 and 1806 he served at Auzterlitz, Saalfeld, Jena, Pultusk and Ostrolenka. He obtained the title of count on March 19, 1808 and soon afterward was ordered to Spain. He annihilated the army of Blake at Maria on June 14, 1809, and on April 22, 1810, defeated O'Donnell at Lerida. Suchet in 1812 conquered Valencia and received the title of duc d'Albufera da Valencia (1812). When the tide set against the French, Suchet defended his conquests step by step till compelled to retire into France. By Louis XVIII he was made a peer of France, but, having rallied to Napoleon during the Hundred Days, he was deprived of his peerage in 1815. He died near Marseilles on Jan. 3, 1826.

SUCHIATE, a river of Pacific coast Central America with a watershed of 712 sq.mi. The river drains Pacific cordillera and coastal plain areas in western Guatemala (San Marcos department) and southern Mexico (Chiapas state). The lower half of its 93-mi. course forms the international boundary between Guatemala and Mexico. (J. T.)

SUCHITEPÉQUEZ, a department in southern Guatemala, extending from the volcanic highlands to the Pacific ocean. Area 969 sq.mi.; pop. (1960 est.) 161,029. The departmental capital is Mazatenango, an agricultural centre. On the south-facing slopes of the highlands there are coffee plantations, and maize, beans and tobacco are also grown. On the lowlands the chief products include bananas, sugar cane, cacao, rice, and such food crops as maize and beans. The department is served by both railroad and all-weather highway. (P. E. J.)

SÜCHOW (Hsü-chou, formerly T'UNG-SHAN), city in northwest Kiangsu province, China. Its site marks the approximate geographic border between the north China plain and the Yangtze delta plain, the meeting place of wheat and rice as staple crops, respectively. Twelve miles north is Wei-shan Hu, one of a series of long, narrow waterways into which the Grand canal, 45 mi. E. of Süchow, leads. Although walled, the city has had its historic struggles with Yellow river floods. Süchow is an important railway crossing where the north-south Tientsin—P'u-k'ou line meets the Lan-chou to Hai-chou (formerly the Lunghai) railway. This made it a strategic target during the Japanese attack on China, and at T'ai-erh-chuang, 30 mi. N.E., the Chinese armies early in 1938 inflicted their first major defeat on the Japanese in a classic battle of manoeuvre that delayed the invaders' advance for two months and greatly raised the morale of the Chinese nation. A decisive battle in the Chinese civil war was also fought near Süchow from Nov. 1948 to Jan. 1949. Current industrial development is mainly in textiles and flour milling. Pop. (1957 est.) 676,000. (TE. H.)

SUCKLING, SIR JOHN (1609–1642), English poet, dramatist and courtier, best known for his lyrics, was born at Whitton, Middlesex, and baptized there on Feb. 10, 1609. He was the son of Sir John Suckling, secretary of state to James I. He was educated at Trinity college, Cambridge, to which he was admitted in 1623, and he inherited his father's considerable estates at the age of 18. He entered Gray's Inn in 1627 and left England to travel in France and Italy in 1628. Knighted on his return in the autumn of 1630, he joined the English volunteer force raised to serve under Gustavus Adolphus and appears to have been present at the battle of Breitenfeld in 1631. He returned to London in May 1632, and became a prominent figure at court with a reputation for being "the greatest gallant of his time, and the greatest gamester both for bowling and cards." He was a gentleman of the privy chamber to Charles I and the friend of the poets Thomas Carew, Richard Lovelace and William Davenant, as well as, rather surprisingly, the famous latitudinarian divine, John Hales. When the war with the Scots broke out in 1639, Suckling raised a troop of horse at his own expense and accompanied Charles I on his ill-fated expedition. The magnificent costumes of Suckling's gaudy warriors and their poor performance in the field were the subjects of much contemporary ridicule.

In May 1641 Suckling took an active part in the plot to rescue the earl of Strafford from the Tower. When the plot was discovered, Suckling fled to France and there is every reason to believe John Aubrey's statement that he committed suicide in Paris in 1642. According to Aubrey he invented the game of cribbage.

Suckling was the author of four plays, the most ambitious of which is the tragedy *Aglaura*, magnificently staged in 1637 and handsomely printed at the author's expense (1638), and the best, the lively comedy *The Goblins* (1638). They all contain echoes of Shakespeare and Beaumont and Fletcher. Richard Flecknoe's criticism of *Aglaura* that it seemed "full of flowers, but rather stuck in than growing there" applies to all Suckling's dramatic efforts.

His reputation as a poet rests on his lyrics, the best of which justifies the description of him by Millamant in Congreve's *The Way of the World* as "natural, easy Suckling." He inherited from Donne the tradition of the "anti-platonic" deflation of high-flown love sentiment and uses it with a charming, gay insouciance in such poems as that which begins with the following lines:

> Out upon it I have loved
> Three whole days together;
> And am like to love three more,
> If it prove fair weather.

"A Session of the Poets" (1637; publ. 1646) is an amusing skit for which he probably took a hint from an Italian work by Traiano Boccalini; it is the prototype of a long line of similar works in the 17th and 18th centuries. He did his best work in the manner of popular poetry. His masterpiece is undoubtedly "A Ballad Upon a Wedding," in the style and metre of the contemporary street ballad. The description of the bride has an unsurpassed delicacy and grace:

> Her feet beneath her petticoat,
> Like little mice, stole in and out,
> As if they feared the light,
> But O, she dances such a way
> No sun upon an Easter-day,
> Is half so fine a sight.

Suckling's extant letters are in a lively, colloquial prose that anticipates that of the Restoration wits.

BIBLIOGRAPHY.—The first published collection of Suckling's works was *Fragmenta Aurea* (1646; enlarged 1658). The so-called *Selections* published by Alfred Inigo Suckling in 1836 is in reality a complete edition of the works, of which W. C. Hazlitt's edition (1874, rev. ed. 1892) is little more than a reprint. A. I. Suckling's edition includes a valuable memoir and a portrait by Van Dyck. The best modern edition is *The Works of Sir John Suckling*, ed. by A. Hamilton Thompson (1910). For anecdotes of Suckling's life *see* Aubrey's *Brief Lives*, ed. by A. Clark, vol. ii., pp. 240–245 (1898) and for "A Session of the Poets" and its successors, *A Journal from Parnassus*, ed. by H. Macdonald (1937). There is important biographical information in "Sir John Suckling and the Cranfields" (*Times Literary Supplement*, Jan. 29, 1960) by Thomas Clayton. (V. DE S. P.)

SUCRE, ANTONIO JOSÉ DE (1795–1830), was one of the most talented, successful and respected leaders of the Latin American war for independence from Spain. Born Feb. 3, 1795, at Cumaná, in eastern Venezuela, Sucre at the age of 15 was swept into the struggle against the Spaniards and loyalists and fought in many campaigns in Venezuela and Colombia. His skill and admirable character early attracted the attention of the revolutionary leader Simón Bolívar, and at the age of 26 with the rank of general, Sucre was assigned by Bolívar the mission of freeing southern Gran Colombia (Ecuador) from Spanish control. This he did by his victory at the battle of Pichincha, near Quito, in 1822.

Moving south to Peru, the central and southern portions of which formed the last stronghold of Spanish and loyalist forces in South America, Bolívar and Sucre organized an army which won the battles of Junín (Aug. 1824) and Ayacucho (Dec. 9, 1824). For his victory as field commander in the decisive engagement at

Ayacucho, Sucre was given the rank of grand marshal of Ayacucho.

Early in 1825 Sucre was ordered by Bolívar to the region of Upper Peru (Bolivia) to occupy and organize the area. Bolívar joined him and the two leaders laboured to create a political structure, which took the form of a complicated constitution written by Bolívar. Under this constitution Sucre in 1826 was elected president of the new nation of Bolivia. Eager to retire from public responsibilities, he agreed to serve only two years of the lifetime tenure which he was offered. During this period Sucre attempted to provide Bolivia with institutions of government, including courts, a revenue system and schools. He sought to avoid excessive use of his great powers, but he became the target of mounting opposition born of political factionalism and ambition and complicated by Peruvian designs to absorb Bolivia. A mutiny at Chuquisaca in 1828, in which Sucre was wounded, and an invasion by Peruvian troops caused him to resign the presidency and retire to his chosen home in Ecuador. There he was called upon immediately to defend Gran Colombia against an invading force of Peruvians, whom he defeated in the battle of Portete de Tarqui early in 1829. The next year Bolívar again called Sucre from retirement, this time to preside over the "Admirable Congress" in Bogotá, a last effort to maintain the unity of Ecuador, Colombia and Venezuela within Bolívar's decade-old Gran Colombian federation. When this congress failed, Sucre started on the long ride home to Quito. At Berruecos, near Pasto in southern Colombia, he was assassinated in June 1830, probably by agents of José María Obando, an ambitious Colombian military chieftain who wished to be rid of the favourite lieutenant and possible successor of Bolívar.

Loyal, brave and modest, Sucre had been like a son or a younger brother to Bolívar, who was 12 years his senior. Among contemporaries the grand marshal of Ayacucho gained the accolade, the "immaculate Sucre," for his high moral qualities.

(T. F. McG.)

SUCRE is a northeastern coastal state of Venezuela, separated from Trinidad by the Gulf of Paria. Area 4,556 sq.mi., pop. (1961) 401,992. It is traversed by the northeastern highlands whose coastal range forms the two peninsulas of Paria and Araya on the east and west respectively. The eastern portion of the state receives abundant rains but the western portion, which supports most of the population, is relatively dry. Cumaná (*q.v.*), the state capital, is a commercial and fishing centre. Mineral resources include asphalt from Lake Guanoco, saltworks on the Araya peninsula and gypsum near Macuro. Predominantly agricultural, Sucre produces bananas, cacao, corn, cotton, rice, sugar cane, tobacco, yuca and coconuts along the coast. The fishing industry is very important. (L. We.)

SUCRE, legal capital of Bolivia, and seat of its supreme court. Pop. (1961 est.) 62,339. The city (elevation 8,532 ft.) is located in a fertile valley crossed by the Cachimayo river which flows into the wide Pilcomayo, and flanked by the majestic Churukella and Sicasica hills. Its climate is mild with an average temperature of 16° C. (61° F.). A variety of fruits and vegetables are grown in the valley.

Known as the City of Four Names because at various times it was called La Plata, Chuquisaca, Charcas and finally Sucre. It was founded by Pedro de Anzúrez in 1538 on the site of a Charcas Indian village. Numerous churches were built including the magnificent 17th-century cathedral the Basilica Metropolitana. In 1561 it became the capital of the Charcas *audiencia* (Upper Peru) and in 1609 the seat of an archdiocese. Its University of St. Francis Xavier, one of the oldest in South America, was founded in 1624. Sucre was an early revolutionary centre; its citizens revolted against Spain on May 25, 1809. In 1839 it became the official capital of Bolivia and in 1840 was renamed Sucre in honour of Marshal Antonio José de Sucre, the first president of the Republic of Bolivia. In 1898 Sucre resisted efforts to move the capital permanently to La Paz and a civil war ensued. The outcome was a compromise; Sucre remained the capital in name and law but the executive and legislature moved to La Paz. (*See* Bolivia: *History.*)

A railroad to Potosí, a road to Cochabamba, a pipeline to the petroleum fields of Camiri and secondary roads to nearby valleys made Sucre a commercial and agricultural centre of growing importance. Industrial establishments include an oil refinery and cement plant. (C. W. A.)

SUCROSE: *see* Sugar.

SUDA LEXICON, Greek lexicon of *c.* A.D. 1000. The title "Suda" is found in the manuscripts and in the earliest reference to it (in the commentary on Aristotle's *Rhetoric* by the 12th-century author Stephanus); the name "Suidas," by which the compiler is frequently known, is due to Eustathius (12th century), who arbitrarily hellenized the apparently barbarian title into a Greek (actually Thessalian) personal name.

The work is rather a combination of lexicon and encyclopaedia than a lexicon in the modern sense. It was compiled from a multitude of sources, lexicographical (principally the *Synagoge* also used by Photius, and the so-called *Lexicon Ambrosianum*), literary, biographical (principally an epitome of Hesychius of Miletus, *q.v.*) philosophical, historical and religious. These sources were often inferior, and the compilation was prepared without adequate care or critical faculty; nevertheless, the lexicon is of immense value and is the most substantial of its kind extant. Especially important are the biographical entries on Greek classical and Byzantine writers, and the numerous quotations (over 5,000) from Aristophanes, for whose text the lexicon is an important authority.

See edition by A. Adler, 5 vol. (1928–38); Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft,* 2nd series, vol. 4, col. 675–717 (1931). (D. M. J.)

SUDAN (Bilad as Sudan, "country of the blacks,") as medieval Arab writers called the area south of the Sahara, stretches for more than 3,500 mi. across Africa from Cape Verde on the Atlantic to the highlands of Ethiopia and the Red sea, between about latitude 8° and 16° N. It borders the Sahara on the north and extends southward to the forests of west Africa and the Congo basin. Mean annual rainfall varies between 10 in. in the north and 60 in. in the south, with the hottest months usually from June to Sept. and a pronounced, and often very prolonged, dry season. Temperatures are generally high throughout the year. Vegetation ranges from semidesert steppe and thorn-scrub near the Sahara through vast grass plains, loosely termed savannas, with numerous varieties of acacia and mimosa, to parkland country where low trees grow among tall grasses, and savanna forest that merges ultimately into equatorial rain forest.

During the dry season the trees shed their leaves, all but the largest rivers run dry, and bush fires which burn up the grass are common. Relative humidity may fall as low as 30%. The rainfall might be adequate for cultivation were it not for the very high rate of evaporation which makes irrigation essential in many areas.

Much of the Sudan is a plateau between 1,000 and 1,500 ft. above sea level, but there are many higher areas, sometimes exceeding 10,000 ft., as in northern Ethiopia and in the west of the Republic of the Sudan (Darfur). The principal rivers include the Senegal and Niger, draining to the Atlantic, and the Nile and its tributaries which derive much of their water from areas beyond the Sudan region. Lake Chad (*q.v.*) is a centre for interior drainage.

The density of population is generally low and many of the people are nomadic or seminomadic. Whereas the Negroes of west Africa are generally cultivators, those of the Sudan are predominantly pastoral. There has been considerable admixture with Hamitic blood, the degree of Hamitic influence decreasing westward and southward toward the Gulf of Guinea. The people are fairly tall, the men averaging 5 ft. 8 in.; they have black or very dark brown skins, wooly hair, broad noses, thick lips and prognathous jaws. Many of them profess Islam, and often wear the flowing robes and fezzes characteristic of that faith.

Movement across the grasslands is generally uninterrupted, especially during the dry season, and the continuity of a similar environment along the southern boundaries of the Sahara for vast distances has encouraged people to move from the north and east from early times. At first they traveled with the aid of horses and oxen but migration was greatly expanded and probably accelerated with the introduction of the camel in about A.D. 300, especially as camel caravans were able to cross the Sahara. Thus the

Sudan was linked with the Mediterranean coastlands whose manufactured articles, together with Saharan salt, were exchanged for the gold, kola nuts and slaves of Guinea.

From Arab historians something is known of some of the powerful states that were established by military rule, the largest and most enduring of which were associated in the western Sudan with the termini of the desert trade routes. In the western Sudan, ancient Ghana was created by Jewish or Berber settlers about A.D. 300 in the area west of Tombouctou (Timbuktu), though its greatest years came when it was ruled by the Negro Sarakolé dynasty (*see* GHANA [ANCIENT]). Almoravid attacks in the 11th century reduced its power and led to its replacement by the Mali or Mandingo empire, centred on the upper Niger. Mali in turn was overthrown during the latter part of the 15th century by the expansion of the Songhai or Gao empire that developed from Berber settlements established on the lower Niger as early as the 7th century. In 1591 the Songhai towns of Gao, Tombouctou and Jenné (Djenné) were occupied by Moroccan troops anxious to control both the lucrative caravan traffic and the long-established trade in gold. The empire was replaced by numerous Negro kingdoms, including the Mossi-Dagomba states, the Bambara kingdoms of Ségou and Kaarta, Bornu and the small Hausa states that were later conquered by the Muslim Fulani early in the 19th century. European penetration was followed by the establishment of political control, principally that of the French and the British, which lasted until the emergence of independent states during the 1950s.

On the eastern side of the continent Egyptian links with the Sudan region were generally strong, notably with Nubia. . After the Nubian empire had been overrun by Muslims it was replaced by kingdoms such as those of Dongola, Darfur and Sennar. Later there was invasion from Egypt and, in 1899, the establishment of the Anglo-Egyptian condominium. The independent Republic of the Sudan was created in 1956.

See CENTRAL AFRICAN REPUBLIC; CHAD, REPUBLIC OF; MAURITANIA, ISLAMIC REPUBLIC OF; NIGER, REPUBLIC OF THE; NIGERIA; MALI, REPUBLIC OF; SUDAN, REPUBLIC OF THE; UPPER VOLTA, REPUBLIC OF THE; NUBIA; *see* also references under "Sudan" in the Index. (R. W. SL.)

SUDAN, REPUBLIC OF THE (JUMHURIYAT AS SUDAN), an independent country of northeast Africa, bounded by Libya on the extreme northwest, by the United Arab Republic (Egypt) on the north, by the Red sea and Ethiopia on the east, by Kenya, Uganda and the Republic of the Congo (formerly Belgian Congo) on the south and by the Central African Republic and the Republic of Chad on the west. Area 967,491 sq.mi. Pop. (1955–56) 10,262,536; (1963 est.) 12,650,000. The Sudan was from 1899 under the joint sovereignty of Great Britain and Egypt and was known as the Anglo-Egyptian Sudan until its independence in 1956. The name Sudan is derived from the Arabic expression *Bilad as Sudan* ("country of the Blacks") which originally applied to a much wider area of Africa. The capital is Khartoum. (X.)

This article contains the following sections and subsections:

I. Physical Geography
  1. Relief, Drainage and Geology
  2. Climate
  3. Vegetation
  4. Animal Life
II. Geographical Regions
III. The People
  1. Tribal Groups
  2. Languages
  3. Religion
  4. Population Distribution
IV. Archaeology
V. History
  1. Early History
  2. Egyptian Occupation, 1820–81
  3. Mahdist Theocracy, 1881–98
  4. Anglo-Egyptian Condominium, 1899–1955
  5. Republic of the Sudan
VI. Administration and Social Conditions
  1. Government
  2. Administration
  3. Living Conditions
  4. Public Health
  5. Justice
  6. Education
  7. Defense
VII. The Economy
  1. Agriculture and Pastoral Production
  2. Forestry
  3. Minerals
  4. Industry and Power
  5. Trade and Finance
  6. Transport and Communications

I. PHYSICAL GEOGRAPHY

1. Relief, Drainage and Geology.—The Sudan is mainly made up of vast plains and plateaus drained mainly by the Nile and its tributaries. A wide basin is bounded on the east by the Red Sea hills (over 7,000 ft.), to the southeast by the Ethiopian highlands and on the west by the Jabal Marra group (highest peak 10,134 ft.) which rises over a lower rounded divide between the Nile and Chad basins and, in the southwest, the Nile and Congo basins.

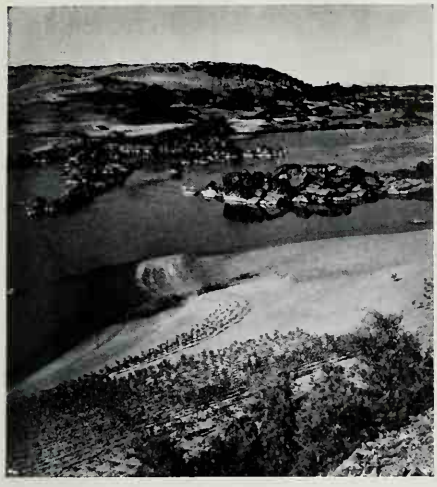
The Nile system is the dominant physical feature. The White Nile (Al Bahr al Abyad), entering the country as the Bahr el Jebel from the south through rapids at Nimule, has a very low gradient until it is joined by the Blue Nile (Al Bahr al Azraq) at Khartoum. The Blue Nile contributes most of the flood water as much of the

SUDAN SURVEYS

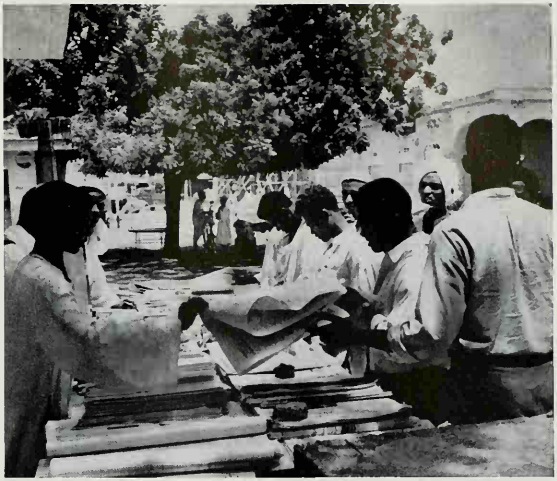
MAP OF THE REPUBLIC OF THE SUDAN SHOWING PROVINCES AND MAIN TOWNS

White Nile flow is lost in the sudd (swampy) region and about half of its discharge to the north is from the Sobat, which, like the Blue Nile, also rises in the Ethiopian highlands. Throughout much of the country drainage does not reach the Niles; the rivers of the southwest infrequently reach the Bahr el Ghazal system, and to the north most hill groups initiate seasonal watercourses which are lost in the surrounding plains. An important inland delta is that of the Gash near Kassala. (*See* NILE.)

Igneous and metamorphic rocks of the Pre-Cambrian Basement Complex series underlie much of the Sudan and appear at the surface in the Red Sea hills, the Nuba mountains (Jibal an Nubah), the southern plateaus and in scattered hills over the plains. Meso-



The second cataract of the Nile river near Wadi Halfa



A newsstand in Khartoum, capital of the Republic of the Sudan

Water-front scene at Port Sudan on the Red sea with mounted policeman in foreground

Irrigating semiarid land with the sakieh (sakia), an ancient water wheel with buckets around its periphery actuated by draft animal

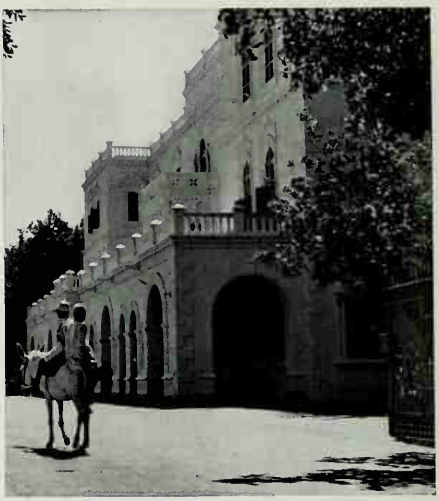
Fishermen throwing their nets into the Nile near Khartoum below the junction of the White and Blue Niles

SCENES OF SUDAN

PHOTOGRAPHS, (TOP LEFT) PAUL ALMASY, (TOP RIGHT, BOTTOM) WAAGENAAR—PIX FROM PUBLIX, (CENTRE LEFT) EWING GALLOWAY, (CENTRE RIGHT) KEATING—BLACK STAR

SUDAN

Typical landscape near Torit in Equatoria province



The palace formerly used by governor generals in Khartoum

Cotton bales awaiting export at a government-owned mill at Anazara

Hadendoa tribesmen winnowing millet, one of the staple diets of the Sudanese, in the traditional way

Campus scene at the University of Khartoum

Nile steamers at Juba. The Nile is a chief means of communication between northern and southern Sudan

SCENES OF SUDAN

PHOTOGRAPHS, (TOP LEFT) SHELL PHOTOGRAPHIC UNIT, (TOP RIGHT) TRIESCHMANN—BLACK STAR, (CENTRE RIGHT) CAMERA PRESS—PIX FROM PUBLIX, (OTHERS) WAAGENAAR—PIX

zoic Nubian sandstones unconformably overlie the older rocks over the northwestern Sudan and in the Gedaref area in the east. In Tertiary times volcanic outbursts gave rise to the Jabal Marrah group and to the many scattered volcanic hills of the northwest. Later events include the formation and alluvial infilling of the large depression (1,000 ft.) in the region of the White Nile (Umm Ruwabah series) and of a smaller depression (200 ft.) along the Blue Nile.

The northern region, between 16° N. and the frontier (22° N.) is a sand- or gravel-covered desert, diversified by the typical flat-topped *jabals* of Nubian sandstone and the islandlike steep-sided granite hills of the Basement Complex. To the south lie the great plains and plateaus of the Sudan. In the west these show evidence of former more arid climates in the relict, now stable, sand dunes and sheets of Kordofan. To the east the desert is replaced by clay plains at about the latitude of Khartoum extending southward to about 6° N. and across the Nile into Bahr el Ghazal and Equatoria provinces. The soils on the clay plains are heavy and sticky during the rains but dry and cracked during the rest of the year. In the extreme south, tropical red earths and ironstone outcrops occur, particularly to the west of the Nile. Little percolation occurs in the clay plains and underground water is not easily obtained. The Nubian sandstone is the best aquifer where rainfall is sufficient or where water percolates from the Nile. Elsewhere local wells and bores are found near the foot of the scattered hill masses. Several streams in the Red Sea hills have an appreciable flow, and Khor (Khawr) Arba'at supplies Port Sudan with surface and underground water.

2. Climate.—North of about latitude 19° N. northerly winds prevail for most of the year and rainfall is rare; to the south of this the seasons are characterized by the oscillation, north and south, of the boundary between moist southerly air and dry northerly air. Abu Hamad is the normal northern limit of the rains; a short wet season at Khartoum provides an average rainfall of 7 in.; there is 15 in. at Wad Medani (Wad Madani) and 50 in. at Gambela (a Sudanese trading station inside Ethiopia). Most of the rainfall occurs in heavy thundery showers. Except for the Red sea coast, where the most rain comes in autumn, 50–75% of the precipitation occurs in the three summer months, although at Juba in the far south there are only two really dry months.

Highest temperatures which usually precede the rains, occur in July in the extreme north, in May–June in Khartoum and in April, March or February progressively farther south. A maximum of 52.2° C. (126° F.) has been recorded at Wadi Halfa and 47.7° C. (118° F.) at Khartoum, but in the south temperatures reach only 42–43° C. (108–109° F.). Haboobs (dust storms) are an unpleasant feature in northern and northeastern Sudan.

3. Vegetation.—Five major vegetation types can be distinguished: (1) the sparse desert vegetation of the north; (2) a zone of semiarid acacia scrub and short grassland farther south; (3) the woodland savanna of most of the southern half of the country; (4) the vegetation of the flood lands and swamps of the Nile; and (5) the flora of the mountain regions.

The desert, with less than 3 in. rainfall, supports permanent vegetation only near watercourses. Date palms are frequent along the Nile, and the doum palm, tamarisk and coarse grasses are found where the banks are not cultivated. The semidesert (3–11 in. rainfall) includes the Red sea coastal plain and a zone running east-west across the country roughly between the latitudes of Atbara ('Atbarah) and Wad Medani. Vegetation there is characteristically a mixture of grasses and herbs with scattered scrub and bare areas. *Acacia tortilis* is common, though not on clay soils, and *Blepharis* species with areas of grasses is the typical climax vegetation. On sands other acacias are more frequent. The woodland savanna—a mixture of grasses, bushes and trees—has several distinct subdivisions. In the areas of the north where the rainfall is less than 30 in. trees are thorny and acacias dominant. On the eastern clays *Acacia mellifera* is found merging into *A. seyal* in the south, while the Qoz (Qawz) sands support *A. senegal* (yielding gum arabic) and, in depressions between the dunes, baobab. With increasing rainfall *Combretum cordofanum* woodland develops with grasses, and in the wetter areas the natural vegetation would be the rain forest now found only in a few remnants. *Khaya,* or African mahogany, *Isoberlinia doka* and *Anogeissus schimperi* are characteristic trees in the drier areas; *Terminalia glaucescens, Albizzia zygea, Vitex doniana* are among the larger trees found in wetter parts. Elephant grass is common. In the Nile valley north of Juba and the widening swamps (sudd) to the north, papyrus is common along the waterways but grasses and the water plant *Pistia* are also important. A new and malignant introduction, the water hyacinth, has spread rapidly since 1957, infilling many of the side channels and lagoons. In each of the mountain groups, Jabal Marrah, Red Sea hills, Imatong, Dongotona and Didinga, distinctive assemblies and vegetation zones are found.

4. Animal Life.—Southern Sudan is one of the areas in Africa where the characteristic wild life may be seen best. Most of the grassland animals (lion, leopard, cheetah, giraffe, zebra, antelope, etc.) are common; buffalo, hippopotamus, chimpanzee and baboon are also abundant among a wide range of other species. Common birds include guinea fowl, bustards, geese, cranes, storks, pelicans, weaver birds, shrikes and starlings. The rare shoebill is found in the Bahr el Ghazal region. Crocodiles are numerous on the Nile and many types of land lizard are found. Snakes are not abundant but many are poisonous. Mosquitoes are abundant on the rivers and blood-sucking horseflies are locally common. The tsetse fly occurs south of about 12° N.

National Parks.—There are 3 national parks, 14 game reserves and 3 sanctuaries in the country. The Dinder (2,470 sq.mi.) contains much game, while the Southern National park (7,700 sq.mi.) and the Nimule National park provide sanctuary for elephant, white rhinoceros, buffalo and the giant eland. (L. Be.)

II. GEOGRAPHICAL REGIONS

Twelve regions may be distinguished:

1. The ironstone plateau of the south and southwest is largely forested. Along its northern margin, west of the Bahr el Jebel (Bahr al Jabal), Nilotic tribes, chiefly the Dinka, occupy their villages during the rainy season, when they graze their cattle and grow food crops. Farther south, and nearer to the international frontiers, agricultural tribes (or parts of tribes), practise shifting cultivation. Much of the centre and west is uninhabited.

2. The plateau of Darfur is remarkable for seasonal watercourses radiating from the Jabal Marrah. These have attracted almost all the permanent settlement in the province because of the proximity of cultivable land and the availability of water in wells when it is not flowing above ground. Between the main valleys the thin infertile soils of the plateau are lightly wooded and during the rainy season thickly carpeted by grass. Nomadic tribes graze their animals widely, but remain near the main watercourses during the dry season.

3. Jabal Marrah, and associated uplands, between 3,000 and nearly 10,000 ft. above sea level, are rugged and difficult of access. They are occupied by the Fur people who live in small villages and scattered farms. The Fur divert mountain streams into small gardens, make terraced fields and keep cattle, which they pen on their arable land after harvest.

4. The great clay plain is divisible into two parts. The southern and larger part is occupied by cattle-owning Nilotic tribes, who retreat to the margins or to small areas of relatively high land during the rainy season, when almost the entire surface is submerged. In the dry season they drive their herds to moist grazing close to the permanent rivers; for lack of such rivers much of the southeast is uninhabited. The northern part receives less rain and therefore does not become a swamp in the rainy season, although at times movement is difficult. Bisected by the Blue Nile and bordered on the west by the White Nile, this northern area is mainly uninhabited in the south, but in the centre and along the two Niles a large population is supported both by cultivation depending on rain and by projects based on irrigation, growing cotton as the chief cash crop. The largest of these irrigation projects is the Gezira (*q.v.*). More developed economically than other parts of the country since 1900, its population has increased rapidly and it has received many immigrants from the north and west. Also, it has a higher proportion of urban population. Far-

ther north, where rainfall is still lower, agriculture gives way to nomadic grazing.

5. The eastern frontier, in principle, lies at the foot of the Ethiopian uplands; but locally, between the Atbara ('Atbarah) river and the Blue Nile near the Sobat river and in the far southeast, foothills or detached small uplands lie within Sudan. All these areas are remote and have small populations.

6. The dunes (Arabic, *qoz*, or *qawz*) form a zone about 750 mi. long, from the White Nile to southern Darfur, and 125–250 mi. wide. Except along the northern fringe, it receives sufficient rainfall for agriculture; but tilled lands are discontinuous, because of lack of water supplies. Southwest Kordofan and much of eastern Darfur are used mainly by nomad cattle-owning pastoralists.

7. The Nuba mountains, with the tableland from which they rise, are comparatively well populated. The Nuba peoples are expert at cutting rock cisterns in the veins of rotted granite, and sink wells at suitable places at the foot of the scarps. They have long been established in hill villages, surrounded by small terraced fields, and have been encouraged to settle in the piedmont areas and to cultivate more land, chiefly for cotton, in areas of silty soil.

8. The Red Sea hills are near-desert except in the extreme south, and are occupied by the nomadic, camel-rearing Beja tribes.

9. The Rea sea coastal plain is also a near-desert and has practically no settlements except for Port Sudan (which depends for its water upon wells sunk in the bed of the Khor (Khawr) Arba'at, nearly 30 mi. N.) and Tokar (Tawkar).

10. The semidesert zone is arid, but the slight rainfall sustains a sparse and discontinuous vegetation (chiefly in watercourses) which is grazed by camel-owning nomads.

11. The desert, in the northwest, is rainless, completely unproductive and uninhabited.

12. The Nile valley below Khartoum is a nearly continuous ribbon of land irrigated by the ancient shadoof (*shaduf*) and sakieh (*saqiya*) methods (*see* EGYPT: *The Economy: Methods of Agriculture*) and the modern pump. It supports a peasant population, living in villages between the sown land and the desert behind, and several large towns, including Shendi, Atbara-Berber ('Atbarah-Barbar) and Wadi Halfa. (J. H. G. L.)

III. THE PEOPLE

The inhabitants of the Sudan are of widely varying ethnic origin. The republic is roughly divisible into the Muslim Arabic-speaking north and the pagan, partly Christianized, south, the boundary being the line of the Bahr al 'Arab (at about 10° N. latitude). The amount of internal migration creates overlaps, but in the north only the pocket of Nuba, in the Nuba mountains, is pagan. Physically the northerners are mostly of the stock generally called Semitic, although people of the types usually classified as Hamitic (the Beja tribes) are found in the east. Southerners are Nilotic (including Nuer, Dinka, Shilluk), Negroes (Azande, etc.) and Nilo-Hamites (Bari, etc.).

1. Tribal Groups.—Apart from the large agglomeration of the Three Towns (Khartoum, Khartoum North and Omdurman) and of Port Sudan, the country is fairly described as rural and tribal although there are sizable market towns and administrative centres throughout.

The northern and central rural populations can be divided into nomadic camel owners, nomadic cattle owners, settled cultivators, riverine peoples and participants in one or another of the rural development projects. The nomadic camel owners are mainly Kababish, Kawahla, Hawawir and Zaghawa Arabs of Kordofan and Darfur provinces and the Beja (*q.v.*; including the Hadendoa, or "fuzzy wuzzies") and the Shukria Arabs of Kassala province. The cattle-owning nomads are the Baggara (*q.v.*), or Baqqarah, Arabs, comprising many different tribes who extend in a central belt westward from the White Nile through Darfur as far west as Lake Chad but who share a similar way of life. The camel nomads and the Baggara belong to tribes whose members regard themselves as being related to one another patrilineally. They move their herds each year over much the same ground, seeking the grass within their circumscribed tribal territories. Between the camel and cattle owners there lies a belt of settled cultivators who depend on millet or the sale of gum arabic for their livelihood. The Nubians and Barabra (*q.v.*) of the Nile valley grow and sell dates, and export labour to the towns, largely in the form of domestic service, although many merchants in the towns and in tribal markets are Nubian in origin. They are forced out of their land by the continual fragmentation on inheritance of the small available riverine estates.

The development of the Gezira plain, as well as the Barakat and the Gash projects and the new Khashm al Qirbah project on the Atbara river, afford poor rural populations some profit from cotton and other crops.

In the south the cattle folk (notably the Nuer and Dinka [*qq.v.*]) are the most numerous. Their climate is wetter and they move about less than the northern husbandmen. The difficulty in gaining access to the swamp land as well as the people's own resistance has delayed the development of this part of the country, although the Dinka are better educated and are more inclined to seek work in towns than are the Nuer. The Bari (*q.v.*), once primarily pastoral, were forced into a greater dependence upon agriculture by Arab depredations in the 19th century.

2. Languages.—Arabic, the official language, was claimed at the 1956 census by more than half of the total population as their native tongue although it is spoken with considerable dialect variations. English is a common alternative. The Beja, Nuba and some peoples of Darfur are large groups in the north with languages of their own, although many of them speak fluent Arabic as well (*see* NUBA, THE; NUBIAN LANGUAGE AND WRITING). The south has a great language mixture, Dinka (over 1,000,000 speakers), Nuer (500,000) and Zande (250,000) being the languages most spoken. The Shilluk are of "Sudanic" linguistic stock. (*See* also SUDANIC LANGUAGES.)

3. Religion.—An estimated two-thirds of the population, mostly in the northern and central areas, profess Islam. There is a broad division on religious grounds between the followers of the Mahdi (*q.v.*) and his successors (largely accounted for by the western nomadic tribes) who follow an austere brand of Islam, and the rest. This also became a political division, the leaders of the two most important religious groups (tariqas, or brotherhoods) becoming respectively leaders of rival parties which emerged at Sudan's independence in 1956 (these were dissolved in 1958).

Pagan religions predominate in the south though there are considerable numbers of Christians (Catholic and Protestant), and since independence Islam has also been making headway. Cattle plays an important part in the religions of the Dinka and Nuer. The Azande (*q.v.*), who inhabit the Congo border, are known for their remarkable system of oracular divination, which regulates much of their daily lives. (I. C.)

4. Population Distribution.—The population of Sudan, which was 10,262,536 in 1956 at the first census ever taken is small in relation to its area. The mean density is slightly more than 10 per square mile. Less than 10% of the population lives in urban centres, the largest of which are the Three Towns (Khartoum, Khartoum North and Omdurman), Wad Medani and El Obeid (Al Ubayyid) in central Sudan; Port Sudan on the Red sea and Atbara in northern Sudan; Kassala in the northeast on the Eritrean border; and Malakal and Juba in southern Sudan. Approximately a quarter of the country is completely uninhabited, including the desert in the north and northwest, a large area in the southeast divided between Upper Nile and eastern Equatoria provinces, and tracts in central and western Bahr el Ghazal. About a third is used only by migratory pastoralists. The areas containing a settled population account for less than a half of the total area, and comprise, in order of importance: (1) the central rain lands, between about 10° and 15° N.; (2) the southwestern margins of the southern clay plain and the adjoining fringe of the ironstone plateau; (3) the agricultural belt in the far south, close to the borders with the Republic of the Congo and Uganda; (4) the Nile valley below Khartoum; (5) the areas containing patches of high land close to the Bahr el Jebel, the White Nile above Melut (Malut) and the Sobat, on which the eastern Dinka, Nuer and Shilluk have their villages.

The extent of migratory modes of living is remarkable. About 14% of the population is wholly nomadic. These people, numbering about 1,300,000, live in northern Sudan, and include the Beja tribes of the Red Sea hills, Arabs of northern Kordofan and Darfur and the Baggara of south Kordofan and southern Darfur. The Nilotic tribes of the southern clay plain, although possessing permanent villages, normally migrate during the dry season for distances of from 5 to 50 mi. Some of the agricultural tribes living close to the southern and eastern frontiers practise shifting cultivation, and abandon their old homesteads for new after a period varying between 5 and 15 years. Many cultivators in the central rain lands migrate during the dry season because local water supplies fail. Altogether, it can be estimated that nearly half of the population is migratory in varying degrees; and only slightly more than half is settled.

Tribal grouping is also very diverse. More than 570 tribes and numerous subtribes were registered at the 1956 census. Arab tribes account for about two-fifths of the population, but some other large groups are numerous. In northern and central Sudan the Arab tribes (Ga'alin, Guhayna, Baggara and others) form 37.9% of the total population. Other major groupings are the Nuba 5.6%, the Beja 6.3% and the tribes of western Darfur 8.8%. In southern Sudan, the Dinka account for 8%, Azande for 2.1% and Bari for 2%. An interesting indicator of the sparse population is the presence of about 500,000 "westerners" from the countries which were formerly part of French Equatorial and West Africa who have settled in Darfur, parts of Kordofan and the Blue Nile province. In the eastern parts, they have been attracted by unused agricultural land and the demand for labour in agricultural projects. Some are Sudanese nationals, and altogether they comprise 6.4% of the total population. (J. H. G. L.)

IV. ARCHAEOLOGY

With minor exceptions, it was only in northern Sudan, where expeditions from many countries, all working under the auspices of UNESCO, were excavating sites doomed to be flooded by the Aswan high dam reservoir, that there had been any scientific excavations by the mid-1960s. (*See* NUBIA.) Elsewhere in the Sudan archaeology remains in its infancy. No thorough archaeological surveys have been made, though from surface finds and from the observations of officials and others certain general lines are beginning to emerge. Roads by which foreigners obtained the products of the Sudan (particularly slaves, gold and ivory) were always of paramount importance.

In the upper valley of the river Atbara in eastern Sudan, fossils and Lower Palaeolithic implements have been found, indicating one route by which the early Stone Age cultures of Egypt and east Africa were connected. In the desert in the northeastern Sudan gold was extensively mined from ancient Egyptian times down to the 11th century A.D. Stone-built huts, stone mortars, hammers and circular grinding mills occur with pots at the ancient mines. The most important route in this area from the earliest historic period has been that which gave access from the Red sea (south of Port Sudan) to the central Sudan by a route which passed north of the Ethiopian massif to the Kassala area. The earliest pottery known in the Sudan, associated in the Khartoum area with barbed bone harpoons has been found as far east as Kassala and as far west as Wanyanga (Ounianga) in Borku. From Qoz Regeb (Qawz Rajab) has come evidence of occupation from the protodynastic to the Christian period; and at Sarsareib and in the Kassala area are sites dating back to protodynastic and possibly earlier times. In the Red Sea hills there are early occupation sites at Agordat in Eritrea (C Group and Pan-Grave) and at Erkowit. At 'Aqiq on the Red sea coast are remains suggesting that this was the site of Ptolemais Therion, whence the Ptolemies obtained elephants for their army and which probably continued as a trade centre until the Aksumite period. In Roman times trade with India diverted attention from the African mainland; but Meroe and Aksum became affected by influences from India as a result of the development of that trade, and it was probably trade rivalry that led Aksum to destroy a degenerate Meroe *c.* A.D. 350. Early in the Muslim era Ptolemais was overshadowed by Badi on Airi Island (Jazirat ar Rih), where there is a ruined town with cisterns and where Muslim tombstones dating from the 10th and 11th centuries have been found. Similar tombstones dating from the 8th to the 10th centuries (the earliest is A.H. 153 = *c.* A.D. 790) have been found in Khawr Nubt, while early Muslim stone-built towerlike tombs occur at Ma'man east of Kassala and elsewhere. Circular stone graves with flat tops are probably those of the Muslim Beja.

In the Blue Nile valley of central Sudan fossils are found below the clay of the Gezira plain, notably at Abu Hugar and Sinjah, where they include a proto-Bushman skull, contemporary with a crude Levalloisian culture. Ancient Egyptian expeditions must have penetrated far south of that part of the Nile valley occupied by them during the New Kingdom. One such expedition to localities previously unidentified is recorded by Amenhotep II, one of whose scarabs was found 100 mi. E. of Khartoum. The Blue Nile valley from Khartoum to Er Ruseres (Er Roseires; Ar Rusayris) was no doubt always the road to Beni Shangul (Ethiopia), whence came slaves and gold. Occupation sites occur all the way to Er Ruseres, and no doubt the Napatan, Meroitic, Christian and Funj (Fung) periods are represented. In the 6th century A.D. Aksum was trading with this area directly, bartering gold for iron at Fazughli. Finds of Napatan date were made at Sennar and Kosti, while at Moya (Jabal Mayyah) Sir Henry Wellcome excavated an occupation site and cemetery showing Negroes' trade connections with Napata. Finds of Meroitic date were made in and near Sennar. Rock pictures from this period occur at Geili (Jabal Jayli) east of Khartoum; and the artificial water reservoirs which are common as far south as the Jabal Mayyah area are probably the work of the kingdom of Meroe. A Meroitic ram, brought from Soba (Sawba) to Khartoum by Gen. C. G. Gordon, indicates that Soba was probably a Meroitic town before becoming the capital of the Christian kingdom of Aloa. Little now remains on the surface at Soba but red-brick fragments and the stone capitals of a Christian church; Old Nubian inscriptions have been found there in the past. Christian sites are reported from the Blue Nile between Khartoum and Sennar and from al Geteina (Qutaynah) on the White Nile. At Sennar the site of the Funj capital (A.D. 1500–1821) is being destroyed by the river. At Abu Jayli near Sennar, a site excavated for Wellcome in 1912, Funj pottery was recognized with considerable probability.

South of Kosti, occupation sites in southern Sudan occur along the bank of the White Nile as far as Malakal. Shards of apparently early date have been found near Malakal, Doleib Hill and Jabal Zaraf al Azraq, and a microlithic culture of Wilton type with pottery on Belli Island. Shards from a pre-Shilluk site near Kaka suggest a connection with Sennar and Meroe. Ancient occupation mounds have been reported from the Bahr el Ghazal in Aweil district, and polished celts of hematite near Yambio. This southern region must long have been the goal of traders seeking slaves and ivory, and copper mining probably began at Hofrat el Nahas (Hufrat an Nahas) at an early date. The Azande show evidence of cultural contact with ancient Egypt, and the Masai and the Kikuyu of east Africa seem to appear more than once, in scenes representing the arrival of the tribute of Cush, in the Tombs of the Nobles at Thebes.

In western Sudan the most important road is that which runs west from the White Nile between the desert and the tropical forest through Darfur and past Lake Chad to west Africa. A few roads running north to south across the desert follow lines of water holes and link this great road with the north African coast. In early times the steppe country extended farther north, and finds of protodynastic date come from the southern Libyan desert. It may be that Tibesti is the source of the amazon stone common to the Neolithic cultures of the Fayyum and Shaheinab. From Wadi Howar come remains which suggest that the cattle-owning C Group people, first known to archaeologists in lower Nubia, once lived there. The 6th-dynasty inscriptions on Harkuf's tomb at Aswan suggest that the ancient Egyptians used donkey caravans on the Darb al Arba'in. The earliest traditional rulers of Darfur were the Dagu (otherwise Daju, Tajo, Tago), and some of the brands used till recently by the Dagu sultans of Dar Sila may have

a hieroglyphic origin. The extent to which the ancient Egyptians and, later, the kingdoms of Nāpata and Meroe controlled the great east-west road was still unknown in the early 1960s.

Meroitic "archer's looses" were found in Darfur and at Jabal Harazah. That, after the fall of Meroe itself, the Meroitic royal family may have carried on in the west is suggested by an Old Nubian graffito at Audun, a tumulus near Kaboija in Jabal Maydub (said to be the grave of a queen who brought the Meidob-Kagiddi people from the Nile) and also by the survival of Kash (Cush), the old name of the kingdoms of Napata and Meroe, in the modern tribal names Kagiddi, Kaja and (Birgid) Kajar.

There are many stone-built ruins in the hills of Darfur. A few are attributed to the Dagu. The red-brick ruins at Ein Fara ('Ayn Fa'rah) in northern Darfur were proved to have been a monastery and two small churches (one, domed), indicating that between the 9th and 13th centuries, if not earlier, Darfur was within the sphere of influence of the Christian kingdom of Mukurra (from which the tribal name Tungur is derived). That sphere probably once extended as far as Lake Chad. The fine date palms and hill-terracing near Ein Fara suggest that the Tungur originated the excellent terracing and good agriculture of the Darfur hills. They still maintain their connections with the Bideyat (Bedyat) of the Ennedi plateau (who were once Christian). The walled stone town containing two royal residences at Jabal Orrei (Uri) probably dates from the same period. Rock pictures of horsemen in the Tungur hills and at Jabal al 'Afarit in Kordofan are probably crude copies of Christian saints on horseback as represented in the church of Abd al Qadir at Wadi Halfa. The red-brick ruins at Abu Sufiyan and Jabal Zankor in northern Kordofan no doubt date from this period, and there are rectangular stone Christian graves in the Wadi Muqaddam. The influence of Christian Nubia in Darfur must have been destroyed by the advancing power of the Muslim kingdom of Kanem, which by A.D. 1240 controlled the trade route between Kanem and Nubia. From that date to the 16th century, the influence of the medieval kingdom of Bornu was predominant in Darfur and Kordofan and is responsible for the many stone compounds between Kordofan and Dongola, and no doubt for some stone buildings in the Darfur hills. The most recent stone forts in Turrah at the northern end of the Marrah massif are traditionally associated with the early Fur sultans. *See* also EGYPT: *Archaeology.* (A. J. AL.)

V. HISTORY

1. Early History.—The southern Sudan has no recorded history until the 19th century A.D. although the equatorial regions were vaguely known to ancient Egyptian, Greek and Arab geographers. The northern Sudan first came under the influence of Egyptian culture *c.* 3000 B.C. and was actively occupied by the Egyptians between 2150 B.C. and 1580 B.C. A native power subsequently arose with its capital at Napata (*q.v.*), near the fourth cataract of the Nile, whose rulers conquered and held Egypt itself from *c.* 730 B.C. to 660 B.C. The capital was transferred *c.* 590 B.C. to Meroe (*q.v.*) near the sixth cataract, and the following centuries saw a growing isolation from Egypt. The Ptolemaic and Roman rule in Egypt made little permanent impression on the Sudan.

By *c.* A.D. 500 three kingdoms had emerged: Nobatia, with its capital on the present Sudanese-Egyptian frontier; Mukurra (Muqarra), the central kingdom, with its capital at Old Dongola (*see* DONGOLA); and Alodia (Aloa), the southernmost kingdom, centred on Soba at the junction of the Blue and White Niles. The people of all three were called Nubians by contemporary writers. The unsettled state of the frontier probably explains why Christianity did not penetrate Nubia until the 6th century, at a time when the eastern Roman empire was beset by theological disputes. Julianus, a missionary from Alexandria, converted the Nobatae to monophysite Christianity *c.* 543 and Longinus, who had been sent as monophysite bishop of Nubia, converted the Alodians, while the middle kingdom seems for a brief time to have received the Orthodox faith.

The conquest of Egypt in 640–641 by an Arab army commanded by 'Amr ibn al-'As brought triumphant Islam to the northern border of the Sudan and isolated the people from the political and religious influence of Constantinople. In 651–652 'Abdullah ibn Sa'd, Arab governor of upper Egypt, invaded Nubia and made a treaty with the Christian ruler.

After 700 years of relative prosperity followed by slow decline, the power, and finally the existence, of the Christian states on the Nile were brought to an end probably in the 14th century. So long as Egypt remained under Arab domination, Muslim relations with the Christian kingdoms were usually friendly, but after the Arab governors had given place to non-Arab mercenaries, and later to the Mamelukes, relations with Egypt deteriorated. Arab tribes, chiefly Guhayna, edged out of Egypt by the foreign yoke and fleeing from uncongenial settled government, infiltrated into the Christian Nile valley and the broad savanna country of the central Sudan. The Arabs intermarried with the local Nubians and Beja and gradually obtained control, though both Nubians and Beja largely retained their own languages. The southernmost kingdom of Alodia, later known as 'Alwa, survived for a longer time, though fixing the time and circumstances of its fall is difficult.

The origin of the Funj, or Fung, sultanate which rose on the ruins of the Christian kingdoms has long been the subject of controversy. Various theories connect the foundation of this state with immigrants from the White Nile, the upper Blue Nile and Bornu. The ruling body seems to have been Muslim from the beginning of the sultanate, though sultans and subjects retained pagan practices. Arabic was the language of administration. The Funj chronicles record that the dynasty was founded by Amara Dunqas who reigned from 1505 to 1534. The sultanate was a loose federation. The sultans directly ruled only the Gezira, the peninsula between the Blue and White Niles, while the country immediately to the north was administered by 'Abdullab Arab viceroys who had their seat at Qarri at the sixth cataract (Shallal as Sablukah). The rest of the sultanate was governed by autonomous tribal chiefs, the more important of whom received the title of *manjil* and were invested with a ceremonial stool and distinctive two-horned headdress, symbols of the sultan's warrant.

The Funj sultanate was brought early into contact with the Ottoman power which conquered Egypt in 1517. The Turks planted garrisons of European Muslim troops in fortresses at Aswan, Ibrim and Sai Island (Jazirat Say) and at an unascertained date defeated a Funj army at Hannek. The Turkish garrisons mingled with the local inhabitants and modified the racial type. It was during the reign of the Ottoman sultan Suleiman I that Özdemir Pasha, beglerbeg of the Yemen, occupied the ports of Suakin and Massawa which became the chief towns of the Ottoman coastal province of Habash. Trade appears to have been an important concern of the Funj sultans who, however, struck no coins. The earlier sultans were munificent patrons of Muslim scholars and saints whose mysticism gave Sudanese Islam its distinctive character. About 1680 the Funj confederacy was weakened by a revolt of the Shayqiya (Shaikia, or Shagia) tribe in the north, and a general decline in the administration set in during the 18th century. In 1789 chiefs of the Hameg (*q.v.*; or Hamaj), a people of indeterminate origin, usurped the government, and the sultanate relapsed into political fragmentation. In 1812 Dongola was seized by Mameluke fugitives from Mohammed Ali's massacres in Egypt. By 1820 the anarchic condition of the northern Sudan made it an obvious objective for foreign invasion.

2. Egyptian Occupation, 1820–81.—For the next 60 years the Sudan was governed as a dependency of Egypt under the nominal suzerainty of the Ottoman empire. The Egyptian penetration was effected in 1820–21 by two invading columns. The first, commanded by Ismail, third son of Mohammed Ali, set out from Wadi Halfa in Sept. 1820, ascended the Nile valley, defeated the Shayqiya and entered Sennar on June 14, 1821. The second column, under Mohammed Bey Khusrau, the daftardar, followed up the Nile in the spring of 1821, left the valley at Dabbah and marched southward over the Bayyudah steppe to Barah in Kordofan where the Egyptians defeated Musallim, the Darfur regent of Kordofan, on Aug. 16. On Aug. 20 they entered El Obeid. Ismail's murder by a Sudanese chieftain at Shendi in Oct. 1822 was the signal for a desperate Sudanese revolt which involved all the tribes from Berber to Sennar. The revolt was bloodily sup-

pressed by Mohammed Bey with a force from Kordofan. In 1824 'Osman Bey, military commander of Sennar, built a fort at the village of Khartoum which became the seat of government.

After five years of military misrule and savage taxation, 'Ali Khurshid Pasha was sent to the Sudan in 1826 as its first governor general with full military and civil powers. He converted Khartoum from a military headquarters to an administrative capital, put the administration on a sound footing and conscientiously attempted to carry out Mohammed Ali's plans for economic development. He persuaded many Sudanese fugitives from Egyptian maladministration to return to their lands, but the Ethiopian ruler of Gondar resented his aggressive frontier policy and there was fighting in 1836–38. Khurshid was recalled in 1838 after proving himself an able administrator by the standards of the time. His successor, Ahmad Pasha Abu Adhan, who held office from 1838 to 1843, ruled with a strong arm but mysteriously died in 1843.

The next considerable extension of Egyptian rule occurred in 1840 when Ahmad Pasha Abu Adhan seized the region of Taka, the future Kassala province, from the Hadendoa. The ports of Suakin and Massawa, leased to Egypt in 1846, were finally ceded to the khedive Ismail (*see* ISMAIL PASHA) in 1865. Ismail sought to develop and expand his Sudanese possessions. In 1869 he commissioned Sir Samuel White Baker to annex the region between Gondokoro and the great lakes. Baker's mission, though heroically accomplished, was an expensive failure, and his successor in the Equatoria province, Col. (afterward Maj. Gen.) Charles George Gordon (*q.v.*) withdrew the Egyptian garrisons from most of the area. Zubayr (Zobeir) Pasha Rahma Mansur, an influential slave and ivory trader in the Bahr el Ghazal basin, was in 1873 made its first governor. The Egyptians extended their authority as far as the Somali coast and would have occupied Kismayu (Chisimaio) in 1875 if Britain had not intervened. The final stage of Egyptian expansion was completed in 1874 with the conquest of Darfur. Ismail's attempts to gain territory from Ethiopia failed when two Egyptian invading armies operating from Massawa were defeated in 1875 and 1876.

Despite various experiments in decentralization by Mohammed Ali, Mohammed Said Pasha and Ismail, the government was for most of the time based on Khartoum. The political structure devised for the Sudan followed the Egyptian model. A governor general and central government services were located at the capital. Each main territorial subdivision was ruled by a governor. The settled parts of each governorate were again divided into a number of districts, each under a district officer. Each tribe was administered by its own sheikh. A cardinal weakness of the Egyptian regime, and one which contributed to its destruction, was its failure to win the obedience, much less the loyalty, of the nomad tribes who were periodically raided for cattle in lieu of taxes. During Ismail's rule technical departments responsible for steamers, posts and telegraphs were added to the government services.

The Egyptian rule brought many changes to the Sudan. For the first time in its history the Egyptians imposed a single government over the whole country. Practices foreign to Sudanese religious traditions were introduced; the Sudanese Muslims followed the Maliki discipline, whereas the Egyptian law courts administered the Hanafite discipline which was official throughout the Ottoman empire. At first the judges and the higher ulema (*'ulama'*), or Muslim scholars, were exclusively Egyptian but as time went on the Sudanese took an increasing part in judicial work and about 1866 a jurisconsult, Mohammed al-Amin al-Darir, became the first Sudanese president of the Sudanese ulema.

The forces which made the first penetration of the Sudan were composed of Turks, Albanians and Arab irregulars. Climate and economy dictated the conscription of Negro soldiers captured in raids. By 1843 most of the troops in the Sudan were composed of Sudanese Negroes. Their officers were at first Turks; then, as the supply of these grew short, selected Sudanese were given commissions until in 1872 a Negro, Adham Pasha al-Taqalawi, became acting governor general. In the higher civil administration there were two Sudanese provincial governors: Husayn Pasha Khalifa and Mohammed Rasikh Bey. The cultural influence of Egypt was confined to the small assimilated class in larger towns where, in the time of Ismail, primary schools were established on the Egyptian model. There was no large-scale immigration of Egyptian settlers, though there was much intermarriage with Sudanese in the garrison towns. The way of life of the mass of the Sudanese people was unaffected by the Egyptian presence.

3. Mahdist Theocracy, 1881–98.—On June 29, 1881, at Aba on the White Nile, Mohammed Ahmed ibn al-Sayyid 'Abdullah proclaimed himself as the expected Mahdi who had been divinely elected to restore Islam (*see* MOHAMMED AHMED). He had long been venerated for his holiness and asceticism and a consciousness of his mission was said to have come to him in progressive revelations. His revolt was partly a protest against Egyptian misgovernment. The Egyptian administration had grave defects; its character was Ottoman; most of the senior public service posts were held by members of the europeanized, Turkish-speaking, Egyptian ruling class, whose way of life scandalized the austere Sudanese; and while there were influential Sudanese who were favourably disposed toward the regime by which they profited, the nomads were universally hostile.

A more potent cause for discontent was the rough-handed attempt to suppress the slave trade. The khedive Ismail's own role as a reformer, as well as the pressure upon him from the European powers, drove him to Draconic measures for its eradication. Islam tolerated slavery, but whereas in Egypt slavery had become an anachronism, the Sudanese economy was still based on it. When Ismail recruited European officers to carry out his policy more effectively, the Sudanese concluded that he had sold himself to the European unbelievers and that his government was undermining Islam. Another cause lay in the nature of Sudanese Islam itself, which found corporate expression in the Sufi religious orders, a religion of the heart, colourful and emotional. The Sudan was affected by the religious revival which had its centre in Arabia in the 18th and 19th centuries. Mohammed 'Uthman al-Mirghani, a disciple of Ahmad ibn Idris al-Fasi at Mecca, came to the Sudan in 1817 and founded the great Mirghaniya, or Khatmiya, order. The Ismailiya order, strong in Kordofan, was founded by Ismail al-Wali of El Obeid (1793–1863), the great-grandfather of the first Sudanese premier, Ismail al-Azhari. Followers of the Sammaniya, a branch of the Khalwatiya order, introduced during the last years of the Funj sultanate, were numerous in the Gezira and counted the Mahdi among their adherents. The Mahdist movement was not, as Egyptian and European writers of the time termed it, a revolt of dervishes against orthodox Sunnite Islam. The Mahdi was no dervish and expressly forbade the use of the term by and for his followers. The movement was intended rather to restore primitive Islam and to cleanse it from innovation and superstition. The Mahdi's immediate enemy was the Egyptian government, which in his view had lapsed into unbelief.

The governor general at the time, Mohammed Ra'uf Pasha, an indecisive man, underestimated the strength of the movement. He was probably influenced by the Egyptian-trained Islamic religious and legal notables in Khartoum who rejected the Mahdi's claims. After massacring a small government force sent to arrest him, the Mahdi and his followers retired to the Nuba mountains where in Dec. 1881 and May 1882, they annihilated two columns sent against them. In 1882–83 they won a series of spectacular victories over the Egyptian garrisons and the expeditions sent for their relief. Early in 1884 the Mahdist horde closed on Khartoum. On Jan. 26, 1885, the capital fell and its redoubtable commander, Gordon, was killed. Four days later the Mahdi solemnly led the Friday prayers in the city mosque. Within the next few months the remaining Egyptian garrisons in the Sudan were evacuated or compelled to surrender, and the Anglo-Egyptian force, which had vainly attempted the relief of Khartoum in the winter of 1884–85, retired from Dongola. Only Wadi Halfa, Suakin and the Equatoria province remained in Egyptian hands.

The Mahdi died at Omdurman, which he had made his capital, on June 22, 1885, and the control of the Mahdist state fell to the Khalifa 'Abdullahi (*see* KHALIFA, THE).

4. Anglo-Egyptian Condominium, 1899–1955.—The Mahdist power was broken by an Anglo-Egyptian army commanded by Sir Herbert (afterward Lord) Kitchener. The occupation of the

Nile valley up to Khartoum was completed by the battle of Omdurman on Sept. 2, 1898, but an extensive area of the country remained either unoccupied or in disputed ownership. However, disagreements with the French which had led to the Fashoda incident (*see* EGYPT: *History*) were settled by an Anglo-French declaration of March 21, 1899, whereby the French recognized the Nile-Congo watershed as the eastern limit of French interest, while the Belgians, who had occupied the country between the Nile-Congo watershed and Lado on the upper White Nile, agreed in 1906 that the Lado enclave (*q.v.*) should revert to the Sudan on the death of King Leopold II (1909). By the Anglo-Egyptian condominium agreements of Jan. 19 and July 10, 1899, the northern border of the Sudan was fixed at latitude 22° N. with enclaves to include some villages north of Wadi Halfa and some Beja grazing grounds in the Red sea hinterland. The Italians who had wrested Kassala from the Mahdists in 1894 handed it back to the Egyptian army in 1897. On Nov. 24, 1899, Sir Francis R. Wingate's forces defeated and killed the Khalifa 'Abdullahi at Umm Diwaikarat, and in December the Egyptian army reoccupied El Obeid. In 1898 a scion of the former royal house of Darfur, 'Ali Dinar, hastened to El Fasher and seized the throne. In 1900 the Khartoum government accorded him recognition as a tributary sultan.

The condominium devised by Lord Cromer established the future form of the government. The British and Egyptian flags were to fly side by side. Supreme civil and military powers were to be vested in a governor general appointed by the khedive on the recommendation of the British government. The Sudan was to be excluded from the capitulary regime and consular jurisdiction then in force in Egypt. Kitchener, who was appointed the first governor general, received an instruction from the British government informing him that the main object of the condominium agreements was to ensure adequate British control, in concert with the Egyptian government, while conferring upon him powers to settle internal matters without reference to the British agent and consul general in Egypt (through whom he was to receive instructions from London). The structure of the condominium was modeled closely on that of the former Egyptian regime. Under the governor general were governors of provinces, who in turn were assisted by inspectors and their Egyptian and Sudanese staffs. In a memorandum of 1899 Kitchener laid down that the most important task of the governors and inspectors was to gain the confidence of the people by friendly dealing and close contact. Legal forms were to be strictly adhered to and Islam respected. Although slavery was not recognized, voluntary service by former slaves to their former masters was not to be interfered with, but cruelty to slaves was to be punished as a breach of law. Judicial officers prepared criminal and civil codes in 1899–1900. The Islamic law, confined as in Egypt to the realm of personal status, was administered by an Egyptian grand *qadi* (Muslim judge) and a system of central and provincial courts. The first governors and inspectors were British officers of the Egyptian army, but, as these were always liable to be recalled, a few graduates of British universities were appointed to form the nucleus of a civil service.

Kitchener was succeeded in Dec. 1899 by Wingate who was governor general until 1916 and who laid the foundations of the administration. The wars had reduced the country to poverty; in 1898 the revenue amounted to £35,000 against expenditure of £235,000. The annual deficit was met from the Egyptian treasury and the Sudan was not solvent until 1913 when the budget was balanced. Land surveys were made and land titles registered. A simple system of taxation was devised.

Education under the Mahdist theocracy had been confined to the teaching of elementary Arabic and religion on traditional lines in koranic schools. In 1898 Kitchener opened an appeal for funds to build the Gordon Memorial college which was opened as a primary school in 1902. By 1907 there were 6 primary and 20 elementary schools functioning and the first government girls' school was opened at Rufa'ah. In 1905 a secondary section was added to Gordon college which by 1913 provided a general secondary education with special courses for future judges of the Islamic law, engineers and teachers. Sudanese communications expanded steadily. The Egyptian telegraph network had been destroyed during the Mahdist rule and had to be entirely reconstructed. The core of the railway system was the 3 ft. 6 in.-gauge military line built from Wadi Halfa to Khartoum North in the 1890s. An extension was then laid to Sennar and El Obeid. The Red Sea railway was built from Atbara to Suakin and extended to Port Sudan where a new port was established. A steamers' department, based on a dockyard at Khartoum North, operated services on the principal navigable reaches of the two Niles and the Bahr el Ghazal (see *Transport and Communications,* below).

During World War I the Sudan enjoyed internal tranquillity and economic prosperity. Though Turkey was at war with Britain, the sultan's claim to the caliphate of Islam was not recognized in the Sudan. The Arab revolt which broke out in the Hejaz (Arabia) in 1916 found much sympathy in the Sudan which formed a base of supply for the Arab insurgents in the first phase of hostilities. In 1916 Sultan 'Ali Dinar of Darfur, who was already troubled by French encroachments on the buffer states on his western border, revoked his nominal allegiance to the Sudan government. His design to invade Kordofan was forestalled by a force commanded by Lieut. Col. (afterward Brig. Gen.) P. J. V. Kelly, which routed the Fur army near El Fasher on May 22. 'Ali Dinar fled to Jabal Marrah where he was killed on Nov. 6. Darfur was brought within the Sudanese administrative system but the western frontier with the French was not finally demarcated until 1924.

Experiments proved that the Egyptian long-staple cotton (which originated in the Sudan) could be profitably grown in the Gezira. The construction of a dam over the Blue Nile at Sennar, started in 1914, was delayed by the war and not brought into use until 1926. The water thus impounded was used to irrigate a vast cotton plantation in the eastern Gezira, which attained about 1,000,000 ac. in 1955 and provided the Sudan with its principal cash export crop and with a great addition to its grain acreage. The cotton plantation at Tokar in the delta of Khor (Khawr) Barakah was reconditioned in 1922, and another plantation was established in the Gash delta near Kassala, a project made possible by the completion of a railway from Hayya on the Red Sea line to Kassala, opened in 1924. A line from Kassala to Sennar, providing an alternative rail link with Port Sudan, was brought into operation in 1929. An international commission was set up in 1925 to report on irrigation works in the Sudan which could be undertaken without detriment to Egyptian rights. Its findings were incorporated in the Anglo-Egyptian Nile Waters agreement of May 7, 1929. A water-storage dam for Egyptian use was built between 1934 and 1937 at Jabal Awliya', 25 mi. S. of Khartoum.

Nationalism and World War II.—The first signs of a Sudanese nationalist movement in the modern sense appeared in May 1922 with the publication of a manifesto entitled *The Claims of the Sudanese Nation.* Its author was a junior officer of a Sudanese battalion of the Egyptian army, 'Ali 'Abd al-Latif, who in 1924 founded the White Flag league. After his arrest and imprisonment in June 1924 cadets of the military school paraded the streets of Khartoum, and at Atbara the Egyptian railway battalion mutinied. On Nov. 19, Sir Lee Stack, governor general of the Sudan and sirdar of the Egyptian army, was assassinated in Cairo. The resulting British ultimatum to the Egyptian government included a demand for the immediate withdrawal from the Sudan of all Egyptian officers and purely Egyptian units of the Egyptian army. After disorders in Khartoum caused by Sudanese mutineers who ran amuck and killed British, Egyptian and Syrian patients in a hospital, the Egyptian withdrawal was carried out. The Sudanese units of the Egyptian army were then embodied in a Sudan Defense force with British and Sudanese officers under a commandant responsible to the governor general.

The negotiations which led to the Anglo-Egyptian treaty of 1936 were closely followed by the educated minority in the Sudan and the next stage of nationalist development was reached in 1938 with the formation of a Graduates' General congress, a body drawn from school-educated Sudanese. Ismail al-Azhari became secretary to the congress in 1939 and its president in 1940. The Sudan was heavily involved in World War II. In 1940 the Sudan Defense force took the strain of the Italian pressure on the frontier and in 1941 assisted in the rout of the Italian forces in Eritrea and

Ethiopia. Sudanese troops also took part in the north African campaigns (1942–43).

The war years saw a further growth of national awareness. An advisory council, created in 1944, presaged further constitutional change. Municipal development was slow and it was not until 1942 that the first town council with full control over its budget was installed, but since that date the number of municipal authorities has increased. An ordinance of 1951 increased the representative element in local government. The development of education was a powerful contributing factor to an awakening sense of nationhood. Newspapers and broadcasting popularized the conception of the Sudan as a national unit. Even the religious orders found themselves taking political paths; the Khatmiya, under the leadership of al-Sayyid Sir 'Ali al-Mirghani, whose family was traditionally loyal to Egypt, adopted a position friendly to Egypt's claims in the Sudan, while the Ansar, the Mahdist confraternity, under al-Sayyid Sir 'Abd al-Rahman, posthumous son of the Mahdi, championed eventual Sudanese nationhood unshackled by political connections with Egypt. Out of the intense political activity after 1946 two main parties, reflecting these two divergent points of view, emerged. In 1947 Anglo-Egyptian negotiations broke down on the issue of the Sudan and in July the Egyptian government referred the dispute to the UN Security council, but got no satisfaction. In the absence of a settlement the governor general in June 1948 announced a program of constitutional advance including the creation of an executive council and a legislative assembly, which met in December. The preceding elections had returned the Umma party to power with the party secretary, Brig. 'Abdullah Khalil, as leader of the assembly. On Oct. 8, 1951, Egypt unilaterally abrogated the treaty of 1936 and proclaimed King Faruk I as king of Egypt and the Sudan, a pronouncement which made no serious impresssion upon Sudanese opinion.

The Egyptian *coup d'état* of July 1952 radically altered Sudanese-Egyptian relations. On Nov. 2, Gen. Mohammed Naguib, himself half-Sudanese, publicly recognized the Sudan's right to self-determination. On Feb. 12, 1953, Britain and Egypt signed an agreement by which the Sudan would attain self-government within three years. In the elections held late in the year the National Unionist party won a victory and Ismail al-Azhari was elected prime minister. Though he came to office on a political program favourable to Egypt, relations with Cairo worsened during 1955 and the prime minister complained of Egyptian interference in the internal affairs of the Sudan. By the end of 1954 almost all the British officials had handed over their duties to their Sudanese colleagues. A mutiny in the Equatorial battalion of the army broke out in Aug. 1955, accompanied by serious disturbances and loss of civilian lives. A commission of inquiry which issued its report on Oct. 19 advanced seven causes for the outbreak, among them fear in the non-Muslim south of the political domination of the Muslim north, the circulation of false rumours and official maladroitness.

5. Republic of the Sudan.—On Jan. 1, 1956, the Sudan proclaimed its independence. The last governor general, Sir Knox Helm, resigned; British and Egyptian troops had already left the country. In spite of the formation (Feb. 1956) of a national government headed by al-Azhari, the ministry fell on July 4 by the defection of ministers, loss of support in the assembly and the intervention of al-Sayyid Sir 'Ali al-Mirghani, who became patron of the seceding party. 'Abdullah Khalil was elected prime minister of a coalition in which his own Umma party was prominent.

A frontier incident arose in Feb. 1958 between Egypt and the Sudan resulting from the disputed ownership of the two enclaves north of the 22nd parallel. The Sudan submitted the dispute to the UN Security council, but the incident was closed by direct agreement between the disputants.

Elections held in March 1958 gave the victory to an alliance between the Umma and the People's Democratic parties, and 'Abdullah Khalil remained in office. By the summer of 1958 the serious state of the national economy was disturbing public opinion. Inability to dispose of the cotton crop, on which the wealth of the country depended, had reduced the foreign currency reserves to a low level. There followed a period of political tension threatening the existence of the government, which was heavily attacked for its handling of the economic crisis and for its reputedly excessive dependence on British and U.S. aid. There were serious divisions in the cabinet. On Oct. 27 all political meetings in Khartoum, Omdurman and Khartoum North were banned.

A military *coup d'état* was carried out bloodlessly on Nov. 17, 1958, when a group of army officers headed by Gen. Ibrahim 'Abboud, took over the government, dismissed parliament, dissolved all political parties and suspended the transitional constitution. The officers who formed the supreme council of the revolutionary government soon began to exhibit the same lack of unity which had weakened the fallen ministry. A counterrevolution on March 4, 1959, by Brigadiers 'Abd al-Rahim Shannan and Muhyi al-Din Ahmed 'Abdullah, whose troops overawed Khartoum, forced the president to include them in the reconstituted supreme council. The two officers opposed alignment with the west and urged a closer identification of the Sudan with the Arab states. The supreme council did not achieve internal harmony and on June 1 the two brigadiers, with other officers, were arrested on a charge of attempting once again to use troops to further their aims. The brigadiers and four others were found guilty and sentenced to death, but the sentences were commuted to imprisonment. A third attempt to overturn the regime occurred on Nov. 10, 1959, when some officers, mostly from the Omdurman infantry school, defied the government. Five of the officers were hanged and five imprisoned.

On the grounds of preventing what it considered a threat to the unity of the Sudan the government abandoned the policy (followed by the condominium authorities since the reoccupation) of segregating the mainly pagan south (which also had small Christian communities) from the Muslim north in favour of a policy of gradual assimilation. Mission schools were nationalized and the teaching of Arabic facilitated. A Sudanese-Egyptian agreement was signed in Cairo on Nov. 8, 1959, which superseded the Nile Waters agreement of 1929 between Britain and Egypt. The Sudan gave formal consent to the building of the high dam at Aswan (*q.v.*) by Egypt (United Arab Republic), while the latter agreed to compensate the Sudanese whose properties would be inundated by the water impounded in the vast reservoir. Thus the Sudan was then free to carry out various irrigation projects including a dam over the Atbara river at Khashm al Qirbah, another dam over the Blue Nile at Er Roseires (Ar Rusayris) and the completion of the Manaqil extension in the central Gezira. The Sudan has also made a notable extension to its communications facilities since 1945.

The former prime ministers al-Azhari and 'Abdullah Khalil were arrested in 1961 and exiled in the south following agitation for the restoration of constitutional government. They were released in 1962, but no concessions were made to their political demands.

(R. L. HE.)

VI. ADMINISTRATION AND SOCIAL CONDITIONS

1. Government.—The Anglo-Egyptian condominium that began in 1899 ended on Dec. 31, 1955. It was terminated after a transitional period under the terms of the Anglo-Egyptian agreement of Feb. 1953. For nearly three years the new republic was governed by a council of ministers based upon a majority in the house of representatives; there was also a senate and a five-man commission which exercised the functions of head of state. On Nov. 17, 1958, the Sudanese army assumed power in the course of a peaceful *coup d'état;* and its commander in chief, Gen. Ibrahim 'Abboud, became head of state, with the title of president of the supreme council for the armed forces, and head of the government as prime minister (see *History*, above). A council of ministers, including civilians and senior army officers (each member of which has departmental responsibility), is responsible to the supreme council through the prime minister. The public service commission, auditor general and chief justice are responsible to the supreme council and not to the council of ministers.

2. Administration.—The Sudan is divided into nine provinces which are related to the two main lines of communication: the Nile river and the east-west "pilgrim" route from west Africa to Suakin on the Red sea. The provinces were formerly divided into districts, at first administered directly by district commissioners;

but devolution to local authorities began in 1918 and passed through further stages in 1937, 1954 and 1960, when a local government ordinance established what may be regarded as the first instalment of a constitution embodying the political theory of the authoritarian military government. The offices of governor and district commissioner were abolished. The governors were replaced by government representatives, who are chairmen of province councils; these councils are composed partly of members elected by local authorities from approved candidates and partly of appointed members chosen from province officials and private citizens. The district commissioners were replaced by local government inspectors who, in 57 local government areas (or districts) out of the total of 86, are chairmen of the district councils. Local government areas without councils, and thus administered directly, are mainly in the three southern provinces. Provinces and districts have powers of taxation, and responsibilities in the spheres of health, education and veterinary services.

Khartoum is the capital of the republic and in 1956 had a population of 93,103.

Provinces of the Republic of Sudan

Province	Capital	Area (sq.mi.)	Population (1956)
Bahr el Ghazal	Wau	82,529	991,022
Blue Nile	Wad Medani	54,879	2,069,646
Darfur	El Fasher	191,648	1,328,765
Equatoria	Juba	76,495	903,503
Kassala	Kassala	131,527	941,039
Khartoum	Khartoum	8,097	504,923
Kordofan	El Obeid	146,929	1,761,968
Northern	Ed Damer	184,198	873,059
Upper Nile	Malakal	91,189	888,611

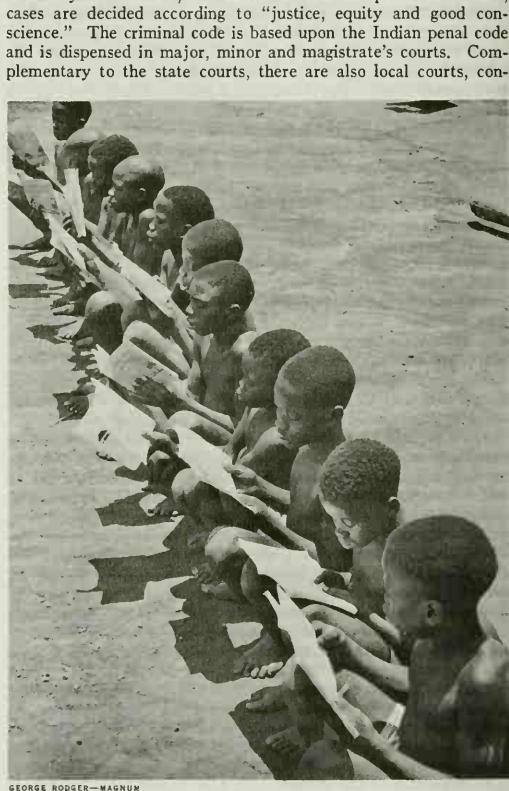
Police and Prisons.—The police force is under the authority of the minister of the interior. The commandants of the province police forces are responsible to the government representatives for administration, organization and discipline. There are branches of the force mounted on camels, horses and mules; and mobile reserve companies provided with motor transport. Prisons are similarly organized; the commissioner for prisons directly controls central prisons at Khartoum North, Port Sudan and Suakin, as well as two reformatories and a training school. Probation officers are appointed in the seven principal northern towns.

3. Living Conditions.—According to the 1956 census the labour force comprised about 3,800,000 (37% of the total population), the great majority of whom were engaged in farming and animal husbandry. In industry, commerce and public administration there is legislation to cover employment exchanges, conditions of employment in factories and in domestic service and workmen's compensation. There are four employment exchanges in the largest towns of northeastern Sudan. Trade unions have been recognized since 1948, and negotiate with employers (including the government) on wages and working conditions. Under an amending ordinance of 1960, provision is made for closer control by the labour commissioner. Another ordinance of 1960 prescribed that disputes should be compulsorily arbitrated when negotiation and conciliation fail. Under legislation applying to co-operative societies, many such bodies have been registered, of which about 150 are agricultural and more than 100 are for marketing and credit.

Towns are planned so that there are separate first-, second- and third-class residential areas, as well as commercial and industrial zones. Buildings must be erected in conformity with regulations governing materials to be used, number of stories, etc. Outside the limits of towns and separated from them by open land, are *deims* (*dayms*) where unregulated houses, *i.e.*, shacks and native huts, may be erected. In rural areas generally, where the population is sedentary, housing conforms to tribal custom and usually comprises square flat-roofed, mud-brick dwellings (in the north) or circular grass huts with spired thatched roofs. The nomads live in tents.

4. Public Health.—Early in the 20th century the Sudan was reputed to be very unhealthy, and epidemics of sleeping sickness, cerebrospinal meningitis and yellow fever were frequent. The progress of medical and sanitary services under the Sudan medical service (later the ministry of health), supplemented in many places by missionary institutions, has reduced mortality. By 1960 there were more than 60 government hospitals of different grades, located in all important centres. Rural areas were served by more than 900 dispensaries and dressing stations under medical assistants. There are doctors in private practice in larger towns. There is also a rural midwife service, provided by more than 1,000 women trained in government schools who practise from their homes.

5. Justice.—Civil justice is administered through the high court of justice in Khartoum, province courts and lower courts. There is a body of civil law; but where there is no special enactment, cases are decided according to "justice, equity and good conscience." The criminal code is based upon the Indian penal code and is dispensed in major, minor and magistrate's courts. Complementary to the state courts, there are also local courts, constituted under the native courts ordinance (1932) and the chiefs' courts ordinance (1931), which try a substantial proportion of both civil and criminal cases. Muslims in the north are subject to the Shari'ah, or religious law of Islam, in questions of inheritance, marriage, divorce, family relationship and charitable trusts.



GEORGE RODGER—MAGNUM

SCHOOLBOYS OF THE SOUTHERN SUDAN IN READING CLASS

6. Education.—In the Muslim areas of the north, boys have long been instructed in religious subjects according to traditional methods. During the condominium, and as the outcome of an appeal made by Lord Kitchener in 1898, modern primary education was begun in northern Sudan, and secondary education commenced in 1913 (see *History*, above). Post-secondary schools in arts and science, agriculture, veterinary science, engineering and law were established between 1938 and 1941. In 1947 these schools were brought together within Gordon Memorial college, which was linked in special relationship with the University of London. Education in medicine began in 1923, at the foundation of the Kitchener School of Medicine. In 1951 this school was united with Gordon Memorial college in the University College of Khartoum (which in 1956 became the University of Khartoum). In the

south, missionaries were responsible for education, with the consent and support of the government, until 1957; then the government assumed direct control, as a result of which Arabic was given greater prominence in the curriculum.

The census of 1955–56 disclosed that over 85% of the total population aged five years and above had never attended any school. In the early 1960s there were about 3,000 government schools with 388,000 pupils, and more than 170 nongovernment schools, of all grades, with about 43,000 pupils. There were also 37 intermediate and secondary religious (Muslim) schools, attended by over 6,300 pupils. The *Maakad Ilmi* in the mosque at Omdurman, attended by 1,400 pupils from Muslim areas, provides theological education and trains the ulema. The principal centre for the training of intermediate school teachers is the Institute of Education at Bakht ar Ruda, near Ed Dueim for men, and at Omdurman for women.

The University of Khartoum, with its faculties of agriculture, arts, science, economics, engineering, law, medicine and veterinary science, has an academic staff of about 150 and about 1,800 students. English is the medium of instruction in the university and the secondary schools, but it is being replaced by Arabic in the secondary schools. A branch of Cairo university in Khartoum provides facilities for evening classes in arts, law and commerce, the medium of instruction being Arabic. The Khartoum Technical institute was founded in 1950.

7. Defense.—The Sudan army had its origin in the Sudan defense force, founded in 1925. The total rifle strength in 1955, when the British officers left, was between 4,000 and 5,000, and it has since remained at about 5,000. Commands are located in El Obeid (Central), Gedaref (Eastern), El Fasher (Western), Shendi (Northern) and Juba (Southern). There are also technical corps based on Omdurman and Khartoum, and artillery is based on Atbara.

VII. THE ECONOMY

Sudan's economy depends almost wholly upon agricultural and pastoral products. Minerals and sources of power are lacking, and industrial development was in its infancy in the early 1960s. Commercial development has been mainly in the north and east.

1. Agriculture and Pastoral Production.—In the west and south both crops and livestock are produced mainly for subsistence, and little is sold. In the north and east surpluses are available and, thanks to better communications, can be moved to towns or for export. The only capitalized agricultural production, applying modern technical knowledge, is from the irrigated areas close to the Nile in the northeast and some areas under rain cultivation near Gedaref.

The chief food crops are the common millet (durra) and bulrush millet (*dukhn*). The latter flourishes on the sands of the dunes. Millet cultivation extends across the country in an east-west zone —the central rain lands—between latitude 10½° N. and 14° N., but is also grown by Nilotic peoples near their permanent villages. In the far south, finger millet (eleusine) and maize (corn) are preferred. The area under durra and *dukhn* has increased in northeastern Sudan concomitantly with the population in the 20th century; this has been true especially in the clay plain (since World War II) as a result of the provision of new roads and perennial water supplies from *hawafir* or excavated reservoirs. Near Gedaref, the introduction of mechanized production has proved profitable. Durra also enters into the rotation of crops in irrigated areas. Yields of millets fluctuate mainly according to rainfall, and in years of poor harvests the supplies to towns and deficiency regions cause the government anxiety; but the cultivator makes use of underground storage in pits and in lean years draws upon his reserves. In good years some grain is exported. Wheat is grown as a winter crop along the Nile in Northern province.

Sesame and peanuts are also important; both are normally exported. A bean known as *lubia affin* (*Dolichos lablab*), a clover (*berseem higazi*) and alfalfa are grown for fodder on irrigated land. Tobacco is widely grown in Darfur and is the only crop which can justify the cost of transport from that remote part of the country. Medium- and low-quality dates are a traditional product of the Nile valley north of Shendi and find a market throughout northern Sudan. Citrus fruits, especially limes and grapefruit, are produced at Kassala, in irrigated gardens along the Nile and near some towns in both north and south. Market gardening is also established in some areas but does not produce an adequate supply of fresh vegetables. In southern Sudan tobacco, coffee, sugar, rice and the oil palm are all cultivated experimentally, but commercial output has been negligible.

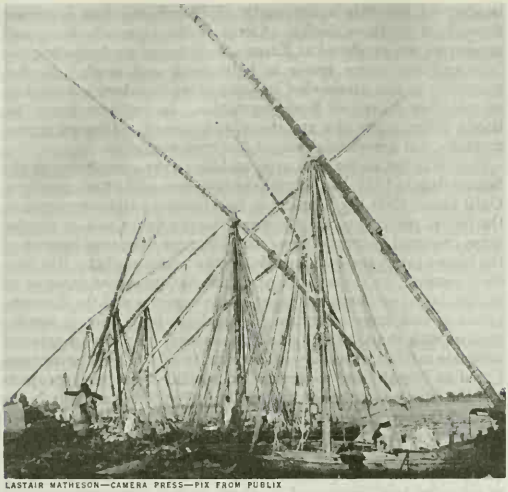
Cotton is the major cash crop and, since the completion of the Sennar dam in 1925, has become the mainstay of Sudan's economy. Until about 1960 the long-staple Sakellarides varieties, grown in the Gezira project, on the Gash and Baraka deltas and on smaller areas (watered by pump) along the Nile, were dominant because of the higher prices they commanded in the world market. But lessened demand for these types of cotton has prompted increase in production of medium- and short-staple varieties and the diversification of cash crops generally. Long-staple production is supplemented with short-staple (chiefly American) varieties in areas of adequate rainfall, especially in the vicinity of the Nuba mountains (Jibal an Nubah) and southwest Equatoria. New pump projects and the Manaqil extension to the Gezira project have greatly increased the area under cotton. In the early 1960s this exceeded 900,000 feddans (1 feddan = 1.47 ac.), of which less than 200,000 were unirrigated, mainly in Kordofan and Equatoria provinces. The production of long-staple cotton, the almost exclusive product of irrigated land, amounted to about 98,000 metric tons and of short-staple (rain-grown) cotton to about 11,000 metric tons. The Sudan Gezira board (established in 1911 and nationalized in 1950) is the country's main cotton-producing body.

Livestock.—Pastoral activities extend throughout the country, except on the ironstone plateau of the southwest, which is infested by the tsetse fly, and on the desert of the north and northwest. There are about 2,000,000 camels, owned by Beja, Arab and nomadic tribes of the north and west, which graze over the desert margin scrub. About 6,900,000 cattle are bred mainly in the central rain lands, especially by the Baggara tribes of southwest Kordofan and southern Darfur. Cattle also form the chief wealth of the Nilotic tribes. There are an estimated 7,000,000 sheep and 5,700,000 goats, associated both with camel and cattle rearing in the north. Goats are also numerous in the towns as milk suppliers. Camels and small numbers of cattle are exported to Egypt, and sheep to Saudi Arabia; there is also sale for slaughter to towns. A corned-beef factory established at Kosti in 1950 was forced to close for lack of regular and adequate supply of animals despite the improvement of drove roads from the west and the possibility of transporting animals by air. Since 1945 the preparation and export of hides has been encouraged, with some success.

The growth of population and the settlement in new areas during the condominium, together with control of animal diseases by veterinary services, have greatly increased the use of cultivable land and natural grazing. In some areas, land is losing its fertility, and the best fodder plants are disappearing. The provision of new water supplies (wells and *hawafir*) has been the chief conservation measure, rendering new areas of pasture and cultivable land available for use. Further methods of control are desirable, and have been under study by the land use department.

Fisheries.—The Nile system, and especially the White Nile above Kosti and the Bahr el Jebel, is an important source of fish, particularly Nile perch, for riverine peoples. Small quantities are sold in the towns, and a little is dried and salted.

2. Forestry.—The Sudan is one of the world's chief suppliers of gum arabic, chiefly *hashab* gum, which is collected from *Acacia senegal* in the central rain land zone from Darfur to Kassala. It is bought by exporting companies at government auctions, held at many places, of which El Obeid is the most important. Timber is produced mainly in the south, apart from the dense hardwood known as sunt (*Acacia nilotica*), exploited in the Blue Nile valley above Sinjah. There are groups of sawmills near Wau, Loka and Katire, producing mahogany and other hard and soft timbers. The forests are the local sources of firewood and charcoal, and cutting is supervised; but severe inroads have been made into the forest resources of northern Sudan in the 20th century. The growing use



LASTAIR MATHESON—CAMERA PRESS—PIX FROM PUBLIX

LATEEN-RIGGED FELUCAS, LOCAL CRAFTS USED TO CARRY FIREWOOD UP THE NILE

of oil fuels is proving to be an ameliorative.

3. Minerals.—No large mineral deposits have been discovered, but gold, iron ore, manganese, limestone and mica are exploited on a small scale. Salterns at Port Sudan supply unrefined salt for the country's needs and for export.

4. Industry and Power.—Oil-milling and soap-making are established in some northern towns; Khartoum and Port Sudan have factories of moderate size. There is a cement works at Atbara. Steel and aluminum are fabricated into utensils in Khartoum and Wad Medani. A modern brewery operates in Khartoum North and there are two distilleries in the Three Towns. Factories for shoes, paints, glass, knitwear, sweets and cigarettes also operate in Khartoum. A cotton-spinning and weaving mill was opened at Nzara, near Yambio, Equatoria, in 1951 as part of a plan for social emergence based upon regional self-sufficiency. Khartoum North and Omdurman each has a cotton-spinning and weaving mill. There are motor-maintenance workshops and tire-retreading works in Khartoum and elsewhere. The railway workshops at Atbara are the largest industrial unit in the country. River steamers and barges are repaired in the dockyards at Khartoum North and Kosti. Local craftsmen make leather goods, clothes, pottery, furniture, household utensils and implements in the bazaars of northern towns.

Small electric-power stations serve domestic and industrial consumers in larger towns. A power station on the Sennar dam has the potential to supply Gezira and Khartoum with 15,000 kw. of electricity.

5. Trade and Finance.—The principal exports of Sudan are cotton and cottonseed, peanuts, gum arabic and sesame. The principal imports are cotton piece goods, petroleum products, machinery, refined sugar, base metals and manufactures, and vehicles. In the early 1960s the chief receiving countries were the United Kingdom, Egypt, India, Pakistan, Italy, France and West Germany. Chief sources of supply were the United Kingdom, India, Pakistan and Egypt.

Indirect sources of revenue include import duties, royalties on exported products, and the profits of trading concerns such as the cotton projects and the sugar monopoly. Direct taxes include those on dates, animals, land, crops, houses and business profits. There is also a poll tax, and tribute is levied upon certain tribes. The monetary unit is the Sudanese pound (£S), divided into 100 piastres and 1,000 milliemes, which was equal to £1 0*s.* 6*d.* sterling in the early 1960s.

Revenue exceeded £S60,000,000 and ordinary expenditure was about £S50,000,000. From 1945 additional development budgets were prepared for financing five-year plans. Subsequently, annual development budgets were approved for projects including the Er Roseires dam, the Khashm al Qirbah (Atbara river) scheme the Manaqil extension to the Gezira scheme, the Sennar hydroelectric scheme and the Guneid (Junayd) sugar factory. A ten-year development plan was prepared for 1961–71.

By about 1961 the external public debt was about £S3,000,000. Loans or credits have been obtained from the International bank, International Monetary fund, the United Kingdom, the Federal Republic of Germany, Yugoslavia, and Kuwait (for railway development). A credit under the American aid agreement (1952–59) was principally for the purpose of establishing a textile industry in northern Sudan and for railway extensions and improvements. Under an agreement with the U.S.S.R. (1961) financial and technical aid was provided for the construction of food-processing factories.

Banking.—Until 1960 banking was provided by foreign concerns, led by Barclays Bank D.C.O. (Dominion, Colonial and Overseas) and the National Bank of Egypt. The latter was acquired by the state by agreement with the Egyptian government, renamed the Bank of Sudan and reorganized as the central bank for the country. The state-owned Agricultural Bank of Sudan provides agricultural credit. Foreign banks, apart from Barclays, include the Arab, Misr, Ottoman, Crédit Lyonnais and the State Bank of Ethiopia. The Industrial Bank of Sudan was officially inaugurated in 1962.

6. Transport and Communications.—Communications in the north are provided by the Sudan railways. The line from Wadi Halfa to Khartoum was built in 1897–99; the Atbara-Port Sudan branch was opened in 1904–06; and the main line from Khartoum (via Sennar and Kosti) to El Obeid was extended in 1906–11. Except for the Kassala-Hayya (Haiya) loop, completed in 1929, railway building ceased until 1950. A branch from Sennar to Roseires was completed in 1956, and a westward extension from Ar Rahad via Abu Zabad and Ed Da'ein (Ad Du'ayn) to Nyala in 1959. A branch line from Babanusah (on the latter extension) southwestward across the Bahr al 'Arab to Wau was completed in 1962. This railway system facilitates the economic advance of the north by providing the means for bulk movement of cotton, oilseeds, hides and gum arabic, as well as the distribution of imported goods. The extensions westward are intended to stimulate commerce in Darfur and Bahr el Ghazal.

In the south the White Nile and the Bahr el Jebel, between Kosti and Juba, and the Bahr el Ghazal as far as Meshra' er Req (Mashra' ar Raqq) are navigable throughout the year and have regular passenger and freight services. The Jur river is navigable from its confluence with the Bahr el Ghazal to Wau from July to October, and the Sobat from May till December. From ports on these rivers an all-weather road system serves the provinces of

CAMERA PRESS—PIX FROM PUBLIX

PORT SUDAN ON THE RED SEA, THE ONLY SEAPORT IN THE REPUBLIC OF THE SUDAN

Bahr el Ghazal and Equatoria (on the ironstone plateau); but Upper Nile, on the clay plain, possesses only dry-weather tracks, and porterage is still necessary away from the rivers except between December and May.

There are no roads in northern Sudan; but motorable tracks (with locally improved stretches to ensure continuity over soft, rocky or marshy ground) are used by trucks and estate and heavier cars, except after heavy rainfall between July and September. The soft sand of the dunes greatly impedes motor transport in central and southwestern Kordofan and in eastern Darfur. Trucks take four days for the journey from El Obeid to El Fasher.

Khartoum is an important stopping place on the main Cairo-Cape Town air route, with feeder services to the east (Eritrea and Aden) and to the west. The government-owned Sudan Airways operates internal air services between the Three Towns and most provincial capitals and some other towns, and external services to Egypt, Aden, Jordan, Eritrea, Lebanon, Athens, Rome and London.

Telecommunications.—The telegraph system serves most settled parts of the country. A telephone service exists in the north and east. In the south, and in Darfur, radio transmission is employed instead of land lines. Communications abroad are by means of cable, land line and radio (telegraphy and telephony). The government-controlled Sudan Broadcasting service at Omdurman broadcasts in Arabic and English.

See also references under "Sudan, Republic of the" in the Index volume. (J. H. G. L.)

BIBLIOGRAPHY.—*Geography:* R. A. Hodgkin, *Sudan Geography* (1951); A. F. Broun and R. E. Massey, *Flora of the Sudan* (1929); F. W. Andrews, *Flowering Plants of the Anglo-Egyptian Sudan,* 3 vol. (1950–56).

The People: H. A. MacMichael, *History of the Arabs in the Sudan,* 2 vol. (1922); J. S. Trimingham, *Islam in the Sudan* (1949); D. Forde (ed.) *Ethnographic Survey of Africa,* East Central Africa, Part IV, *Nilotes;* Part VI, *Northern Nilo-Hamites;* Part VII, *Central Nilo-Hamites;* Part IX, *Azande;* this series contains full and annotated bibliographies.

Archaeology: A. J. Arkell *et al., The Pleistocene Fauna of Two Blue Nile Sites* (1951); C. R. Lepsius, *Denkmäler aus Aegypten and Aethiopien,* Ab. vi (1849); Frederic Cailliaud, *Voyage à Méroé, au Fleuve Blanc,* etc. (1826–27); E. A. Wallis Budge, *The Egyptian Sudan,* 2 vol. (1907); J. W. Crowfoot, *The Island of Meroe* (1911); "Some Red Sea Ports in the Anglo-Egyptian Sudan," *Geogr. J.* (May 1911); A. J. Arkell, "Meroe and India," *Aspects of Archaeology,* pp. 32–38 (1951); A. Llewellyn, *Report on a Mining Commission in the Egyptian Sudan* (1903); *The Wellcome Excavations in the Sudan,* vol. i and ii by F. Addison, *Jebel Moya* (1949), and vol. iii by O. G. S. Crawford and F. Addison, *Abu Geili and Saqadi and Dar el Mek* (1951); O. G. S. Crawford, *The Fung Kingdom of Sennar* (1951); G. A. Wainwright, "Cosmas and the Gold Trade of Fazogli," *Man,* vol. xlii, p. 30 (1942). *See* also works cited under NUBIA.

History: R. L. Hill, *A Biographical Dictionary of the Anglo-Egyptian Sudan* (1951), and *Egypt in the Sudan, 1820–1881* (1959); H. A. MacMichael, *The Sudan* (1954); K. M. Barbour, *The Republic of the Sudan* (1961); Sudan Notes and Records (1918–); P. L. Shinnie, *Medieval Nubia,* Sudan Antiquities Service Museum Pamphlet, 2 (1954); A. J. Arkell, *A History of the Sudan: From the Earliest Times to 1821* (1961), and *Kush* (1953 *et seq.*); G. Douin, *Histoire du règne du Khédive Ismaïl,* 3 vol. (1936–40); *Histoire du Soudan égyptien,* vol. i (1944); M. F. Shukry, *The Khedive Ismail and Slavery in the Sudan, 1863–1879* (1937); M. Sabry, *Le Soudan égyptien, 1821–98* (1947); F. R. Wingate, *Mahdiism and the Egyptian Soudan* (1891); B. M. Allen, *Gordon and the Sudan* (1931); A. B. Theobald, *The Mahdiya* (1951); M. E. T. Shibeika, *The Independent Sudan* (1959), *British Policy in the Sudan, 1882–1902* (1952); P. M. Holt, *The Mahdist State in the Sudan, 1881–1898* (1958), *A Modern History of the Sudan* (1961); R. A. Gray, *A History of the Southern Sudan, 1839–1889* (1961); R. O. Collins, *The Southern Sudan, 1883–1898* (1962); M. Abbas, *The Sudan Question: the Dispute Over the Anglo-Egyptian Condominium, 1884–1951* (1952); J. S. R. Duncan, *The Sudan: a Record of Achievement* (1952); K. D. D. Henderson (ed.), *The Making of the Modern Sudan: the Life and Letters of Sir Douglas Newbold* (1953).

Administration and Economy: Annual Reports on the Administration, Finances and Condition of the Sudan (1899–1956); V. L. Griffiths, *An Experiment in Education* (1953); F. Saad al-Din Fawzi, *The Labour Movement in the Sudan, 1946–1955* (1957); H. C. Squires, *The Sudan Medical Service* (1958); J. D. Tothill (ed.), *Agriculture in the Sudan* (1948); A. Gaitskell, *Gezira: a Story of Development in the Sudan* (1959); Central Office of Information, The Republic of Khartoum, *Sudan Almanac* (annually).

Current history and statistics are summarized annually in *Britannica Book of the Year.*

SUDANIC LANGUAGES. The term Sudanic has been used to designate many African languages (*q.v.*) spoken from Ethiopia and Eritrea in the east to Nigeria in the west. Totally unrelated languages have been included in the various groupings listed as Sudanic, either on geographic or other nonlinguistic grounds.

A more satisfactory classification is that of Joseph H. Greenberg, which is based upon strictly linguistic relationships.

Chari-Nile (or Macro-Sudanic) includes Central Sudanic, Eastern Sudanic, Berta and Kunama. Even this broad grouping is tentative. Central Sudanic includes the Moru-Madi group (in Sudan, northeast Republic of the Congo and northwest Uganda) and the Bongo-Baka-Bagirmi group (largely in Cameroun, Chad and the Central African Republic but extending east into Sudan). Eastern Sudanic (on the upper Nile, largely in Sudan, Uganda and Kenya but extending into Tanganyika) has seven branches: Nubian, Beir-Didinga, Barea, Tabi, Merarit, Dagu and Southern. Old Nubian is known from medieval records ranging from the 8th to the 11th centuries. The Southern branch includes (1) Nilotic proper or West Nilotic, including Shilluk, Nuer, Dinka, Lango and Acholi; (2) East Nilotic, including Bari, Masai, Turkana and Lotuko; and (3) South Nilotic (Nandi-Suk).

BIBLIOGRAPHY.—A. N. Tucker, *The Eastern Sudanic Languages* (1940); Joseph H. Greenberg, *Studies in African Linguistic Classification* (1955). For the older type of classification *see* A. Meillet and M. Cohen, *Les Langues du Monde,* rev. ed. (1952). (C. T. HE.)

SUDBURY, SIMON OF (sometimes called SIMON THEOBALD or TYBALD) (d. 1381), archbishop of Canterbury, was probably born at Sudbury, in Suffolk, Eng. He was a doctor of civil law. Pope Innocent VI, who made him a papal chaplain, sent him in 1356 on a mission to Edward III of England, and in 1361 appointed him bishop of London. He was translated to Canterbury in 1375.

He crowned Richard II in July 1377. John Wycliffe appeared before him at Lambeth in 1378. Chancellor of England from 1380, Sudbury was regarded by the peasants who were in revolt as one of the authors of their distress. The Kentish rebels damaged his property at Canterbury and Lambeth, and, dragging him from the Tower of London, beheaded him on Tower hill on June 14, 1381.

SUDBURY, a city in northern Ontario, Can., 264 mi. N.W. of Toronto and the seat of Sudbury district. Pop. (1961) 80,120. In modern times the chief retail and wholesale centre for northern Ontario, Sudbury began with the arbitrary location of a station on the new Canadian Pacific railway. Copper-bearing ores containing large quantities of nickel were unearthed in 1883 during construction of the railway, but it was not until the metallurgy of copper-nickel separation was solved that development started and the major growth of the city began.

By the end of 1945 over $1,000,000,000 worth of nickel and $500,000,000 worth of copper had been produced. In the early 1960s Sudbury still produced about 90% of the world's nickel and almost all of Ontario's copper. The International Nickel company, with a huge concentrator and smelter at Copper Cliff, was at that time the world's largest producer of nickel; it operated five underground mines, two open pits and a quartz quarry. Falconbridge Nickel Mines Ltd. operated a nickel-copper mine, concentrator and smelter at Falconbridge, about 12 mi. N.E. of Sudbury. More than three-quarters of those gainfully employed worked in mining and metallurgy, earning a higher average wage than the workers of any other Canadian city.

Sudbury was incorporated as a town in 1893 and as a city in 1930. Modern Sudbury is a well-built city with more than 20 mi. of paved streets. It has an extensive business section with many stores, hotels and large office buildings, as well as many fine churches and schools.

The University of Sudbury, a bilingual Roman Catholic university, was established in 1957. (F. A. CK.)

SUDBURY, a municipal borough of West Suffolk, Eng., in the parliamentary division of Sudbury and Woodbridge. Pop. (1961) 6,643. Area 2.9 sq.mi. It lies on the Stour river (navigable up to the town), 22 mi. W. of Ipswich by road. One of the medieval wool towns, it has many beautiful half-timbered build-

ings and three Perpendicular churches: All Saints, with a fine oaken pulpit dating from 1490; St. Peter's, with a carved nave roof; and St. Gregory's, once collegiate, which contains a rich, spire-shaped font cover of wood, gilded and painted.

The grammar school was founded by William Wood in 1491 (rebuilt 1857).

Thomas Gainsborough (1727–88) was born there, and an earlier native was Simon Tybald, archbishop of Canterbury, who was murdered in Wat Tyler's rebellion and whose head is preserved in St. Gregory's.

Before the conquest the borough was owned by the mother of Earl Morcar of the Northumbrians, from whom it was taken by William I. It was alienated from the crown to an ancestor of Gilbert de Clare, earl of Gloucester, who in 1271 gave the burgesses their first charter confirming to them all their ancient liberties and customs.

Sudbury was incorporated in 1554.

SUDD, a huge, peaty mass of decaying vegetation forming obstructions on the upper Nile. It is composed chiefly of a grass, *Vossia procera,* with *Saccharum spontaneum* and papyrus, which cover a large area of the Ghazal swamps. Loosened by storms, these plants reach the main channel near the Sobat river junction, lodge on some obstruction and form a dam, sometimes 25 mi. in length and nearly 20 ft. below the surface. Sudds are compressed by the current so that in parts they can support an elephant. At length the pressure of the water forms a side channel or causes the sudd to burst. *See* Nile: *White Nile.*

SUDERMANN, HERMANN (1857–1928), German writer, was born at Matziken, East Prussia, on Sept. 30, 1857, of a Mennonite family. His father owned a small brewery, but owing to a financial crisis was obliged to apprentice his son, at 14, to a chemist. Young Sudermann was able, however, to study at Tilsit and afterward at Königsberg university. He became a tutor in Berlin, and worked as a journalist on the *Deutsches Reichsblatt* (1881–82), afterward turning to novel writing. His books *Im Zwielicht* (1886), *Frau Sorge* (1887, Eng. trans. 1917), *Geschwister* (1888) and *Der Katzensteg* (1890) revealed neither beauty nor emotional power, but invariably showed keen observation, a vivid touch and dramatic technique. The tale is his chief concern, and he shows a masterly control of tension, as in *Iolanthes Hochzeit* (1892); *Es War* (1894) and *Das Hohe Lied* (1908) showed a falling off, but the old mastery reappeared in *Litauische Geschichten* (1917).

More instantly popular, and later more bitterly decried, were plays, of which *Die Ehre* (1880, Eng. trans. 1915) brought him immediate fame, and *Heimat* (1893), better known as *Magda,* made him known throughout Europe. Part of the great success of *Magda* was no doubt due to the fact that the title role was interpreted by Helena Modjeska, Sarah Bernhardt, Eleonora Duse and Mrs. Patrick Campbell. His dramas, *Sodoms Ende* (1891), *Johannisfeuer* (1900), *Es lebe das Leben* (1902), *Der Sturmgeselle Sokrates* (1903), *Stein unter Steinen* (1905) and numerous others ending with *Der Hasenfellhändler* (1925) were uniformly successful on the stage. During World War I he wrote the cycle of three plays collectively entitled *Die Entogötterte Welt.* His last novels were *Der tolle Professor* (1926) and *Purzelchen* (1928). He belonged to the Naturalist movement of the last quarter of the 19th century, and was not sympathetic to the methods and outlook of postwar literature. He died at Berlin-Grunewald on Nov. 21, 1928.

Bibliography.—Collections of Sudermann's dramatic works (Stuttgart, 1923) and narrative works (Stuttgart and Berlin, 1919; Stuttgart, 1930); H. Sudermann, *Das Bilderbuch meiner Jugend* (autobiography) (Stuttgart and Berlin, 1922; Eng. trans., London and New York, 1923); H. Schoen, *Hermann Sudermann, Poète dramatique et romancier* (Paris, 1905); I. Axelrod, *Sudermann* (Stuttgart, 1907); K. Busse, *Hermann Sudermann, sein Werk und sein Wesen* (Stuttgart and Berlin, 1927); I. Leux, *Hermann Sudermann* (Leipzig, 1931).

SUE, EUGÈNE [Joseph Marie] (1804–1857), French novelist, was born in Paris on Jan. 20, 1804. He was the son of a distinguished surgeon in Napoleon's army, and is said to have had the empress Josephine for godmother. Sue himself acted as surgeon both in the Spanish campaign undertaken by France in 1823 and at the battle of Navarino (1828). In 1829 his father left him a fortune, and he settled in Paris. His naval experiences supplied much of the materials of his first novels. In the quasi-historical style he wrote *Jean Cavalier, ou Les Fanatiques des Cevennes,* 4 vol. (1840) and *Latréaumont,* 2 vol. (1837). He was strongly affected by the Socialist ideas of the day, and these prompted his most famous works: *Les Mystères de Paris,* 10 vol. (1842–1843) and *Le Juif errant,* 10 vol. (1844–1845), which were among the most popular specimens of the *romanfeuilleton.* He followed these up with some singular and not very edifying books: *Les Sept péchés capitaux,* 16 vol. (1847–1849), which contained stories to illustrate each sin, *Les Mystères du peuple* (1849–1856), which was suppressed by the censor in 1857.

Some of his books, among them the *Juif errant* and the *Mystères de Paris,* were dramatized. After the revolution of 1848 he sat for Paris (the Seine) in the Assembly, and was exiled after his protest against the *coup d'état* of Dec. 2, 1851. His later works were inferior. Sue died at Annecy (Savoy) on Aug. 3, 1857.

SUEBI or Suevi, the name used by Latin writers for certain Germanic tribes; in English Sueves, in Spanish Suevos. For the Suebi in central Europe *see* Germany: *History;* Germanic Peoples; Alamanni; Marcomanni. Their name was preserved in that of Swabia (*q.v.*).

When the Huns disturbed the Alani and the Goths *c.* A.D. 370, they set in motion westward two hordes of Vandals together with the Alani and the Sueves. These reached Mainz in 406, crossed the Rhine and entered Spain in 409. After the campaigns of the Visigothic king Wallia (415–419) in Spain and the defeat of a Suevic army by the Vandals near Mérida in 428 or 429, the Sueves were mainly confined to Galicia. The Roman generals might then have crushed them, but Aëtius could not spare enough troops; and by 447 the Sueves under their king Rechila had spread over Lusitania and Baetica. Rechiari, who came to the throne as a Catholic in 448 (the Sueves had arrived in Spain as pagans), attacked the Romans in Tarraconensis but was attacked in turn and overthrown by the Visigoths under Theoderic II in 456. This was described by the contemporary Idatius, bishop of Aquae Flaviae (Chaves), as the destruction and end of the Suevic kingdom, but remnants of the Sueves continued the fight under Maldras (456–460) and rival kings. Remismund, who became sole king of the Sueves in 464, converted his people to Arianism during a temporary *rapprochement* with the Visigothic king Euric. For the 90 years following 468 the history of the Sueves is obscure. King Theodemir (558–570), perhaps identifiable with the Chararic mentioned by Gregory of Tours, was converted to Catholicism by Martin, bishop of Braga (d. 579). The reigns of his two successors, however, were disturbed, and *c.* 585 the kingdom was annexed to the Visigothic state by Leovigild.

Bibliography.—J. P. Migne, *Patrologia Latina,* lxxiv and lxxxiii (Paris, 1850), reprints the *Chronicon* of Idatius and the *Historia de Regibus Gothorum, Wandalorum et Suevorum* of Isidore of Seville; for a Spanish translation with extensive commentary *see* M. Macias y Garcia, *Galicia y el Reino de los Suevos* (Orense, 1921). *See* also C. Barlow (ed.), *Martini Episcopi Bracarensis opera omnia* (New Haven and London, 1950), for the long homily *De Correctione rusticorum,* which instances the survival of pagan practices among the 6th-century Sueves. (M. Dy.)

SUECA, a town of eastern Spain, in Valencia. Pop. (1960), 20,612 (mun.). Sueca is separated from the Mediterranean (7 mi. E.) by the Sierra de Cullera.

It is a modern town, although many of the houses have the flat roofs, view-turrets (*miradores*) and horseshoe arches characteristic of Moorish architecture. Sueca has a thriving trade in grain and fruit from the Júcar valley, which is irrigated by waterways created by the Moors.

SUESS, EDUARD (1831–1914), Austrian geologist, was born at London on Aug. 20, 1831. He was educated at the universities of Prague and Vienna, specializing in invertebrate paleontology. In 1852 he became assistant in the Hofmuseum, Vienna, and in 1857 extraordinary professor of geology at the University of Vienna, becoming ordinary (or full) professor in 1867, a position he held until 1901.

While connected with the Hofmuseum, Suess published papers on the anatomy and classification of brachipods, established the

classification of ammonites, and described fossils belonging to several other groups.

After 1857 he turned to geology proper. In 1875 he published *Die Enstehung der Alpen,* a small book which denied vertical uplift of mountains and continents, limited the role of vulcanism in mountain building, and declared that the Alps and related mountains had been formed by thrusting movements which crumpled and broke outer portions of the earth's crust. He also emphasized the existence of unyielding continental areas composed of very old mountain masses, now generally called shields.

Between 1885 and 1907 Suess published *Das Antlitz der Erde,* an interpretation of the earth's surface in terms of ancient changes in continents and seas. He also dealt with fluctuations of ancient seas, the rise of magmas, and other subjects related to his main thesis. *Das Antlitz der Erde* achieved great influence and was translated into Italian and French, and into English as *The Face of the Earth.*

Suess became a member of the *Landtag* of Lower Austria in 1869 and in 1873 entered the lower house of the *Reichsrath,* where for more than 30 years he was Liberal deputy from Vienna. He died at Vienna on April 26, 1914. (M. A. F.; C. L. Fe.)

SUETONIUS TRANQUILLUS, GAIUS (b. *c.* A.D. 70, d. after A.D. 122), Roman biographer and antiquarian, is best known for his *De vita Caesarum* (*Lives of the Caesars*). Probably born in Rome, he was of equestrian family, a friend and protégé of the younger Pliny. He seems to have toyed with the law as a career but abandoned it; and having secured through his friend a military tribunate, he relinquished that also. It is possible, though not certain, that he accompanied Pliny on his mission to Bithynia (*c.* 110–113). After Pliny's death Suetonius found another patron, Septicius Clarus (to whom Pliny had dedicated his letters, and to whom Suetonius himself was later to dedicate the *De vita Caesarum*); upon the accession of Hadrian, Septicius became praetorian prefect, and, either then or a little before, Suetonius entered the imperial service. He held, probably simultaneously, the posts *a studiis* and *a bibliothecis,* controlling the Roman libraries, keeping archives and advising the emperor on cultural matters. In 121 Hadrian began his first grand tour of the empire, and probably at this time Suetonius was promoted to be *ab epistulis,* secretary of the imperial correspondence, in which capacity he traveled (along with his patron) in the emperor's suite. Both men were dismissed for neglect of court formality; the incident is recorded in the context of Hadrian's visit to Britain in 122, but may have taken place somewhat later. Presumably Suetonius devoted himself thenceforward to literary pursuits. The date of his death is unknown.

Works.—A list of writings of Suetonius is given by the "Suda" lexicon. Most of them were antiquarian; their titles indicate that they dealt with Greek pastimes, the history of Roman spectacles and shows, the Roman year, critical signs in books, the terminology of clothing, oaths and imprecations and their origin, Roman customs and a defense of Cicero's *De re publica* (probably of its style, not its politics). Other sources refer to similar works on famous courtesans, physical defects, the growth of the civil service and on kings. An encyclopaedia like Pliny's *Natural History,* called *Prata* ("Meadows"), was attributed to him and often quoted in late antiquity; some of the works mentioned above may have been subsections of this.

However, Suetonius' main works were biographical: the *De vita Caesarum* and the *De viris illustribus.* The latter was a collection of short biographies of celebrated literary figures of Rome, divided into books on poets, orators, historians, teachers (*grammatici* and *rhetores*) and perhaps philosophers. There survive the whole of the *De grammaticis* and the preface and first five lives of the *De rhetoribus;* in addition, the lives attached to manuscripts of several famous poets derive from Suetonius' book, especially those of Horace and Lucan. The life of Terence that begins Aelius Donatus' commentary on Terence comes from the same source, and so do the facts (but not the padding) in Donatus' life of Virgil.

Since, in addition, the notices about writers inserted by St. Jerome in his Latin version of Eusebius' chronicle come from Suetonius, it is fair to say that almost all of what little is known about the lives of Rome's eminent authors is owed to him.

Lives of the Caesars.—Nevertheless, it is for the *Lives of the Caesars,* from Julius Caesar to Domitian, that Suetonius is famous. Upon them (along with the works of Tacitus) is largely based that vivid picture of Roman society and its leaders, morally and politically decadent, which dominated historical thought until modern times when the discovery of nonliterary sources of evidence did something to modify it. Moreover, under the influence of the *Lives,* much of later Roman historical writing was directed, to its detriment, into the biographical channel. They are organized not chronologically but by topics (the family background of the emperor, his career before accession, his public actions, his private life, appearance and personality and his death), a procedure more antiquarian than historical; they are full of scandalous gossip; and on the empire in general—its growth, administration and defense—they are largely silent. However, Suetonius exhibits compensating merits, quite apart from the fact that he was not aiming to write history and that the gossip and the lampoons are, in their way, valuable historical evidence. He is free from the bias of the senatorial class which distorts much Roman historical writing, and he is commendably willing to quote documents verbatim.

The earlier *Lives* down to Nero (especially the first two) are much the fullest; for that period a mass of sources was available, already collected and published, whereas for the Flavian period original research had to be done.

It is surprising that Suetonius, who, in his position *a studiis,* was ideally placed for the necessary research, should have failed just in this (he quotes no original sources in his account of the Flavians); and the opinion is held by some that he made no attempt to investigate any sources himself but took everything from existing collections. However, it may be that as an antiquarian he was drawn to the documentary byways of the earlier age and had less interest in the Flavian emperors, and it seems likely that posterity owes to his brief tenure of the post *a studiis* such immensely valuable pieces of historical material as imperial wills and private letters.

Suetonius was not incapable of critical analysis of his sources, but he exercised it only on minutiae. He was trying to evaluate the personalities of individuals in terms of the standards and traditional interests of Roman public life and society; and he was ready, like Plutarch, to press into service the "characteristic anecdote," without exhaustive inquiry into its authenticity. His pen pictures of the habits and appearance of the emperors are invaluable; but the "rubric" method of composition made it impossible for him to analyze the development of their different characters.

The *Lives* are exciting to read. Suetonius avoids the excesses of archaism and preciosity to which many authors of his day were prone. He writes with firmness and brevity; he loves the *mot juste,* the Greek word for a Greek thing, the correct technicality (this was the direction in which his interests lay, as can be seen from the titles of his lost works); and his use of this precise vocabulary enhances his pictorial vividness.

Above all, he is unrhetorical and unpretentious, not shrinking from bald lists of facts where they are appropriate, but capable of molding complex events into lucid, well-designed periodic sentences.

Bibliography.—*Editions:* For the whole extant works, *see* K. L. Roth (ed.), in the "Teubner Series" (1858); for the *Lives of the Caesars,* see M. Ihm (ed.), in the "Teubner Series" (1907; minor ed., 1923); J. C. Rolfe (ed.), in the "Loeb Series," with Eng. trans., 2 vol. (1914); H. Ailloud (ed.), in the "Budé Series," with French trans. (1931–32). All the minor works were edited by A. Reifferscheid (1860); the *De grammaticis* and *De rhetoribus* by R. P. Robinson (1925); the *De poetis* by A. Rostagni (1956); *Index verborum* by A. A. Howard and C. N. Jackson (1922).

English Translations: The *Lives of the Caesars* were translated by Philemon Holland (1606; often reprinted) and R. Graves (1957).

General: On Suetonius' life, works and place in the history of ancient biography *see* A. Macé, *Essai sur Suétone* (1900); G. Funaioli in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft,* vol. iv, 2nd series (1931); W. Steidle, "Sueton und die antike Biographie," in *Zetemata,* vol. i (1951); R. Syme, *Tacitus,* pp. 501–502, 778–782 (1958). (J. A. Cr.)

SUEZ, a port of Egypt on the Red sea and southern terminus of the Suez canal. Pop. (1957) 152,300. Suez is supplied with water by the fresh-water canal, which starts from the Nile at Cairo and was opened in 1863. Before this, water had to be brought from "the Wells of Moses," a small oasis 3 mi. distant. About 2 mi. S. of the town are the quays constructed on the canal and connected with the town by an embankment. On one of the quays is a statue to Thomas Waghorn, the organizer of the "overland route" to India.

The port is built on land reclaimed from the sea. The accommodation provided includes a dry dock 410 ft. long, 100 ft. broad and nearly 36 ft. deep. There are separate basins for warships and merchant ships, and in the roadstead at the mouth of the canal is ample room for shipping. Suez is a quarantine station for pilgrims from Mecca.

In the 7th century a town called Kolzum stood, on a site adjacent to that of Suez, at the southern end of the canal which then joined the Red sea to the Nile. On the Ottoman conquest of Egypt in the 16th century Suez became a naval as well as a trading station, and there fleets were equipped which for a time disputed the mastery of the Indian ocean with the Portuguese. According to Niebuhr, in the 18th century a fleet of nearly 20 vessels sailed yearly from Suez to Jidda, the port of Mecca and the place of correspondence with India. The overland mail route from England to India by way of Suez was opened in 1837. The regular Peninsular and Oriental steamer service began a few years later, and in 1857 a railway was opened from Cairo through the desert. The present railway follows the canal from Suez to Ismailia and Zagazig, from which branches diverge to Cairo and Alexandria.

SUEZ CANAL, an artificial waterway about 100 mi. long, connecting Port Said on the Mediterranean sea with Suez on the Red sea, thus enabling shipping between Europe and the east to avoid the long sea passage by the Cape of Good Hope.

The Isthmus of Suez is a principal nodal point in world communications. It joins the African to the Asian continent. It separates the Mediterranean from the sea routes that lead most directly to the far east. At its narrowest point it is barely 75 mi. across.

The Earliest Canal in History.—Plans to traverse Egypt with a navigable waterway, or at any rate to connect lower Egypt with the Red sea, go back to very ancient times. The need in those distant days was to link the important commercial centres of the Nile valley with the Red sea trade routes. The Nile itself assured exit at need to the Mediterranean. And so it is that the earliest canals of which history has mention were constructed to link the Nile valley to the Red sea and not to pierce the narrow neck of land which separated the latter from the Mediterranean. There has been much argument among scholars about the dates and exact alignments of these earliest works.

The available evidence is meagre enough but it seems to be established that, perhaps as early as 2000 B.C., a canal linked the Pelusiac branch of the Nile, via the Wadi Tumilat, with the Bitter lakes, whence another channel was dug to the Red sea. Following the vicissitudes of Egyptian political history this canal was sometimes allowed to fall into disuse, and again, depending on the vigour of individual rulers, it was re-excavated to serve the southward-looking commerce of the day. After the Persian conquest Darius the Great is credited with the restoration and enlargement of the canal of the Pharaohs. The Ptolemies, followed by the Romans and later still by the Arabs, are known to have restored the waterway till finally trade between Europe and the east passed to other routes and into the hands of other people.

Maritime Canal Projects.—The Cape route to India and the east was discovered in 1498. This gave to the countries on Europe's western seaboard an immense advantage in the competition for Indian trade. France, however, looks out not only on to the Atlantic but on to the Mediterranean and this, combined with its growing political and economic rivalry with England, explains why the thoughts of Frenchmen turned early to the possibility of piercing the Isthmus of Suez with a navigable waterway. This possibility was often debated in French writings during the 17th and 18th centuries, the discussions all being coloured by the then prevalent misconception that an important difference in levels existed between the two seas that the contemplated channel was intended to unite. It was against this background that, in the struggle against England, Napoleon was sent by the Directory (1798) to occupy Egypt. With him went a distinguished French engineer, Charles le Père, with orders to study and report on the possibility of constructing a sea-to-sea canal across the Isthmus of Suez. Le Père, labouring under the still prevalent misconception that the level of the Red sea was about 30 ft. higher than the level of the Mediterranean, reported against a canal to join the two seas and advocated instead the reopening of the old canal of the Pharaohs. It was left to M. A. Linant de Bellefonds, another notable French engineer, to prove (1853) that the difference in levels between the two seas was unimportant and so to remove what was until then regarded as an almost insuperable difficulty in the way of excavating a navigable channel to cut the Suez isthmus.

De Lesseps.—The stage was now set for the arrival on the scene of a French personality of the first order. Born in 1805, Ferdinand de Lesseps (*q.v.*), following in his father's footsteps, entered the French diplomatic service. At the age of 27 he was appointed vice-consul at Alexandria, where later (1835) he was promoted consul general. From the five years he spent in Egypt at this time may be dated the first stirrings of the great ambition which was presently to dominate his life. He read Le Père's reports with fascinated interest while he also enjoyed the friendship of Linant de Bellefonds.

His social gifts and the charm of his personality were such that he gained the particular favour of the viceroy, Mohammed Ali. With the latter's younger son, Mohammed Said, he was on terms of affectionate intimacy. After service in Rotterdam, Barcelona, Madrid and Rome he returned to private life. In his new-found leisure his thoughts turned again to Egypt and to the possibility of piercing the Suez isthmus with a navigable canal.

An unexpected chance favoured him. Abbas, who had succeeded Mohammed Ali as viceroy, was violently suspicious of western, and particularly of French, influence. He was murdered in 1854 and was succeeded by Said, De Lesseps' close friend. On hearing of Said's succession De Lesseps started immediately for Egypt.

His hopes of enlisting the sympathy of the new viceroy for the project he had at heart were well founded. On Nov. 30, 1854, Said signed what is known as the first act of concession. It authorized De Lesseps to form the Compagnie Universelle du Canal Maritime de Suez, the purpose of the company being to pierce the Isthmus of Suez by a navigable canal, to exploit the waterway when constructed and to build one or two ports at its extremities. The concession was to last for 99 years.

The Company Formed.—This was followed by a second and more precise act signed by Said on Jan. 5, 1856. The statutes of the new company formed an annex to this act which also recited that, subject to the approval of the viceroy's suzerain, the sultan of Turkey, the canal and its dependent ports should be open at all times as a neutral passageway to all merchant ships without distinction of individuals or of nations, provided that the transit dues were duly paid. The company itself in a succeeding article was expressly prohibited from according to any individual, company or organization any particular or preferential treatment whatsoever. The period for which the concession was to last was confirmed at 99 years, to count from the date on which the canal was opened for navigation. At the end of that period, and in default of other arrangements, the canal was to pass into the hands of the Egyptian government. In addition to the maritime canal the company was made responsible for excavating a fresh-water canal from the Nile near Cairo to Lake Timsah, from which branches were to run parallel to the new waterway, one north to Pelusium and the other south to Suez. The company was to have the right to fix and levy transit dues on ships and passengers subject to the restriction that such dues might not exceed 10 fr. per ton or per passenger. The company's central offices were to be located at Alexandria with administrative headquarters in Paris.

The Egyptian government was to receive as a first charge 15% of the net profits, while a further 10% was to be distributed to those persons who had helped most in the fulfillment of the enterprise. The company's capital was to consist of 200,000,000 fr. divided into 400,000 shares of 500 fr. each.

British Opposition.—The concession had hardly been signed when political complications arose. The British government, from fear of French pretensions in Egypt and anxious about the safety of communications with the British Indian empire, set itself to oppose the construction of the canal by every means in its power. Not content with making its displeasure felt in Cairo, the British government used all its influence, considerable at this time, in Constantinople to prevent the sultan's government from giving its approval to the acts of concession signed by the viceroy. In the result it was not until March 19, 1866, that the sultan's sanction was forthcoming, by which time the work of constructing the canal was already under way.

For De Lesseps, undeterred by Palmerston's threats and abuse or by the vacillations of the Sublime Porte, the moment the second act of concession was signed set the work in train. The shares of the new company were placed on the market in Nov. 1858. With the vision that inspired all his actions, De Lesseps had decided that all the western powers should participate in financing the enterprise. Blocks of shares were therefore set aside for each of these countries, including a block of 80,000 shares, one-fifth the whole issue, for Great Britain.

His hopes were, however, disappointed. Neither Great Britain nor the United States (the latter had been offered a block of 20,000 shares) subscribed at all. More than half the shares (207,111) were taken up in France, while the Egyptian government, which had been allotted a block of 64,000, actually subscribed for 177,642 shares. His capital being thus assured, De Lesseps decided not to wait for the sultan's approval and on April 25, 1859, he himself turned the first spadeful of sand at the site where later was to rise the town of Port Said.

Soon, however, he ran into difficulties. Under British pressure a formal order emanated from the sultan requiring all operations on the spot to cease forthwith. Expenses were heavy as well, and the actual progress made was disappointing by comparison. Nevertheless, by 1862 the fresh-water canal had reached Lake Timsah and toward the end of the same year a narrow channel had been excavated between Lake Timsah and the Mediterranean. In 1863 the fresh-water canal was continued to Suez. In the meantime Port Said itself was rising on the former flats and marshes of Lake Menzala.

So far the work had been done by Egyptian labour. The second act of concession provided that at least four-fifths of the labourers employed by the company should be Egyptians, and later the viceroy undertook to supply labourers as required by De Lesseps' engineers, who were to house and feed them and also to pay them at certain specified rates. Although wages and conditions were both better than the men could obtain elsewhere in the Egypt of that day, the khedive Ismail, who had succeeded Said on the latter's death in 1863, allowed himself to be persuaded that for the good of the country the men would be better employed elsewhere. The British government in its tireless resistance to the whole project made representations to the same effect in Constantinople. As a result the Sublime Porte ordered (April 1863) not only that "forced labour" should cease but that the company should retrocede to the Egyptian government the bulk of the lands granted to them by Said in 1856.

On Jan. 16, 1864, the company appealed to the emperor Napoleon III asking him to arbitrate, ostensibly in order to protect the considerable French capital already sunk in the enterprise. By his award, made in July 1864, the company was allowed 38,000,000 fr. as compensation for the abolition of "forced labour," 16,000,000 fr. for the retrocession of the Wadi-Lake Timsah-Suez section of the fresh-water canal (the upstream portion between the Nile and Wadi had already been handed back by an agreement dated March 18, 1863) and 30,000,000 fr. in respect of certain lands granted to the company by Said. The compensation, 84,000,000 fr. in all, was to be paid to the company over a period of 15 years.

The emperor's award was accepted by both parties, and from this date onward the work made rapid progress. More machinery was imported to compensate for reduction in labour force. Political opposition to the project had also by this time died away. Finally in the summer of 1869 the waters of the Mediterranean and the Red sea were united in the Bitter lakes.

Altogether up to the end of 1869, when it was possible to open the canal for traffic, the accounts of the company showed a total expenditure of 432,807,882 fr.

The Opening of the Canal.—The canal was formally opened on Nov. 17, 1869, amid scenes of great splendour, the procession of ships being headed by the French imperial yacht "Aigle" with the empress Eugénie herself on board. De Lesseps' dream, so long cherished, thus ended in triumph. Great Britain, tardily generous, rendered him toward the end of this same year all the honour and all the credit that were his due. He was received in London by Queen Victoria, who conferred upon him the insignia of knight grand commander of the Star of India. He was made a freeman of the city of London and a magnificent fete was held in his honour at the Crystal palace.

As originally constructed the canal had a depth of 8 m. with a bottom width of 22 m. Over the years after 1869 the Suez Canal company was diligent in improving, widening and deepening the waterway. In the early days, except in the Bitter lakes, vessels could pass each other only at certain crossing places or *gares*. As a result of successive widenings passing became possible at any point over the greater part of the canal, one vessel stopping while the other proceeds on her way. The width of the canal had by 1953 been increased to 120 m., while by 1956 the central channel (nowhere less than 60 m. wide) had been deepened to permit the passage of ships with no less than 35 ft. draft. By the early months of 1957 it was expected that further deepening would permit the passage of ships with a draft of 36 ft. In 1951 there was opened to traffic a bypass channel 11 km. long in the northern section of the canal. Ships making the passage of the canal were marshaled in convoys and this new channel made it possible for north- and southbound convoys to pass each other without meeting in the main channel. In addition to all this the Suez Canal company's board sanctioned early in 1955 the inception of an "eighth program of improvements" which contemplated the excavation of two additional bypasses and a widening of the channel throughout its length, at an estimated cost of about 19,600,000,000 fr. By July 1956 these two bypasses were well on the way to completion and it was expected that the whole program would have been carried out by 1959. In the meantime an active study of likely future traffic trends had also been undertaken and the first outlines of a "ninth program of improvements," designed to permit by 1965 the passage annually of at least 170,000,000 tons of crude oil from the middle east to Europe, had been drawn out.

For all these improvements and for annual maintenance, the latter in itself no mean task, the company maintained a staff of highly qualified engineers. In addition, all proposed major works were scrutinized each year by an International Consultative commission composed of distinguished engineers, with specialized marine experience and recruited from various countries, including three from Great Britain.

Judged by the number of transits made, traffic through the canal in 1955 was almost three times greater than it was in 1913 while the net tonnage of shipping using the canal had increased more than five times in this interval. (*See* Table I.) And not

TABLE I.—*Ships in Transit Through the Suez Canal*

Item	1913	1938	1951	1955
Number of transits	5,085	6,171	11,694	14,666
Net tonnage (in 000s)	20,034	34,418	80,356	115,756

only was the number of transits so much greater, but the average size of ships using the canal was also constantly on the increase. This was especially true of oil tankers. In 1955, 805 ships of over 20,000 tons made the passage of the canal compared with 496 in 1954 and 296 in 1953. In 1955, 67,000,000 tons of crude oil passed through the canal from the middle east for destinations

in western Europe and America.

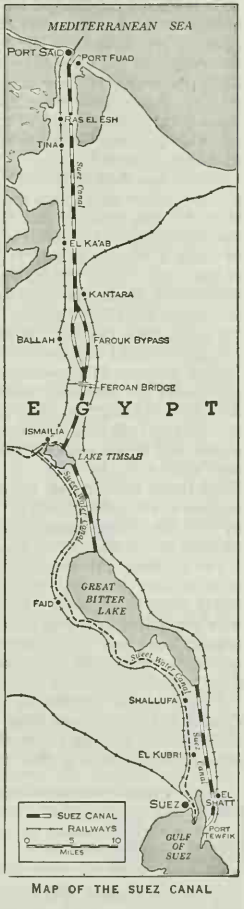
Goods in transit through the canal in 1955 exceeded the 100,000,000-ton mark for the first time in its history. British ships as usual headed the list. British shipping in 1955 accounted for

TABLE II.—*Goods in Transit Through the Suez Canal*
(in 000 metric tons)

Direction	1913	1938	1951	1955
Southward	11,320	7,768	17,420	20,082
Northward	14,455	21,011	59,333	87,426
Total	25,775	28,779	76,753	107,508

4,358 transits, a net tonnage of 32,789,874 or 28.3% of the entire traffic. The gross receipts from transit dues of the Suez Canal company in 1955 came to 34,539,000,000 fr., an increase of 2,083,000,000 fr. over the previous year. In pursuit of their traditional policy, however, the board of directors placed the most part of these extra earnings to reserves, designed to cover the cost of future improvements to the channel. The increase in the dividend granted to the shareholders was by comparison unimportant.

International Status.—As already mentioned, art. 14 of the second act of concession prescribed that the canal should always be open as a neutral passageway to merchant ships of all nations. Neither this nor the sultan's firman of 1866 approving the grant of the concession sufficed however to "neutralize" the canal. For a status of neutrality can be conferred only by international agreement. De Lesseps himself had always been in favour of the canal's being proclaimed neutral, and the issue assumed acute form at the time of the Arabi revolt (1881–82). As a result a convention was signed at Constantinople on Oct. 29, 1888, by representatives of France, Great Britain, Germany, Austria-Hungary, Italy, Russia, Spain, Turkey and the Netherlands. Art. 1 of this convention recites that the canal shall always "be free and open, in time of war as in time of peace, to every vessel of commerce or of war, without distinction of flag." But though the principle of free user was confirmed by art. 1 of the convention of 1888 it does not seem possible to maintain that the canal enjoyed a status of neutrality. No warlike acts could, it is true, be committed within its limits without a violation of the convention. It was not, on the other hand, like a neutralized territory over which passage of armed forces is denied to belligerents. Equally the territorial waters of a neutral state provide no exact parallel, for in their case there are no limitations as to time of arrival and departure.



MAP OF THE SUEZ CANAL

Organization of the Suez Company.—Management of the Suez Canal company was entrusted to a board of 32 directors. By 1956 this board had come to be made up as follows: French 16, British 9 (including 3 representatives of the British government, the company's most important shareholder), Egyptian 5, Dutch 1 and U.S. 1. The board met once a month in Paris. A committee of eight directors chosen by the board met once a week to conduct the current affairs of the company. Two of the British government directors were given seats on the committee of management.

In 1924 the company's shares were doubled in number, to 800,000. Of these the British government held 353,204. The shares were subject to redemption by annual drawings ending in 1968 when the company's concession was due to expire. All holders of 25 or more shares were entitled to vote at the annual general assembly of shareholders. Each shareholder so qualified was entitled to cast ten, but not more than ten, votes.

The company's policy through the years was to revise the dues in a downward direction whenever its financial position made it possible. In 1956 the rates of transit dues were 36.50 Egyptian piastres per Suez net ton for laden ships and 17 piastres per Suez net ton for ships in ballast. Dues on passengers were abolished by a decision of the board taken in 1951.

Such was the state of affairs when on July 26, 1956, Gamal Abdel Nasser, the president of Egypt, announced that the Suez Canal company had been nationalized, that martial law had been declared in the canal zone and that his officers were at that moment seizing control of the company's buildings and office installations. (*See* also *Suez Conflict*, below.)

BIBLIOGRAPHY.—J. Charles-Roux, *L'Isthme et le canal de Suez*, 2 vol. (Paris, 1901); A. T. Wilson, *The Suez Canal: Its Past, Present and Future* (London, New York, 1933); A. Siegfried, *Suez and Panama* (London, New York, 1940); G. Mehrlin, *Der Suezkanal als Konjunkturanzeiger der Weltwirtschaft* (Zürich, 1945); Moustapha el-Hefnaoui, *Les problèmes contemporains posés par le canal de Suez* (Paris, 1951); H. J. Schonfield, *The Suez Canal in World Affairs* (London, 1952).
(F. V. W.)

Suez Conflict.—For the history of the Palestine war and events leading up to the crisis of 1956, *see* ARAB LEAGUE; EGYPT; ISRAEL; etc.

Despite the armistice agreement of 1949 between Egypt and Israel, border incidents continued, and in 1950 Egypt prohibited ships bound for Israel from passing through the Suez canal. Under an agreement of 1954, Great Britain evacuated its forces from the canal zone, the last troops leaving in June 1956. Col. Gamal Abdel Nasser, who became head of the Egyptian government in 1954, announced on July 26, 1956, that his government had seized full control of the Suez canal. It planned to use the profits from its operation under nationalization to finance the Aswan high dam project. Nasser rejected the protests of the western powers and proposals to create an international board to operate the canal.

Military Action.—On Oct. 29, 1956, Israeli forces invaded the Sinai peninsula of Egypt and drove westward across the desert toward the canal. Within 24 hours Great Britain and France demanded that all troops be withdrawn a distance of 10 mi. from the canal and asked Egypt to agree to a temporary Anglo-French occupation of key points in the canal zone. Egypt rejected the ultimatum, and aircraft from Cyprus began to bomb Egyptian bases on Oct. 31.

The Israeli attack moved swiftly. By Nov. 2 the city of Gaza had fallen, and the Gaza strip on the Mediterranean was soon cut off. The Egyptian army east of the canal was routed, and Israeli forces won key positions throughout the Sinai peninsula. In the lightning action, Israel also captured the island of Tiran at the entrance to the Gulf of Aqaba.

French and British forces invaded the north end of the canal zone on Nov. 5. (On the same day the United States rejected as "unthinkable" a Soviet proposal that U.S.S.R. and U.S. troops intervene.) Port Said and the smaller city of Port Fuad soon fell. The allies ordered a cease-fire, and major fighting ended on Nov. 7. Ismailia, at the midpoint of the canal, and Suez, at the southern terminus, remained in Egyptian hands. Sunken ships, wrecked cranes and other equipment blocked the canal. The U.S.S.R. warned that "volunteers" would intervene if Anglo-French and Israeli forces did not withdraw from Egypt.

UN Action and Evacuation.—In emergency sessions on Nov. 2 and Nov. 7 the United Nations general assembly called for the withdrawal of the three invaders from Egypt. The assembly also voted to create an international UN command force to supervise

the evacuation and to patrol the Israeli-Egyptian frontier. Great Britain, France and Israel agreed to withdraw as soon as an adequate UN force occupied the territory. The UN force supervised the withdrawal of Anglo-French troops from the northern end of the canal and of Israeli forces to the 1949 armistice lines. Great Britain and France complied with the general assembly's further demand on Nov. 24 to withdraw "forthwith," and their last units left Egypt on Dec. 22. Clearance operations began in Jan. 1957, and in April the canal was reopened for traffic.

SUFFOLK, EARLS AND DUKES OF. The 1st earl of Suffolk, ROBERT DE UFFORD (1298–1369), was one of six new earls created by Edward III in 1337. After taking part in the arrest of Roger Mortimer, earl of March, at Nottingham Castle in October 1330, he became one of the king's most trusted commanders, diplomats, and councilors. The Uffords were a Suffolk landowning family recently risen to prominence; the 1st earl's father, Robert (1279–1316), who was summoned to Parliament as a baron in 1309, was the son of Robert de Ufford (d. 1298), twice justiciar of Ireland in Edward I's reign. Earl Robert fought at Crécy (1346) and Poitiers (1356) and served in almost every campaign between 1340 and 1360. His heir, WILLIAM (1339–82), 2nd earl, narrowly escaped capture by the East Anglian rebels in the Peasants' Revolt of 1381 and later took the lead in suppressing the insurrection in Suffolk. All Earl William's five children predeceased him; on his death the earldom became extinct, and the bulk of the de Ufford estates escheated to the crown.

The extinction of the de Ufford earls made possible the rapid rise of the de la Poles to magnate status; in August 1385, Richard II created MICHAEL DE LA POLE (*c.* 1330–89) earl of Suffolk. His grandson, WILLIAM (1396–1450), 4th earl, was made a marquess in 1444 and duke of Suffolk in 1448 (*see* POLE; SUFFOLK, WILLIAM DE LA POLE, Duke of). JOHN DE LA POLE (1442–91/92), 2nd duke of Suffolk, was the son of the 1st duke by his wife, Alice Chaucer, granddaughter of the poet; he married Elizabeth (d. *c.* 1503), sister of the future Edward IV and Richard III, and survived the Wars of the Roses, having supported in turn Edward IV, Richard III, and Henry VII. His second but first surviving son, EDMUND (1471/72–1513) surrendered the dukedom in 1493 and became earl of Suffolk; he was attainted in 1504 (*see* POLE, RICHARD DE LA). The male line ended when William, the last survivor of Duke John's seven sons, died in the Tower of London in 1539.

The title of duke of Suffolk, revived by Henry VIII in 1514 for his favourite, CHARLES BRANDON (*c.* 1484–1545; *see* SUFFOLK, CHARLES BRANDON, 1st Duke of), became extinct on the death from the sweating sickness of Brandon's sons (by his fourth wife, Katharine Willoughby), Henry and Charles, on July 14, 1551. Frances, Brandon's elder daughter by his third wife, the king's younger sister Mary Tudor (d. 1533), had married HENRY GREY (1517–54), 3rd marquess of Dorset, a great-grandson of Edward IV's queen, Elizabeth Woodville. Well-educated and a strong Protestant, Dorset was a weak and ambitious man; he belonged to the party of John Dudley, duke of Northumberland, and was created duke of Suffolk on Oct. 11, 1551. He supported Northumberland's attempt to make his daughter Lady Jane Grey (1537–54) queen in July 1553, but he was quick to abandon her cause when the plot failed and was pardoned by Queen Mary. His opposition to the queen's plans for her Spanish marriage involved him in Sir Thomas Wyat's Rebellion in 1554; the dukedom again became extinct after his execution for treason on Feb. 23, 1554.

In 1603 THOMAS HOWARD (1561–1626), Lord Howard de Walden, second son of Thomas Howard (1538–72), 4th duke of Norfolk, was created earl of Suffolk by James I (*see* SUFFOLK, THOMAS HOWARD, 1st Earl of), and that branch of the Howard family has held the title ever since.

BIBLIOGRAPHY.—R. E. Chester Waters, *Genealogical Memoirs of the Extinct Family of Chester of Chicheley* (1878); G. H. Orpen, *Ireland Under the Normans*, vol. iii–iv (1911); J. G. Nichols (ed.), *The Chronicle of Queen Jane and of Two Years of Queen Mary* (1850).

(T. B. P.)

SUFFOLK, CHARLES BRANDON, 1st Duke of (*c.* 1484–1545), was the son of William Brandon, standard-bearer to Henry VII, who was slain by Richard III in person on Bosworth Field. He was high in Henry VIII's favour and held a succession of offices in the royal household. On May 15, 1513, he was created Viscount Lisle, having entered into a marriage contract with his ward, Elizabeth Grey, Viscountess Lisle in her own right, who, however, refused to marry him when she came of age. He distinguished himself at the sieges of Térouenne and Tournai in the French campaign of 1513. At that time Henry VIII was secretly urging Margaret of Savoy, regent of the Netherlands, to marry Brandon, whom he created duke of Suffolk on Feb. 1, 1514. Brandon took part in the jousts that celebrated the marriage of Mary Tudor, Henry's sister, with Louis XII of France, and on Louis' death early in 1515 he was sent to congratulate the new king Francis I. An affection between Suffolk and Mary had existed before her marriage, and Francis charged him with intending to marry her, having extracted a confession from Mary herself. Francis and Henry both professed a friendly attitude toward the marriage of the lovers, but Suffolk had many political enemies, and Mary feared that she might again be sacrificed to political considerations. The truth was that Henry was anxious to obtain from Francis the gold plate and jewels that had been given or promised to the queen by Louis, and he practically made his acquiescence in Suffolk's suit dependent on his obtaining them. The pair cut short the difficulties by a private marriage.

Suffolk was perhaps saved from Henry's anger by Thomas Wolsey. Mary and he eventually agreed to pay to Henry £24,000 in yearly installments of £1,000, and the whole of Mary's dowry from Louis of 200,000 crowns, together with her plate and jewels. They were openly married at Greenwich on May 13, 1515. The duke had been twice married already, to Margaret Mortimer and to Anne Browne. Anne Browne died in 1511, but Margaret Mortimer, from whom he had obtained a divorce on the ground of consanguinity, was still living. He secured in 1528 a bull from Pope Clement VII assuring the legitimacy of his marriage with Mary Tudor, and of the daughters of Anne Browne. After Mary's death in 1533, Suffolk married (1534) Katharine, Baroness Willoughby de Eresby in her own right, who had been granted to him in ward as the intended wife of his eldest son.

Suffolk was present at the Field of the Cloth of Gold in 1520, and in 1523 he was sent to command the English operations in the north of France. Suffolk was in favour of Henry's divorce from Catherine of Aragon, and after Wolsey's disgrace his influence increased daily. He was sent with the duke of Norfolk to demand the great seal from Wolsey; the same noblemen conveyed to Queen Catherine the news of Anne Boleyn's marriage, and Suffolk acted as high steward at the new queen's coronation. He was commissioned by Henry to dismiss Catherine's household. He received a large share of the plunder after suppression of the monasteries. In 1544 he again commanded an English army for the invasion of France. Suffolk died at Guildford on Aug. 24, 1545.

SUFFOLK, THOMAS HOWARD, 1st Earl of (1561–1626), second son of Thomas Howard, 4th duke of Norfolk, was born on Aug. 24, 1561. He behaved very gallantly during the attack on the Spanish Armada. He commanded the expedition on which Sir Richard Grenville and the "Revenge" were lost (1591) and held commands in the Cadiz and Islands voyages (1596–97). Created Baron Howard de Walden in 1597 and earl of Suffolk in July 1603, he was lord chamberlain of the royal household from 1603 to 1614 and lord high treasurer from 1614 to 1618, when he was deprived of his office on a charge of misappropriating money. He was tried in the Star Chamber and was heavily fined.

Suffolk's second wife was Catherine (d. 1633), widow of Richard Rich, a woman whose avarice was partly responsible for her husband's downfall. She shared his trial and was certainly guilty of taking bribes from Spain. One of his three daughters was the notorious Frances Howard, who instigated the poisoning of Sir Thomas Overbury (*q.v.*). The earl died on May 28, 1626, in London.

(R. B. WM.)

SUFFOLK, WILLIAM DE LA POLE, Duke of (1396–1450), second son of Michael de la Pole, 2nd earl of Suffolk, served the early part of his career in the wars against France and the latter in the government seeking peace with that country. He was born on Oct. 16, 1396, at Cotton, Suffolk. Suffolk served in all the later French campaigns of the reign of Henry V, and in spite of

his youth held high command on the marches of Normandy in 1421–22. In 1423 he joined the earl of Salisbury in Champagne and shared his victory at Crevant. He fought under John, duke of Bedford, at Verneuil on Aug. 17, 1424, and throughout the next four years was Salisbury's principal lieutenant in the direction of the war. When Salisbury was killed before Orléans on Nov. 3, 1428, Suffolk succeeded to the command. After the siege was raised, Suffolk was defeated and taken prisoner by Jeanne d'Arc at Jargeau on June 12, 1429. He was soon ransomed, and during the next two years was again in command on the Norman frontier. He returned to England in Nov. 1431, after more than 14 years' continuous service in the field.

In the next decade, as a royal household official and a supporter of the predominant faction of Cardinal Beaufort, Suffolk acquired a considerable influence in the government, which he used mainly to further his own interests. Greedy and unscrupulous, he showed himself unusually grasping even in that mercenary age. He acquired a large number of wardships and offices of profit in the king's gift, and his agents made the heavy hand of his lordship felt throughout East Anglia, where his principal estates lay. It was not until 1443 that the retirement of Cardinal Beaufort brought Suffolk reluctantly to the forefront of politics. Like Beaufort, Suffolk had a genuine desire to achieve a peaceful settlement with France, but he had no clear and practical plan for obtaining peace without disgrace, and he shirked necessary but unpopular decisions. His first success (for which he was made a marquess) was the securing in 1444 of a two years' truce and the hand of Margaret of Anjou for Henry VI. Yet soon afterward the English government was forced to surrender Anjou and Maine in return for a further extension of the truce, a concession which increased Suffolk's growing unpopularity at home. Rumour, though without sufficient reason, made Suffolk responsible for the death of his rival, Humphrey, duke of Gloucester, in Feb. 1447. Another critic of the government, Richard, duke of York, was replaced in his French command by the incompetent Edmund Beaufort, duke of Somerset. Suffolk's promotion to a dukedom in 1448 marked the height of his power.

The difficulties of his position may have led him to give some countenance to a treacherous attack on Fougères during the time of truce (March 1449). The renewal of the war and the loss of all Normandy were its direct consequences. When parliament met in Nov. 1449, the whole Suffolk administration came under attack. The treasurer, Adam Moleyns, bishop of Chichester, was forced to resign, and on Feb. 7, 1450, the commons presented articles of accusation against Suffolk himself, dealing chiefly with alleged maladministration and the ill success of the French policy; there was a charge of aiming at the throne by the betrothal of his son to the little Margaret Beaufort. Suffolk denied the accusations as false, untrue and too horrible to speak more of. Ultimately the king sentenced him to banishment for five years. Suffolk left England on May 1. He was intercepted in the channel by the ship "Nicholas of the Tower," and next morning was beheaded in a little boat alongside.

Popular opinion at the time, followed by Yorkist chroniclers and Tudor historians, judged him a traitor, and later legend made him a paramour of Margaret of Anjou. These charges are not supported by any reliable evidence, and probably Suffolk was unfortunate in being made a scapegoat for an unpopular and corrupt administration and policies for which others were as much responsible as he.

Suffolk's wife, Alice, was widow of Thomas, earl of Salisbury, and granddaughter of Geoffrey Chaucer. By her he had an only son John, 2nd duke of Suffolk.

BIBLIOGRAPHY.—Suffolk's character and policy have been the subject of much discussion; for modern opinions *see* W. Stubbs, *Constitutional History of England* (Oxford, 1880); C. Oman, *The Political History of England*, vol. vi (London, 1906); C. L. Kingsford, *Prejudice and Promise in Fifteenth Century England* (Oxford, 1925); K. B. McFarlane in *Cambridge Medieval History*, vol. viii (Cambridge, 1936). Besides the authorities for his career listed in the *Dictionary of National Biography*, reference should be made to the series of *Calendars* of Patent and Close Rolls. (C. D. R.)

SUFFOLK, an eastern county of England, bounded north by Norfolk, south by Essex, west by Cambridgeshire and east by the North sea. Its area is 1,468.0 sq.mi., with a population (1951) of 337,776. In the west, the Upper Chalk escarpment of the Chilterns, its structure modified by glacial action, continues through Suffolk and on to the Lincolnshire Wolds. To the east it is overlaid by the London Clay and the Crag deposits of the Pliocene, the rich and specialized fauna of which attracts the attention of paleontologists throughout the world. These basic formations are largely overlaid by glacial boulder clays, sands and gravels, the general contour being that of an undulating plateau rising in places to over 400 ft. Alluvial deposits border the river valleys of the Orwell, Stour and Deben, the subsidence begun in Neolithic times, and possibly still continuing, having extended the tidal range of the sea far inland. Alluvium is also found in the lowlands of the northwest. The last millennium has seen the formation of the ten-mile-long spit of shingle extending from the mouth of the Alde to Weir point.

The scenery is varied. The porous chalk of the west and the sandlings of the east, with their light vegetation, lent themselves to cultivation from Neolithic times. The sands of the Breckland are so light that they would be blown away but for their covering of heath, giving place in the 1950s to coniferous plantations. High Suffolk, with its boulder clay once blanketed by forest, is now fertile farmland.

History.—Evidences of man's occupation date from the very early stages of the Paleolithic period, even if the controversial pre-Crag flints are ignored. Work in the mid-20th century at Hoxne, where undoubted implements were first recognized in 1797, establishes the period as the Great Interglacial (*c.* 250,000 B.C.), the culture as Acheulean. Upper Paleolithic, Mesolithic and Neolithic are found at various sites. At Creeting St. Mary, Neolithic Rinyo-Clactonian ware was associated with cremations. The Beaker folk are found both near Ipswich and in the Breckland area of the west. The Bronze Age saw a penetration from the Wessex area along the Icknield way; barrows of the Middle Bronze Age have been excavated at Brightwell, and Late Bronze Age urnfields demonstrate the spread from the east by way of the valleys. Iron Age "A" settlements were naturally confined principally to the light lands. Iron Age "B" is represented by a chariot burial near Mildenhall. Notable Belgic finds come from a cemetery at Boxford and a settlement at Burgh, near Otley.

Along with Norfolk, Suffolk formed part of the kingdom of the Iceni. A bronze head of the emperor Claudius, dredged from the river Alde at Rendham, may represent jettisoned spoil from the sack of Colchester during the rebellion of Boadicea in A.D. 61. Excavations near Baylham have revealed a Claudian settlement beneath the Roman road from Colchester to Venta Icenorum; air photography disclosed a camp, possibly of this period, near Ixworth, in an area containing several villa sites. A statuette of Nero from Baylham and a plaque of Mithras from Whitton (Ipswich) have a religious connotation. The notable Mildenhall treasure (in the British museum, London) was buried at the time of the Saxon raids. Burgh castle still stands, a fort of the Saxon shore, and a print of the 18th century shows such a fort at Walton (Felixstowe).

The name Suffolk represents the south-folk of the Angles who settled in this part of England. Of this period, the cemeteries at Lackford and Ipswich are noteworthy. The remarkable ship burial at Sutton Hoo, near Rendlesham, seat of the East Anglian monarchs, on the Deben, throws new light on the aesthetic standards of the time. It appears to have been a cenotaph to the pagan Aethelhere, his body possibly lost in the floods at the battle of Winwaed in Yorkshire in A.D. 655. Suffolk suffered during the Danish invasions and the treaty of Wedmore made it a part of the Danelaw.

After the Norman conquest the political history of the county revolves round the castles. Two of these, Clare and Eye, were the headquarters of great honours, and of the remainder the most important were Framlingham, Bungay, Haughley and Orford. The castles of Mettingham and Wingfield were built during the 14th century. The history of the county was also strongly influenced by the two pre-Conquest ecclesiastical liberties of Bury St. Ed-

munds and St. Etheldreda. The first of these consisted of the eight-and-a-half western hundreds of the county which belonged to the Benedictine abbey of Bury St. Edmunds; the second, of five-and-a-half hundreds in east Suffolk, which belonged to the prior and convent of Ely, and after 1541 to the dean and chapter of Ely cathedral. Together these two liberties comprised almost two-thirds of the county. The remaining hundreds, known as the "Geldable," were within the jurisdiction of the sheriff. During the 14th century serious uprisings occurred in west Suffolk against Bury abbey. Until 1575 only one sheriff was appointed for both Norfolk and Suffolk. During the 15th century the county was relatively unaffected by the Wars of the Roses, in spite of the presence of one of the chief seats of the dukes of Norfolk at Framlingham, which gave it a strongly Yorkist character.

Though strong continental Protestant influences penetrated the county through the port of Ipswich in the early 16th century, Mary Tudor derived much of the support for her claim to the throne from Suffolk. Puritanism did, however, establish itself as a dominant force by the early 17th century, and many Suffolk families emigrated to the New England colonies in America, for religious as well as economic reasons. During the Great Rebellion the county was, for the most part, parliamentarian, and the local gentry gave strong support to the Association of the Eastern Counties.

Until 1914 Suffolk was within the East Anglian diocese, the seat of which was, after 1070, at Norwich. In 1837 the archdeaconry of Sudbury was moved, with some modifications, from the Norwich to the Ely diocese. The present diocese of St. Edmundsbury and Ipswich was founded in 1914 and comprises the whole of Suffolk, less the deaneries of Lothingland and Flegg.

Architecture.—Besides the great abbey at Bury St. Edmunds, Suffolk had many smaller religious foundations, particularly in the eastern half of the county. The most important remains are those of the college at Clare, founded in 1248, dissolved in the 16th century and, after 400 years, occupied again in 1953 by the Order of St. Augustine; the Augustinian priory of Butley; the Grey Friars monastery at Dunwich; the Augustinian priory at St. Olaves Herringfleet; and Leiston abbey.

The churches of Suffolk are, in the main, constructed of flint, and in some parts of the county they have early round flint towers, of which type there are 41, the most notable being at Little Saxham and Wortham. The majority of the churches are, however, of the Decorated and Perpendicular periods, with open roofs, good woodwork and fine fonts. Among many magnificent churches may be mentioned Long Melford with its remarkable Lady chapel, Lavenham, Blythburgh, Southwold, Clare and Framlingham.

The most interesting castle remains are at Framlingham, Orford, Wingfield and Mettingham. Other important sites are to be found at Eye, Clare, Bungay and Haughley. Martello towers are to be found on the coast, for example, at Aldeburgh and Felixstowe.

There are some fine examples of domestic architecture in the county, particularly Little Wenham hall, a 13th-century fortified dwelling house, also Hengrave and Helmingham halls of the early 16th century, the latter with a moat and drawbridge, and Heveningham hall by James Wyatt, which has beautiful interior decorations.

Suffolk is also exceptionally rich in fine villages. The most beautiful of these are in the Stour valley, immortalized in the paintings of John Constable, such as Chelsworth, Kersey, Lavenham and East Bergholt. Thomas Gainsborough, another great Suffolk painter, was born at Sudbury in the Stour valley in 1727. Other lovely villages and small towns include Beccles and Bungay, Mildenhall, Framlingham and Woodbridge.

Industries, Agriculture and Communications.—In early times fishing was among the most important Suffolk industries. Fishing fleets left the ports for cod and ling from Iceland and herring and mackerel from the North sea, and a big fish trade was carried on with Flanders. Fishing was later concentrated on Lowestoft.

From the 14th to the 17th centuries Suffolk was among the chief cloth-weaving counties and the industry was at its height in the 15th century. This industry declined rapidly in the 18th and 19th centuries, and was replaced by textile manufactures such as sailcloth, coconut fibre, horsehair and clothing. Silk weavers also migrated from Spitalfields to the county.

Besides textiles, the most important manufactures relate to agriculture. Agricultural machinery is especially important at Ipswich, Bury St. Edmunds and Stowmarket. Grain milling and malting are extensive and fertilizers are made at Ipswich. There are sugar-beet factories at Sproughton and Bury St. Edmunds. Heavy engineering is also found at Ipswich. Brandon was the centre of a flint-knapping industry for centuries, now obsolescent because of lack of demand. Shipbuilding is found at both Lowestoft and Ipswich, and Beccles and Bungay are important for printing. Paint is manufactured at Stowmarket, and at Brantham a large plastic works has been established.

Agriculture was and still is the greatest employer of labour in Suffolk, and after the breaking up of much of the pasture during the Napoleonic Wars, it became one of the principal grain-growing counties. More than half the land is arable, and the plowing of grasslands during World War II led to considerable fluctuations in the number of cows, sheep and pigs. The main crops in order of importance are barley, wheat, oats, sugar beet, clover and grass (for hay), mangolds (with turnips and swedes), peas and beans. Horse breeding is declining—including the famous Suffolk Punch—but Newmarket retains its pre-eminence as the centre of the bloodstock and race-horse industry.

Considerable areas of woodlands include 16,000 ac. of state forest in the Breckland and 10,000 ac. near the coast. Many plants and insects of the Breckland are found nowhere else, and extensive areas of the county are protected as nature reserves or areas of special scientific or historic interest such as Minsmere; Staverton Thicks, a remnant of the primeval forest; Havergate Island, the only British home of the avocets; and Cavenham heath. The National trust owned 521 ac. in 1955 and some small properties in the county.

British railways (Eastern region) serve the county with a main line from London (Liverpool street) to Ipswich and branch lines to other parts. There are civil aerodromes at Ipswich and Newmarket Heath (a private landing ground).

Administration and Population.—Prior to the Reform act of 1832, the county returned 2 members to parliament and 14 members from its seven boroughs. By this act the boroughs of Orford, Aldeburgh and Dunwich were disfranchised. In 1956 the county constituted five divisions, each returning one member. Until 1872 the county had a single commission of the peace, and quarter sessions were held in four divisions, one each for the liberties of Bury St. Edmunds and St. Etheldreda, and one each for the two parts of the "Geldable," based on Beccles and Ipswich respectively. After 1872 the number of divisions was reduced to two, namely East and West Suffolk. Since 1951 separate commissions of the peace have been in force for East and West Suffolk. There are separate commissions for Ipswich and Bury St. Edmunds. Since 1888 Suffolk has been divided for administrative purposes into East Suffolk, covering 857.2 sq.mi. (excluding Ipswich) and having a population (1951) of 217,124, and West Suffolk, covering 610.8 sq.mi. and with a population (1951) of 120,652.

The following are the principal boroughs of East Suffolk: Aldeburgh (pop. 1951 2,689), Beccles (6,870), Eye (1,631), Ipswich, a county borough and the county town (104,785), Lowestoft (42,834) and Southwold (2,473); there are seven urban and seven rural districts.

The boroughs of West Suffolk are Bury St. Edmunds (20,056) and Sudbury (6,620). It has three urban and six rural districts.

BIBLIOGRAPHY.—A. Suckling, *The History and Antiquities . . . of Suffolk*, 2 vol. (1846–48), an index was published in 1952; W. White, *History, Gazetteer & Directory of Suffolk* . . . (1844, and subsequent editions); J. Kirby, *The Suffolk Traveller* . . . (1735); A. Page, *A Supplement to the Suffolk Traveller*, 5 vol. (1843); *Victoria County History of Suffolk*, 2 vol. (1907–11); L. J. Redstone, *Suffolk* (1930); Ipswich and East Suffolk Record Office, *East Suffolk County Plan 1951: the Report of the Survey; West Suffolk County Plan 1951: the Analysis of the Survey.* (D. Cx.)

SUFFOLK, a city of southeastern Virginia, U.S., about 20 mi. S.W. of Norfolk at the head of navigation of the Nansemond river. Settled in the early 18th century, Suffolk was incorporated as a town in 1808 and as a city in 1910. In 1919 it adopted a

council-manager form of government. During the American Revolution Suffolk was burned by the British, and during the American Civil War it was occupied by Union troops in 1862. Industries include peanut processing, meat packing, cotton ginning, tea processing and the manufacture of lumber, building materials and textiles. For comparative population figures *see* table in VIRGINIA: *Population*.

SUFFRAGAN, (1) a diocesan bishop in his relation to the metropolitan; (2) an assistant bishop. (*See* BISHOP.)

SUFFRAGE, in political affairs, the right or the exercise of the right to vote; in ecclesiastical usage, the short intercessory prayers in litanies spoken or sung by the people as distinguished from those of the priest or minister.

See ELECTORAL SYSTEMS; REPRESENTATION; VOTING BEHAVIOUR; and, for the women's suffrage movement, WOMEN'S SUFFRAGE.

SUFFREN SAINT TROPEZ, PIERRE ANDRÉ DE (1729–1788), French admiral, third son of the marquis de Saint Tropez, was born in the Château de Saint Canat (Aix) on July 17, 1729. The French navy and the Order of Malta offered the usual careers for the younger sons of noble families of the south of France who did not elect to go into the church. Suffren entered the close and aristocratic corps of French naval officers as a "garde de la marine," the equivalent of a cadet or midshipman, in Oct. 1743, in the "Solide," one of the line of battleships which took part in the confused engagement off Toulon in 1744. He was then in the "Pauline" in the squadron of M. Macnémara on a cruise in the West Indies. In 1746 he went through the duc D'Anville's disastrous expedition to retake Cape Breton, which was ruined by shipwreck and plague.

Next year (1747) he was taken prisoner by Hawke in the action with the French convoy in the Bay of Biscay. When peace was made in 1748 he went to Malta to perform the cruises with the galleys of the Order of Malta technically called "caravans," a reminiscence of the days when the knights protected the pilgrims going from Saint Jean d'Acre to Jerusalem. In Suffren's time this service rarely went beyond a peaceful tour among the Greek islands. During the Seven Years' War he was present as lieutenant in the "Orphée" in the action with Admiral Byng, which, if not properly speaking a victory, was at least not a defeat for the French, and was followed by the surrender of the English garrison of Minorca. In 1757 he was again taken prisoner, when his ship the "Océan" was captured by Boscawen off Lagos. On the return of peace in 1763 he intended again to do the service in the caravans which was required to qualify him to hold the high and lucrative posts of the Order. He was, however, named to the command of the "Caméléon," a zebec—a vessel of mixed square and lateen rig peculiar to the Mediterranean—in which he cruised against the pirates of the Barbary coast. Between 1767 and 1771 he performed his caravans, and was promoted from knight to commander of the Order. From that time till the beginning of the War of American Independence he commanded vessels in the squadron of evolution which the French government had established for the purpose of giving practice to its officers. His nerve and skill in handling his ship were highly commended by his chiefs.

In 1778–79 he served with the squadron of D'Estaing (*see* ESTAING, CHARLES HECTOR) on the coast of North America and in the West Indies. He led the line in the action with Adm. John Byron off Grenada, and his ship, the "Fantasque" (64), lost 62 men. His letters to his admiral show that he strongly disapproved of D'Estaing's half-hearted methods. In 1780 he was captain of the "Zèle" (74) in the combined French and Spanish fleets which captured a great English convoy in the Atlantic. His candour towards his chief had done him no harm in the opinion of D'Estaing. It is said to have been largely by the advice of this admiral that he was chosen to command a squadron of five ships of the line sent to help the Dutch, who had joined France and Spain to defend the Cape against an expected English attack, and then to go on to the East Indies.

He sailed from Brest on March 22 on the cruise which has made him unique among French admirals, and puts him in the front rank of sea commanders. He was by nature even more vehement than able. The disasters which had befallen the navy of his country during the last two wars, and which, as he knew, were due to bad administration and timid leadership, had filled him with a burning desire to retrieve its honour. He was by experience as well as by temperament impatient with the formal manoeuvring of his colleagues, which aimed at preserving their own ships rather than at taking the English, and though he did not dream of restoring the French power in India, he did hope to gain some such success as would enable his country to make an honourable peace. On April 16, 1781, he found the English expedition on its way to the Cape under the command of George Johnstone (1730–1787), at anchor in Porto Praya, Cape Verde islands. Remembering how little respect Boscawen had shown for the neutrality of Portugal at Lagos, he attacked at once. Though he was indifferently supported, he inflicted as much injury as he suffered, and proved to the English that in him they had to deal with an admiral of quite a different type from the Frenchmen they had been accustomed to. He pushed on to the Cape, which he saved from capture by Johnstone, and then made his way to the Île de France (Mauritius), then held by the French. D'Orves, his superior officer, died as the united squadrons, now 11 sail of the line, were on their way to the Bay of Bengal.

The campaign, which Suffren now conducted against the English admiral Sir Edward Hughes (1720?–1794), is famous for the number and severity of the encounters between them. Four actions took place in 1782: south of Madras (Feb. 17); near Trincomalee (April 12); off Cuddalore (July 6), after which Suffren seized upon the anchorage of Trincomalee, compelling the small British garrison to surrender; and again off Trincomalee (Sept. 3). No ship was lost by the British in any of these actions, but none was taken by them. Suffren attacked with unprecedented vigour on every occasion, and if he had not been ill-supported by some of his captains he would undoubtedly have gained a distinct victory. As it was, he maintained his squadron without the help of a port to refit, and provided himself with an anchorage at Trincomalee.

His activity encouraged Hyder Ali, who was then at war with the company. He refused to return to the islands to escort the troops coming out under command of Bussy, maintaining that his proper purpose was to cripple the squadron of Sir Edward Hughes.

During the northeast monsoon he would not go to the islands but refitted in the Malay ports in Sumatra, and returned with the southwest monsoon in 1783. Hyder Ali was dead, but Tippoo Sultan, his son, was still at war with the company. Bussy arrived and landed. The operations on shore were slackly conducted by him, and Suffren was much hampered, but when he fought his last battle against Hughes (April 20, 1783), with fourteen ships to eighteen he forced the English admiral to retire to Madras, leaving the army then besieging Cuddalore in a very dangerous position.

The arrival of the news that peace had been made in Europe put a stop to hostilities, and Suffren returned to France. While refitting at the Cape on his way home, several of the vessels also returning put in, and the captains waited on him. Suffren said in one of his letters that their praise gave him more pleasure than any other compliment paid him.

In France he was received with enthusiasm, and an additional office of vice-admiral of France was created for him. He had been promoted bailli in the Order of Malta during his absence. He died on Dec. 8, 1788, when he was about to take command of a fleet collected in Brest. The official version of the cause of death was apoplexy, and as Suffren was a very corpulent man it appeared plausible. Many years afterward, however, his body servant told Augustin Jal, the historiographer of the French navy, that he had been killed in a duel by the prince de Mirepoix. The cause of the encounter, according to the servant, was that Suffren had refused in very strong language to use his influence to secure the restoration to the navy of two of the prince's relatives who had been dismissed for misconduct.

Suffren was crippled to a large extent by the want of loyal and capable co-operation on the part of his captains, and the vehemence of his own temperament sometimes led him to disregard prudence, yet he had an indefatigable energy, a wealth of re-

source, and a thorough understanding of the fact—so habitually disregarded by French naval officers—that success at sea is won by defeating an enemy and not by merely outmanoeuvring him, and this made him a most formidable enemy.

The standard authority for the life of Suffren is the *Histoire du Bailli de Suffren* by Ch. Cunat (1852). The *Journal de Bord du Bailli de Suffren dans l'Inde*, edited by M. Mores, was published in 1888. There is an appreciative study in Captain Mahan's *The Influence of Sea Power upon History, 1660–1783* (1890).

ṢŪFĪISM (*taṣawwuf*) is formed from the Arabic word *Ṣūfī*, which was applied, in the 2nd century of Islam, to men or women who adopted an ascetic or quietistic way of life. The word *Ṣūfī* from *ṣūf* (wool) refers to garments worn by such persons.

Mysticism in Islam goes back to Mohammed who, on one side of his nature, was an ascetic and in some degree a mystic. Notwithstanding his condemnation of the solitary life and celibacy of Christian monkery, the example of the Ḥanīfs, with some of whom he was acquainted, and of the Christian hermits made a deep impression on his mind and led him to preach the efficacy of ascetic exercises, such as prayer, vigils and fasting. Again, while Allah is described in the Koran as the one God working his arbitrary will in unapproachable supremacy, other passages lay stress on his all-pervading presence and intimate relation to his creatures, *e.g.*, "Wherever ye turn, there is the face of Allah" (ii, 109), "We (God) are nearer to him (man) than his neck-vein" (l, 15). The germs of mysticism latent in Islam from the first were rapidly developed by the political, social and intellectual conditions which prevailed in the two centuries following the prophet's death. Devastating civil wars, a ruthless military despotism caring only for the things of this world, Messianic hopes and presages, the luxury of the upper classes, the hard mechanical piety of the orthodox creed, the spread of rationalism and free thought, all induced a revolt toward asceticism, quietism, spiritual feeling and emotional faith. Thousands, wearied and disgusted with worldly vanities, devoted themselves to God. The terrors of hell, so vividly depicted in the Koran, awakened in them an intense consciousness of sin, which drove them to seek salvation in ascetic practices.

Ṣūfiism was originally a practical religion, not a speculative system. It arose, as Junayd of Baghdad says "from hunger and taking leave of the world and breaking familiar ties and renouncing what men deem good, not from disputation." The early Ṣūfīs were closely attached to the Muslim church. It is said that Abū Hāshim of Kūfa (d. before A.D. 800) founded a monastery for Ṣūfīs at Ramleh in Palestine, but such fraternities seem to have been exceptional. Many ascetics of this period used to wander from place to place, either alone or in small parties, sometimes living by alms and sometimes by their own labour. They took up and emphasized certain Koranic terms, such as *dhikr* (praise of God), consisting of recitation of the Koran, repetition of the Divine names and the like, and *tawakkul* (trust in God), now defined as renunciation of all personal initiative and volition, leaving one's self entirely in God's hands.

Quietism soon passed into mysticism. The attainment of salvation ceased to be the first object, and every aspiration was centred in the inward life of dying to self and living in God. Toward the end of the 2nd century the doctrine of mystical love was set forth in the sayings of a female ascetic, Rābi'a of Basra, the first of a long line of saintly women who played an important role in the history of Ṣūfiism. Henceforward the use of symbolical expressions, borrowed from the vocabulary of love and wine, becomes increasingly frequent as a means of indicating holy mysteries which must not be divulged. This was not an unnecessary precaution, for in the course of the 3rd century, Ṣūfiism assumed a new character. Side by side with the quietistic devotional mysticism of the early period there now sprang up a speculative and pantheistic movement which was essentially anti-Islamic and rapidly came into conflict with the orthodox *ulemā*. It is significant that the oldest representative of this tendency, Ma'rūf of Baghdad, was the son of Christian parents and a Persian by race. He defined Ṣūfiism as a theosophy and his aim was "to apprehend the Divine realities." A little later Abū Sulaimān al-Dārānī in Syria and Dhu'l-Nūn in Egypt developed the doctrine of gnosis (*ma'rifat*) through illumination and ecstasy. The step to pantheism was first decisively taken by the Persian Abū Yazīd (Bāyezīd) of Bisṭām (d. A.D. 874), who introduced the doctrine of annihilation (*fanā*), *i.e.*, the passing away of individual consciousness in the will of God.

ITS EVOLUTION

In the evolution of Ṣūfiism, influences outside of Islam made themselves powerfully felt. Christian influence had its source, not in the church, but in the hermits and unorthodox sects, to whose ideal of unworldliness Ṣūfiism owed much. More than one Ṣūfī doctrine—that of *tawakkul* in particular—show traces of Christian teaching. The monastic strain which insinuated itself into Ṣūfiism in spite of Mohammed's prohibition was derived, partially at any rate, from Christianity. But Buddhistic influence may also have been at work. Buddhism flourished in Balkh, Transoxiana and Turkestan before the Mohammedan conquest, and in later times Buddhist monks carried their religious practices and philosophy among the Muslims who had settled in these countries. The use of rosaries, the doctrine of *fanā*, which is probably a form of Nirvāna, and the system of "stations" (*maqāmāt*) on the road thereto, would seem to be Buddhistic in their origin. The third great foreign influence on Ṣūfiism is the Neoplatonic philosophy. Between A.D. 800 and 860 the tide of Greek learning, then at its height, streamed into Islam. The so-called "Theology of Aristotle," which was translated into Arabic about A.D. 840, is full of Neoplatonic theories, and the mystical writings of the pseudo-Dionysius were widely known throughout western Asia. It is not mere coincidence that the doctrine of gnosis was first worked out in detail by the Egyptian Ṣūfī, Dhu'l-Nūn (d. A.D. 859), for Ṣūfiism on its theosophical side was largely a product of Alexandrian speculation.

By the end of the 3rd century Ṣūfī mysticism was fast becoming an organized system, with rules of discipline and devotion which the novice was bound to learn from his spiritual director, to whose guidance he submitted himself absolutely, as to one regarded as being in intimate communion with God. At the head of these directors stood a mysterious personage called the *Quṭb* (Axis); on the hierarchy of saints over which he presided the whole order of the universe was believed to depend. During the next two hundred years (A.D. 900–1100), various manuals of theory and practice were compiled: the *Kitāb al Luma'* by Abū Naṣr al-Sarrāj, the *Qūt al-Qulūb* by Abū Ṭālib al-Makkī, the *Risāla* of Qushairī, the Persian *Kashf al-Maḥjūb* by 'Alī ibn 'Uthmān al-Hujwīrī, and the famous *Ihyā* by Ghazālī. They all expatiate on the discipline of the soul and describe the process of purgation which it must undergo before entering on the contemplative life. The traveller journeying towards God passes through a series of ascending "stations"; in the oldest extant treatise these are: (1) repentance, (2) abstinence, (3) renunciation, (4) poverty, (5) patience, (6) trust in God, (7) acquiescence in the will of God. After the "stations" comes a parallel scale of "states" of spiritual feeling (*aḥwāl*), such as fear, hope, love, etc., leading up to contemplation (*mushāhadat*) and intuition (*yaqīn*). It only remained to provide Ṣūfiism with a metaphysical basis, and to reconcile it with orthodox Islam. The double task was finally accomplished by Ghazālī (*q.v.*). He made Islamic theology mystical, and since his time the revelation (*kashf*) of the mystic has taken its place beside tradition (*naql*) and reason ('*aql*) as a source and fundamental principle of the faith.

MANY SHADES OF OPINION

The Ṣūfīs comprise many shades of opinion—from asceticism and quietism to pantheism. The pantheistic type which prevails in Persia throws the transcendental and visionary aspects of Ṣūfiism into undue relief, as in the sayings attributed to Bāyezīd (d. A.D. 874), *e.g.*, "I am the winedrinker and the wine and the cup-bearer"; "I went from God to God, until they cried from me in me, 'O Thou I.'" The peculiar imagery which distinguishes the poetry of the Persian Ṣūfīs was developed by Abū Sa'īd ibn Abi'l-Khair (d. A.D. 1049) in his quatrains which express the relation between God and the soul by glowing and fantastic alle-

gories of earthly love, beauty and intoxication. Henceforward, the great poets of Persia, with few exceptions, adopted this symbolic language. The whole doctrine of Persian Sūfiism is expounded in the celebrated *Mathnawi* of Jalāl-uddīn Rūmī (*q.v.*) but in a discursive and unscientific manner; its leading principles may be stated briefly as follows:

God is the sole reality (al-Ḥaqq) and is above all names and definitions. He is not only absolute being, but also absolute good, and therefore absolute beauty. It is the nature of beauty to desire manifestation; the phenomenal universe is the result of this desire, according to the famous tradition in which God says, "I was a hidden treasure, and I desired to be known, so I created the creatures in order that I might be known." As things can be known only through their opposites, being can only be known through not-being, wherein as in a mirror being is reflected; and this reflection is the phenomenal universe, which accordingly has no more reality than a shadow cast by the sun.

The Ṣūfī theosophy as it appears in Persian and Turkish poetry tends to abolish the distinction between good and evil—the latter is nothing but an aspect of not-being and has no real existence—and it leads to the deification of the hierophant who can say, like Ḥusain b. Manṣur al-Ḥallāj, "I am the Truth." Ṣūfī fraternities, living in a convent under the direction of a shaykh, became widely spread before A.D. 1100 and gave rise to dervish orders, most of which indulge in the practice of exciting ecstasy by music, dancing, drugs and various kinds of hypnotic suggestion (*see* DERVISH).

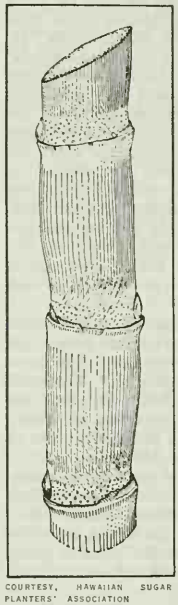
BIBLIOGRAPHY.—The foremost authority is R. A. Nicholson, whose writings (*Selected Poems from the Dīvāni Shamsi Tabriz*, translated with an introduction, etc. [Cambridge, 1898]; *Enquiry Concerning the Origin and Development of Ṣūfiism* [J.R.A.S., 1906, 303 *sqq.*]; *Translation of the Kashf al-Mahjūb, the Oldest Persian Treatise on Sūfiism* [London, 1911]; *Studies in Islamic Mysticism* [Cambridge, 1921]; *The Idea of Personality in Ṣūfiism* [Cambridge, 1923]; together with his more popular treatise *The Mystics of Islam* [London, 1914]) have caused most previous works to have merely an antiquarian interest. He also published the text and translation of *The Mathnawi of Jalálu'ddín Rúmí* (Books i and ii, London, 1925, 1926). An excellent introduction to the study of Islamic mysticism is given in E. J. W. Gibb's *History of Ottoman Poetry*, 1, 33 *sqq.* (London, 1900). L. Massignon's *La Passion d'al-Ḥallaj, martyr mystique de l'Islam*, followed by *Essai sur les origines du lexique technique de la mystique musulmane* (Paris, 1922), is a mine of erudition. Some account of Sūfiism is also given by D. B. Macdonald, *The Religious Attitude and Life in Islam* (Chicago, 1909) and I. Goldziher, *Vorlesungen über den Islam* (Heidelberg, 1925).

SUGAR. Though sugar did not become a major carbohydrate food until the 19th century, even neolithic man found its sweet taste palatable. In prehistoric and classical periods some fruits and honey apparently were the major sweet foods. The present major source, sugar cane, is believed to have originated in the South Pacific or Monsoon Asia. It has been widely distributed for at least two or three millenniums, but seems to have arrived in the Mediterranean part of Europe relatively late, perhaps with the Arabs after A.D. 636. As late as the 14th century its product was a scarce luxury in the western world. The beet, now a second major source, was known as a sweetish vegetable during the middle ages, but was not known as a source of crystalline sugar until late in the 18th century. In some areas, the coco palm, the aloe, the maples and the sorghums, as well as some other plants have served as minor but locally important sources of sweetish foods or sugars. Technologists have learned to convert starch, particularly those of grains, to glucose by treatment with dilute acid.

While the name "sugar" was originally applied to the sweet, crystalline substance, sucrose, derived from the sugar-cane plant, its meaning was broadened to cover many organic compounds of somewhat similar construction. Gathering energy from sunlight by means of their chlorophyll, green plants make sugars out of carbon dioxide and water, such as sucrose or its components, glucose and fructose. By condensing several molecules of glucose into one large molecule, many plants make starch; and by combining one molecule of glucose with one molecule of fructose, many plants make the compound sugar, sucrose, which is the true sugar of commerce. Glucose, fructose, starch and sucrose are all important constituents of the human diet. Starch, sucrose and fructose revert to glucose in the course of digestion and assimilation. The relative sweetness of fructose, sucrose and glucose is as 17:10:7. Crystalline sucrose, the familiar table sugar, is derived on a commercial scale from the juices of the sugar cane and the sugar beet.

SUGAR CANE

The sugar cane is a giant perennial grass producing clumps of solid, prominently jointed stalks, each bearing two ranks of sword-shaped but gracefully arching leaves. Diverse varieties of sugar cane are grown in the widely separated sugar-cane regions of the world. Some of these produce stalks only ½ in. in diameter, while others make stalks up to 3 in. in diameter. Leaves vary in width as the stalks do in diameter. Stalks of all possible colours may be found, but those in red, yellow, purple and green predominate. Varieties with striped stalks are not uncommon. Well-developed stalks reach 10 to 20 ft. in length, but stalks 42 ft. long have been measured. While some sugar cane plants produce real seeds, these are never used for renewing commercial cultures, the cane growers planting and replanting their fields with pieces of stalks. These pieces or cuttings are usually placed lengthwise in furrows and lightly covered with soil. A proper cutting includes several joints, each of which bears a dormant eye or bud. Under favourable conditions, the eyes on a cutting sprout within a few days after planting, each sprout developing roots at its base and growing into a new plant. New shoots spring up from the stubble after a crop of sugar cane has been cut, so it is customary in most sugar-cane growing regions to harvest several crops from a field before it is replanted.



COURTESY, HAWAIIAN SUGAR PLANTERS' ASSOCIATION

FIG. 1. — PORTION OF CANE STALK (CUTTING) SUCH AS IS USED FOR PLANTING

Sugar cane grows to best advantage on a rich, moist soil under sunny skies in a tropical climate. A uniform high temperature, strong sunlight and frequent showers during the growing season are desirable. Winter frosts may do much damage in marginal areas. The cane likes plenty of water and is a voracious feeder, responding quickly when plant foods are applied to the soil about its roots.

Cultural practices are designed to meet local conditions. Rainfall is a major factor and makes for nonirrigated or irrigated types of agriculture which have different costs and result in different yields of sugar per acre. Heavy, rugged equipment is required to prepare the soil. Plowing is usually deep, but a fine-surfaced seedbed is not essential. Cuttings or seed pieces 12 to 24 in. long, cut from the upper ends of stalks, are placed end-to-end in rows spaced 4 to 6 ft. apart and covered with soil 2 to 4 in. deep. Plantings can be made at any time when suitable soil temperature and moisture are obtained.

Subsequent cultivation is primarily for weed control. Maximum cane yields depend on the maintenance of an adequate soil moisture supply from either rainfall or artificial irrigation. The crop transpires an enormous amount of water while growing, and irrigation records show crops requiring as many as 3,000,000 gal. (more than 100 in. per acre) annually for maximum growth. Also, since the crops remove large amounts of plant nutrients from the soils, a permanent soil fertility is maintained through heavy applications of fertilizers. A carefully controlled technique determines what, how much and when fertilizer is needed.

A field of sugar cane ready for harvesting will consist of many primary stalks and subsequent tillers. If the crop is a heavy one, some of these will be erect, but many will be recumbent. Harvesting takes place at from 8 to 30 months after the crop is started. The length of time in which a crop can be grown depends upon the

duration of favourable growing conditions; cold weather causes suspension of normal growth and makes early harvesting necessary.

In Australia, Florida, Louisiana and Hawaii considerable progress has been made in developing cane harvesting machinery, but in most countries sugar cane is still harvested by hand, the stalks being chopped off at or near the surface of the ground and cut up into suitable lengths for convenient handling by men who wield sharp knives especially designed for the purpose. Harvested cane is conveyed from the field to the factory by the methods most conveniently and economically employed on each plantation. On some plantations all of the cane is transported on animal-drawn carts and wagons; on others these have been replaced by motor trucks, while on many of the larger and more modern plantations the cane stalks are loaded onto cars and hauled over railroad tracks by locomotives. On some plantations situated on mountain slopes, the cane is transported in streams of water diverted at a high elevation from a river and conducted through the fields in a series of flumes into which the cane stalks are thrown. These flumes converge at the factory and deliver the cane directly to the crushing plant.

Yields as high as six and seven tons of cane stalks per acre per month of growing time have been harvested, and in the better sugar producing regions somewhat more than a half ton of sugar per acre per month of age at harvest has frequently been recovered.

Sugar Cane Breeding.—The genus *Saccharum* to which sugar cane belongs includes wild as well as cultivated forms. *Saccharum robustum*, one of the wild species, grows in profusion in the New Guinea region. In appearance, it resembles closely certain of the cultivated sugar canes, but lacks the juiciness and the high sucrose content of the latter. A more slender wild relative, *Saccharum spontaneum*, is widely distributed throughout India, Malaya, the Philippines and the East Indian archipelago. Like *S. robustum*, it is pithy and contains little sucrose.

The sweet, juicy, domesticated types are apparently unable to perpetuate themselves in the wild. It seems probable that they owe their origin and development to chance mutation and perhaps also to hybridization among the wild species, followed by long continued selection and propagation at the hand of man.

Both cultivated and wild forms of sugar cane come into flower under the influence of shortening days—in the northern hemisphere from September to December, in the southern hemisphere from March to June. The "tassel" or "arrow" is a handsome plumelike panicle which bears many hundreds of small spikelets. The seed, when seen under magnification, resembles a grain of wheat in miniature; it is so small that nearly 100 sugar-cane seeds would be required to equal the weight of a single wheat grain.

In 1858, J. W. Parris, a planter in Barbados, succeeded in growing to maturity sugar-cane seedlings which had germinated along the borders of a field of sugar cane which had flowered profusely. However, this achievement was not widely publicized and was soon forgotten. When, in 1888, J. B. Harrison and J. R. Bovell in Barbados and, independently, F. Soltwedel in Java reported success in their attempts to grow seedlings of sugar cane, their findings were hailed as a new discovery, affording opportunities to create new cane varieties of exceptional merit.

In Java, interest in sugar-cane breeding was stimulated from the beginning by the occurrence of disease epidemics among the existing varieties which emphasized the need for new disease-resistant types. Resistant varieties were imported from India and elsewhere, and were hybridized with the local varieties. A wild Javanese cane was also found to be disease resistant and was subsequently used extensively as a parent in crosses with the sweet but susceptible cultivated varieties.

Sugar-cane breeding became an important function of experiment stations in many cane-growing countries. New varieties were developed which far surpassed the old varieties in vigour, disease resistance and yield of sucrose. These superior varieties are, for the most part, complex hybrids in which have been combined the desirable qualities of many parent strains.

Insect Pests of Sugar Cane.—There are many insect pests of sugar cane. Several species of moth borers badly damage cane in Louisiana, the West Indies, South America, Mexico, India, Java and the Philippines. The giant moth borer (*Castnia licoides*) is particularly destructive in Trinidad and the Guianas. The frog hopper (*Tomaspis saccharina*) is an important pest in Trinidad. In Australia, heavy losses are caused by the white grubs of several species of Scarabaeid beetles. Other species of the same family are sugar-cane pests in the Philippines, Java, Mauritius and the West Indies. Another conspicuous pest of sugar cane in the Malay archipelago and in Formosa is the woolly aphis (*Oregma lanigera*). Sap-sucking leaf hoppers of the genus *Pyrilla* are especially injurious in India.

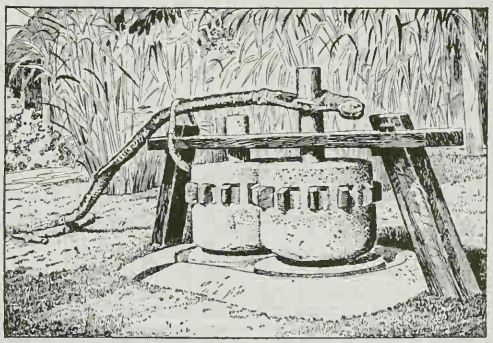
Very little can be done to control these pests, except by biological or natural methods. In Hawaii, this method has been applied with great success, where about a dozen different insect pests attack sugar cane. These include a leaf hopper (*Perkinsiella saccharicida*), a beetle borer (*Rhabdocnemis obscura*), a root grub (*Anomala orientalis*), a leaf roller (*Omiodes accepta*), an aphis (*Aphis sacchari*), two army worms (*Laphygma exempta* and *Cirphis unipuncta*), two species of mealy bugs, an oriental grasshopper, a mole cricket and a bud worm. Most are of oriental, Malayan or South Pacific origin. Satisfactory control of them was permanently effected following a thorough study and importation of the beneficial insects found naturally operating on them in their home countries.

Sugar Cane Diseases.—The sugar-cane plant is attacked by many diseases, a number of which may cause serious losses. Crop damage by disease depends principally on the susceptibility of the variety grown and existing environmental conditions. A disease of major importance in one country may be of minor importance in another.

Sereh disease (cause unknown) in Java once caused great reductions in sugar yield, yet with new varieties it became of little significance. Gumming disease (*Bacterium vasculorum*) has exacted heavy tolls on sugar yields in Brazil, Australia, Fiji, Mauritius and the West Indies. Leaf scald (*Phytomonas rubrilineans*) occurs in Fiji, Australia, Philippines, Java and Hawaii and has, at times, caused severe losses. Fiji disease (a virus) has greatly reduced sugar yields in Fiji, Australia and the Philippines. Mosaic, a serious virus disease, has at one time or another become epidemic in almost every sugar-cane country of the world; in some instances the absolute failure of the industry was averted only by resorting to resistant varieties. Other major cane diseases are eyespot (*Helminthosporium sacchari*), Pythium root rot (*Pythium arrhenomanes*), smut (*Ustilago sacchari*), chlorotic streak (cause unknown), and red rot (*Colletotrichum falcatum*).

The physiological diseases (nonparasitic) caused by unfavourable environmental factors such as climatic conditions, physical nature and chemical composition of the soil, are often responsible for unsatisfactory sugar yields.

Sugar-cane diseases have been most effectively controlled through the development of resistant varieties by breeding and selection. The application of fungicides as dusts and sprays has



COURTESY, EXPERIMENT STATION, HAWAIIAN SUGAR PLANTERS' ASSOCIATION

FIG. 2.—ANCIENT STONE-ROLLER CANE MILL

been of little or no value in the control of cane diseases. Rigid quarantines governing the interchange of cane cuttings and seed has aided greatly in preventing the spread of cane diseases.

Cane Juice Extraction.—The earliest cane mills consisted of two vertical rollers (fig. 2) of wood or stone, geared to a long lever or sweep and turned by hand or animal power. The cane stalks were fed into the narrow opening between the rollers, but only a small amount of the total juice in the cane was extracted. This type of mill was followed by two rollers set horizontally one above the other and later by the three-roller mill (fig. 3) with two bottom rollers and a top roller. By means of pinion gears fitted to extensions of the shafts of each roller, the power that turns the top roller is transmitted by the pinions to the two lower rollers, so that the three rollers rotate slowly at the same speed. The cane feed enters the mill between the top and front, or feed, roller and emerges between the top and back, or discharge, roller, thus receiving two separate crushings in each three-roller unit, the juice from each of these units flowing into separate receivers.

Modern mill roller sizes range from 34 in. to 37 in. in diameter and 78 in. to 84 in. in length and have suitable surface and juice grooves cut into their surfaces to improve their grip on the feed and allow easy drainage of the juice. They have from $2\frac{1}{2}$ to $3\frac{1}{2}$ r.p.m. with surface speeds of 25 to 30 ft. per minute.

Hydraulic pressure from an accumulator is applied to jacks acting downwardly on each end of the top roller, which is free to move vertically, thus exerting a heavy pressure on the feed of crushed cane passing between the rollers. The top roller thus floats on the cane blanket, changing its vertical position slightly with changes in thickness of the blanket, under a pressure of 70 to 80 tons per foot of its length or a total pressure of 500 to 550 tons. A modern two-roller unit with coarse grooves is known as a crusher, and a coarse grooved three-roller unit set at the head of a train of mills is also called a crusher.

Cane milling plants in modern factories consist of the slat carrier upon which the cane is unloaded from the cars or trucks, one or two sets of knives along the carrier to act as levellers to compact the feed, a crusher or a double crusher (two two-roller units), and then four to six units of three-roller mills in tandem, with suitable bagasse conveyors between each unit of the train. In some factories a shredder of the high-speed revolving-hammer type is placed between the crusher and the three-roller mills to disintegrate the cane further. A conveyor then carries the fibrous residue or final bagasse to the boiler furnaces or to storage. This bagasse usually supplies an excess of fuel for normal steam requirements. Larger factories have two complete tandems and a few have three tandems, crushing up to 6,000 or more tons of cane in 24 hours.

To facilitate extraction of the sugar-containing juice of the cane, the bagasse on the conveyor before entering the last mill unit receives and absorbs a spraying of very hot water equal to one-fourth or one-third the original weight of the cane (*see* fig. 3). This system of dilution is known as compound imbibition, in which the last mill juice, which includes the water, is returned and absorbed in the bagasse entering the next to last mill. This resulting dilute juice is again returned to the bagasse blanket before the next earlier unit and is repeated until all mills except the first three-roller unit receive bagasse containing diluted juice. The combined juice of the crusher, first and second mills flows into a single receiving tank, and this mixed juice is pumped to the boiling house. If compound clarification is practised (*see* below), the mill juices are divided; the crusher with first mill juice (undiluted) is primary juice and the second mill juice containing all the dilute juices is secondary juice. The percentage of sucrose extracted in modern milling plants ranges from 96 to as high as 99 and varies with the amount of field trash and soil accompanying the cane, and with the tonnage rate, number of mill units, hydraulic pressures, dilution used and the skill of the engineering staff. The final bagasse contains from 38% to 45% water, from 1% to 4% sucrose, about 50% fibre and around 2% ash. Fresh bagasse, as burned, has a fuel value of 2,700 to 3,000 British thermal units (B.T.U.) per pound and on a dry, ash-free basis about 8,300 B.T.U. per pound.

The strained mixed juice from the milling plant contains from .2% to nearly 1% suspended matter, is of dark brown colour, full of air bubbles and acid in reaction with a *p*H of 4.5 to 5.4. The density range is 12% to 14% dissolved solids, sucrose 10% to 12% and purity (sucrose per cent solids) 80 to 86; these figures vary greatly in different localities. There are small percentages of glucose, fructose, organic acids, ash, nitrogen compounds, pectin, gum and colouring matters. The suspended matter consists of field soil, bits of cane fibre and a small amount of cane wax. The crushed cane residue, bagasse or megasse is used as fuel under the boilers, as wall board and, more potentially than actually, as cellulose for paper.

THE SUGAR BEET

The beet, *Beta vulgaris*, long used as a garden vegetable and for cattle fodder, was not prized for its sugar content until a full half century after the discovery in 1747 by A. S. Marggraf that the sugar in beet juice was the same as that in cane. Experimental and promotional work by Franz Carl Achard about 1799 in which the king of Prussia became interested and work by J. P. B. Delessert was followed by a decree by Napoleon that 70,000 ac. be planted to beets as soon as possible. Meanwhile the development of the White Silesian beetroot, from which all modern strains have come, was carried on by Moritz von Koppy. By 1840 the sugar beet provided about 5% of the world sugar production and by 1880 exceeded cane as a producer, in which leading position it remained in most years until World War I.

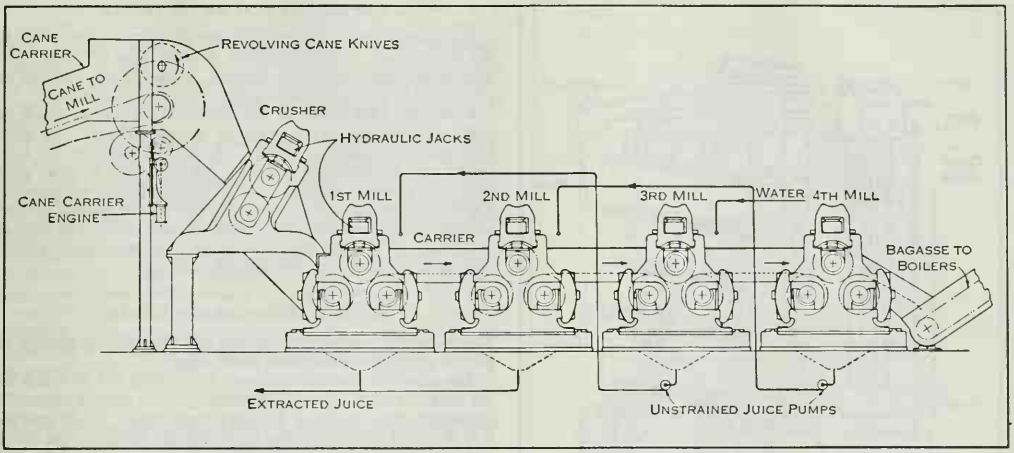
The sugar beet has a wide range in the temperate mid-latitudes but sugar content develops best in areas with moderate summer temperatures averaging 63° F. to 73° F. for the three summer months and with well-distributed growing-season rain-

TABLE I.—*Sugar Beets: Acreage and Production in Specified Countries*

Continent and country	Acreage (in 000 ac.)*					Production (in 000 short tons)				
	Average		1950	1951	1952†	Average		1950	1951	1952†
	1935–39	1945–49				1935–39	1945–49			
North America										
Canada	50	66	102	93	93	504	690	1,128	966	1,009
United States	827	755	925	691	678	9,505	10,262	13,535	10,485	10,334
Total	877	821	1,027	784	771	10,009	10,952	14,663	11,451	11,343
Europe										
Austria	102	49	69	95	101	1,202	322	938	1,136	1,213
Belgium	124	119	155	161	163	1,603	1,626	2,948	2,048	2,179
Denmark	98	117‡	180‡	177§	178	1,657	1,790‖	2,890‖	2,706¶	2,720
Finland	10	13	24	25	26	96	116	308	168	193
France	791	715	976	1,005	1,030	9,976	7,942	14,965	13,041	11,375
Germany, Western	333	355	477	551	550	4,900	4,010	8,334	8,444	8,377
Ireland	55	70	60	60	50	590	672	659	644	550
Italy	312	231	429	490	554	3,422	2,640	4,925	6,567	6,200
Netherlands	106	111	165	166	154	1,700	1,888	3,417	2,700	2,866
Spain	161	190	230	264	207	1,245	1,515	1,525	2,026	3,307
Sweden	128	126	129	134	134	2,089	1,909	2,124	1,909	2,001
Switzerland	6	14	15	15	15	91	206	205	231	226
United Kingdom	344#	415	425	421	411▽	3,395	4,401	5,822	5,080	4,816
Yugoslavia	77	159	243	251	200	658	1,041	938	2,135	1,280
Total above	2,649	2,684	3,577	3,815	3,803	32,684	30,078	50,128	48,855	47,303
Total other European countries	1,520	1,750	2,248	2,303	2,308	18,402	13,331	20,843	20,633	19,100
Total Europe	4,169	4,434	5,825	6,118	6,171	51,086	43,409	70,971	69,488	66,523
U.S.S.R.	3,096	2,492	3,200	3,300	3,300	19,982	12,830	19,000	17,600	20,500
Other Countries										
Iran△	35	63	85	104	111	144	265	416	546	584
Japan	42	34	35	33	38	310	105	193	236	266
Turkey△	63	115	126	125	123	434	724	943	1,500	1,100
Uruguay	3	6	7	10	10	16	20	60	99	100
Total	143	218	253	272	282	904	1,114	1,612	2,381	2,050
Grand total	8,285	7,965	10,305	10,474	10,524	82,071	68,305	106,246	100,920	100,416

*Area estimates are for harvested areas unless otherwise stated. †Preliminary. ‡Includes about 5,000 acres sown in Denmark to be processed in Sweden. §Includes about 7,000 acres sown in Denmark to be processed in Sweden. ‖Includes about 57,000 short tons processed in Sweden. ¶Includes about 65,000 short tons processed in Sweden. #Topped beets. △Harvested for sugar only.

Source: *Foreign Crops and Markets*, vol. 65, no. 16, p. 348, U.S. Department of Agriculture (Washington, Oct. 1952).



COURTESY, HONOLULU IRON WORKS

FIG. 3.—CANE MILLING PLANT

fall, unless irrigation is practised. In the United States it is most favoured by conditions on the northern margin of the Corn Belt. Soils should be fertile, deep, well drained and without hardpan, for the plant is a deep-rooted one.

Production, shown in Table I, is predominantly European, a result more perhaps of national policies than of favourable natural conditions. Three areas are outstanding; the Kiev and Podolia section of southwestern U.S.S.R., the Magdeburg area of Germany and the chalky plains of northern France.

Growing and Harvesting the Beet.—Sugar-beet seed must be sown each year to produce the annual beet crop. Since 1942 not whole seed but treated seed has been largely used, the treatment consisting of removal of corky outer layers, reduction to uniform size and in the number of seed germs per unit and treatment with fungicides and insect repellents. The time of planting depends on the weather and area but is mostly done in the spring. Planting is in rows spaced 12 to 28 in. apart. The seedbed is carefully prepared and in some areas will have been prepared for irrigation as well. The stand in the row is later thinned, or blocked by hand hoe or mechanically so that the young beet plants are spaced about 10 in. apart in the row, giving a density of as many as 30,000 to 40,000 plants per acre. Weeds must be severely controlled with hoe, cultivator or chemical spray. Fertilizer is applied to the seedbed or as a side dressing. In all except humid areas irrigation is generally practised, amounting to as much as 30 in. of water in dry areas; each plant may require as much as 20 gal. of water during its growth. Harvesting, formerly done by hand but now considerably mechanized in some areas, consists of a process whereby each beet is lifted and topped. Topping consists of cutting away a top layer of the beet and the attached leaves which are used for livestock feed, being almost as nutritious as alfalfa hay. The decapitated roots are then picked up, perhaps mechanically after windrowing, and hauled by truck or wagon, depending on the area, to the receiving station of a beet sugar factory if one is near, or to a concentration station for shipment to a more distant factory. There the beets are stored in piles 15 to 20 ft. high and more than 100 ft. wide, with perhaps controlled ventilation and protection against freezing. The accumulation of such a stockpile is necessary to prolong the processing season much beyond the short harvest season.

Beet Seed Production.—The breeding of the sugar beet and the production of high-grade seed is a scientific story in itself. Much of the research, hybridization and multiplication of desirable strains had been carried on in Europe, though the U.S. moved forward in that respect particularly after the 1930s. The seed beets are taken in from the field over winter and tested for sugar bearing qualities; the desirable ones are replanted in the spring, so that seed may be harvested in the late summer or autumn of the second year. Trade in sugar beet seed has been an important part of world seed trade.

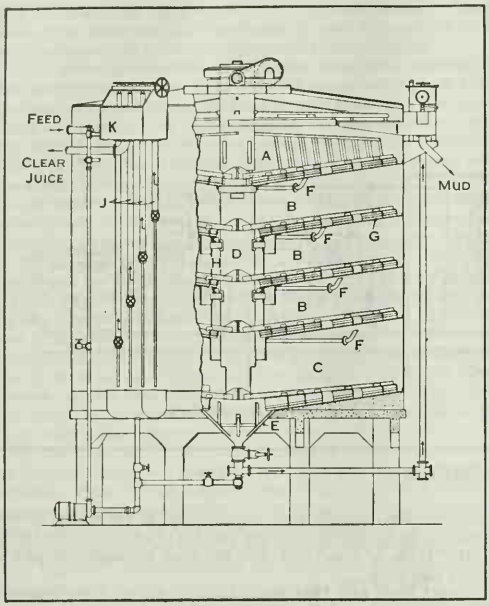
Insect Pests and Diseases.—About 40 insects cause economic damage to the sugar beet. The beet webworm causes damage over wide areas unless controlled with DDT, arsenicals or some other new compounds. The beet army worm and others are destructive but may be controlled by ditching and the use of crude oil. Wireworms may be controlled by fumigating the soil with ethylene dibromide and by DDT. The crown borer causes damage in the west. Thrips, mining flies, several kinds of beetles and grasshoppers damage the leaves; nematodes, centipedes and cutworms are sometimes destructive.

Seedlings may be damaged or destroyed by damping off or black root fungi and several other organisms. Foliage may be severely damaged by curly top, a virus disease carried by a small leafhopper. Leafspot is a most serious disease in the middle west; mosaic and mildew take a toll in some areas at times. Root rots, if conditions favour them, may cause significant loss.

Juice Extraction.—As the beets are flumed into the factory, they may carry some dirt, small stones and other field trash with them. These beets are fed, through a series of trash- and stone-catchers, into a tumbling washer for removal of adhering dirt. The cleaned beets are then weighed and delivered to the slicing machines.

Unlike sugar cane, which is crushed to extract the juice, sugar beets are sliced with a minimum rupture of plant cells. This procedure permits later extraction of the juice (by diffusion through the walls of the plant cells) in a higher degree of purity than would be obtainable from crushed beets. The first operation in the extraction of the juice is slicing the beets. The modern slicer is a revolving drum with a series of inwardly slanting grooved knives forming part of its periphery. The beets are fed into the drum and the V-shaped chips or cossettes (which somewhat resemble shoestring potatoes) cut from them are discharged from the outside of the drum, and conveyed to the diffusion battery for juice extraction.

This diffusion battery consists of an endless connected series of 10 to 14 vertical cylindrical tanks (commonly called cells) each fitted with top and bottom doors and having a capacity of two to five tons each. Inside of each cell is a network of loose chains to hold the cossettes from packing too tightly in the cell. Between each cell and the next in the series is a heater for maintaining the temperature of the extracted juice, together with necessary pipes, valves and fittings. This endless chain of cells is some-



COURTESY, PETREE & DORR ENGINEERS, INC.

FIG. 4.—CANE JUICE CLARIFIER

A. Flocculating cell for entering juice. B. Shallow settling cells. C. Mud-thickening cell. D. Rotary central feed-tube. E. Thickened mud discharge cone. F. Clear juice draw-off. G. Rotary squeegee sediment scrapers. H. Annular mud discharge opening. I. Foam canal and mud-pump discharge. J. Clear juice to discharge box K. K. Discharge box.

times arranged as a circular, sometimes as a straight, battery. The operation of the battery is relatively simple. To start operating, any cell in the battery is filled with the sliced beets, or cossettes. That cell is closed and hot water is introduced until all voids are filled.

While the first cell is being filled with water, the next cell is being filled with cossettes, and when this latter cell is closed it is connected to the first cell, and so on around the series. In this fashion hot water enters any chosen unit, leaches or diffuses sugar from the contained beet cossettes, and this solution proceeds, through a heater, into the next filled unit. The water always enters the oldest or earliest filled unit of the series thus passing from the unit containing exhausted cossettes toward the most recently filled one. This solution of sugar, called raw juice, which leaves the most recently filled cell, has a sugar concentration approximating that within the plant cells of the newly sliced beets, and is periodically withdrawn, through a measuring device, for purification and sugar recovery.

At the end of a specified extraction period, the exhausted cossettes are discharged from the oldest, or exhausted, unit, and the hot water supply transferred to the next unit in the series. This empty unit is then refilled and becomes the newest unit.

The diffusion battery is thus an endless chain of units or cells, within which hot water is passed over sliced beets, starting with nearly exhausted cossettes and moving progressively over cossettes of increasing sugar content, and from which the resulting raw juice is periodically withdrawn.

By-Products.—Beet pulp, either moist or dried, perhaps mixed with molasses, is used as a livestock food. Pulp drying is highly developed, and pelletizing is done with some of the material to meet specialized requirements of sheep feeders. The wet pulp is stored in silos until used. Several kinds of waste may be processed particularly to recover glutamic acid and other amino acids when it proves to be commercially feasible to do so.

THE MANUFACTURE OF SUGAR

The early stages of the manufacture of sugar are carried on in the general area of the production of sugar cane or beets. The later stages, particularly the refining, is much more likely to be carried on in port areas near great consuming populations.

Chemical Treatment of the Juice.—This discussion will be pointed particularly to the treatment of cane juice but beet juice treatment differs only in details. As part of the dissolved non-sugars are colloidal (*see* COLLOID), the juice must undergo some form of chemical treatment before it can be filtered satisfactorily. The natural acidity of the juice must also be neutralized, because sucrose decomposes on heating with the acid juices, the sucrose combining with water and breaking up into two simpler sugars, glucose and fructose, as shown in the equation:

Sucrose Glucose Fructose

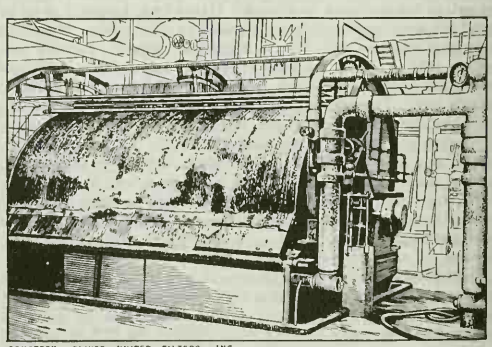
$$C_{12}H_{22}O_{11}+H_2O=C_6H_{12}O_6+C_6H_{12}O_6.$$

This is termed inversion of sucrose, and the resulting mixture of glucose and fructose is termed invert sugar.

The object of chemical treatment is fourfold: (1) to neutralize the natural acidity (as just explained); (2) to remove suspended impurities, thus rendering the juice transparent; (3) to precipitate dissolved nonsugars, thus increasing the purity of the juice; and (4) to decolourize the juice in the manufacture of white sugar.

The chemical agents used are: quicklime (CaO), usually added in the form of lime cream or milk ($Ca(OH)_2$); sulphur dioxide gas (SO_2); carbon dioxide gas (CO_2); and, occasionally, phosphoric acid (H_3PO_4). The effect of these chemical agents is completed by heating the juice to the boiling point in an apparatus called a juice heater, the juice being pumped at high speed through several series of long brass tubes, heated externally by low-pressure steam or evaporator "vapour." Certain impurities, previously dissolved in the juice, are thus rendered insoluble and form a dark coloured precipitate which occludes most of the colloidal material and which can be separated from the juice in various ways: (1) by causing it to rise to the surface (with the escaping air bubbles), forming a scum which can be removed by skimming ladles; (2) by allowing it to subside (after separation of the air) and removing it as a sediment; (3) by filtration through cloth or other filtering media; (4) by centrifugal action.

In the manufacture of raw sugar, a dilute milk of lime solution of 10° to 12° Brix (6° Baumé) to facilitate control of the reaction, is mixed with the juice to approximately p_H 8.0 which is equivalent to about 1½ lb. of available CaO per ton of cane. The mixture is heated to the boiling point and discharged into continuous clarifiers or into settling tanks (subsiders or clarifiers), each holding 1,000 gal. or more and filled in rotation. After subsidence for about an hour, the clear juice is drawn off into one gutter and the sediment or settlings into another. The above intermittent system of settling has been largely changed to continuous settling (*see* fig. 4) in one to three units of much larger size, giving clearer



COURTESY, OLIVER UNITED FILTERS, INC.

FIG. 5.—ROTARY DRUM FILTER

juices and more compact settlings. If compound clarification is practised, the high density and high purity primary juice and the lower density, lower purity secondary juice are clarified separately through continuous clarifiers. The latter juice is then reclarified by mixing with the juice fed to the primary clarifier and the primary settlings mixed and resettled with the secondary juice in the secondary unit.

The settlings, in modern practice, are filtered by rotary drum vacuum filters (*see* fig. 5) of large capacity and high efficiency. The older practice was through filter presses (*see* FILTRATION). Clear, settled juice and the clear portion of the filtered juice are mixed and passed to the multiple-effect evaporator (described below).

In the manufacture of white sugar, a more thorough purification of the juice is essential. Two methods are used, acid sulphitation and carbonation. In the former, from 4 to 5 gal. of lime cream about 27° Brix (15° Baumé) are added per 1,000 gal. of juice, and sufficient sulphur dioxide gas to precipitate this lime, with the formation of a granular precipitate of calcium sulphite ($CaSO_3$), as shown in the equation

$$CaO + SO_2 = CaSO_3.$$

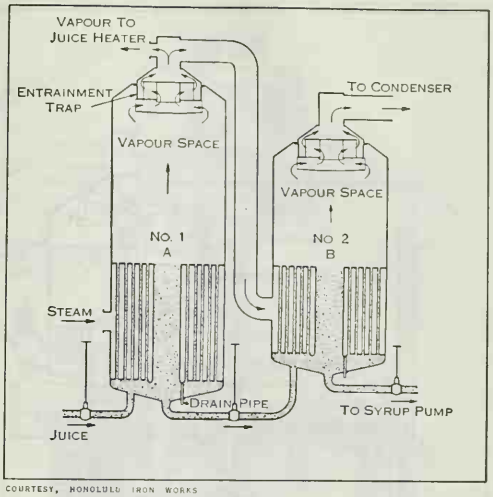
The gas may be added either before or after liming, the final result being the same. The juice is then heated and subsided.

The diffusion juice of the sugar beet is purified by a process known as carbonation. This consists essentially in defecation of the hot raw juice with either milk of lime or a saccharate-of-lime milk followed by precipitation of this lime by a gas rich in carbon dioxide obtained from the factory limekiln. The amount of lime used is around 3% of the weight of the beets sliced. By this treatment much of the colloidal and other impurities are removed with the precipitated carbonate of lime. This carbonation procedure is customarily conducted in two stages, with an intermediate removal of the first precipitate. Such divided procedure prevents possible resolution of some of the first-precipitated impurities, which would occur if the process were conducted in only one stage. Removal of the precipitated impurities from both the first and second carbonation processes is accomplished by filtration or analogous means, commonly by a combination of settling and filtration. The clarified juice from these carbonation procedures is then customarily treated with sulphur dioxide gas in order to bleach some of the residual colouring matters and to remove small remaining amounts of lime. This treatment is again followed by filtration, yielding a pale yellow solution of sugar known as thin juice.

Formerly the juice was carbonated in batches. An important improvement was the adoption of continuous carbonation whereby the juice, lime and gas are all introduced continuously into the carbonator, and the treated juice is removed continuously. This procedure results in a more uniform precipitate than did the older methods and permits use of settling devices instead of filters for separating much of the precipitate.

Evaporation of Water.—A large quantity of water must be evaporated from the purified juice before the dissolved sucrose can begin to crystallize. To avoid destruction of sucrose at high temperatures the juice is boiled in closed vessels under vacuum, thus lowering the boiling point. Further, in order to economize fuel, multiple-effect evaporators are used.

Fig. 6 is a sectional view of a double-effect evaporator with an enlarged first effect, for vapour heating of mixed juice and boiling vacuum pans at certain stages. The general principles are the same with three or more effects, and vapour can be removed from any effect except the last for juice heating. The lower part of each vessel contains numerous vertical brass tubes, their open ends passing through perforations in two horizontal plates. The space between these plates forms a steam drum having a centre well open at each end, a steam inlet and condensate drain. The centre well increases evaporator capacity by causing rapid circulation of the juice up through the tubes and down the well. The large space above the drum is filled with steam or vapour generated by the boiling juice. Steam entering the drum of No. 1 condenses on the exterior of the tubes, transmitting its latent heat to



COURTESY, HONOLULU IRON WORKS

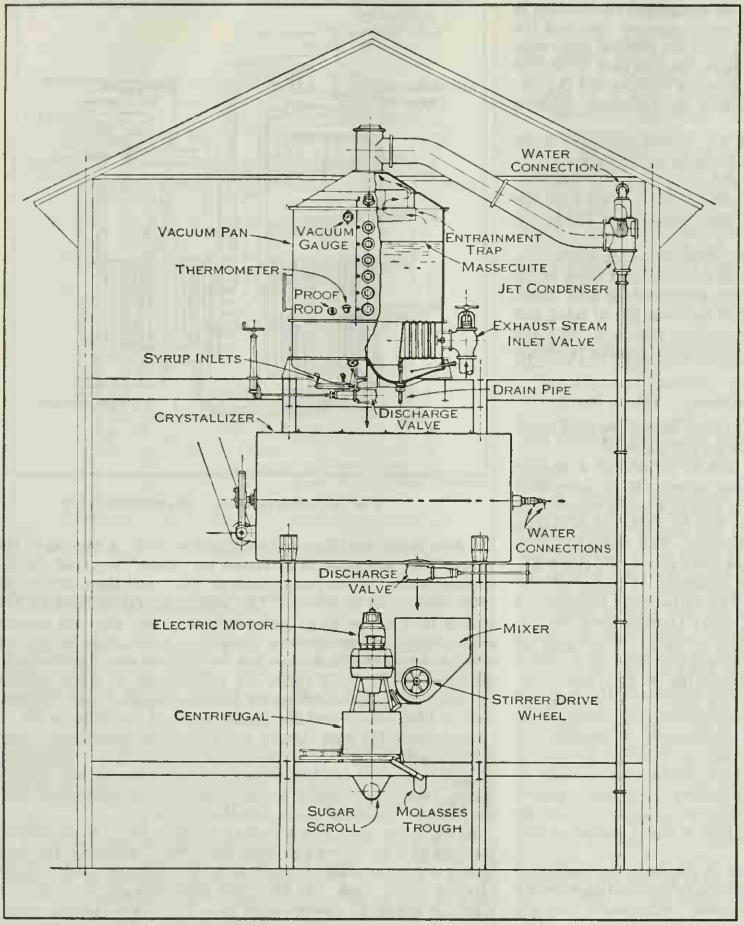
FIG. 6.—DOUBLE-EFFECT EVAPORATOR

the juice inside the tubes and causing it to boil. A portion of the vapour thus generated is extracted for "extra" use, and the remainder passes to the steam drum of No. 2 and there acts in the same manner as in No. 1. The vapour generated here by the boiling juice passes to a "jet" condenser which does not require an air pump, and maintains a vacuum of 25 to 26 in. in the last effect which in this instance is No. 2. This vacuum fixes the boiling point of the juice there, the temperature at which vapour (leaving No. 1) can condense on the tubes in No. 2, and the pressure of that vapour and the boiling point of the juice in No. 1. Consequently, the high vacuum and low boiling point in No. 2 produce a lower vacuum and higher boiling point in No. 1. Quadruple- or quintuple-effect evaporators are in general use, giving approximately four or five pounds respectively of water evaporated from the juice, per pound of steam admitted to No. 1.

In fig. 6 the level of the boiling juice in No. 2 is maintained constant at B by drawing in juice from No. 1; similarly, the level in No. 1 is maintained constant at A by drawing in juice from an external supply tank. At the same time, concentrated juice, or syrup, is drawn off continuously from No. 2 by a pump, and at such a rate that the density of the syrup in No. 2 and of that drawn off remains constant at between 60° and 65° Brix (32° and 35° Baumé). This syrup contains about 60% of dissolved solids (including sucrose). It is, therefore, an unsaturated solution, free from crystals.

Crystallization of the Syrup.—The final stage of evaporation, accompanied by the crystallization of the syrup, is carried out in vacuum pans. This apparatus is shown in fig. 7. Part of the interior, the steam or vapour-heated tubes forming the "calandria," centre well and cone bottom are shown on the right. Above and to the right, is the jet condenser (shown in section) wherein the vapour generated by the boiling syrup is condensed by a series of water jets. A vacuum of about 26 in. is maintained inside the pan by the suction of the condenser, thereby lowering the boiling point of the syrup to about 130° F. Cane and beet sugar factories are equipped with three or more pans, each discharging from 10 to 30 tons, or more, of crystals after each boiling operation is completed.

Sufficient syrup is first drawn into the pan to occupy only a part of its total capacity, and steam is admitted to the calandria which is covered by the syrup. After a period of rapid boiling, the syrup becomes supersaturated and microscopic crystals begin to form. This is called the graining point. It is important to produce a sufficient quantity of grains as rapidly as possible in order



COURTESY, HONOLULU IRON WORKS

FIG. 7.—CROSS-SECTION THROUGH RAW SUGAR BOILING-HOUSE

that the crystals may be of nearly uniform size during all stages of growth. To obtain a more rapid formation of grain in a minimum of time, shocking is commonly employed by introducing into the supersaturated syrup, just prior to the natural graining point, a few ounces of powdered white sugar. The number of grain introduced is immaterial: they induce formation of new crystals. Having formed sufficient grain by either method, the operator stops the formation of new crystals by quickly reducing the supersaturation with a little water or a quantity of unsaturated syrup from the supply tank. The boiling then proceeds on syrup with either intermittent or set feed, holding the amount of supersaturation within narrow limits, causing the previously formed grain to grow until the pan is filled.

The boiling time for high purity syrup is from 3 to 4 hours, and 8 to 12 hours for the lowest purity products. To make a large-grained final product, the operator starts crystal growth with a relatively small quantity of grain, thereby increasing the rate of growth of each grain. Conversely, to make a small-grained final product, he starts with a larger quantity of grain, thereby reducing the rate of growth of each grain.

The time available for growth can be extended by transferring half the contents of a full pan to a second, empty pan, the two halves then being boiled simultaneously with further additions of syrup to each, causing continuous growth of the original grain until both pans are filled. During the process of boiling, the operator is guided by the appearance of small samples of the crystallized syrup which he removes from the pan at short intervals by means of a proof stick. The samples are spread upon a sheet of glass and examined by transmitted light from a lamp. When the pan is full, the mixture of crystals and mother liquor (called massecuite) is finally concentrated until it contains from 9% to 11% of water. It is then semifluid and contains about 65% by weight of crystals and 35% of mother liquor. Air is now admitted to the pan, the bottom gate opened and the massecuite discharged. There are several boiling-control devices in use. They operate by control of vacuum, elevation of boiling point (massecuite temperature above water temperature) and electrical conductivity, and are either automatic or of visual aid to the pan operator.

The mother liquor separated in the centrifugals, described below, contains a high percentage of sucrose. It is reboiled in the vacuum pan where it crystallizes, forming second-grade massecuite. This is treated in the centrifugals, where it separates into second-grade crystals and second-grade mother liquor. The last is reboiled in the vacuum pan to form third-grade or final massecuite which is discharged into large cylindrical crystallizers, each fitted with a slowly rotating central shaft carrying water coils or pipes of ample surface ratio, which control the massecuite temperature and act as stirrers. Fig. 8 shows one of several types for rapid cooling of high-density massecuite to a predetermined minimum temperature, such as 120° F., holding it at this temperature for a time interval with warm water, then reheating with warmer water to a predetermined optimum saturation temperature. This reduces the sucrose content and purity of the final mother liquor or final molasses to a much lower minimum and in far less time than by the older type of air-cooled crystallizer. The final massecuite is purged at its original high density in modern high-speed centrifugals running 1,800 to 2,200 r.p.m. (*see* fig. 7 and 9). Molasses true purities (sucrose per cent true solids) of 30 to 35 are obtainable. Besides sucrose and reducing sugars it contains various organic and inorganic impurities derived from the original cane juice.

To save time and fuel, it is a common custom to grain, usually by shocking, only the final massecuite. The purged final crystals are mixed with syrup to a heavy magma and used for seeding the two higher grade massecuites. The first two grades produce raw sugar, and any third-grade sugar not required for seed is melted and returned to process in one of the high-grade pans.

Separation of Crystals from Mother Liquor.—This operation, called curing or purging the sugar, is effected by centrifugal action in a number of machines operating simultaneously. Each

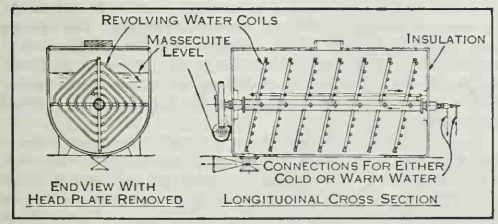
(fig. 9) consists of a vertical metal basket or drum which is suspended on a central vertical spindle and for high grades is rotated at speeds of 1,200 to 1,400 r.p.m.; low grades, 1,800 to 2,200. The cylindrical wall of the basket is made of perforated steel plate and fitted with an inner lining of brass wire, $\frac{1}{8}$-in. mesh screen, over which is fitted a very finely perforated round-hole or slotted screen. After the basket is set in motion, a charge of massecuite enters through a central opening at the top, and centrifugal action forces the massecuite against the perforated lining which acts as a strainer, allowing the fluid mother liquor to pass through but retaining the crystals. The second grade is washed with a few pints of warm water to improve purging. Separation in first and second grades is complete in about two minutes, but the lowest grade requires a much longer time. Fig. 7 shows positions of vacuum pan, crystallizer used for final massecuite, massecuite mixer, centrifugals and scroll conveyor for the sugar.

In the manufacture of raw sugar, the above treatment yields crystals containing from 96% to 98% of sucrose and, inversely with the sucrose, should not contain over 1.0% to 0.5% of water, to avoid deterioration in storage. The colour, moisture and odour of this sugar are largely due to residual mother liquor adhering to the crystals, which cannot be completely separated by centrifugal action alone. The conveyer carries the sugar from the centrifugals to a storage bin from which the sugar gravitates through chutes and weighing machines into bags or into bulk sugar storage awaiting shipment in bags or bulk form to the refinery.

In the manufacture of direct process white sugar, the separation of the mother liquor is carried much further by spraying a larger quantity of water over the layer of sugar in the centrifugal basket while rotating at full speed. Former practices of repurging the sugar in a second set of centrifugals and of spraying sugar with bluing water, to conceal yellowish tint, were largely abandoned with improvements in the technique of manufacture. After passing through a rotary drier the dry white sugar contains from 99.5% to 99.9% of sucrose.

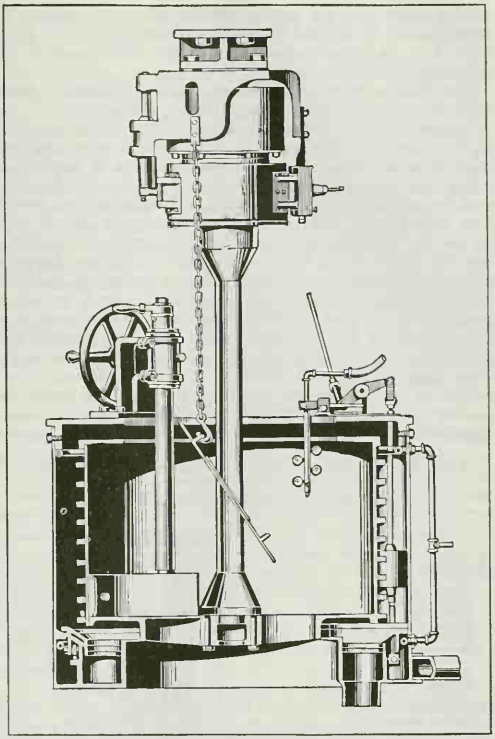
Desugarization of Molasses.—After all possible sugar has been crystallized from the purified beet juices, there remains an end product, molasses, containing about 50% sugar. Since the molasses produced in a factory may retain 15% to 20% of the sugar of the original beets, many factories recover this sugar by a special desugarization process which consists essentially of the following steps: finely powdered quicklime is added to chilled diluted molasses and combines with most of its sugar to form an insoluble saccharate of lime. This precipitate is filtered off; most of the molasses impurities remain in the filtrate from which small additional amounts of saccharate may be recovered by heating it. The saccharate is mixed with factory sweet waters to form a saccharate milk which is then introduced into the carbonators in place of the milk of lime usually added at that point. The carbon dioxide gas, added here, breaks up the lime-sugar combination, releasing the sugar which is recovered in the usual operations of the factory. This method of desugarization is known as the Steffen lime process, after its inventor Carl Steffen, and is much used in the United States. Plants for the desugarization of beet molasses by means of baryta, or of strontia, are in operation in a few localities. (H. L. Ln.; A. A. Bt; M. H. St.; J. K. R.)

Refining of Raw Sugar.—The objects of refining raw cane sugar are to remove impurities and to prepare the pure sugar in the various forms required by consumers.



COURTESY, HONOLULU IRON WORKS

FIG. 8.—TEMPERATURE CONTROLLED CRYSTALLIZER



COURTESY, WESTERN STATES MACHINE CO.

FIG. 9.—SUGAR CENTRIFUGAL BASKET SECTION

The cane sugar refining process consists essentially of three principal operations: (1) removing adhering molasses from the raw sugar crystals and melting the crystals in water to form a solution; (2) filtering and decolourizing the solution; and (3) recrystallizing the sugar from the decolourized solution.

Dissolving or Melting the Raw Sugar.—In modern practice, dissolving is usually accomplished in three steps, the first two of which are for the purpose of eliminating at least 75% of the impurities which accompany the raw sugar. First, the raw sugar is mingled with a hot syrup, previously removed from raw sugar, to form a magma which is fed into centrifugal machines. Second, the magma is spun in the centrifugals to remove the syrup, and the crystals are sprayed with water to wash off adhering molasses. Third, the washed raw sugar is discharged from the centrifugals and is dissolved in hot water to produce what is termed washed raw sugar liquor.

The mingling, washing and separation of washed sugar crystals from the syrup constitute what is called the affination process, and the resulting syrup is known as affination syrup or wash syrup. This process results in the production, from 100 tons of raw sugar (96% sucrose), of about 85 tons of washed raw sugar crystals (99% sucrose) and about 15 tons of dissolved solids (85% sucrose and 15% impurities) in the affination syrup.

The washed raw sugar is melted in large cylindrical melting tanks with hot water or, preferably, with high-purity, low-density sweet water, a by-product resulting from other refinery operations. The melting tanks are usually capable of holding from 5 to 10 tons of sugar together with the necessary quantity of solvent to give a washed raw sugar liquor containing 60% to 65% sugar in solution. Tanks are equipped with mechanical stirrers, per-

forated steam coils for heating and the necessary facilities for regulating the density and temperature of the liquor, which properties have an important bearing upon the effectiveness of the subsequent filtration process.

The affination syrup is handled separately, its impurities eventually being eliminated by recrystallization of its sucrose.

Filtering and Decolourizing the Liquor.—The washed raw sugar liquor ranges from light brown to dark brown in colour and contains variable amounts of insoluble impurities in suspension. These impurities are removed by forcing the liquor under pressure through cotton or metal cloth in leaf filters or filter presses (*see* FILTRATION), this apparatus having almost entirely supplanted the old style bag filters. A small quantity of filter aid is regularly added to the liquor to maintain satisfactory flow rate and clarity of filtrate. The filter aid most commonly employed is diatomaceous earth, or kieselguhr, although a small amount of paper pulp is also used. This filter aid forms a porous layer upon the cloth of the filter leaves or plates, and the suspended impurities entering with the liquor are retained in this layer as the liquor passes through under pressure. The filtered liquor is transparent but has a colour that may range from straw to brown, depending upon the character of the raw sugar being processed. Kieselguhr filter aid is generally discarded after use, but in some refineries is regenerated by the application of heat in a kiln, where the organic impurities removed from the liquor are burned out. Removal of similar impurities from used paper pulp is accomplished by thorough washing with hot water in pulp washers.

Decolourization of the filtered liquor is accomplished in most refineries by its percolation at a relatively slow rate through a deep bed of bone char in char filters—large cylindrical vessels, usually of cast iron, measuring from 15 to 50 ft. in height and from 4 to 10 ft. in diameter. Each filter is fitted at the top with a door for the admission of the bone char and with one or more similar doors near the bottom for the removal of the used char. A perforated metal false bottom, which is carefully covered with a specially woven cloth blanket before each filling, supports the column of char. An inlet pipe is provided at the top for admission of the liquor to be decolourized, while an outlet pipe is connected to the filter bottom for the efflux of the decolourized liquor.

In operation, a filter is filled to the top either with dry char or a mixture of char and liquor, then filtered washed raw sugar liquor is admitted through the inlet pipe. The rate of liquor flow is regulated by a valve in the outlet pipe. Upon completing its flow through the char, the decolourized liquor passes, by way of an extension of the outlet pipe, to a liquor classification and distribution gallery, commonly termed the liquor gallery. There the liquors from the various char filters are segregated according to quality and are directed to the proper places for further processing.

Following a specified period of washed raw sugar liquor filtration, lower purity syrups resulting from other phases of the refining process are applied to the char for decolourization, in order to utilize the maximum decolourizing ability of the char. This causes the appearance of colour in the effluent, which is then directed to the second-liquor tank and, subsequently, upon the appearance of still more colour, may be directed to a third-liquor tank.

After several hours, the decolourizing power of the char begins to deteriorate, because of the cumulative effect of the colouring matter and soluble impurities adsorbed from the liquor. At a predetermined point, the syrup supply is shut off and hot water is applied in its place. This water displaces the syrup remaining in the char and gradually dilutes it, producing what is termed sweet water. The operation is known as sweetening off the char. When the sugar content of the sweet water has fallen to a fraction of 1%, the effluent is run to waste and the flow of hot water through the char is continued for several hours in order to wash from it the maximum possible quantity of impurities. Following this, the char is removed from the filter, sent to a drier and thence to a char kiln where it is heated to a high temperature; this burns off some of the organic matter and restores the decolourizing ability of the char, thus rendering it fit for reuse. As this cycle of operations is repeated over and over, the char suffers gradual deterioration because of the cumulative effects of imperfect regeneration and of attrition due to handling. Consequently, it is common practice to maintain a policy of periodical discard of spent char and addition of new char.

Some refining has been carried on by the use of activated vegetable carbons which decolourize liquors by adsorption, and also by the use of certain chemical compounds which liberate nascent chlorine and have a bleaching action.

Crystallization.—The colourless and slightly coloured liquors are then recrystallized by boiling in vacuum pans which are heated by steam and within which a rather high vacuum is maintained by means of suitable condensers and vacuum pumps. The vacuum pans are equipped with the necessary facilities for control of sugar crystal formation and growth, so that at the completion of a boiling period the crystals discharged from a pan are of the proper size to meet the trade requirements for the particular type of sugar under production.

The massecuite discharged from a vacuum pan upon completion of boiling is a mixture of crystals and mother liquor or syrup. This massecuite is then directed to centrifugal machines in which the syrup is separated from the crystals and the crystals are sprayed with water to remove all possible traces of syrup from their surfaces. In this operation the purity of the crystals is

TABLE II.—*Production of Centrifugal* Sugar by Areas and Major Producing Countries†*
(Raw value, thousands of short tons)

	1935–39 Average	1945–49 Average	1951	1952‡
North America	8,737	11,925	15,055	12,805
Cuba	3,183	5,807	7,964	5,680§
United States (beet)	1,518	1,513	1,552	1,532
United States (cane)	474	455	418	557
Puerto Rico	974	1,134	1,360	1,100‖
Hawaii	980	801	1,045	1,075
Mexico	353	636	807	865
Dominican Republic	491	506	659	700
Jamaica	119	235	298	320
Barbados	114	121	176	180
Trinidad	149	144	154	160
Canada (beet)	76	99	133	160
Europe (beet)	7,274	5,949	9,680	9,583
Eastern Europe	2,926	2,056	3,175	3,179
France	1,078	823	1,396	1,350
Germany, Western	610	524	1,168	1,100
Italy	419	329	853	800
United Kingdom	514	613	753	740
Spain¶	202	200	304	535
Netherlands	261	270	383	400
Belgium	259	247	300	365
Denmark	260	266	397	335
Sweden	340	311	339	389
Asia (cane)	5,237	2,508	4,646	4,858
India	1,300	1,304	1,900	1,700
Philippines, Republic of	1,058	382	1,073	1,245
Formosa	1,240	372	610	800
Indonesia	1,207	102	466	500
Turkey (beet)	76	131	228	185
Pakistan	30	36	96	125
China including Manchuria	100	80	95	100
South America (cane)	2,115	2,098	3,817	3,087
Brazil	830	1,418	1,865	2,044
Argentina	510	654	700	700
Peru	444	483	509	520
British Guiana	210	198	260	280
Colombia	51	135	220	250
U.S.S.R. (beet)	2,761	1,643	2,300	2,500
Africa (cane)	1,203	1,452	1,710	1,791
Union of South Africa	498	542	532	600
Mauritius	320	351	535	560
Egypt	166	211	215	215
Reunion	91	81	128	150
Oceania (cane)	1,113	961	961	1,175
Australia	804	830	834	1,030
Fiji	150	131	127	145
World total (cane)	16,754	18,054	24,158	22,644
World total (beet)	11,776	9,382	14,011	14,055
World total (beet & cane)	28,530	27,436	38,169	36,699

*Centrifugal sugar, as distinguished from noncentrifugal, includes cane and beet sugar produced by the centrifugal process, which is the principal kind moving in international trade.

†Years shown are for crop years; generally the harvesting season begins in the fall months of the year shown or in the early months of the following year, except in certain cane sugar producing countries in the southern hemisphere, such as Australia, Argentina, Mauritius, Union of South Africa, etc., where the season begins in May or June of the year shown.

‡Preliminary.

§Production probably would slightly exceed that of 1951–52 except for restrictions decreed by the Cuban government.

‖Production probably would slightly exceed that in 1951–52 except for restrictions imposed under the Sugar act. On Dec. 31, 1952, the department of agriculture announced that Puerto Rican sugar production from the 1952–53 crop had been set at 1,190,000 tons.

¶Includes a small amount of cane.

☆Includes sugar from Danish beets processed in Sweden.

Source: *The World Sugar Situation*, Bureau of Agricultural Economics, U.S. Department of Agriculture (Washington, Jan. 19, 1953).

TABLE III.—*Production of Noncentrifugal* Sugar, by Areas and Specified Countries†*

	1935–39 Average	1945–49 Average	1951	1952‡
Asia	4,340	4,606	5,148	5,211
Burma	86	74	75	75
China	350	305	485	520
Formosa	32	6	8	8
India	3,008	3,366	3,675	3,700
Japan	4	14	15	15
Pakistan	600	771	800	800
Philippines, Republic of	63	44	50	50
Ryukyu Island	90	4	19	20
Thailand	17	22	21	21
South America	874	1,302	1,177	1,175
Brazil	370	395	300	300
Colombia	420	755	720	720
Ecuador	15	17	23	23
Peru	9	25	24	22
Venezuela	60	110	110	110
North America	163	303	272	272
Costa Rica	15	36	33	33
El Salvador	17	27	10	10
Guatemala	31	35	30	30
Hooduras	8	8	8	8
Mexico	83	174	165	165
Nicaragua	7	14	15	15
Panamá	2	9	11	11
World total	5,377	6,211	6,597	6,658

*Noncentrifugal sugar includes all types of sugar produced by other than the centrifugal process which is largely for consumption in the relatively few areas where produced. The estimates include such kinds known as piloncillo, panelo, papeloa, chaocaca, rapadura, jaggery, gur, muscovado, panocha, etc.

†Years shown are for crop years; generally the harvesting seasoo begins io the fall months of the year shown or in the early months of the followiq year except in certain cane sugar producing countries in the southern hemisphere, where the season begins in May or Juoe of the year shown.

‡Preliminary.

raised to about 99.9% sucrose. The syrup is reboiled two or three times in order to secure maximum sucrose extraction before returning it to char for further decolourization. Magnetic filtration by an ion exchange process gives promise of displacing this technique. The moist white sugar leaving the centrifugals is carried by conveyers to sugar driers in which the moisture content of the sugar is reduced, by means of a current of hot air, to a very small fraction of 1%. The granulated sugar drier, or granulator, consists of a large revolving cylindrical drum, the axis of which is slightly inclined from the horizontal. The interior is fitted with a number of narrow longitudinal shelves which alternately lift the sugar and shower it down through the current of heated air as the drum revolves.

From the driers the dried sugar passes to sets of vibrating or gyrating screens which remove undesirable coarse particles and dust and classify the sugar into the sizes demanded by the trade. The sifted sugar is then spouted to bins, from which it is drawn into automatic weighing and filling machines. The modern refinery is equipped with many such machines, designed to pack a wide variety of types and sizes of containers, generally ranging from 1-lb. paperboard cartons to 100-lb. cloth or paper bags and, to a limited extent, 300- to 350-lb. barrels.

The several varieties of cube, loaf and tablet sugars are prepared by mixing specially boiled white sugar with a controlled amount of heavy white sugar syrup. The resulting moist magma is then pressed or moulded in various types of machines in which either pressure or centrifugal force is utilized to produce moist cubes, tablets or slabs. These moist forms are dried in hot-air ovens, where the sugar in the binding syrup crystallizes, thus bonding together rigidly all the crystals in the cube, tablet or slab. Cubes and moulded tablets are packed directly after drying. Loaf sugar and chipped tablets are produced either by direct cutting of the slabs with special chipping knives or by first sawing the slabs into bars and then chipping the bars as desired.

Powdered or icing sugar is made from granulated sugar by grinding in various types of attrition mills or impact pulverizers. In some cases the pulverized material is also passed through bolting cloth before packing.

Soft, or brown, sugars are boiled from specially selected, dark-coloured, char-filtered liquors which result from the char filtration of affination syrup. Soft sugars are not washed in the centrifugals, but are allowed to retain a certain amount of syrup on the crystals; the quantity of syrup retained is regulated by the centrifuging process to give the desired sugar colour. Soft sugars are characterized by very small crystal size which gives them a soft, spongy texture, and by their colour and pleasing flavour.

Various grades of edible syrups are prepared by blending light- and dark-coloured char-filtered liquors and syrups to meet desired specifications.

By-products.—By-products of a sugar refinery are restricted to: (1) a relatively small amount of molasses which is usually fermented for the production of alcohol or used in the manufacture of cattle food and (2) the discarded spent bone char, which is used as poultry feed or as fertilizer and soil conditioner.

(A. H. Hn.; X.)

PRODUCTION, TRADE AND PRICES

Cane sugar can be and is produced in nearly all tropical areas and beet sugar in many temperate countries. Beet sugar makes up roughly one-third of the total (*see* Table II) and is primarily a product of Europe, the U.S.S.R. and the United States. Caribbean America is outstanding in centrifugal cane-sugar production, whereas India and Pakistan contribute nearly three-fourths of the less highly refined type of cane sugar. Because of intimate relationship with war areas, the beet sugar production, which was slightly in excess of cane sugar production as late as 1912 was hard hit twice during the next 40 years, whereas cane-sugar production suffered no such hazards. Both generally had an upward trend but cane sugar greatly increased its margin over beet-sugar production.

Comprehensive information on beet-sugar production in the United States is shown in Table IV and cane sugar production for continental United States and territories supplying the United States in Table V. Information on production of miscellaneous sugars is harder to obtain. However, such harvests are not insignificant locally. The palm-sugar season as a village or forest industry in some areas lasts about four months during which time a yield of as much as 40 to 80 lbs. of raw sugar is obtained per tree, or more than one ton per acre. The yield of the long-used maple yard of northeastern United States, though highly variable from season to season, averages about 3 lb. per tree but may go as high as 20 lbs.

Following World War I, sugar was uncommonly scarce and prices very high. This encouraged restoration and expansion of beet-sugar production, and national self-sufficiency became the order of the day. Cane sugar areas, too, increased production

TABLE IV.—*Sugar Beets, Beet Sugar and Pulp; Acreage, Yield per Acre, Production, Season Average Price per Ton Received by Farmers, and Value; Production of Beet Sugar and Other Products, United States*

Year of harvest	Sugar beets						Beet sugar and pulp (in 000 tons)			
	Acreage (in 000 ac.)		Yield per acre (in tons)	Production (in 000 tons)	Price*	Farm value (in $000)	Sugar produced (refined)	Molasses pulp	Moist pulp	Dried pulp
	Planted	Harvested								
1935	809	763	10.4	7,908	$5.76	45,565	1,185	125	1,454	74
1936	855	776	11.6	8,028	6.05	54,636	1,304	157	1,543	73
1937	813	753	11.6	8,759	5.26	46,101	1,283	166	1,600	51
1938	985	925	12.4	11,497	4.65	53,478	1,676	219	1,858	105
1939	993	918	11.7	10,781	4.76	51,342	1,644	175	1,711	98
1940	791	912	13.4	12,194	5.13	62,540	1,756	189	1,625	114
1941	796	755	13.7	10,342	6.47	66,928	1,491	181	1,556	101
1942	1,048	954	12.2	11,685	6.87	80,279	1,619	151	1,688	138
1943	619	550	11.9	6,547	8.87	58,089	935	84	1,063	77
1944	633	555	12.1	6,718	10.60	71,354	979	77	1,008	101
1945	775	713	12.1	8,616	10.20	87,636	1,183	115	1,173	101
1946	905	802	13.2	10,590	11.20	117,835	1,423	165	1,431	130
1947	968	879	14.2	12,503	11.90	148,377	1,715	217	1,847	91
1948	800	694	13.6	9,424	10.60	99,935	1,280	207	1,316	76
1949	768	687	14.8	10,196	10.80	110,380	1,467	271	1,282	118
1950	1,014	925	14.6	13,535	11.20	151,293	1,880	293	1,761	113
1951	757	691	15.2	10,485	11.70	122,507	1,450			
1952†	720	667	15.3	10,217	11.90	121,582	1,432			

*Iocludes price support payments, hut excludes Sugar act payments. †Preliminary.

Source: *The World Sugar Situation*, Bureau of Agricultural Economics, U.S. Department of Agriculture (Washington, Jan. 1953).

TABLE V.—*Sugar: Production in Continental United States and Principal Areas Supplying United States with Sugar**
(In 000 tons)

Crop year†	Cane								Beet‖	Cane and beet	
	Continental U.S.‡	Hawaii§	Philippine Islands	Puerto Rico	Virgin Islands	Total of preceding columns	Cuba	Total including Cuba	Continental U.S.	Excluding Cuba	Including Cuba
1935	383	1,042	979	926	4	3,334	2,904	6,238	1,268	4,602	7,506
1936	437	1,044	1,118	996	8	3,503	3,379	6,882	1,395	4,898	8,277
1937	462	941	1,055	1,077	4	3,539	3,380	6,919	1,373	4,912	8,292
1938	583	994	1,002	852	6	3,527	3,094	6,621	1,793	5,320	8,414
1939	504	977	1,044	1,019	7	3,551	3,157	6,708	1,750	5,310	8,467
1940	332	947	1,035	932	8	3,254	2,734	5,988	1,879	5,133	7,867
1941	419	870	550	1,148	1	2,988	3,799	6,787	1,595	4,583	8,382
1942	466	886	5	1,038	4	2,393	3,229	5,622	1,732	4,125	7,354
1943	498	875	5	723	3	2,104	4,741	6,845	1,000	3,104	7,845
1944	437	821	5	964	4	2,231	3,924	6,155	1,048	3,279	7,203
1945	475	680	13	908	5	2,081	4,476	6,557	1,266	3,347	7,823
1946	425	872	78	1,088	3	2,466	6,448	8,914	1,523	3,989	10,437
1947	376	835	308	1,108	4	2,721	6,675	9,396	1,835	4,559	11,231
1948	477	956	730	1,277	5	3,445	5,761	9,206	1,370	4,815	10,576
1949	521	961	685	1,286	11	3,464	6,127	9,591	1,570	5,034	11,161
1950	564	996	935	1,228	7	3,730	6,349	10,079	2,012	5,742	12,091
1951	418	1,045	1,073	1,360	12	3,908	7,964	11,872	1,552	5,460	13,424
1952¶	557	1,075	1,245	1,190	12	4,079	5,680	9,759	1,530	5,609	11,289

*Includes centrifugal only. †Crop year with harvesting season beginning in the fall months or in the early months of the following year. ‡Includes Louisiana and Florida. §Normal grinding season begins Oct. 1. In 1933 production was from grindings of Oct. 1933 to Dec. 31, 1934. For crop year 1934 and subsequent years production is from grindings of the next calendar year. ‖Refined beet sugar reduced to raw basis by multiplying by 1.07. ¶Preliminary.
Source: *The World Sugar Situation*, Bureau of Agricultural Economics, U.S. Department of Agriculture (Washington, Jan. 1953).

partly because of favourable prices but also because of new cane varieties and the availability of new venture capital. So, after 1927, until the outbreak of World War II, there was world overproduction of sugar in relation to effective demand. Stocks accumulated; the carry-over in 1929, for example, was 9,000,000 tons. Attempts at finding a solution, some of which will be discussed presently, were generally unsuccessful. In the years just before World War II, consumption was approximately equal to production, yet the world price for unprotected raw sugar was about one cent per pound.

The war changed all that; production was interrupted in some areas, trade was badly disrupted, and rationing became general in many major consuming countries. The Sugar Committee of the Combined Food Board, membership consisting of the United States, Great Britain and Canada, reviewed the supply and distribution in the free world. Production at the war's end was the smallest in relation to requirements in 25 years. The International Emergency Food committee indicated that available supplies would provide that all claimant countries could receive only about 70% of prewar consumption.

One of the outstanding developments in the world food picture after World War II was the rapid recovery of sugar production. As the sugar supply situation began to relax somewhat in 1947 and prospects for the new crop appeared favourable, the International Emergency Food council first notified importing countries that they might exceed previously recommended allocations; then late in 1947 allocation recommendations were discontinued on sugar. Though demand continued strong, production expanded to new record levels and outbreak of the Korean war in 1950 did not result in a shortage of supplies or any major price changes.

Approximately one-third of the world's sugar enters international trade. Cuba is the major exporter whereas the United States, western Europe and Japan are major importers. In very large part, this is cane sugar; even before World War II the trade in beet sugar was small, mostly from Poland and Czechoslovakia. The economic and political problems of the sugar industry, such as high overhead costs, regional specialization and one-crop economies, technological obsolescence and multiple competition—not only between countries but between cane and beet production, and preferential arrangements have noticeably affected trade even more than production. Only about a tenth of the sugar entering world trade is "free" sugar, not regulated by bilateral or preferential arrangements.

Intergovernmental attempts to solve some of the problems of sugar have been in evidence for nearly 100 years, beginning with a convention signed by four western European countries in 1864. Later major events were the Brussels Sugar convention of 1902, the International Sugar agreement of 1931 (including the Chadbourne agreement) and the International Sugar agreement of 1937, later extended. Some students concluded that these agreements were sought as an instrument for maintaining the market position of the lowest cost producers, an exceptional situation.

The 1931 International Sugar agreement, something between an intergovernmental commercial treaty and a private cartel, was an undertaking by exporter countries to set export quotas and to allocate the American market. It finally lapsed in 1935. The International Sugar agreement of 1937 was signed by 21 countries, only to remain largely inoperative because of World War II. It provided for delineation of the free market, set export quotas and introduced some flexibility whereby export quotas might be gradually reduced if necessary.

Cuba as the major exporter, first in the 1920s and again in 1952, attempted to bring the surplus situation under control by requiring reduction in its own production and setting part of the surplus up as a reserve to be deliberately marketed over several years instead of forced on the depressed market.

In 1953, the British Commonwealth Sugar agreement was extended through 1960, increasing the negotiated price to an equivalent of 5.29 cents per pound.

Governmental programs and policies have been much concerned with sugar, somewhat as regards production (particularly beet sugar), but more often with control of trade, processing and sale of the refined product. Beginning in the 17th century differential tariffs as between raw and refined sugar tended to aid the growth of a sugar refining industry in western Europe. Revenue tariffs, drawbacks, export subsidies, quotas, licences, rationing and price fixing were all explored and used. Some details may be noted for one country.

In the United States the precedent of raising federal revenues by taxing sugar was set in 1789 with a sugar tariff. The industry from 1865 to 1920 was protected by direct subsidy and by tariffs and during war periods was rather rigidly controlled. Tariff duties on sugar were modified 14 times between 1897 and the 1950s, reaching a high during the 1930–34 period of 2 cents per pound on Cuba 96° sugar and 2.5 cents per pound on sugar from other countries. In 1948 the duty on Cuban sugar was set at 0.5 cent per pound and in June 1951 the duty was made .625 cents per pound on sugar from other countries.

The Jones-Costigan act of 1934 provided for a processing tax on sugar and gave the secretary of agriculture power to estimate consumption needs and allocate quotas to four groups of producing areas. Benefit payments and the tax were declared unconstitutional in 1936, but the quota provisions were extended and included in the Sugar act of 1937. That law established a per-

TABLE VI.—*United States Sugar Deliveries by Type of Product or Business of Buyer*
(In 000 cwt. refined)

Product or business of buyer	1950	1951	1952
Bakery and allied products; cereals and cereal products	12,724	12,444	13,617
Confectionery and related products	13,019	12,450	12,933
Ice Cream and dairy products	5,074	5,140	5,541
Beverages	15,047	14,404	15,365
Canned, bottled, frozen foods, jams, jellies, preserves, etc.	11,010	11,165	10,956
Multiple and all other food uses	4,993	4,926	5,230
Nonfood products	780	794	915
Hotels, restaurants, institutions	554	576	521
Wholesale grocers, jobbers, sugar dealers	60,775	54,218	57,006
Retail grocers, chain stores, supermarkets	23,441	20,515	21,401
All other deliveries, including deliveries to government agencies	1,759	2,247	2,650
Total deliveries	150,076	138,879	146,195

Source: *Sugar Statistics*, vol. 1, pp. 38–57, Production and Marketing Association, U.S. Department of Agriculture (Washington, June 1953).

centage quota of 55.9% of estimated requirements, but not less than 3,715,000 short tons, for domestic sugar producing areas, including United States territories. Suspension, reinstatement and extensions followed during and after World War II. The Sugar act of 1948 (as amended and extended in 1951) differed from that of 1937 in that it established fixed, instead of percentage, quotas of 1,800,000 short tons raw value for domestic beet areas, 500,000 tons for mainland cane, 1,052,000 tons for Hawaii, 1,080,000 tons for Puerto Rico and 12,000 tons for the Virgin Islands. The Philippine quota was set at 952,000 short tons of commercial sugar. Additional amounts needed to meet consumption estimates would be allocated, 96% to Cuba and 4% to full duty countries, with provision for supplemental allocation of short falls. Also the steps whereby the secretary determines annual requirements of sugar consumers were rather fully spelled out. These determinations ranged between 7,000,000 and 8,700,000 tons, and were sometimes amended during most consuming years.

Trade in sugar is so largely under agreement of one sort or another that a true world price is difficult to define. The price of 96° raw sugar f.o.b. Cuba is an approximation. When transportation and duty are taken into account the world price is sometimes above and sometimes less than the raw sugar prices at New York. Markets in both spot sugar and futures are available in New York and spot markets are maintained in most producing countries.

Studies indicate that the elasticity of demand for sugar is low and has remained fairly constant from period to period. A reduction in price of 1% tends to be accompanied by an increase of 0.3% to 0.4% in consumption, with some tendency for the demand curve to shift downward after the early 1920s.

Consumption and Cookery.—Sugar provides a high energy value per unit of weight (about 1,794 kilocalories per pound) in addition to providing a degree of sweetness that is highly valuable for improving the palatability of other foods. Furthermore, by virtue of other important properties, sugar serves as a preservative agent for other foods (*e.g.*, fruit, fruit products) and also provides a physical basis for fabrication of various food products (*e.g.*, confectionery). It is also valuable for making desirable modification in the properties of food products, such as controlling texture and retention of moisture, as in the baking industry. Finally, sugar has many accessory features of value (in addition to its specific nutritive value) in the preparation of other foods (*e.g.*, icings, coatings and fillings).

The requirements for sugar in the diet of virtually all people have been better met as a result of constant increases in production which have permitted sugar to be removed gradually from the class of a medicine and an expensive luxury and allowing it to become established as one of the more economical sources of nourishment for the human body.

Relatively little sugar, particularly in the high-consuming countries, is consumed directly and by itself, though the population of such countries may obtain one-sixth or more of their total calories from that food. Instead, it is used to increase nutritive value and to impart a satisfaction to the taste of a long list of other foods and beverages. In the United States of an annual average consumption of about 95 lb. per person something like half is used directly in the home or in restaurants, and the remainder is first processed with other foods or drinks. Table VI indicates a general breakdown in respect to United States consumption.

In other countries wide variations in the amount of sugar consumed have been reported. The variations are associated with changes and differences in the level of national income, the retail prices established for sugar, the general standard of living, living conditions in the country, etc.

As sugar is an important soluble carbohydrate and is easily assimilable, it is valuable in the diet. Candy contains a high proportion of sugar (or glucose), but no dish for the table should do so. In adding sugar to fruit, more is needed if it is cooked with the fruit than if added when the cooking is completed.

Cane and beet sugars (sucrose) are the principal sugars used in cookery. They dissolve in one-third their own bulk of cold water, and in a smaller amount of hot water. Their melting point is 320° F. If cooled at that point, they become barley sugar; if cooked beyond that point they become burnt sugar or caramel.

Fine granulated or castor sugar is generally used in cookery. For icings, meringues and whips, powdered or pulverized (icing) sugar is used, the finest form being confectioners' sugar. All powdered sugars must be rolled and sifted fine before using. These sugars are all pure white refined products. Brown and yellow sugars, which are produced by evaporation, are also used in cookery when there is no objection to a dark colour, and some of them have a delicious flavour. They vary in both colour and flavour, the best known being Demerara or coffee sugar, and dark brown or Barbadoes sugar.

Other sugars such as glucose are less sweet than sucrose, and the commercial form is cheaper. It is used in many manufactured candies, jams and syrups to economize, but is not much used in the home. Commercial grape sugar is a dextrose. Maple sugar (sucrose) is much prized for its special flavour, but is made in comparatively small quantity, as is also palm sugar. Ame is a Japanese sweetening material that contains dextrin and maltose. Saccharin is not a sugar, and has no food value, but it is very sweet and is sometimes substituted for sugar, usually on a physician's orders.

Caramel or burnt sugar is marketed ready prepared, for use as "browning" for sauces, gravies, stews and the like, and also as a flavouring in such dishes as caramel custard (custard with caramel below or around it). It can be prepared at home by melting granulated sugar in a saucepan over a low fire, with constant stirring.

It may be used at once as it is, or after removing from the fire, boiling water ($\frac{1}{4}$ cup water to 1 cup sugar) may be added very slowly, again with constant stirring. Return this mixture to the fire and cook to a rich dark brown (eight to ten minutes). This hardens on standing, but can be melted over hot water. (*See* also TREACLE; HONEY.)

BIBLIOGRAPHY.—Noel Deerr, *The History of Sugar*, 2 vol. (London, 1949 and 1950); R. A. McGinnis, *Beet-Sugar Technology* (New York and London, 1951); A. Van Hook, *Sugar; Its Production, Technology and Uses* (New York, 1949); E. Roboz, *Sugar Beet By-Products* (Sugar Research Foundation, New York, 1950); O. Lyle, *Technology for Sugar Refinery Workers*, 2nd ed. (London, 1950); Roy A. Ballinger, "Sugar During World War II," U.S. Department of Agriculture, *War Records Monograph* 3 (June, 1946); B. C. Swerling, *International Control of Sugar, 1918–41*, "Commodity Policy Studies No. 7" (Stanford, 1949). (B. F. L.; J. K. R.)

SUGAR BEET: *see* BEET.

SUGAR-BIRD, the English name commonly given in the West Indian Islands to birds of the genus *Certhiola* (belonging to the Passerine family *Coerebidae*) for their habit of frequenting the curing-houses where sugar is kept, attracted thither by the flies. In their figure and motions they resemble a nuthatch, while their colouration—black, yellow, olive, gray and white—recalls a titmouse.

SUGER (*c.* 1081–1151), French ecclesiastic, statesman and historian, was born of poor parents either in Flanders, at St. Denis near Paris or at Toury in Beauce. About 1091 he entered the abbey of St. Denis. Until about 1104 he was educated at the priory of St. Denis de l'Estrée and there first met his pupil, King Louis VI. From 1104–06 Suger attended another school, perhaps that attached to the abbey of St. Benoît-sur-Loire. In 1106 he became secretary to the abbot of St. Denis. In 1107 he was made provost of Berneval in Normandy, and in 1109 of Toury.

In 1118 he was sent by Louis VI to the court of Pope Gelasius II at Maguelonne and lived from 1121–22 at the court of his successor, Calixtus II. On his return from Italy Suger was appointed abbot of St. Denis. In 1137 he accompanied the future king, Louis VII, into Aquitaine on the occasion of that prince's marriage to Eleanor of Aquitaine, and during the second crusade was one of the regents of the kingdom (1147–49). He was bitterly opposed to the king's divorce, having himself advised the marriage. Although he disapproved of the second crusade, at the time of his death, on Jan. 31, 1151, he was preaching a new crusade.

Suger was the friend and counsellor both of Louis VI and Louis VII. He urged the king to destroy the feudal bandits, was responsible for the royal tactics in dealing with the communal movements and endeavoured to regularize the administration of

justice. He left his abbey, which possessed considerable property, enriched and embellished by the construction of a new church built in the nascent Gothic style.

Suger was generally considered to be the foremost historian of his day. He wrote a panegyric on Louis VI (*Vita Ludovici regis*), and contributed to *Historia gloriosi regis Ludovici* (Louis VII). In his *Liber de rebus in administratione sua gestis*, and its supplement, *Libellus de consecratione ecclesiae S. Dionysii*, he treats of the improvements he had made to St. Denis, describes the treasure of the church and gives an account of the rebuilding. Suger's works served to imbue the monks of St. Denis with a taste for history, frequently resulting in quasi-official chronicles.

Bibliography.—O. Cartellieri, *Abt Suger von Saint-Denis* (Berlin, 1898); A. Luchaire, *Louis le Gros* (Paris, 1890); F. A. Gervaise, *Histoire de Suger* (Paris, 1721).

SUGGESTION AND SUGGESTIBILITY. The word suggestion applies to the stimulus, suggestibility to the personality organization that determines the extent to which this suggestion or stimulus will be accepted in the absence of logically adequate proof. The suggestion, while usually verbal, may be visual or may come through any of the other sense organs. The suggestion may even be symbolic: a person allergic to roses may develop an attack of asthma from looking through a seed catalogue. The trait is not generalized: a man may be very logical in his business judgments, very illogical and suggestible as regards his health.

Suggestions.—Suggestions may be considered under several headings. They may be either direct or indirect. In the former the recipient is aware of the intention conveyed by the suggestion. Most of the advertisements on television fall in this class; they are direct and positive. In indirect suggestion, the aim and the source remain hidden. Such suggestions are very important in politics and propaganda dissemination. Subliminal suggestion, the presentation of stimuli too faint to be consciously seen or heard, is a dramatic but uncommon example of indirect suggestion; there is evidence that sale of, for example, popcorn in a movie house has been dramatically increased by subliminal advertising material projected on the motion-picture screen.

Prestige suggestion may be contrasted with nonprestige suggestion. In the former, the effectiveness of the suggestion depends largely on the importance of its author. Unfortunately, it is too often assumed that an authority in the field of, say, nuclear physics is *ipso facto* an authority on politics or town planning; his prestige in one field gives undue importance to his opinions in an unrelated field. Nonprestige suggestion, which can be either direct or indirect, is characteristic of newspaper advertising and political editorials.

Suggestions also can be positive or negative. An invitation for a drink extended to a teetotaler will evoke a negative response. Similarly statements on, for example, racial issues will touch off positive or negative reactions according to the background of the person receiving them. Such reactions are seen in many situations involving interpersonal relations.

Heterosuggestion is frequently contrasted with autosuggestion. The former, coming from outside the person, may be either direct or indirect, prestige or nonprestige. The dictator speaking through a controlled press provides an example of direct heterosuggestion. Actually, autosuggestion (in which the person himself is the source of the suggestion) is probably a form of heterosuggestion also; the person identifies himself, consciously or unconsciously, with or against the said dictator or other person and gives himself the appropriate suggestions.

Suggestibility.—Psychology stresses the sociocultural background of the recipient as the principal causative factor in suggestibility. Heredity probably plays a part, but this has not been clearly demonstrated. Undoubtedly the physiological condition of the organism at any particular moment is of some importance. Fatigue, malnutrition and drugs may increase suggestibility but are of relative unimportance in most life situations. Brainwashing (*q.v.*) techniques are good examples of the part unusual physiological conditions may play in suggestibility.

Subject to such limitations, suggestibility is grounded in the psychodynamics of the individual, his sociocultural background. The factors involved are largely unconscious and largely emotional. The ardent political partisan feels that his neighbour who belongs to an opposing party and who is actually his equal in intelligence and integrity is both ignorant and unethical. The attitude is taken in the absence of logically adequate proof.

The drive for such irrational behaviour arises from the unconscious; consequently, the term unconscious motivation, meaning the activation of behaviour without conscious insight, becomes important in any discussion of suggestibility. This is well illustrated in posthypnotic suggestion, wherein the subject will perform the most ridiculous acts with no conscious realization of why he so behaves.

Suggestibility is of great importance in the consideration of hypnosis. H. Bernheim, in fact, claimed that hypnosis was merely a state of exaggerated suggestibility, and while this is open to question, suggestibility remains a very important phenomenon in hypnosis (*see* Hypnosis).

The defense mechanisms (*q.v.*) also must be considered in the psychodynamics of suggestibility, one such mechanism being rationalization, the justification of behaviour on partial and selected evidence.

Bibliography.—C. Baudouin, *Suggestion and Autosuggestion*, trans. by E. Paul and C. Paul (1921); H. Bernheim, *Suggestive Therapeutics*, trans. by Christian A. Herter (1889); A. Binet, *La Suggestibilité* (1900); D. O. Hebb, "The Motivating Effects of Exteroceptive Stimulation," *American Psychologist*, 13:109–113 (1958); Karen Horney, *Our Inner Conflicts* (1945); C. L. Hull, *Hypnosis and Suggestibility* (1933); M. V. Kline (ed.), *Hypnodynamic Psychology* (1955); W. Kaufman, "Self-Inflicted, Food-Induced Allergic Illness," *Ann. Allergy*, 10:308–323 (1952); Leslie M. Le Cron (ed.), *Experimental Hypnosis* (1952); J. V. McConnell, R. L. Cutler, E. B. McNeil, "Subliminal Stimulation," *American Psychologist*, 13:229–242 (1958); G. Murphy, L. B. Murphy and T. M. Newcomb, *Experimental Social Psychology*, rev. ed. (1937); S. A. Portis "Life Situations, Emotions, and Hyperinsulinism," *J.A.M.A.*, 142:1281–86 (1950); J. Stolzenberg, *Psychosomatics and Suggestion Therapy in Dentistry* (1950); Griffith W. Williams, *Suggestibility in the Normal and Hypnotic States* (1931).

(G. H. E.)

SUHARD, EMMANUEL (1874–1949), archbishop of Paris during the German occupation of World War II, was born at Brains-sur-les-Marches in eastern Brittany, April 5, 1874. His father, an illiterate farmer, died that year and the widow struggled to send her son to school. Attending seminaries at Mayenne, Laval and Rome, he was ordained in 1897. His brilliant scholastic record returned him to Laval where he taught philosophy and theology (1899–1928). Consecrated bishop of Bayeux-Lisieux in 1930, he was made archbishop of Reims, then cardinal (1935).

Appointed archbishop of Paris on May 11, 1940, as the Germans were breaking through at Sedan, Suhard accepted a governmental suggestion to visit Spain to gain support for France. He returned to Paris during the confusion of evacuation. Harassed by the Nazis during the occupation, he confined himself exclusively to religious matters and private appeals for Jews and deportees, a course that diminished his popularity at liberation. His principal concern was to establish contact with the masses, and he often met workers in their homes. In Oct. 1942 he launched the Mission of France, with a special seminary at Lisieux, to bring religion to neglected areas and groups. In 1943, after reading Abbé Godin's *France, Pagan?*, he established the Mission of Paris, popularly known as priest-workers. He defended this novel experiment of sending clergy to live and work as labourers until his death on May 30, 1949.

Suhard's pastorals as archbishop of Paris, particularly *Growth or Decline?* (1948), *The Meaning of God* (1948) and *Priests Among Men* (1949), won world acclaim. (J. N. M.)

SUHL, a town in Suhl district and of the former *Land* Thuringia, Germany, on the Lauter, in the Thuringian forest. Pop. (1950) 24,020. Suhl, which obtained civic rights in 1527, formed a part of Saxony assigned to Prussia by the congress of Vienna in 1815. The armourers of Suhl, mentioned in the 9th century, enjoyed their highest vogue from 1550 to 1634. Later, in the 19th and 20th centuries, Suhl became famous for its sporting guns.

SUICIDE. Suicide is the act of voluntary and intentional self-destruction. This definition, however, is subject to legal

qualification. Thus, in England, the term covers an act of self-destruction committed only while the deceased was of unsound mind; otherwise, suicide, termed *felo-de-se*, is a crime which entails penalties against the deceased's estate. In the United States, only a few states regard suicide as a criminal act, although it may be criminal to influence or assist in a suicide.

Causes.—Various psychological motivations have been advanced as conducive to suicide: frustration leading to hostility which is turned inward upon oneself instead of outward toward another person; loss of love, feeling of rejection, feelings of guilt and desire for vengeance; and desire to escape from a physically painful situation. Whatever the dynamics in the individual case,

TABLE I.—*Death Rates from Suicide per 100,000 Total Population in Certain Countries for Specified Years*

Country	Year	Suicide rate	Country	Year	Suicide rate
North America			Italy	1953	6.6
United States	1955	10.2	Netherlands	1953	6.5
Canada	1954	7.3	Norway	1953	7.7
Europe			Portugal	1954	1.5
Austria	1954	23.1	Scotland	1954	5.9
Belgium	1953	13.5	Spain	1950	5.4
Denmark	1954	23.3	Sweden	1953	18.6
England and Wales	1953	10.8	Switzerland	1953	21.8
Finland	1953	17.4	Asia		
France	1954	15.8	Japan	1953	20.5
Germany (W. Berlin)	1954	29.6	Other countries		
Germany, West	1953	18.2	Australia	1953	10.9
Ireland, Northern	1954	3.5	Egypt	1952	0.3
Ireland, Republic of	1953	2.3	New Zealand	1953	9.9
Israel	1949	7.1	Union of South Africa	1951	10.1

Primary source: *Demographic Yearbook, 1955*, Statistical Office of the United Nations, Department of Economic and Social Affairs (New York, 1955).

suicide almost always occurs in a sociological setting; thus, suicide is never automatically the result of any of the above situations. The potential suicide is influenced not only by his immediate perplexity, but also by his religious beliefs, the customs of his country, his sense of duty and various other cultural values. It must not be overlooked, also, that emotional stresses and the privations of life differ from time to time and place to place; some people are subjected to greater pressures than others.

TABLE II.—*Annual Suicide Rates per 100,000 Total Population in Selected Countries from 1910 to 1953*

Year	United States	England and Wales	France	Germany	Italy	Sweden
1910	16.0	10.0	21.8	21.6	8.4	18.1
1912	16.0	9.9	22.9	22.5	8.5	18.3
1914	16.6	10.0	21.8	21.2	8.9	15.9
1916	14.2	7.3	15.5	17.5	6.9	13.2
1918	12.2	7.5	15.8	15.3	7.5	10.0
1920	10.2	9.0	18.9	21.3	7.3	14.7
1922	11.9	10.0	*	21.5	8.1	14.4
1924	12.2	9.6	*	22.8	9.5	14.5
1926	12.8	11.4	19.2	25.9	9.6	13.8
1928	13.6	12.4	18.9	24.9	9.5	14.0
1930	15.6	12.7	19.9	27.8	9.6	15.8
1932	17.4	14.3	20.6	29.2	9.7	17.7
1934	14.9	13.7	21.4	28.7	8.7	15.3
1936	14.2	12.4	19.5	28.6	7.9	16.7
1938	15.3	12.9	*	*	7.2	*
1940	14.3	*	*	*	*	*
1942	12.1	9.2	*	*	5.2	*
1944	10.0	8.9	*	*	*	*
1946	11.6	10.5	*	*	*	*
1948	11.2	11.0	14.2	*	6.3	14.5
1950	11.4	*	*	19.2	6.5	*
1953	10.1	10.8	15.8†	18.2	6.6	18.6

*Not available. †1954.

Finally, some countries provide more adequate guidance and counseling services than others, to which the disturbed person may turn for help.

Religious and Cultural Factors.—Many religions are strongly opposed to suicide as a moral wrong. The Koran of the Moslems severely censures suicide as a graver crime than homicide. The Talmud, or laws, of Judaism emphasizes the sacredness of life and condemns suicide as sinful. After St. Augustine denounced suicide, Christianity regarded it as a form of murder and in the middle ages some Christian countries permitted mutilation of the body and confiscation of property. Although these laws have been moderated to some extent, suicide is still generally condemned by the religions of Europe and the Americas.

In India, however, Brahmans honoured the person who voluntarily freed himself from his body; hence suttee (death of the Hindu widow on the funeral pyre of her husband) was an honourable death. In Japan, hara-kiri (*q.v.*; disemboweling) was practiced as a ceremonial rite. Both types have declined in the 20th century. In general, Catholic countries have low suicide rates. Table I shows low rates for Ireland, Italy and Spain; France, however, has a relatively high rate. Some Protestant countries also have low rates (Netherlands, Norway, Scotland); others have extremely high rates (Denmark, for example). Obviously other factors are also operative.

TABLE III.—*Suicide Rates per 100,000 Population for the United States by Age, 1954*

Age	Rate	Age	Rate
1–9	0	50–54	20.6
10–14	0.3	55–59	21.9
15–19	2.4	60–64	26.3
20–24	6.0	65–69	25.8
25–29	8.5	70–74	25.9
30–34	9.0	75–79	26.5
35–39	10.9	80–84	25.3
40–44	14.1	85 and over	24.1
45–49	17.9		

Source: *Vital Statistics*, vol. 44, no. 12, Sept. 14, 1956, pp. 288–289.

Trends.—In all countries, suicide rates tend to increase during economic depressions and to decrease during times of war and prosperity, as Table II shows. Long-term trends vary from nation to nation and can probably be accounted for only in terms of internal social, economic and political stresses. In the New England states, the only part of the United States for which long-term rates are known, the rates have tended to increase.

Age, Sex, Marital Condition.—Studies made in Europe and the United States at widely different periods of time all show that virtually no children below the age of 15 commit suicide, and that the rate for the later teens is low. The rate tends to increase steadily into old age. Table III for the United States is illustrative; the rate in old age is approximately four times the rate in young adulthood.

The rates for males are consistently higher than the rates for females, in different countries, at widely different periods of time and at all ages. Usually the male rate is approximately four times the female rate. Women, however, make more unsuccessful attempts at suicide than do men.

As early as the 1860s and 1870s, data from a number of European countries showed that divorced persons had the highest suicide rates and the widowed second highest. Married people and those who never married had much lower rates, with sometimes one group and then the other having the lowest. The same relation of marital status to suicide rates was repeatedly found true in the United States.

Urban-Rural Rates.—In both Europe and the United States, cities have higher rates than adjacent rural areas. Within large cities, according to studies made in the United States, suicide rates are highest in disorganized slum areas and lowest in well-organized residential areas.

Means of Suicide.—Means of suicide vary with time and place. In the United States, after 1901–05, the use of firearms, hanging and gas increased in frequency while the use of poison decreased, perhaps because of greater restrictions on its sale. Table IV shows the trends in the United States.

TABLE IV.—*Changing Importance of Means of Suicide in the United States from 1901 to 1952*

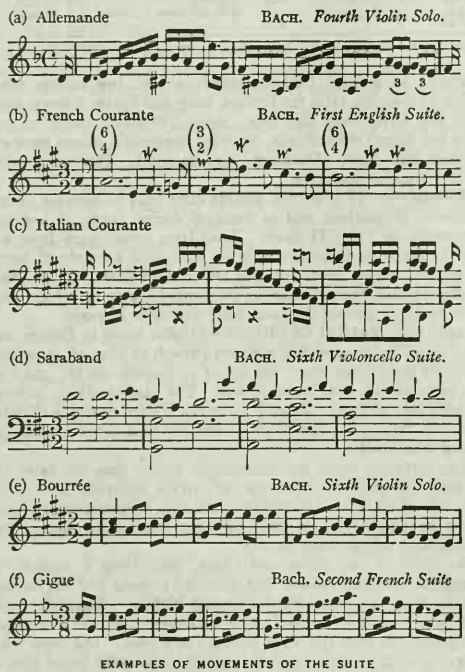
Means	1952	1936	1926–30	1911–15	1901–05
Total Suicides	100.0%	100.0%	100.0%	100.0%	100.0%
Firearms	42.5	37.0	35.1	30.0	24.4
Hanging	21.4	19.3	18.1	14.6	15.0
Poisons	10.9	15.4	16.3	27.3	34.1
Poisonous gas	11.8	13.3	14.8	12.0	8.0
Cutting or piercing instruments	3.7	4.4	5.4	6.4	5.7
Drowning	4.5	5.0	5.2	5.6	5.1
Jumping from high places	3.2	3.8	3.1	1.9	1.2
Crushing	—	.8	1.0	.8	.4
Other means	2.0	1.0	1.0	.8	6.1

Suicide and Other Types of Violent Death.—Suicide outranks homicide as a cause of death, being anywhere from 2 to 25

times as high. No positive correlation exists between suicide and homicide rates; in fact, there is some indication of an inverse correlation. In the U.S. motor vehicle deaths are slightly more than double suicides. (RH. S. C.)

SUINA, a group of ungulate mammals typified by the domestic pig and the peccary, but also including the hippopotamus. Suina is an infraorder of the suborder Suiformes of the order Artiodactyla. *See* ARTIODACTYLA; HIPPOPOTAMUS; PECCARY; SWINE.

SUITE (*Suite de pièces; Ordre; Partita*), in music, a group of dance tunes in melodic forms (*see* SONATA FORMS). It consists essentially of four principal movements with the insertion of one

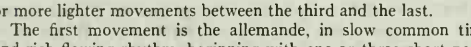


EXAMPLES OF MOVEMENTS OF THE SUITE

or more lighter movements between the third and the last.

The first movement is the allemande, in slow common time and rich flowing rhythm, beginning with one or three short notes before the first full bar. The second movement is the courante, of which there are two kinds. The French courante begins with one or three notes before the main beat, and is in a triple time ($\frac{3}{2}$) which, invariably at the cadences and sometimes elsewhere, drops into a crossing triple rhythm of twice the pace ($\frac{6}{4}$). In homage to Couperin, Bach often uses the French courante, but he is happier with the Italian type of *corrente,* a brilliant continuously running piece in quick triple time ($\frac{3}{4}$ or $\frac{3}{8}$).

The saraband is a slow movement in triple time beginning on the full bar, and with at least a tendency to the rhythm



of which Handel's aria *Lascia ch'io pianga* is a familiar example. Bach's sarabands are among the most simply eloquent and characteristic of his smaller compositions.

Then come the *galanteries,* from one to three in number. These are the only suite movements (except some of Couperin's courantes) which can have an alternative section and a *da capo.* The commonest *galanteries* are: (1) The minuet, often with a second minuet which is called trio only when it is in real three-part writing; it is a little faster than the stately minuet in Mozart's *Don Giovanni,* and it always begins on the full bar. (2) The gavotte, a lively dance in a not too rapid *alla breve* time; the gavotte always begins on the half bar. A second alternating gavotte is frequently founded on a pedal or drone bass, and is then called musette. (3) The *bourrée,* which is not unlike the gavotte, but quicker and beginning on the last quarter of the bar. (4) The *passepied,* a lively dance in quick triple time, beginning on the third beat. These dances are not always cast in binary form, and there are famous examples of gavottes and *passepieds en rondeau.* Other less common *galanteries* are: (5) the *loure,* a slow dance in $\frac{6}{4}$ time and dotted rhythm; (6) the polonaise, a leisurely triple-time piece, with cadences on the second instead of (as in later polonaises) the third beat of the bar; (7) the air, a short movement, quietly flowing, in a more florid style than its name would suggest. It sometimes precedes the saraband.

The suite concludes with a *gigue,* in the finest examples of which the melodic binary form is combined with a light fugue style. The *gigue* is generally in some triplet rhythm, *e.g.,* $\frac{3}{8}$, $\frac{6}{8}$, $\frac{9}{8}$, $\frac{12}{8}$; but examples in a graver style may be found in slow square time with dotted rhythms, as in Bach's first French suite and the sixth *partita* of the *Klavierübung.* In Couperin's first volume of *Ordres,* the *gigue* is followed by an enormous number of pieces which cannot have been intended to be all played on the same occasion, though they were all in the same key.

Suites on a large scale begin with a prelude in some larger form. Bach's French suites have no preludes; his English suites all have a great first movement which, except in the first suite, is in full *da capo* concerto form. His *partitas* cover a wider range both in their preludes and their other contents. Some large suites have finales after the *gigue,* the great chaconne for violin solo being the finale of a *partita* (*see* VARIATIONS).

The later uses of the word suite comprise almost all sets of pieces mainly in forms smaller than those of the sonata, especially such pieces as have been selected from ballets or from incidental music to plays. (D. F. T.)

SUIYÜAN, province of Inner Mongolia, north China, from 1929 until June 1954, when it was absorbed into the Inner Mongolian Autonomous Region. During the Chinese-Japanese War (1937–45) Suiyüan was amalgamated with Chahar and Ningsia to form the puppet Japanese state of Meng-chiang. It lies outside the Great Wall and includes the Ordos desert, the irrigated Ho-t'ao and Kuei-sui plains of the Yellow river and the southern rim of the Mongolian plateau leading to the Gobi desert. Around the deserts are broad grasslands increasingly planted in wheat and sugar beet by Chinese and Mongols alike. Kuei-sui (now Huhehot), the former provincial capital, became the capital of the Inner Mongolian region. Pao-t'ou (*q.v.*) is a new integrated steel and railway centre; Tsining (Chi-ning) is the junction of the railway to Ulan Bator. *See* also MONGOLIA. (TE. H.)

SUK, JOSEF (1874–1935), Czech violinist and composer, was born at Krecovice, Bohemia, Jan. 4, 1874. He studied at Prague conservatory. A member of the celebrated Bohemian quartet, he became (1924) director of the conservatory at Prague. His principal works are the symphonic poems: *Prague,* op. 26, *Asrael,* op. 27, *Summer Tale,* op. 29, and *Zrani,* op. 34 (1924); incidental music to the dramatized fairy tale, *Raduz and Mahulena; Meditation on the choral St. Wenceslaus, Legend of Dead Victors;* piano pieces, settings of Serbian folk songs and chamber music, including a string quartet op. 35a.

SUKHOMLINOV, VLADIMIR (1848–1926), Russian general and war minister, was born in 1848. He took part in the Russo-Turkish war as a staff officer (1877–78), and was head of the officers' cavalry school in St. Petersburg from 1886–97, being promoted general. He was Russian war minister from 1909–16, and it was under him that the Russian orders for mobilization were given at the outbreak of World War I. He played a disastrous role in the administration of the Russian army, but it was only in 1916, under pressure of public opinion, that Nicholas II dismissed him from office. Tried and sentenced to penal servitude on the charge of treason, he was freed by an amnesty granted by the Bolsheviks and went to Finland. In 1921 he began

publication of his memoirs, which were completed in 1924.

SUKHOTHAI, the first independent Thai kingdom. The Thais originally peopled the Yangtze valley, where they founded the empire of Nanchao. Long before the fall of Nanchao, they began to drift southward along the rivers of the Indochinese peninsula, forming small principalities that came under the suzerainty of the Khmers from Cambodia. In 1238 two Thai chieftains from Chiang Rai and Chiengmai waged a successful war of independence against the Khmers. One of them assumed the Khmer title of Sri Indraditya and proclaimed himself king. He established at Sukhothai the capital city of what was to become a mighty and vigorous, though short-lived, Thai state.

Indraditya's third son, Ramkhaeng or Rama the Great (1275–1315), gave Sukhothai its period of greatest glory. He extended his dominion to the Bay of Bengal, Luang Prabang and down the Malay peninsula as far as Ligor. He maintained relations with the Mongol court in China, receiving emissaries from Kubla Khan and sending numerous tributary missions to Peking. Twice he himself visited the Chinese emperor. Political disintegration of Sukhothai followed the death of Ramakhaeng, or Phra Ruang as he is popularly known in Thai history, and the centre of power shifted southward to the succeeding Thai kingdom of Ayudhya.

From many peoples and civilizations, the Sukhothai kings drew the elements that went into the making of their own society. Religious missions, objects of Buddhist art and the basic religious beliefs of Theravada Buddhism came from Ceylon. The pattern of political organization, the Brahmanistic ceremonials associated with kingship and the alphabet first used to write the Thai language were adopted from Cambodia. Art, sculpture and literature were introduced from India and highly skilled artisans in ceramics were imported from China. Sukhothai has been called the cradle of Siamese civilization, and the Sukhothai period is known as the golden age in the development of Thai culture.

No written history of Sukhothai has been discovered. Historians depend for their source materials upon a few ancient manuscripts, some stone inscriptions (chief of which are in the National museum in Bangkok), and the chronicles of neighbouring countries. *See* also THAILAND. (C. A. B.)

SUKHOTHAI, a province in north-central Thailand and the first great Siamese kingdom (13th–14th century). The province, drained by the Yom river, is a major producer of cotton, tobacco, rice, beans and sugar cane. Area 2,641 sq.mi., pop. (1960) 315,948. Sawankhalok (pop. [1960], 7,880), the terminus of a rail spur, and Sukhothai, the provincial capital (pop. [1960], 8,627), are situated at new river sites in the vicinity of ancient Khmer and early Thai ruins.

The names Sukhothai and Sawankhalok also designate, respectively, the exquisite Buddhist sculpture and the prized glazed pottery of the classical period. (G. W. Sk.)

SUKHUMI (SUKHUM-KALEH), a seaport on the Black sea. It is the capital of the Abkhaz Autonomous Soviet Socialist Republic, Georgian Soviet Socialist Republic, U.S.S.R. Pop. (1959) 64,730. It is the ancient Greek Dioskurias, and was in Turkish possession until 1829, after which it was Russian except for temporary Turkish occupation in 1854 and 1877.

SUKKUR, a town, with taluka (subdivision), and a district in the Khairpur division of West Pakistan. The town, situated on the right bank of the Indus is an important trade and industrial centre with biscuit, oil and other factories; pop. (1961) town 103,216; taluka 182,833. It is also noted for cotton hand-loom weaving. The Sukkur industrial trading estate was developed in the 1950s to provide facilities for factories utilizing local materials—wool, oilseeds and hides and skins. It has a degree college.

SUKKUR DISTRICT was created in 1901 out of two subdivisions of Shikarpur district, the remaining two being formed into Larkana district. Area 5,531 sq.mi.; pop. (1961) 836,867. It consists of an alluvial plain broken only at Sukkur and Rohri by limestone hills which provide firm banks to the river Indus. The principal crops are rice, wheat, gram and millets. Earthen, leathern and metal ware, cotton cloth and tussore silk are manufactured. The only other town of importance is Shikarpur (1961) 53,910, 26 mi. N.W. of Sukkur. Brassware and locks are manufactured; and the town contains a degree college affiliated with Sind university.

SULA ISLANDS, a chain of islands (Sulla, Xulla, Dutch *Soela*), east of Celebes, Indonesia, really an extension of the eastern peninsula of that island. There are three large islands, Taliabu, Mangola and Sula Besi, and several smaller ones. Taliabu and Mangola, very long and narrow, extend west and east (with Lisamatula), 135 mi. Both are mountainous, thickly forested, and thinly populated. Taliabu, largest in area and little known, is said to contain mountains (3,000–4,000 ft. high) and hot springs. Mangola, much narrower than Taliabu, has Mt. Buja in its western part.

Sula Besi, the smallest island of the three, is well populated and cultivated. The islands produce good timber for shipbuilding. Damar is collected in the forests; rice, maize, tobacco and sugar cane are grown on Sula Besi, and the sago palm is common, forming the staple food on Taliabu and Mangola. Coal of inferior quality is found on Sula Besi.

The population of the islands in 1956 was 30,596. The people resemble the natives of Buru and Ceram.

The Banggai or Peling Islands, lie off the western end of Taliabu. Peling is mountainous and forested, with bays. They furnish trepang and turtle. The people are mostly pagan and quite poor; some Moslems live along the coast. (E. E. L.; X.)

SULAYMANIYAH, an Iraqi town on the Iranian frontier, on the caravan route from Baghdad to Tabriz. Pop. (1957) 48,450. There was formerly a large trade with Iran. The town gives its name to the administrative division.

SULEIMAN I (1494–1566; reckoned as Suleiman II by the minority who acknowledge the rule of Bayezid I's eldest son Suleiman at Adrianople, 1402–10), Ottoman sultan, called THE MAGNIFICENT, succeeded his father Selim I in 1520. Suleiman was outstanding in an age which produced many great sovereigns in Europe and Asia. He was the most fortunate of the sultans and had no rival worthy of the name.

From his father Suleiman inherited a well-organized country, a disciplined army and a full treasury. He united in his person the best qualities of his predecessors and wisely took full advantage of the talents of the able generals, admirals and viziers who made his reign illustrious. He took Belgrade less than a year after his accession and Rhodes the next year. Budapest, Tabriz, Baghdad, Aden and Algiers were fruits of the other campaigns of his reign, in 13 of which he himself took part. An indication of his political insight is afforded by the alliance concluded with Francis I against Charles V, which marked the beginning of the Franco-Turkish friendship that lasted, albeit with some interruptions, until late in the 19th century.

Under Suleiman the Ottoman empire attained the zenith of its glory, yet even in his lifetime the seeds of decay were present. Historians do not deny that harem intrigues and bribery were factors in running of affairs. The contemporary poet Fuzuli tells of a visit to a government office on a matter of business: "I gave them greeting; they would not take it, as it was not a bribe."

In later life Suleiman became more and more detached from public life; but he was well served, and the state continued to flourish. Turkish tradition tends to ascribe the blame for Suleiman's few misdeeds to his Russian wife Khurrem Sultan, famed in the west as Roxelana. Generally he was just, courteous and wise. To the Turks he is known as Qanuni, "Lawgiver," mainly for his organization of the ulema or learned class, at the head of which he set the Sheikh ul-Islam, the mufti of Constantinople.

He died in early Sept. 1566, at the age of 72, while conducting the siege of Szigetvár, and was buried in Constantinople in the precincts of the splendid mosque that bears his name. *See* TURKEY: *History*. (G. L. L.)

SULEIMAN II (1642–1691), Ottoman sultan, was a son of Sultan Ibrahim and succeeded his brother Mohammed IV in 1687. An enforced retirement of 46 years qualified him for the cloister rather than for the throne, and his first feeling when notified of his accession was one of terror for his brother's vengeance. Nor were the circumstances following his elevation to the throne of a nature to reassure him, as one of the most violent of the revolts of the

janissaries ended in the murder of the grand vizier and the brutal mutilation of his family, with general massacre and pillage throughout Constantinople. Yet Suleiman possessed a martial spirit, ill-matched with a poor physique, and twice took the field with his troops.

The war with Austria was for Turkey a succession of disasters. At this time, fortunately for the Ottoman empire, a third great Köprülü or Kuprili (Mustafa), arose and re-established order in the sorely tried state. (*See* KOPRULU.)

In the reforms which followed, whereby the situation of the Christian subjects of the Porte was greatly improved, Suleiman is at least to be given the credit of having allowed Mustafa Köprülü a free hand. With an improved administration Turkey's fortunes in the war began to revive, and the reconquest of Belgrade late in 1690 was the last important event before Suleiman's death in 1691.

See also TURKEY: *History*. (G. L. L.)

SULEYMAN CHELEBI (d. 1422), Turkish poet, was the author of one work only, the great religious poem *Mevlid*. The son of a vizier, he was born in Bursa and became imam of the great mosque there. The *Mevlid* tells the story of the Prophet's life and death, his miracles and his heavenly journey. Though written in simple language it is inspired with religious fervour and is often recited at religious ceremonies, particularly funerals. Suleyman Chelebi seems to have made use of the legends on Mohammed which spread through Islamic literature. Further, a few verses of Ashik Pasha's *Garibname* (1329) are reproduced in the *Mevlid*, and the passage on the birth of the Prophet is closely inspired by a corresponding passage in Mustafa Darir's verse and prose translation of Ibn Ishak's *Sira*.

See F. Lyman MacCallum, *The Mevlidi Sherif*, with Eng. trans. (1943) and edition by Ahmed Atesh (1954). (F. I.)

SULFA DRUGS: *see* SULFONAMIDES.

SULFONAL GROUP. The group of compounds known as ethylsulfanes were first prepared in 1885 and introduced into therapeutics three years later by Eugen Baumann and Alfred Kast. Formerly popular as hypnotics, these agents are now seldom used. Sulfonmethane (sulfonal) is slowly absorbed from the intestinal tract, and its effects are slow in appearing. Depression after awakening is common and may be severe. Because the drug is slowly excreted, cumulation may occur during continuous administration producing toxic effects and poisoning. Death from overdosage is a result of failure of respiration and circulation. Chronic poisoning causes mental and gastrointestinal symptoms and urinary changes. Other manifestations may make their appearance in the skin, liver and kidneys.

Sulfonethylmethane (Trional) and Sulfondiethylmethane (Tetronal) are representatives of the group. Trional is absorbed more rapidly than sulfonal. Tetronal more slowly. (A. E. SH.)

SULFONAMIDES, a group of chemicals that have as their basic structure *p*-aminobenzenesulfonamide. From the parent substance, designated sulfanilamide, various related compounds are made, usually by substituting or adding various groups to the sulfone radical. New sulfonamides are being developed constantly with the aim of increasing their effectiveness and decreasing their toxicity. Attempts are made to insure the maximum therapeutic efficiency at minimum risk to the patient. Increased solubility, prolonged action and more powerful effects against bacteria are common goals. Initially, sulfonamides were produced for treating bacterial infections, but they have come to be used in treating histoplasmosis, South American blastomycosis and diabetes mellitus and as antihypertensive and diuretic agents.

History.—In 1932 Gerhard Domagk (*q.v.*) reported to the chemists Fritz Mietzsch and Joseph Klarer that the compound Prontosil, which they had prepared for study, had a curative action in many experimental streptococcal infections. Several years later, Domagk described experiments which demonstrated that Prontosil was a dye of low toxicity, and that when it was administered to mice infected with streptococci, it prevented death. This dye was also effective in controlling staphylococcic infections in rabbits, but it was ineffective in pneumococcic and other experimental infections. These studies by Domagk excited interest in chemotherapy in France, England and the United States. Within a relatively short period of time it was demonstrated that Prontosil was not only effective in experimental infections in animals but that it also had a powerful effect upon the course of streptococcic diseases in man, including meningitis and puerperal sepsis. It remained for French investigators to point out that the azo dye Prontosil opened up in the tissues, and that it was broken at the azo linkage to *p*-aminobenzenesulfonamide. Moreover, they suggested and later proved that the active principle in Prontosil was sulfanilamide.

In the United States the first important observations were made by P. H. Long and E. A. Bliss in Baltimore; they were soon followed by those of others. Experimental and clinical observations on patients were placed on a sound basis by E. K. Marshall and his associates through the development of chemical methods for determining the amount and distribution of these drugs in the body after they are given by mouth or injected into the blood stream. To Marshall and his colleagues belongs the credit for placing bacterial chemotherapy on a sound and rational basis, since clinicians, by using their methods of chemical analysis and knowing the pharmacology of these substances, could use objective methods in following patients undergoing such treatment. The first paper on this subject to arouse widespread interest and excitement in the U.S. medical profession was that of Long and Bliss. They described their experimental work in animals and their results in the treatment of patients with various infections, and these results were as remarkable as those reported in England and elsewhere. All these observations were soon confirmed and later extended.

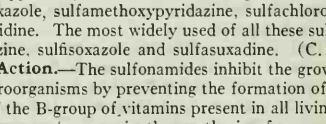
As new sulfonamides have been synthesized and applied in the study of infections and other diseases in man and animals, more effective and less toxic agents have been discovered. Also, some compounds have been developed that exert their main effect locally in the gastrointestinal tract and are not absorbed. Others are absorbed slowly or excreted slowly, so that they are longer acting.

The sulfonamides have come to be used most widely in the treatment of urinary tract infections, for the treatment and prevention of meningococcic meningitis, for bacillary dysentery, for pneumococcic pneumonia and for prophylaxis of streptococcic infections. They are combined with antibiotics to broaden the spectrum of activity and to delay the emergence of resistant bacteria, especially in *H. influenzae* and *Ps. aeruginosa* infections of the meninges.

Scarcely a year goes by that new sulfonamides are not synthesized and tested. A large number of drugs have been tested and found to be effective, among them sulfapyridine, sulfathiazol, sulfadiazine, sulfapyrazine, sulfamerazine, sulfaguanidine, sulfasuxadine, sulfisoxazole, sulfamethoxypyridazine, sulfachloropyridazine and sulfisomidine. The most widely used of all these sulfonamides are sulfadiazine, sulfisoxazole and sulfasuxadine. (C. S. KR.)

Mode of Action.—The sulfonamides inhibit the growth of susceptible microorganisms by preventing the formation of folic acid, a member of the B-group of vitamins present in all living cells and functioning as a coenzyme in the synthesis of many important constituents of protoplasm. Most bacteria make their own folic acid from simpler starting materials, while man and other higher animals must have folic acid supplied in the diet. Thus, the sulfonamides can inhibit the growth of invading microorganisms without harming the host, since the enzymatic reaction these drugs inhibit does not take place in the host.

T. C. Stamp provided the first clue to the mode of action of the



(A) (B)

sulfonamides by discovering in 1939 an unknown substance in yeast that antagonized the antibacterial effect of sulfanilamide.

A year later, D. D. Woods identified this substance as *p*-amino-

benzoic acid (A), which closely resembles sulfanilamide (B) in structure. Woods's explanation of how *p*-aminobenzoic acid reverses the inhibitory effect of sulfonamides had far-reaching effects in chemotherapy, nutrition and enzymology. Although *p*-aminobenzoic acid had never before been found in living cells, Woods postulated that it was essential for the multiplication of the bacterial cell and that the sulfonamides inhibited enzymes involved in its utilization by virtue of their close structural resemblance to *p*-aminobenzoic acid.

His prediction was soon verified by the discovery of *p*-aminobenzoic acid as a growth factor for several microorganisms and as a portion of the folic acid molecule. Thus, it is apparent that the sulfonamides act by being absorbed to folic acid-synthesizing enzymes in place of *p*-aminobenzoic acid and thus preventing folic acid formation.

Soon after Woods's discovery, Sir Paul Fildes predicted that still other compounds closely resembling vitamins and other vital cell constituents would inhibit the growth of microorganisms and that they might prove to be useful chemotherapeutic agents. Fildes' suggestion offered a rational approach to the chemotherapy of infectious diseases, and within a few years hundreds of structural analogues of vital cell components were synthesized and tested. Many specifically inhibited the growth of microbes and of higher organisms as well and thus provided valuable tools for the study of metabolism and nutrition. However, only a few proved useful as chemotherapeutic agents, chiefly because most structural analogues of important cell components were equally toxic to both host and parasite; that is to say, they did not have the selective toxicity for microorganisms which makes the sulfonamides so useful.

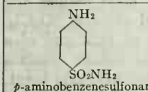
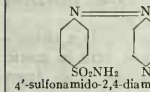
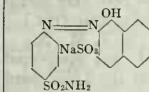
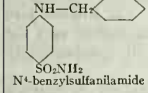
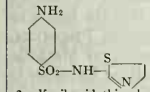
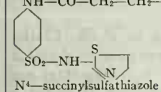
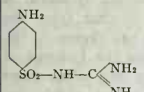
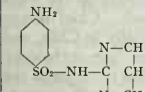
However, the theory of Woods and Fildes not only satisfactorily explains how the sulfonamides act, but it still offers the only approach to a rational search for chemical agents which effectively combat infectious diseases. (J. W. Mr.)

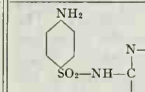
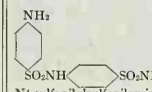
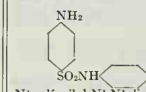
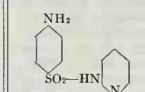
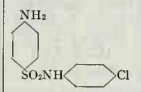
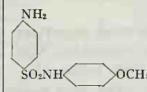
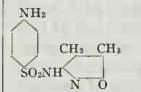
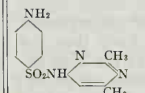
Sulfonamide Resistance.—Bacteria become increasingly resistant to the sulfonamides after exposure to them. These agents are principally bacteriostatic, and with the body's defense mechanism they act to control infection. When the action of the sulfonamides is interfered with by pus or chemicals such as procaine or para-aminobenzoic, then strains of bacteria that are resistant to their action emerge and survive.

Also, the bacteriostatic action of sulfonamides is diminished by acetylation, a process in which an acetyl group is substituted for one of the hydrogens of the para-amino group of the sulfonamide molecule.

The problem of sulfonamide resistance is of importance in the

Formulas, Medical Designations and U.S. Trade Names of the Sulfonamides

Chemical formula	Medical designation	Trade name
 *p*-aminobenzenesulfonamide	Sulfanilamide	Prontylin Streptocide Prontosil Album Lysococcine
 4′-sulfonamido-2,4-diaminoazobenzene	Prontosil	Prontosil Prontosil Flavum
 sodium-4′-sulfonamidophenyl-2-azo-7-acetylamino-1-hydroxynaphthalene-3,6-disulfonate	Neoprontosil	Neoprontosil Prontosil Soluble
 N^4-benzylsulfanilamide	Benzyl Sulfanilamide	Septazine Proseptasine
 2-sulfanilamidothiazole	Sulfathiazole	
 N^4—succinylsulfathiazole	Sulfasuxidine	
 Sulfanilyl—guanidine	Sulfaguanidine	
 2-sulfanilamidopyrimidine	Sulfadiazine	

Chemical formula	Medical designation	Trade name
 2-sulfanilamido-4-methylpyrimidine	Sulfamerazine	
 N^4-sulfanilylsulfanilamide	Sulfanilyl Sulfanilamide	Disulon
 N^4-sulfanilyl-N^1-N^1-dimethylsulfanilamide	Sulfanilyl Dimethyl Sulfanilamide	Uliron Ulcron
 2-sulfanilamidopyridine	Sulfapyridine	M & B 693 Dagenan
	Sulfachloropyridazine	
	Sulfamethoxypyridazine Sulfacetamide	Kynex
	Sulfisoxazole	Gantrisin
	Sulfisomidine	Elkosin

management of infections. When sulfadiazine was used in mass prophylaxis during World War II in an attempt to control the spread of hemolytic streptococcic infections and thus prevent tonsillitis, scarlet fever and their consequences (rheumatic fever and acute nephritis), it was found that resistant strains soon appeared in the population being treated. From patients harbouring sulfonamide-resistant strains the organisms spread to others in the community; thus new outbreaks of sulfonamide-resistant streptococcic disease were observed.

Triple Sulfonamides.—In an attempt to increase the solubility of sulfonamides and decrease their toxicity and their capacity to produce hypersensitivity, sulfonamide mixtures have been advocated and widely used. O. Lehr pointed out that several sulfonamides can be dissolved simultaneously in the same medium almost to the full extent of their saturation level without the occurrence of precipitation. Moreover, he showed that mixtures of partial dosages of several sulfonamides showed substantially less tendency toward the development of renal obstruction by intratubular deposition of crystals than do any of the single components of the mixture in equal total dosage.

Highly soluble sulfonamides, or sulfonamide mixtures that are more soluble than individual members of the mixtures, have been found to be highly effective in the prevention of renal complications as well as therapeutically.

Toxic Effects.—All the sulfonamides are capable of causing drug intoxication, and some patients are hypersensitive to them. It is advisable that these drugs be taken only under the direction of a physician and for no longer a period than necessary. The commonest side effects are nausea, vomiting and mental confusion. The signs of hypersensitivity are fever and skin eruptions. Other signs of intoxication are anemia resulting from destruction of red blood cells and leukopenia from destruction of white blood corpuscles. Irritation of the kidneys and obstruction of the free flow of urine are undesirable reactions that can be avoided by using the triple sulfonamides and keeping the reaction of the urine alkaline or by using the highly soluble bacterocyclic derivatives of sulfanilamide and maintaining urinary output at a quart to a quart and a half a day.

All the signs of drug intoxication should be watched for in patients receiving sulfonamide drugs, and if any signs occur the drug should be discontinued at once.

See CHEMOTHERAPY. *See* also Index references under "Sulfonamides" in the Index volume.

BIBLIOGRAPHY.—"Second Conference on Sulfonamides," *Ann. N.Y. Acad. Science*, 69:377–524 (Oct. 12, 1957); A.M.A. Council on Drugs, "Sulfonamide Compounds," *New and Non-Official Remedies*, pp. 45–63 (1958); "Toxic and Other Special Features of Sulfonamide Therapy" in Harry Beckman, *Treatment in General Practice*, 6th ed., pp. 977–89 (1948); F. Hawking and J. S. Lawrence, *The Sulphonamides* (1950); P. H. Long and E. A. Bliss, *The Clinical and Experimental Use of Sulfanilamide, Sulfapyridine and Allied Compounds* (1939); T. S. Work and F. Work, *The Basis of Chemotherapy* (1948); "The Sulfonamides," *Merck Manual*, 9th ed., pp. 1016–20 (1956); *Sulfanilamide and Related Compounds in General Practice* (1942); A. Smith and P. L. Wermer, *Modern Treatment*, pp. 51–74 (1952); D. W. Woolley, *A Study of Antimetabolites* (1952). (C. S. KR.)

SULFONIC ACIDS, organic acids containing sulfur and having the general formula $R.SO_3H$, where R is either an alkyl or an aryl radical.

Aromatic Sulfonic Acids.—The acids of this more important group are obtained generally by treating aromatic compounds with concentrated sulfuric acid, the process being called sulfonation. Such operations are frequently conducted on a manufacturing scale as in the preparation of the sulfonic acids of benzene (*q.v.*), naphthalene (*q.v.*) and anthraquinone (*q.v.*). The sodium salt of benzenesulfonic acid is of importance since on fusion with sodium hydroxide it yields sodium phenoxide, from which phenol is liberated by a strong acid. During World War I large quantities of synthetic phenol were made by this process in order to eke out the supply of coal-tar phenol, which was then insufficient for the wholesale production of the explosive, picric acid (*q.v.*). From then, the process retained its importance in consequence of the large demand for phenol in the manufacture of synthetic resins and plastics (*q.v.*). Benzenedisulfonic acid, made from oleum (fuming sulfuric acid) and benzene, yields by alkali fusion the important dye intermediate, resorcinol.

The sulfonic acids of naphthalene are manufactured on an extensive scale. Anthraquinone, the oxidation product of anthracene (*q.v.*), when sulfonated with oleum at high temperatures yields anthraquinone-β-sulfonic acid, the sodium salt of which ("silver salt") is an important intermediate used in the manufacture of alizarin (*q.v.*) and the indanthrene vat dyes. Sulfonation of anthraquinone is facilitated by a mercuric sulfate catalyst. Sulfonation may be effected also by chlorosulfonic acid whereby sulfonyl chlorides are formed. (*See* SACCHARIN.)

Sulfonic acids of aromatic hydrocarbons are very soluble in water and when obtained crystalline are extremely hygroscopic. With phosphorus pentachloride they yield sulfonyl chlorides, which are reducible to thiophenols.

Sulfanilic acid, a sulfonic acid of aniline, is sparingly soluble in cold water and is to be regarded as an internal salt,

$$\overset{+}{N}H_3.C_6H_4.\overset{-}{S}O_3.$$

Its amide is important as the simplest member of the medically important sulfonamides (*q.v.*). Sulfanilic acid and the similar naphthylamine sulfonic acids are important dye intermediates.

Aliphatic Sulfonic Acids.—These are less important than those of the aromatic series. They are obtained generally (1) by oxidation of mercaptans, alkyl disulfides, alkyl thiocyanates or alkyl sulfinic acids; (2) by the interaction of alkyl iodides and metallic sulfites; (3) by the action of sulfuryl chloride SO_2Cl_2, upon a saturated hydrocarbon in the presence of light and of a catalyst such as pyridine; (4) by the action of sodium bisulfite upon an olefine in the presence of oxygen or other oxidizing agent. They are either viscous liquids or crystalline solids, converted into sulfonyl chlorides by phosphorus pentachloride.

Hydroaromatic Sulfonic Acids.—Camphor yields a characteristic sulfonic acid on sulfonation with sulfuric acid in acetic anhydride. The optically active camphor sulfonic acids from dextro- and laevo-modifications of camphor have been very useful in the resolution of racemoid bases. (*See* STEREOCHEMISTRY.)

(G. T. M.; X.)

SULFUR, a nonmetallic chemical element, was discovered in prehistoric times. Its name is derived from the Latin word *sulphurum*. It is often called brimstone (burning stone), and ignited sulfur is mentioned in the earlier records of many countries as having been used for religious ceremonies, for purifying (fumigating) buildings and for bleaching cloth. The Romans also used it in medicine and in warfare. During the middle ages the alchemists regarded sulfur as the principle of combustibility (phlogiston).

This article, after discussion of the classification, occurrence and methods of production of sulfur, is divided into the following main sections: I. Physical Properties; II. Chemical Properties; III. Commercial Forms and Uses; and IV. Analytical Methods.

Sulfur was first classified as an element by Antoine Lavoisier in 1777 but many of his contemporaries considered it a compound of hydrogen and oxygen. Joseph Gay-Lussac and Louis Thénard finally proved it an element in 1809.

Occurrence.—The earth's crust, according to F. W. Clarke, contains 0.06% of sulfur as sulfides, sulfates and elemental sulfur. As elemental or native sulfur, it occurs chiefly in sedimentary and volcanic deposits. Most of the world's production of sulfur is from sedimentary deposits and the majority of that from the calcareous horizons of the cap rock of shallow salt domes in the coastal region of Texas and Louisiana, U.S., and the Isthmus of Tehuantepec, Mex. Italy, chiefly Sicily, is the only other large producer from sedimentary deposits. Volcanic sulfur, widely distributed throughout the world, occurs as a sublimate on the walls of vents or permeating porous lava and ash beds. Chief occurrences are in Japan, the Andes mountains, especially in Chile, and Indonesia. Smaller deposits, formed at or near the surface by atmospheric oxidation, are also widely distributed throughout the world in almost any type of rock, but these are of minor commercial importance.

In combination, the element occurs chiefly as sulfides and sulfates. The former are of importance as a source of sulfur; *e.g.*, pyrite (FeS_2), marcasite (FeS_2), pyrrhotite (Fe_nS_{n+1}), chalcopyrite ($CuFeS_2$), galena (PbS) and sphalerite (ZnS). The chief sulfates are anhydrite ($CaSO_4$), gypsum ($CaSO_4+2H_2O$) and barite ($BaSO_4$) also known as barytes or heavy spar. Sulfur also occurs in organic substances; *e.g.*, hair, wool, albumen, garlic, mustard, horse-radish, cabbage and many proteins.

Methods of Production.—Ancient producers described their method of obtaining sulfur from its ores as purification by fire. Similar methods are still used in Sicily where native sulfur ore, mined by conventional methods, is piled in open cylindrical kilns (*calcaroni*), covered with spent ore and ignited. Heat from burning about one-third of the sulfur melts the remainder which drains from the kilns and is cast into blocks. Somewhat better sulfur recovery is obtained in other kilns (Gill furnaces) constructed so that combustion gases from one kiln can be used to preheat the ore in another. In Japan, sulfur is usually distilled from the ore in retorts. Sulfur ores may also be heated with steam or superheated water in autoclaves. The sulfur melts, drains from the gangue (worthless rock) and is recovered in the liquid state. Some sulfur is recovered by this method in the western part of the United States but after 1903 nearly all U.S. sulfur was produced by the Frasch process from the deposits located near the Gulf of Mexico in Texas and Louisiana.

All the sulfur deposits mined by the Frasch process occur at depths of 200 ft. to 2,000 ft. in the cap rock of salt domes. In horizontal section, the domes are circular to elliptical in shape. Surrounding and overlying the domes are clay, sand and gumbo. The upper portion of the dome consists of a porous cap rock which is largely limestone and calcite, with an occasional trace of the carbonates and sulfates of the other alkaline earths. The sulfur occurs in the porous limestone in the form of rhombic crystals. Below the limestone is a stratum of massive anhydrite which in turn overlies the rock-salt core of the dome.

Crystalline sulfur melts at about 116° C. The Frasch process takes advantage of this fact and fuses the sulfur underground by pumping water heated above this temperature into the deposit. The melted sulfur, which is almost chemically pure, flows away from the gangue and is pumped to the surface.

In mining, wells are drilled into the sulfur formation using rotary rigs such as those employed in the petroleum industry (*see* DRILLING, PETROLEUM). The wells are equipped with pipes of various sizes placed concentrically, which reach from the surface of the ground into the sulfur deposit. A 6-in. pipe extends through the sulfur-bearing stratum and rests in the upper portion of the underlying calcium sulfate or gypsum. A 3-in. pipe is placed inside the 6-in. pipe, reaching nearly to the bottom of the sulfur-bearing rock but resting on a collar that is set within a 6-in. pipe and seals the annular space between the two. Finally, a 1-in. air pipe, inside the 3-in. pipe, extends to a depth slightly above the collar. The 6-in. pipe is perforated at two levels, separated by the collar, the upper set of holes permitting the escape of hot water and the lower the entrance of molten sulfur.

Hot water is pumped down the annular space between the 6-in. and 3-in. pipes and discharged into the porous formation through the perforations already described. The entire region through which the water circulates is raised to a temperature above the melting point of sulfur. The melted sulfur, being heavier than water, settles and forms a pool around the foot of the well and, entering the lower perforations, rises in the 3-in. pipe. Compressed air, released at the bottom of the central 1-in. pipe, raises the liquid sulfur to the surface. Then the molten sulfur is pumped from the wells to near-by sumps, and from them to vats (bins) to solidify. A completed vat, consisting of a solid block of sulfur, may be 1,200 ft. long, 50 ft. high and 160 ft. to 200 ft. wide and contain 500,000 tons of sulfur.

The sulfur production described is that obtained in elemental form from natural deposits. In addition, increasing quantities of elemental sulfur are recovered from pyrites and from such industrial gases as coke-oven gas, "sour" natural gas and petroleum refinery gases. By-product sulfur, as this type of sulfur is called, was being produced in Canada, Egypt, France, Germany, Iran, Iraq, Mexico, Netherlands, Norway, Sweden, Trinidad, United Kingdom, Portugal and the United States by the 1950s.

I. PHYSICAL PROPERTIES

The element sulfur, chemical symbol S, exhibits nonmetallic properties, has an atomic number of 16 and chemical atomic weight 32.066. It occupies the position immediately below oxygen in Group VI of the periodic table. Four stable isotopes are known, each with a nucleus containing 16 protons, but differing by having 16, 17, 18 and 20 neutrons respectively. The respective mass numbers are 32, 33, 34 and 36, and the normal composition of sulfur is given as S^{32} (95%), S^{33} (0.75%), S^{34} (4.2%), S^{36} (0.017%). Three radioactive isotopes, S^{31} (half life 2.6 sec.), S^{35} (half life 87 days) and S^{37} (half life 5.0 min.), have been prepared. The arrangement of the electrons in the outer shell of all atoms is represented as follows: $1S^2$, $2S^2$, $2p^6$, $3s^2$, $3p^4$.

Pure solid sulfur has a pale yellow colour but its naturally occurring forms may be sulfur yellow, straw or honey yellow, yellowish brown, greenish or reddish to yellowish gray. Just above the melting point sulfur is a transparent yellow liquid which, with increasing temperature, turns dark red and becomes black at 230° C. At the boiling point (444.6° C.) vapours are generated which turn deep red at 500° C. and a straw yellow at 850° C.

Sulfur is tasteless and odourless. It has no action on the skin. It is insoluble in water, highly soluble in carbon disulfide, moderately so in aromatic solvents and indifferently so in aliphatics. The liquid density varies from 1.808 g. per millilitre at 115° C. to 1.599–1.614 g. per millilitre at the boiling point. At 200° C. (reduced pressure) the vapour density corresponds to S_8 mole-

TABLE I.—*World Production of Elemental Sulfur*
(long tons)

Country	Source	1952	1953	1954	1955	1956
Argentina	sulfur ores	15,000	16,000	17,000	17,651	23,038
Bolivia	sulfur ores	5,497	2,458	2,565	3,975	3,418
Canada	pyrites	—	—	—	10,000*	15,000*
	gas	3,773	14,352	16,667	25,980	29,222
Chile	sulfur ores	47,821	32,275	39,075	54,132	45,000*
Colombia	sulfur ores	2,974	2,657	5,118	5,413	4,921
Egypt	gas	2,800	3,250	3,640	4,000	3,000
France	gas	—	—	—	2,850	5,000
	sulfur ores	17,871	10,818	10,000	—	—
Germany	gas	51,100	59,600	67,800	70,800	77,110
Greece	sulfur ores	—	1,200	2,507	3,600	3,600*
Iran, Iraq	gas	6,000*	6,000*	6,000*	6,000*	10,000*
Italy	sulfur ores	232,706	224,761	200,215	179,755	170,094
Japan	sulfur ores	176,652	186,556	184,234	199,219	243,312
Mexico	sulfur ores	11,784	5,000	5,000*	5,000*	5,000*
	gas	36,300	27,588	26,514	20,062	25,723
	frasch mines	—	—	47,497	470,487	753,415
Netherlands	gas	3,000	11,713	9,252	6,900	12,000*
Netherlands Antilles	gas	4,000*	10,000*	15,000*	20,000	12,000
Norway	pyrites	103,143	100,112	90,002	98,871	95,423
Peru	sulfur ores	5,066	4,916	5,000*	2,000*	2,000*
Portugal	pyrites	15,627	16,340	16,100	15,466	16,879
Spain	sulfur ores	4,800	5,100	5,400	6,500	5,000
	pyrites	34,500	33,445	35,000*	31,888	45,000*
Sweden	gas	15,365	23,426	27,274	28,434	30,342
Taiwan (Formosa)	sulfur ores	5,001	3,423	5,873	4,854	7,864
Trinidad	gas	—	3,000*	4,300	5,000*	5,000*
Turkey	sulfur ores	8,232	9,626	9,862	11,318	3,722
United Kingdom	gas	14,800	28,300	40,850	45,850	53,000
United States	sulfur ores	2,197	38,257	64,333	60,902	60,402
	gas	251,198	341,660	359,271	398,780	464,758
	frasch mines	5,293,145	5,155,342	5,314,640	5,738,978	6,423,883
Total		6,370,552	6,374,750	6,845,299	7,554,665	8,655,026

*Estimated.
Source: From Table 7, "World Production of Native Sulfur, by Countries," compiled by Helen L. Hunt, and part of *Mineral Market Report MMS. No. 2701* prepared by Leonard P. Larson and Annie L. Mattila, with additional data supplied by E. Wick Eddy.

cules, at 600° C. to S_6, at 1,000° C. to S_2 and at 4,000° C. to monatomic sulfur. Increased pressure decreases the dissociation of the vapour. The critical temperature is 1,040° C. The latent heat of vaporization decreases from 84.8 cal. per gram at 120° C. to 68.1 cal. at 360° C. and then rises to 73.9 cal. at 540° C. The latent heat of fusion is 9.21 cal. per gram (118.9° C.).

Sulfur is a poor conductor of both heat and electricity and becomes charged negatively if rubbed with glass, fur, silk, wool or hard rubber. On Mohs' scale, its hardness varies from 1.5 to 2.5. The vapour pressure varies from 0.0002 mm. of mercury at 50° C. to 3,824 mm. at 570° C. The coefficient of thermal expansion continuously increases with temperature for the solid while that for the liquid passes through a minimum value. The bond angle, the interatomic distance and the atomic radius for crystalline S_8 are 105°, 2.12Å (Å, the angstrom unit $= 10^{-8}$ cm.) and 1.04 Å, respectively. For the S_8 gas the bond angle is 100° but the other two values are the same as for the solid. The ignition temperature of sulfur in air is 261° C.

Allotropic Modifications.—A number of crystalline forms of sulfur have been identified by their external appearance and optical properties. These, according to D. M. Yost and H. Russell, Jr., arranged in their suggested order of stability are:

S_θ (tetrahedral)$\rightarrow S_\delta$ (monoclinic)$\rightarrow S_\eta$ (monoclinic)$\rightarrow$ S_ε (monoclinic)$\rightarrow S_\gamma$ (monoclinic)$\rightarrow S_\beta$ (the usual monoclinic sulfur)$\rightleftharpoons S_\alpha$ (rhombic sulfur).

These forms differ in external appearance, crystal structure and physical properties; *e.g.*, density, solubility and melting point. Of the allotropic forms, the two most important are rhombic (α, S_α, or S_1) and monoclinic (β, S_β, or S_{11}).

Thermodynamically, rhombic sulfur is stable at room temperature and is readily obtained by crystallization from solution in carbon disulfide. It is a yellow solid with a density of 2.07 g. per millilitre and melting point 112.8° C. (if heated rapidly). Above 95.5° C., the transition temperature, the crystals become opaque and consist of aggregates of the monoclinic form. An X-ray investigation of rhombic crystals has shown that the unit cell is composed of 16 puckered rings of S_8 molecules arranged approximately in layers perpendicular to the C-axis.

Monoclinic sulfur is obtained when molten sulfur is cooled slowly. On breaking the surface crust and pouring off the still liquid sulfur, the interior of the cooled mass will be found to be filled with needlelike, monoclinic crystals. The density of the monoclinic sulfur is 1.96 g. per millilitre and its melting point is 119° C. Under higher pressures, the two forms can only coexist at a definite temperature for any particular pressure (*i.e.*, the transition point is raised), and if the pressure is raised while temperature is kept constant, the rhombic form is produced since it is the more dense. This balanced state of affairs continues to 151° C. and 1,288 atm.; above this temperature and pressure the monoclinic form cannot exist. The atomic heat of transition is 95 cal.

On standing at room temperature, monoclinic sulfur slowly changes into rhombic sulfur but the outward monoclinic crystalline form is preserved, each monoclinic crystal consisting of an aggregate of very small crystals of rhombic sulfur. Commercial varieties of sulfur such as ordinary brimstone, roll sulfur and flowers of sulfur consist mainly of the rhombic form and a small amount of insoluble or amorphous sulfur.

Carbon disulfide dissolves both rhombic and monoclinic sulfur, 100 g. of it dissolving 50.4 g. of rhombic sulfur at 25° C., 100 g. at 40° C. and 376.1 g. at 80° C. At 0° C., however, the monoclinic form is 1.28 times as soluble as the rhombic and at 25.3° C. it is 1.4 times as soluble. The molecular weight of both forms in this solvent corresponds to S_8. Another very good solvent for rhombic or monoclinic sulfur is sulfur monochloride, S_2Cl_2.

Rhombic sulfur melts at 112.8° C.; monoclinic sulfur melts at 119° C., but the natural melting point of commercial sulfur is a few degrees lower. This is because liquid sulfur consists of a mixture of two or three molecular species and time is required before equilibrium is established between them. At 120° C. liquid sulfur has a viscosity of approximately 10 centipoises; further heating causes the viscosity to drop to a minimum at about 157° C. At 159° C. to 160° C. the viscosity increases rapidly to a maximum of 932 poises at 186° C. to 188° C. after which it again decreases.

Various explanations have been given for the peculiar viscosity of liquid sulfur. At lower temperatures near the melting point it is supposed that the S_λ or S_8 molecule predominates. As the temperature is increased to 160° C., the eight-membered ring molecules rupture, giving an equilibrium mixture of rings and eight-membered chains. The increasing concentration of these eight-membered chains causes polymerization to long complicated chains. In addition to the long straight chain, it is highly possible that side chains of varying lengths exist. A maximum chain length of 12,000 atoms to 27,000 atoms is attained at approximately 187° C., above which the polysulfur chains break into shorter chains. In addition to S_λ or S_8, and S_μ or S_6, still another molecular species of sulfur, S_π or S_4, is believed to exist in both the liquid and in the rapidly cooled solid. The percentage of S_π at 120° C. is approximately 3.5% and at 170° C., close to the temperature of maximum viscosity, it reaches a maximum of approximately 6.8%. S_λ is soluble in carbon disulfide while S_μ and S_π are not.

If hot, molten sulfur is poured into water it forms a clear, soft, sticky, elastic mass known as plastic sulfur. X-ray analyses show a fibre diagram and it is believed that long chains of sulfur atoms are arranged parallel to the direction of stretching. The stress-strain curves for these elastic strings appear very similar to those given by raw rubber. Plastic sulfur becomes opaque and brittle after a few days because of the conversion of most of the S_μ and some of the S_π into crystalline rhombic sulfur. When treated with carbon disulfide, only part (S_λ) of this hardened sulfur will dissolve, leaving amorphous or insoluble S_μ and S_π.

Other allotropic forms have been reported, such as gamma sulfur (γS, Muthman's S_{III}, nacreous sulfur). This form, mixed with S_α and S_β, is prepared by the crystallization of molten sulfur as it is cooled to 90° C. It is also formed by the slow diffusion of aqueous solutions of sodium thiosulfate and potassium hydrosulfate. The yellowish-white monoclinic crystals have a mother-of-pearl lustre with axial ratios different from those of beta sulfur. Colloidal sulfur (delta sulfur, Muthman's S_{IV}), showing no sign of crystal structure, is formed when hydrogen-sulfide gas is passed into a saturated aqueous solution of sulfur dioxide at 0° C. or when an alcoholic solution of sulfur is poured into water. Other forms reported are theta sulfur (θS), which is an unstable pale-yellow tetragonal form; zeta sulfur (ζS), which is a colourless rhombic plate crystal, and eta sulfur (ηS), which is colourless, double refracting hexagonal plates.

II. CHEMICAL PROPERTIES

All the metals except gold and platinum combine with sulfur to form sulfides and, in most cases, heat is given off during the union. Moist sulfur is slowly oxidized at ordinary temperatures to sulfuric acid. In combination with metals and hydrogen, sulfur has a valence of -2, forming compounds like H_2S, FeS, CuS and HgS. In combination with nonmetals, however, the valence or oxidation state may be $+4$ or $+6$ as in SO_2 and SO_3 respectively. Sulfur is therefore an oxidizing agent as well as a fair reducing agent. Its oxides are acid-forming.

Sulfides.—Hydrogen sulfide (H_2S) is formed by the reaction of hydrogen with molten sulfur at 200° C. to 400° C. and with sulfur vapour at temperatures up to 800° C. It is also formed by the action of acids on metallic sulfides such as ferrous sulfide, $FeS + 2HCl \rightarrow FeCl_2 + H_2S$; by putrefaction of organic substances containing sulfur; by the destructive distillation of coal; and by heating sulfur with hydrocarbons such as paraffin, petroleum oil, methane, etc.

Hydrogen sulfide, a colourless gas having the odour of rotten eggs, is fairly soluble in water, giving a feebly acidic solution which slowly deposits sulfur when exposed to air. This gas, poisonous when inhaled, burns in air with a pale-blue flame to give sulfur dioxide and water. It can be liquefied at 18° C. by a pressure of 17 atm., the liquid boiling at -60.1° C. under atmospheric pressure and solidifying at -82.9° C. Hydrogen sulfide, often used

as a reducing agent, is oxidized by the halogens to sulfur, excess halogen oxidizing the sulfur to sulfuric acid; by hydrogen peroxide to thiosulfate and finally, to sulfate; and by concentrated sulfuric acid to sulfur and sulfur dioxide.

Hydrogen persulfide, H_2S_2, density 1.3339 (20° C.), melting point −89.6° C., boiling point 70.7° C. (± .5), is prepared by composing a polysulfide with a strong acid, as by treating $Na_2S_{2.5}$ with hydrochloric acid. Hydrogen trisulfide, H_2S_3, may be obtained as a by-product in the reaction. The compounds H_2S_5 and H_2S_6 have also been described. The persulfides all dissolve readily in organic solvents.

The hydrogen atoms in hydrogen sulfide can be replaced by metals to form the normal sulfide, M_2S, or the acid sulfide, MHS. Sulfides of the alkali metals dissolve readily in alcohol and in water and are used in the paper and leather industries. Many sulfides, those with high heats of formation, can be made by grinding the metal and sulfur. Others require heating; in some instances a high temperature is necessary to start the reaction. Sulfates, especially those of the alkali and alkaline earth metals, can be reduced to the sulfides by heating with carbon. Heavy-metal sulfides, which are all highly insoluble in water (being among the most insoluble of all substances), can be made also by passing hydrogen sulfide through solutions of metal salts. This precipitation with hydrogen sulfide is used in qualitative analysis.

The sulfides of the heavy metals have a characteristic appearance and colour. Some are volatile at moderate temperatures, mercuric sulfide subliming at 446° C. They can be iodized in air, and many are hydrolyzed by water with the evolution of hydrogen sulfide. Some metallic sulfides containing small amounts of heavy-metal salts are phosphorescent. Zinc sulfide is used in luminous paints and, combined with traces of a radium salt, it is used for painting and figures on luminous watch dials.

Thio Salts.—Metallic sulfides react with sulfides of the more negative elements ($3Na_2S+P_2S_5 \rightarrow 2Na_3PS_4$) to give thio salts or sulfo-salts. This reaction is used analytically to separate As, Sb, Sn from Cu, Pb, Bi, Cd and Hg.

Polysulfides.—Sulfur combines with simple sulfides like Na_2S or NaSH to form polysulfides having the general formula Na_2S_x in which x varies from 2 to 5. (Some Na_2S_6 is formed in liquid ammonia solution.) The polysulfides, well-crystallized hygroscopic solids which are oxidized by air to form thiosulfates, are used by the tanning industry. Lime-sulfur solution, consisting of calcium sulfide and polysulfides, is made by boiling sulfur with a suspension of lime and is used both as an insecticide and as a fungicide.

The sulfides of carbon, silicon and boron are obtained by direct union of sulfur with these elements at high temperatures. CS_2 is unimolecular. It boils at 46.25° C., while SiS_2 and B_2S_3 are solid compounds which are difficult to volatilize. Both of the latter, in contrast to CS_2, burn in oxygen with difficulty. SiS_2 and B_2S_3 are hydrolyzed by water. CS_2 is important commercially, being used in the preparation of viscose rayon and carbon tetrachloride and as a solvent. Other sulfides such as CS, C_3S_2 and SiS are also known. The former, a solid, is prepared by the action of a silent electric discharge on the vapour of CS_2. Carbon and silicon also form sulfides which contain oxygen or a halogen. Carbon oxysulfide, COS, is always formed when carbon, oxygen and sulfur are brought together at high temperatures. In its physical properties, carbon oxysulfide is intermediate between CO_2 and CS_2. It is a colourless, odourless gas which condenses to a liquid at −47.5° C.

Nitrogen Sulfides.—Another series of compounds is the sulfides of nitrogen. Nitrogen sulfide, NS_2, is a red liquid. The tetrasulfide, N_4S_4, a yellow or orange-yellow solid, melting point 179° C., is prepared by passing ammonia through SCl_2 dissolved in benzene or by dissolving sulfur in liquid ammonia at −15.5° C. The material explodes when struck. The pentasulfide, N_2S_5, a red liquid with a melting point of 10° C., is obtained when the tetrasulfide is boiled with CS_2. Nitrogen tetrasulfide has been suggested for use as an ignition promoter in diesel fuels and as an insecticide and fungicide.

Oxides of Sulfur.—Sulfur burns readily in air to form SO_2 and traces of SO_3. SO_2 may also be easily prepared by treating sulfites or bisulfites with a strong acid: $Na_2SO_3+2HCl \rightarrow 2NaCl+H_2O+SO_2$. It is available in the liquid form, boiling point −8° C., melting point −72.7° C., at low cost in steel cylinders. It has many important commercial uses, being used as a bleaching or reducing agent, a refrigerant and a solvent.

Sulfur monoxide, known only in the gas phase, may be prepared by subjecting a mixture of SO_2 and sulfur vapour at a pressure of from 1 mm. to 10 mm. to an electric discharge. It may be condensed to a solid at liquid-air temperatures but on melting it converts into SO_2 and sulfur.

Sulfur trioxide, SO_3, is formed by heating SO_2 and oxygen in the presence of a catalyst such as finely divided platinum, nickel or cobalt sulfates, or the oxides of vanadium, tungsten, molybdenum, chromium or iron. This reaction is the basis of the contact process for the manufacture of sulfuric acid (*q.v.*). SO_3 may also be prepared by heating anhydrous ferric sulfate, $Fe_2(SO_4)_3$, to 600° C. or 700° C. or by allowing SO_2 and ozone, O_3, to come together at room temperature. Two other oxides of sulfur, S_2O_3 and SO_4, are known, and S_2O_7 has been described. The sesquioxide, S_2O_3, is formed on adding powdered sulfur to liquid sulfur trioxide at 15° C. It is a blue to blue-green solid which decomposes slowly into SO_3, SO_2 and S. Sulfur tetroxide, SO_4, a white solid, is prepared by subjecting a mixture of SO_2 and O_2 to a silent electric discharge. In the gas phase, SO_3 consists of monomeric molecules but in the liquid state, boiling point 46° C., it consists of a mixture of SO_3 and S_3O_9 molecules and perhaps others. It also exists in several solid modifications: the alpha or icelike modification, the beta or low-melting asbestoslike modifications and the gamma or high-melting asbestoslike form. The molecular composition of liquid and solid SO_3 has not been completely investigated. SO_3 reacts energetically with organic matter to form sulfonic-acid derivatives of carbon and SO_2, and with basic oxides to form sulfates. In liquid form it combines with NO_2 to form oxynitroso sulfuric anhydride, $(NO_2SO_3)_2$, and with water to form a series of hydrates of sulfuric acid.

The oxides of sulfur dissolve in water to form acids. The following list from Yost and Russell's *Systematic Inorganic Chemistry of the Fifth-and-Sixth-Group Nonmetallic Elements* (copyright, 1944, by Prentice-Hall, Inc., New York) contains all the definitely established oxyacids of sulfur. Where the free acid is not known, the fact is indicated by giving the formula of a known salt of the acid in parentheses. In each case a method of preparation of the acid or salt is given:

H_2SO_2	($ZnSO_2$) Sulfoxylic. SO_2Cl_2+Zn in ether solution
$H_2S_2O_4$	Hyposulfurous. $NaHSO_3$ (aq.) + Zn
H_2SO_3	Sulfurous. SO_2 + water
$H_2S_2O_5$	($K_2S_2O_5$) Pyrosulfurous. From solution $KHSO_3+SO_2$
$H_2S_2O_6$	Dithionic. SO_2 + aq. suspension MnO_2
H_2SO_4	Sulfuric. SO_3 + water
$H_2S_2O_7$	Pyrosulfuric. $SO_3+H_2SO_4$ (100%). Salts, *e.g.* $Na_2S_2O_7$, formed by strongly heating $NaHSO_4$
$H_2S_2O_3$	($Na_2S_2O_3$) Thiosulfuric. Na_2SO_3+S
H_2SO_5	Peroxymonosulfuric. 30% H_2O_2 + conc. H_2SO_4
$H_2S_2O_8$	Peroxydisulfuric. Electrolysis of $KHSO_4$ solution
$H_2S_nO_6$	Polythionic, n = 3, 4, 5, 6. In Wackenroder's solution, SO_2+H_2S + water.

The sodium salt of hyposulfurous acid, sodium hyposulfite, is an article of commerce and is used as a reducing agent in the dye industry. Sulfurous acid and its salts are excellent and cheap reducing agents and are also used as disinfectants and food preservatives. Calcium bisulfite and magnesium bisulfite are used to extract lignin from wood in the preparation of wood pulp. Thiosulfates are used extensively in photography.

The Halides and Oxyhalides of Sulfur.—The halides and oxyhalides of sulfur, as described by Fritz Ephraim, possess certain properties in common. They are all readily prepared, usually by several methods. All are either gases or volatile liquids. Except for the fluorides, which exhibit great stability, all hydrolyze readily to give acid products which are irritating to mucous membranes. Again, except the colourless, tasteless fluorides, these compounds possess a pungent odour. As one might expect, the physical properties, reactivity and stability of these materials vary regularly with the position of the halide in the periodic table as illustrated in Tables II and III. Sulfur monochloride and

TABLE II.—*Properties of Sulfur Halides*

Compounds		Molecular weight	Melting point ° C.	Boiling point ° C.
Name	Formula			
Sulfur monofluoride	S_2F_2	102.12	−105.5	−99
Sulfur monochloride	S_2Cl_2	135.04	− 80	138
Sulfur monobromide	S_2Br_2	223.95	− 46	54 at 0.18 mm.
Sulfur dichloride	SCl_2	102.97	− 78	59 decomposes
Sulfur tetrafluoride	SF_4	108.06	−124	−40
Sulfur tetrachloride	SCl_4	173.89	− 31	−15 decomposes
Sulfur pentafluoride	S_2F_{10}		− 92	29
Sulfur hexafluoride	SF_6	146.06	− 50.8	−63.8 sublimes

Sources: Compiled in part from Don M. Yost and Horace Russell, Jr., *Systematic Inorganic Chemistry* (copyright, 1944, by Prentice-Hall, Inc., New York), pp. 299, 302, 307, by permission of the publisher; C. D. Hodgman, *Handbook of Chemistry and Physics*, 30th ed. revised, pp. 500, 502, 506 (Chemical Rubber Publishing Co., Cleveland, 1948).

sulfur dichloride are used to vulcanize rubber. The oxychlorides are useful reagents for the introduction of sulfoxide, sulfone or chlorine groups into organic molecules. Chlorosulfonic acid, HSO_3Cl, prepared by reacting HCl directly with sulfur trioxide, reacts readily with water to form sulfuric and hydrosulfuric acid. It fumes in air and has found application as a smoke producer in military operations. Its most useful application is in organic chemistry where it serves as a sulfonating agent.

TABLE III.—*Properties of Sulfur Oxyhalides*

Compounds		Molecular weight	Melting point ° C.	Boiling point ° C.
Name	Formula			
Thionyl fluoride	SOF_2	86.06	−129.5	−43.8 ± 0.1
Thionyl chlorofluoride	$SOClF$	102.5	−139.5	12.2 ± 0.1
Thionyl chloride	$SOCl_2$	118.97	−104.5	78.8
Thionyl bromochloride	$SOBrCl$	163.43		115 decomposes
Thionyl bromide	$SOBr_2$	207.89	−50	137
			−52 to −50	138 at 733 mm. pressure
				68 at 40 mm. pressure
Thionyl tetrafluoride	SOF_4	124.1	−107 ± 0.5	−48.5
Sulfuryl fluoride	SO_2F_2		−136.7	−55.4
		102.06	−120	−52
Sulfuryl chlorofluoride	SO_2ClF		−124.7	7.1
		118.5		7.6
Sulfuryl chloride	SO_2Cl_2		− 54.1	69.1
		134.97	− 54.1	69.1
Pyrosulfuryl chloride	$S_2O_5Cl_2$		− 37.5	153
		215.03	−39.10 − 37	140
Sulfuryl bromofluoride	SO_2BrF	163.0	−86 ± 0.5	40

Sources: Compiled in part from Don M. Yost and Horace Russell, Jr., *Systematic Inorganic Chemistry* (copyright, 1944, by Prentice-Hall, Inc., New York), pp. 299, 302, 307, by permission of the publisher; C. D. Hodgman, *Handbook of Chemistry and Physics*, 30th ed. revised, pp. 500, 502, 506 (Chemical Rubber Publishing Co., Cleveland, 1948); H. S. Booth and F. C. Mericola, in *J. Amer. Chem. Soc.* 62:640–642 (1940); Heinz Jonas in *Z. anorg. allg. Chem.*, 265:273 (1951); H. J. Emeleus and J. F. Wood, *Journal of Chemical Society*, pp. 2183–2188 (Burlington House, London, 1948).

Organic Chemistry of Sulfur.—The organic chemistry of sulfur may be divided into reactions which occur under reducing conditions and reactions which occur under oxidizing conditions. The first type is encountered when elemental sulfur is heated with petroleum hydrocarbons, rubber or other organic substances. The organic material is sulfurized and hydrogen sulfide is formed as a consequence of sulfur's action as a dehydrogenation agent. The hydrogen sulfide, being a reducing agent, may induce further reactions, particularly with unsaturated compounds, to yield mercaptans and thiophenols which in turn may be alkylated to sulfides; *e.g.*, mustard gas. All such sulfur compounds are analogous to similar compounds of oxygen, but differ from them because of dissimilarities in the respective molecular weights and electronic structures. The additional shell of electrons possessed by sulfur may lead to the formation of polysulfides such as the thiokol-type synthetic resins. The resistance of these resins to attack by gasoline has led to their use as linings for aircraft fuel tanks. The aromatic sulfur compounds, usually nitrogen derivatives, find extensive uses as accelerators in the vulcanization of rubber. The second type of reaction, oxidation of organic sulfur compounds, yields materials which, with the exception of the disulfides and the peroxides, are not analogous to oxygen compounds with respect to structure. The most important of these compounds are the sulfonic acids and their derivatives which are used extensively as medicinals (sulfa drugs), detergents, lubricants and dyestuff intermediates. Sulfones, sulfoxides and sulfinic acids are less useful but give promise of greater utility in the future.

III. COMMERCIAL FORMS AND USES

Crude Sulfur or Brimstone.—The form of crude sulfur on the United States market is 99.5% to 99.9% pure and free from arsenic, selenium and tellurium. As shipped, it contains about 50% fines with lumps 8 in. or more in diameter. Italian sulfur was marketed in four grades known as best seconds, best thirds, good thirds and current thirds and ranging in sulfur content from 99.5% down to 96.0%. Japanese sulfur was usually marketed in the form of cylindrical bars, each containing slightly more than one cu.ft. of 99.8% to 99% sulfur and weighing 130 lb. to 150 lb.

Broken-rock brimstone is sublimed or refined sulfur broken and sold as a mixture of lumps and fines. It has uses similar to roll sulfur.

Colloidal sulfur is sulfur in such small particles that it remains suspended in water. Colloidal sulfur may be prepared by intensive grinding but is usually prepared by chemical means. The uses are mainly pharmaceutical.

Commercial flour sulfur is produced by grinding crude sulfur. Its fineness is 50% to 99% through 200-mesh or 99% through 300-mesh. Uses: vulcanizing rubber, dyes, gunpowder, insecticides and fungicides.

Flowers of sulfur is crude sulfur refined by sublimation. It contains from 25% to 30% "amorphous" sulfur, and 90% of it will pass through a 100-mesh sieve. Uses: chemicals, vulcanizing rubber, insecticides, fungicides and pharmaceuticals (especially after washing with ammonia to remove all traces of acid).

Lac sulfur, precipitated from polysulfide solutions by sulfuric acid, contains up to 45% calcium sulfate. The uses are mainly pharmaceutical.

Precipitated sulfur is precipitated from polysulfide solution by hydrochloric acid and washed to remove all calcium chloride. The uses are mainly pharmaceutical.

Roll sulfur is refined sulfur cast into convenient sizes, also known as stick or cannon sulfur. Uses: chemical, burning for curing, fumigating and preserving or bleaching effects.

Uses.—The uses of sulfur are many and varied. Often it is considered, along with salt, coal and limestone, as being one of the four basic raw materials of the chemical industry. The major portion of it is burned to form sulfur dioxide for use in producing wood pulp, for bleaching and for conversion to sulfuric acid. But sulfur, because of its interesting properties, is also applied to many other uses. Combined with an aggregate, it is used to make acid-resistant cements. It is used to impregnate paper and other porous materials to produce useful articles of commerce. Sulfur ointments have been used in treating skin affections and sulfur compounds are used to treat a variety of diseases. It is used as an insecticide or fungicide in the control of insects and fungus diseases of plants and animals. A further use, not commonly recognized, is as a plant food. Sulfur and its compounds are used both as direct and indirect fertilizers. As indirect fertilizers, sulfur and its compounds have various functions, including the correction of alkali soils, reaction with soil constituents and release of the nutrient elements, soil amelioration resulting from the flocculating effect and the control of the reaction of the soil.

IV. ANALYTICAL METHODS

Free Sulfur is determined by extraction with carbon disulfide. Amorphous sulfur, present in sublimed sulfur, is not soluble in carbon disulfide but is rendered soluble by heating samples at 100° C. to 105° C. for 18 hr.

Moisture in Commercial Sulfur is determined by drying a 50 g. sample at 80° C. for 16 hr. in a wide-mouth weighing bottle.

Ash in Commercial Sulfur is determined by burning 10 g. to 50 g. of the sulfur in a porcelain crucible using sufficient auxiliary heat to keep the sulfur burning and finally igniting the crucible to destroy all organic matter.

Arsenic, Selenium and Tellurium in Sulfur are determined by first oxidizing the sulfur to sulfuric acid through the use of bromine and nitric acid. Arsenic is then determined by the Gutzeit method, selenium by reduction with hydrazine sulfate and tellurium by reduction with stannous-chloride solution.

Combined Sulfur may be determined by converting it to a soluble sulfate and precipitating it as barium sulfate by the addition of barium chloride.

BIBLIOGRAPHY.—F. W. Clarke, *The Data of Geochemistry*, Bull. 770, 5th ed., U.S. Geological Survey, Dept. of the Interior (1924); Williams Haynes, *The Stone That Burns* (New York, 1942); Don M. Yost and Horace Russell, Jr., *Systematic Inorganic Chemistry of the Fifth-and-Sixth-Group Nonmetallic Elements* (New York, 1944); Freeport Sulphur Company, *Sulphur Data Book* (New York, 1954); Fritz Ephraim, *Inorganic Chemistry*, ed. by P. C. L. Thorne and E. R. Roberts, 6th rev. ed. (Edinburgh, 1954; New York, 1955). (W. W. D.)

SULFURIC ACID, H_2SO_4, is one of the most widely used of all commercial chemical substances. It is made from sulfur, or compounds of sulfur, water and oxygen from the air. Although rarely used or even seen by the layman, sulfuric acid is employed in some stage of the manufacture of practically all industrial products. The annual consumption of the product in the United States is approximately 15,000,000 tons annually. Because of its extensive use, the quantity consumed from year to year serves as an accurate and sensitive index of general business conditions.

After a discussion of the widespread use of the acid, this article is divided into sections on properties; sulfates; grades and terminology; materials of construction of containers; manufacture; concentration; and first aid.

Sulfuric acid occupies its unique position of usefulness because it can be employed for an unusually wide range of purposes and because it is one of the cheapest of the industrial chemicals to produce. Its great affinity for water, *i.e.*, its great tendency to absorb water, makes it a powerful dehydrating agent. It is a valuable electrolyte, a strong oxidizing agent, and a solvent in which many important reactions can be brought about. It is also an effective catalyst for various chemical reactions. It is used in varying concentrations from the extremely dilute solutions for control of acidity or alkalinity to the highly concentrated fuming acids prepared for the manufacture of drugs, dyes and explosives. It is made in various grades from those of high purity for the rayon and pharmaceutical industries to those of relatively low grade for the manufacture of superphosphate and ammonium sulfate for the fertilizer industries. The same batch of acid is often used for several purposes in succession. After an initial use in the petroleum, explosives, drug or dye industries, for instance, it is frequently recovered in a solution too dilute or impure for further use in the same process but still usable for the pickling of steel or manufacture of a fertilizer. Sulfuric acid is made by two very different processes: the contact process and the lead-chamber process. (See *Manufacture*, below.) For use in either process a large variety of sulfur-containing raw materials is available.

Properties.—Pure (100%) sulfuric acid is a colourless liquid which is sometimes called monohydrate of sulfur trioxide because its anhydride, sulfur trioxide (SO_3), and water are combined in equimolal quantities ($SO_3+H_2O\rightarrow H_2SO_4$). Its specific gravity is 1.830 at 25° C. (77° F.), freezing point 10.37° C. (50.7° F.) and at one atmosphere it boils at about 340° C. (644° F.) but undergoes decomposition into sulfur trioxide and water. The former escapes as a vapour until the concentration has fallen to 98.54%, which is the composition of the constant-boiling mixture of sulfuric acid and water at atmospheric pressure. The solution then distills at 317° C. (603° F.), the boiling point of the mixture. The freezing point of the acid is lowered markedly by the addition of either sulfur trioxide or water.

The existence of the following crystalline phases has been established by the study of freezing temperatures and other properties: SO_3, $(SO_3)_2.H_2O$; *i.e.*, $H_2S_2O_7$, H_2SO_4, $H_2SO_4.H_2O$, $H_2SO_4.2H_2O$, $H_2SO_4.4H_2O$, $H_2SO_4.6.5H_2O$ and $H_2SO_4.8H_2O$. The hydrate containing 6.5 molecules of water per molecule of anhydrous acid (13 molecules of water for 2 molecules of acid) was believed, prior to publication of careful measurements by J. E. Kunzler and W. F. Giauque, to be a hexahydrate. It is possible (though not likely) that the octahydrate may not have exactly the composition ascribed to it. It is known, however, that the ratio of H_2O molecules to H_2SO_4 molecules is less than nine.

Sulfuric acid is dibasic; that is, its molecules each contain two hydrogen atoms and ionize in two steps: $H_2SO_4=H^++HSO_4^-$ and $HSO_4^-=H^++SO_4^=$. The second stage of the dissociation is moderately weak, the constant being 0.010 at 25° C. Solutions of various concentrations of the acid therefore may contain in addition to hydrogen ions (H^+ or H_3O^+), molecules of the acid (H_2SO_4), bisulfate ions (HSO_4^-) and sulfate ions ($SO_4^=$) in proportions which vary with the dilution. In very dilute solutions the concentration of the $SO_4^=$ ion is equal to the stoichiometric concentration, that is, there is an $SO_4^=$ ion for practically every molecule of H_2SO_4 which was originally added to the water. As the concentration of the solution is increased, the concentration of the HSO_4^- rises rapidly; that of the $SO_4^=$ increases more slowly and reaches a maximum value of two molar at a stoichiometric concentration of about seven molar. It then decreases to a very small value, virtually zero, in a solution of high concentration (14 molar). At this same high concentration (14 molar) the HSO_4^- is at a maximum value of 14 molar, that is, practically all of the acid has undergone primary ionization but no secondary ionization. At still higher molarities the HSO_4^- ion concentration diminishes as the H_2SO_4 increases.

As a result of these variations in the concentrations of the various molecular species, the properties of aqueous sulfuric acid solutions undergo interesting changes; *i.e.*, the specific conductance reaches a maximum in the 4-molar acid, the specific conductance-viscosity product attains a maximum in the 9-molar solution and the surface tension reaches a maximum when the stoichiometric concentration is about 6.5 molar. At low temperatures (0° C.) there is a minimum in the surface tension-concentration curve at about one molar concentration. Other properties such as density, heat of dilution and heat capacity are also affected by the changes in the concentration of the various components which result from the dissociation of the H_2SO_4 molecules.

Dilute sulfuric acid reacts with the more electropositive metals such as zinc, magnesium and iron with the evolution of hydrogen. It also reacts with metal oxides and hydroxides forming sulfates. Because of its high boiling point the concentrated acid is used to manufacture volatile acids, some of them stronger than sulfuric acid, from their salts. It liberates hydrogen chloride (hydrochloric acid in water solution) from sodium chloride, nitric acid from nitrates, and hydrofluoric acid from calcium fluoride. It is also used to convert calcium phosphate to a more soluble acid phosphate for fertilizer and in solution displaces numerous weak acids; *e.g.*, acetic and boric acids from their respective salts. Since bromide ion and iodide ion are oxidized by the concentrated acid, it cannot be used for the manufacture of hydrobromic and hydriodic acids.

In sulfuric acid the sulfur is at its maximum oxidation state of +6; it may therefore function as an oxidizing agent and be reduced to the $+4(SO_2)$, zero (S) or $-2(S^=)$ state. The dilute acid is not an effective oxidizer but the hot concentrated acid oxidizes less active metals, *e.g.*, copper and silver, as well as the nonmetals carbon and sulfur, and is itself reduced to sulfur dioxide. With stronger reducing agents as zinc or iodide it is reduced to hydrogen sulfide, which under favourable conditions may escape before being oxidized by more of the sulfuric acid.

It removes by oxidation the tars and organic sulfides present in petroleum and serves as a catalyst in the sulfuric acid alkylation processes involved in the manufacture of high-octane aviation gasoline.

The concentrated acid has great avidity for water and serves as an excellent drying agent and as a dehydrating agent, removing not only moisture from gases and liquids which do not react with the acid, but extracting water from compounds. Thus it chars many organic compounds such as wood, paper, cloth, starch and sugars. For the same reason the acid becomes brown in colour if it comes in contact with traces of oil, grease or atmospheric dust. It is essential in the manufacture of explosives, celluloid, lacquers, etc., for the nitration processes involved in these processes liberate water which is removed by the sulfuric acid, thereby permitting the reactions to go to completion. Its value to organic chemistry

is due also to its reaction with many aromatic compounds to form sulfonic acids. (*See* DYES; NAPHTHALENE; PHENOLS; SULFONIC ACIDS.)

Sulfates.—Since sulfuric acid is dibasic, it forms both normal or neutral salts (Na_2SO_4) and acid salts ($NaHSO_4$). The latter show acid properties in solution because of the ionization of the HSO_4^- ion and are sometimes employed for purposes requiring moderate acidity. The common sulfates are soluble in water with the exception of those of lead, strontium, barium and calcium. Many of the sulfates have important industrial uses: barium and lead sulfates in the manufacture of paints; sodium sulfate in the manufacture of paper, and its hydrate, $Na_2SO_4.10H_2O$ (Glauber's salt), in medicine; copper sulfate, $CuSO_4.5H_2O$, in insecticides and its solution as an electrolyte in the electrolytic refining of copper; gypsum, $CaSO_4.2H_2O$, and plaster of paris, $(CaSO_4)_2.H_2O$, in the building industries. The sulfates of potassium and aluminum crystallize from solution as a double salt, $K_2SO_4.Al_2(SO_4)_3.24H_2O$, which is called alum (*q.v.*). The term is also applicable to other double sulfates of similar crystalline form in which the potassium ions are replaced by other univalent ions (*e.g.*, those of sodium or ammonium) and the aluminum ions by other trivalent ions (*e.g.*, those of chromium or iron).

Grades and Terminology.—Since there is a linear relationship between the concentration and density of sulfuric acid in the range from 0% to 93.19%, the concentration may be determined fairly accurately with the aid of a hydrometer. The Baumé (Bé.) hydrometer, American modulus, is used by the industry in the U.S., while in continental Europe it is employed with a different modulus. The Twaddell hydrometer is used in the United Kingdom although there is an increasing tendency to express the concentration directly in per cent of H_2SO_4 or of SO_3 rather than in the degrees of the hydrometer scale. Temperature corrections must be made for all hydrometer readings.

The 100% H_2SO_4 is rarely shipped because of its high freezing point (*see* above); the 98%–99% acid which is produced in the contact process is frequently used and because of its somewhat lower freezing point may be shipped if care is exercised.

The 93.19% acid is a sirupy, oily liquid whose freezing point is −34° C. (−29° F.) and boiling point 281° C. (538° F.); at 15.6° C. (60° F.) its specific gravity is 1.835 and its viscosity 24 times that of water at the same temperature. It is one of the most popular shipping concentrations because of its low freezing point and the fact that it does not substantially react on and corrode the economical steel containers at atmospheric temperatures. The name oil of vitriol (o.v.) is used in the U.S. to designate this concentration of the acid because of its appearance and "biting" characteristics. In the United Kingdom and Europe, however, the term is used more generally to mean simply sulfuric acid. Thus, in the British Isles the following designations are used: brown oil of vitriol (b.o.v.), ordinary commercial 77%–80% acid; best brown oil of vitriol (b.b.o.v.), the commercial 77%–80% acid relatively free from arsenic and other impurities; concentrated or rectified oil of vitriol (c.o.v. or r.o.v.), 93%–96% acid now usually made by the contact process rather than by concentration or rectification of the more dilute chamber acid. The term chamber acid may refer either to any acid produced by the lead-chamber process or to the acid of the particular concentration 62%–68% otherwise produced. This more dilute acid reacts with steel containers and is therefore rarely shipped in quantity. The term 50° Bé. acid (62.18%) is frequently used for statistical purposes in the U.S. Acids of still lower concentrations are sometimes prepared by dilution but must be shipped in the more expensive glass carboys or bottles.

Another popular shipping strength is the approximately 78% acid because it is the maximum strength made by the chamber process and because its freezing point, −10.8° C. (12.6° F.), is low. In cold climates the shipping concentration is often changed to 76%, which acid has a still lower freezing point. The 76%–78% acid is not seriously corrosive to steel tank cars or tank trucks. The fertilizer industry, one of the largest consumers of sulfuric acid, requires acid of only 68% concentration for the manufacture of superphosphate.

Oleum or fuming sulfuric acid (also known as Nordhausen acid) is a solution of sulfur trioxide, SO_3, in 100% sulfuric acid. Its concentration is usually expressed as "per cent oleum," *i.e.*, the percentage of free SO_3 dissolved in the acid, and occasionally as the total percentage of SO_3 which includes both the free oxide and that combined with water in the H_2SO_4. Thus 100 lb. of 40% oleum contains 40 lb. of SO_3 and 60 lb. of 100% H_2SO_4, or a total of 89 lb. (89%) of SO_3 free and combined. Addition of 9 lb. of water to the 100 lb. of 40% oleum yields 109 lb. of 100% H_2SO_4. For this reason oleum has an equivalent acid concentration greater than 100% H_2SO_4, that of the 40% oleum being 109% acid. As sulfur trioxide dissolves in sulfuric acid in all proportions, the resulting oleums may have a wide range of concentrations. When the two combine in equimolal proportions, pyrosulfuric acid, $H_2S_2O_7$, is formed. This acid has the maximum melting point, 36° C. (96.8° F.), of all the compounds of the H_2O–SO_3 system.

Sulfuric acid may also be designated by terms which refer to its purity, origin or use rather than to its concentration. Thus battery acid and electrolyte refer to acid of a quality suitable for accumulators or storage batteries; in the U.S. the concentrations sold are either 50% H_2SO_4 (specific gravity 1.40) or 93.19% (specific gravity 1.835). Brimstone or sulfur acid refers usually to acid produced from sulfur rather than from less pure raw materials such as pyrites, spent oxide, etc.

Mixed acid consists of nitric acid and concentrated sulfuric acid. Spent acid is that which has been used once and is usually dilute and impure. Sludge acid is the residue of the acid used in the refining of petroleum. It is contaminated with organic impurities and is frequently made usable by further dilution to separate as many as possible of the impurities followed by concentration; or it may be decomposed by heat, the organic material oxidized and the sulfur dioxide recovered for use in the manufacture of more acid. Spent alkylation acid is a by-product of sulfuric acid alkylation processes for the manufacture of high-octane aviation gasoline. It is of a higher purity than sludge acid, usually containing 85%–90% H_2SO_4, and is frequently reused in the petroleum refinery.

Like other chemicals, sulfuric acid is graded by standards stipulated by the appropriate scientific societies or industries. The technical grade is usually intended for large-scale industrial use and need not be of high purity. The U.S.P. label designates acid which meets the specifications of the *United States Pharmacopeia*, while C.P., chemically pure, refers to acid of a higher grade usually prepared by distillation of a commercial grade acid. The purest grades, suitable for analytical and research purposes, are designated as reagent, analyzed, etc.

Materials of Construction of Containers.—Of the common metals, only lead is satisfactorily resistant to cold sulfuric acid of nearly all concentrations up to 100% and to hot acid of concentrations up to about 70%. Generally the metals are less resistant to the hot and the more concentrated acids, but steel and cast iron are exceptional in that they are resistant to cold acid of concentrations above approximately 78% because of the formation of a protective scale and are attacked with increasing vigour as the concentration decreases. Cast iron is more resistant than mild steel to acid of most concentrations and has a satisfactory life even with very hot acid, provided its concentration is above 90%. However, cast iron should never be used as a container for oleum or for sulfur trioxide gases for it may crack with explosive violence when in contact with sulfur trioxide. Mild steel should not be used at temperatures in excess of approximately 49° C. (120° F.) regardless of the concentration of the acid but it has a satisfactory life for acid of concentrations above 20% oleum even though hot. Cast steel, malleable iron and many resistant alloys are available for specific concentrations, temperatures and applications; in some instances acid-resistant brick linings are used, *e.g.*, for hot oleum below 20%.

The U.S. Interstate Commerce commission issued regulations in 1946 permitting the use of rubber-lined metal containers for acid of not over 42° Bé. (about 50%). Lead-lined containers may be used for acid up to and including 93.2%. Steel tank cars and tank trucks may be used for acid of not less than 52° Bé. (about

65%), but steel drums or barrels which are made of thinner metal may not be used for acid below 60° Bé. (about 78%) and even then only if the acid has no greater corrosive effect than 93% acid at 37.7° C. (100° F.).

Moisture should be carefully excluded from iron or steel containers as the dilution of the acid at any point causes rapid corrosion. Sparks or open flames must be kept away from any closed vessel, which should not be opened unless it is previously washed and the residual acid neutralized. The vapour space over the acid may contain explosive hydrogen and poisonous arsenic, nitrogen or sulfur compounds.

Manufacture.—The date of the discovery of sulfuric acid is lost in antiquity; it was probably first prepared by Arabian chemists in the 8th century and was apparently well known to the alchemists. Near the end of the 15th century Basil Valentine described a method of preparation in which the acid was distilled from a mixture of green vitriol ($FeSO_4.7H_2O$) and silica and another method whereby the acid was obtained by burning sulfur with saltpetre ($NaNO_3$). The older of the two commercial methods, the lead-chamber process, was introduced in England in 1746 although the principles of the process had been used much earlier. It was first manufactured on a commercial scale from weathered pyritic shales in Bohemia and at Nordhausen in Germany.

The first step in the manufacture of sulfuric acid is to prepare sulfur dioxide (SO_2) gas by burning sulfur or sulfur-bearing materials such as iron pyrites, copper and zinc ores, spent oxide from gas plants and coke ovens, hydrogen sulfide from natural gas and petroleum refinery gas. It is also obtained by decomposition of various sulfates: gypsum ($CaSO_4.2H_2O$), anhydrite ($CaSO_4$), ferrous sulfate ($FeSO_4$) from waste pickling liquors, sludge acid and spent acid. The raw material used by a particular plant is determined by a number of considerations of great complexity which may include the location, the kind of process employed, the grade and concentration of acid to be produced and the cost. In the U.S. about 75% of the reported production of new acid at mid-20th century came from sulfur, 15% from pyrites, pyrrhotite and coal brasses, 9% from copper and zinc smelters and 1% from hydrogen sulfide. In the U.K. approximately 40% of the production came from sulfur, 28% from pyrites, 20% from spent oxide, 12% from zinc blende and minor quantities from anhydrite and hydrogen sulfide.

The next step in the production of the acid is to oxidize the SO_2 to SO_3. This is done by two different processes both employing catalysts. The lead-chamber process yields acid of a concentration of not more than 78%, and unless a pure raw material, *e.g.*, sulfur, is used, the product is impure. The newer contact process was developed on a commercial scale during the last decade of the 19th century and is used to produce acid of any desired concentration and of high purity regardless of the quality of the raw materials.

Because of the increasing demand for acid of high concentration and purity, the latter process has steadily gained and by mid-20th century about 80% of the acid produced in the U.S. and more than one-half of that produced in the U.K. were manufactured by the contact process. The two methods, however, remained competitive under certain circumstances.

Contact Process.—Contact plants are of two main types. The newer and simpler type, often known as a raw gas or hot gas purification unit, was developed especially for use with vanadium pentoxide (V_2O_5) or other catalysts immune to impurities in the gas stream, and with sulfur as the raw material. The other kind, known as the cold gas purification or simply the purification type, is used when the SO_2 gas contains dust and other impurities from the ores which constitute the raw material.

Sulfur-Burning Contact Plants.—In these plants the sulfur is melted to remove moisture and to permit the major portion of the dust or ash which may be present to settle. The molten sulfur with or without subsequent filtration is pumped to the burner. Combustion air (*i.e.*, air which has been dried by concentrated sulfuric acid) is supplied to the burner by a blower operating under a pressure between 1.5 and 3.5 lb. per square inch greater than that of the atmosphere, thus preventing the infiltration of moist air into the plant. The SO_2-air mixture containing 8%–11% of SO_2 leaves the sulfur burner at temperatures which range from about 760° to 1,200° C. (1,400° to 2,200° F.), the temperature being proportional to the percentage of SO_2 in the gases and to the amount of heat in the combustion air. Except in very small plants the gas is usually cooled by means of a waste-heat boiler to approximately 400° C. (750° F.), and in some instances is passed through a hot-gas filter to remove any remaining dust or ash before it comes into contact with the catalyst.

The catalyst is placed in layers on shelves or packed in vertical tubes in an insulated steel tank or tower called a converter. The reaction, $2SO_2 + O_2 \rightleftarrows 2SO_3 + 45$ kilocalories, takes place at an almost imperceptible rate without a catalyst. However, the rate is greatly increased by the surface catalysts, vanadium pentoxide or finely divided platinum and further accelerated at elevated temperatures. Practically all substances have some degree of catalytic action at high temperatures but the two mentioned are the most effective and, because of its immunity to impurities, the use of vanadium pentoxide after 1925 gradually replaced that of platinum in nearly all plants. The fraction of the SO_2 converted to SO_3 decreases at excessively high temperatures. The temperatures at which the catalysts function effectively to give quickly the maximum yield of SO_3 are above about 400° C. (750° F.) for vanadium pentoxide and 360° C. (680° F.) for platinum. To prevent excessively high temperatures from the exothermic reaction, which would diminish the yield of SO_3, the conversion is carried out in stages with removal of part of the heat of reaction by waste-heat boilers or other means between the stages. The gas leaves the final layer of catalyst usually at temperatures between 438° C. (820° F.) and 460° C. (860° F.). In the development of the many types of converter with various arrangements and amounts of catalyst a high efficiency has been achieved whereby 95% to more than 98% of the SO_2 is converted to SO_3.

The SO_3 gas from the converter is cooled and passed into the absorption system where it is absorbed in 98%–99% sulfuric acid. It combines with the free water in the acid solution to form more H_2SO_4, thus increasing the concentration to any desired value up to 100% or to higher than 40% oleum. The SO_3 is more readily absorbed by concentrated sulfuric acid than by water, for with the latter fog particles consisting of minute drops of H_2SO_4 are formed and can be recovered only with great difficulty and expense. Lower concentrations of the acid may be prepared by dilution of the 99%–98% acid with water. The contact process may also be used to fortify spent acid or the more dilute acid produced in a chamber plant.

Sulfur-burning contact plants with capacities in single units up to 600 tons per 24 hours and requiring only one operator have been built. Considerably more than a ton of high pressure, saturated or superheated steam may be generated for every ton of 100% acid produced. The required electric power is usually less than 35 kw.hr. per ton of 100% acid and the cooling water about 4,000 U.S. gallons per ton of acid. No buildings are needed for the housing of equipment.

Purification Type Contact Plants.—For the production of SO_2 from low grade sulfur-bearing materials a large variety of furnaces or roasters is available. It is desirable to select a roaster which will remove the maximum quantity of sulfur from the raw materials with a minimum carry-over of dust into the effluent gas and which will produce gas with a maximum SO_2 concentration at minimum operating and repair costs. It is also desirable to avoid the oxidation of SO_2 to SO_3 in the burner and the consequent formation of the acid mist. The concentration and purity of the SO_2 gas from the furnace determines the size, quantity and cost of subsequent gas purification equipment.

The gas usually leaves the roaster at temperatures between 650° C. (1,200° F.) and 1,100° C. (about 2,000° F.) and may be cooled by cast-iron pipe or by waste-heat boilers. It is important to remove the dust at this stage to prevent its deposit in subsequent equipment. This is done by various methods: about 80%–90% may be removed in cyclone centrifugal collectors and 90%–99% of the remaining dust in electrical precipitators which operate at

50,000 to 70,000 v. and at 150°–430° C. (about 300°–800° F.). Precipitators have the advantage of low pressure drop, low power consumption and of recovery of dust in dry form but at the disadvantage of relatively high cost. This equipment may be built of steel or of steel lined with firebrick, as the gas is normally so hot that no corrosive acid can condense. Some plants have replaced precipitators with proprietary gas washers which have as high or higher efficiencies in removal of dust and have the additional advantages of removing a large portion of other impurities such as metallic fume, arsenic and acid mist while simultaneously cooling the gas. They have the disadvantage of relatively high pressure drop, as do cyclone dust collectors. They necessitate recovery of dust in the wet condition but in many cases this is considered an advantage. Smaller and older plants occasionally use dust-settling chambers or scrubbing towers of various types instead of precipitators or gas washers.

After removal of the dust the gas must be cooled to permit removal of other impurities such as metallic fume, acid mist and fluorine, chlorine or arsenic compounds. They must be removed to the point where they do not clog or damage subsequent equipment and do not adversely affect the quality of the product acid for the purpose for which it is to be used. Purification is necessary regardless of the type of catalyst used but the purification equipment need not be quite so carefully designed and operated if a catalyst like vanadium pentoxide which is immune to all common impurities except fluorine compounds is used. Cooling of the gas is necessary not only to remove impurities but also to condense and thus remove part of the water vapour in the gases. If all of the water vapour were allowed to remain in the gas, it would cause the product acid in many cases to be more dilute than desired.

After cooling and removal of impurities, the SO_2 gas is dried with concentrated sulfuric acid. Equipment after the drying tower is much the same as in sulfur-burning plants with the single exception that waste-heat boilers cannot be employed to recover the heat evolved in the converter. As a result of its purification, the gas in this process is cold, instead of hot as in sulfur-burning plants, and must be heated to the temperature at which the catalyst begins to function. The heat evolved in the converter is utilized for this heating by passing the hot SO_3 gas and the cold SO_2 gas through a series of insulated tubular heat exchangers built of steel. The SO_2 gas is thus heated and the SO_3 gas cooled.

Contact plants of this type have been built with a single purification line capable of producing 300 tons of 100% acid per 24 hours and serving two 300-ton conversion and absorption lines. The labour, cooling water, power and maintenance requirements are much higher than in sulfur-burning plants because of the additional roasting and purification equipment. However, in many cases the SO_2 gas is obtained at little or no cost from copper and zinc smelters or from other low-cost raw materials and the cost of the product acid may be substantially less than that from sulfur-burning plants.

The Lead-Chamber Process.—This process derives its name from the use of large chambers of sheet lead and differs radically from the contact process, particularly in its use of gaseous oxides of nitrogen whose function may be considered to be catalytic. The chemical reactions which occur in the lead-chamber process are more complex than those in the contact process, and there is not unanimity of opinion as to their mechanisms. It is thought that the catalyst reacts with SO_2 gas, oxygen and water vapour to form the intermediate compound, nitrososulfuric acid ($ONHSO_4$), which in turn reacts with water to form sulfuric acid and the oxides of nitrogen (NO and NO_2). Thus the catalyst goes through repeated cycles, combining with the SO_2, O_2 and H_2O to form the intermediate compound and being set free again with the simultaneous formation of H_2SO_4.

The SO_2 gas is generally produced in the same manner as in the contact process. However, if it is impure, *i.e.*, obtained by the roasting of ores or by the burning of spent oxide, the purification is carried out only to the extent where the dust and other impurities are not present in quantities sufficient to clog the equipment. This less efficient purification obviates the necessity of cooling the gas which is needed in the next stage of the process at a high temperature, preferably in excess of approximately 600° C. (1,100° F.), and is necessarily at the expense of the quality of the product acid.

After the hot sulfur dioxide-air mixture has undergone the necessary purification, it is passed in succession upward through the Glover tower, through the lead chambers and upward through the Gay-Lussac towers, nearly all of the SO_2 being removed in the first two steps. The Glover tower is a large masonry structure in some instances sheathed outside with lead. It is packed with lumps of loose quartz, acid-resisting brick or ceramic shapes so that the hot gases are brought into intimate contact with acid which trickles downward through the packing.

The Glover tower has two main functions: (1) to concentrate the approximately 65% acid made in the chambers to 78% acid, the water vapour thus released entering the chambers with the gas stream; (2) to remove the nitrogen oxides from the 78% acid (nitrous vitriol) which has been used in the Gay-Lussac tower to absorb these oxides released at the end of the process and to return them to the cycle. The denitrification is accomplished by dilution of the 78% nitrous vitriol with the less concentrated chamber acid and by the elevation of the temperature of the mixture by the entering hot gases. The gases also begin to react and a substantial amount of sulfuric acid is formed in the Glover tower. The 78% acid drawn off at this point is hot and contains only traces of the oxides of nitrogen. After cooling, part of it is transferred to the Gay-Lussac towers where it is used to absorb the oxides of nitrogen released from the chambers. The balance of it represents the production of the plant and is delivered to steel storage tanks.

The cooled gases leaving the Glover tower proceed to the rectangular boxlike lead chambers which may vary in size from 5,000 to 500,000 cu.ft. and in number from 1 to 20. The large chambers afford sufficient space and time for the gases which do not mix rapidly to react completely. The additional water necessary for the reactions is supplied through atomizing nozzles in the tops of the chambers. The sulfuric acid forms as a rain, drops to the bottom of the chambers and is pumped to the Glover tower. A large amount of heat is evolved by the reactions and is dissipated by the natural circulation of air around the chambers which are set well above the ground. The chambers are usually housed in a building to lessen the effect of varying atmospheric conditions on the operation. As the gases pass through the Glover tower and the chambers, their SO_2 content gradually decreases to 0.1% or less, this amount being desirable for the optimum recovery of the nitrogen oxides in the Gay-Lussac towers.

The Gay-Lussac towers, usually two arranged in series, are similar in construction to the Glover tower but are taller and of smaller diameter. They have thinner walls and linings because they receive relatively cool gases. Their only function is to recover the nitrogen oxides in the spent gases and for this purpose cool 78% acid is used. The recovery is incomplete, and usually 10%–15% of the oxides is lost, constituting an appreciable item of expense. Hence, fresh oxides of nitrogen must be constantly supplied to the process. In some plants the fresh nitrogen oxides are made from sodium nitrate and 78% sulfuric acid which is heated in an iron pot either by the hot SO_2 burner gases or by fuel. This method is obsolete at most plants and "nitre pots" have been superseded by ammonia oxidation units. Such units produce oxides of nitrogen by catalytic action of an incandescent platinum alloy gauze on a mixture of air and ammonia (about 9%). The oxides of nitrogen are then introduced into the process usually at the Glover tower inlet or outlet. The gases may be propelled through the chamber plant by the natural draft created by the final Gay-Lussac tower and its exit stack but usually one or more lead or alloy fans are employed. Such a fan may be located between the Glover tower and the first chamber, after the last chamber or at both points.

Because of the considerable lead and space requirements of the chambers, the lack of efficient cooling and the necessity for a large building, many variations of the process are in existence in which the boxlike chambers have been superseded by towers or by chambers of other shapes designed to promote more intensive

gas mixing. Some are water cooled externally while some depend on cooling by the internal circulation of large quantities of acid. The principal successor in the U.S. is the Mills-Packard water-cooled chamber constructed in the shape of the frustum of a cone and first developed in England in 1914. By this means the chamber space required was reduced from 6 to 9 cu.ft. per pound of sulfur burned, in the case of air-cooled box chambers, to 2.0–3.5 cu.ft. No building is required for the Mills-Packard chambers. In the U.K. the Gaillard-Parrish chamber replaced the box chamber in some instances. It is relatively tall but cylindrical in shape and is cooled by a spray of 65% acid from a horizontal rotating disk (disperser) through the closed top. The Gaillard disperser has also been used in the tops of box chambers to increase their acid-making capacity. Various processes have also been evolved in which a series of packed towers take the place of the box chambers and in which the oxidation of SO_2 takes place largely in the liquid phase. Among these may be mentioned the Opl, the Kachkaroff-Guareschi and the Petersen processes, much used in Europe but not in America.

Concentration.—Concentration of sulfuric acid by evaporation of the water is practised intensively only at explosives and dye works or in petroleum refineries, where a dilute acid appears as a by-product which must be recovered and concentrated for reuse. A variety of types of equipment is used for this purpose, among which are the Cascade, Kessler, Gaillard, Pot, Simonson-Mantius and Chemico concentrators. They may be divided into three classes: (1) those like the Cascade process in which the acid is heated to its boiling point in acid-resisting basins by heat applied to the exterior of the vessel; (2) those like the Chemico drum concentrator in which the acid is brought into direct contact with a mixture of hot air and gases from the combustion of a fuel; and (3) those like the Simonson-Mantius concentrator in which the acid in a closed vessel under high vacuum is heated indirectly by steam.

Concentrators in the first two classes operate at substantially atmospheric pressure and consequently the acid temperatures are high, especially where hot air in direct contact with the acid is used. The exit gases may contain appreciable quantities of acid mist, which is recovered in coke-packed towers and electrical precipitators. Concentrators of the third class operate at relatively low temperatures because of the high vacuum and create so little nuisance that precipitators are not required.

First Aid.—The best first-aid measure for a sulfuric acid burn is "a lot of water quickly." Speed in removing the acid is of utmost importance, but small amounts of water must not be used because much heat is generated by dilution of the acid and the burn is thus made more severe. Neutralizing agents and ointments are dangerous and should never be used until the burned area has been washed completely free of acid with water. After this has been done the burn may be treated in the same manner as a heat burn.

BIBLIOGRAPHY.—T. J. Kreps, *The Economics of the Sulfuric Acid Industry* (1938); A. M. Fairlie, *Sulfuric Acid Manufacture* (1936); Don M. Yost and Horace Russell, Jr., *Systematic Inorganic Chemistry of the Fifth-and-Sixth-Group Nonmetallic Elements* (1944); H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions*, 2nd ed. (1950); R. A. Robinson and R. H. Stokes, *Electrolyte Solutions* (1955); *American Institute of Physics Handbook* (1957).
(C. M. D.; T. F. Y.)

SULGRAVE MANOR, the early English home of the ancestors of George Washington. Situated 2½ mi. N.W. of Helmdon in Northamptonshire, it is a notable shrine. Laurence Washington, twice mayor of Northampton, bought the property from King Henry VIII in 1539, on the dissolution of the smaller religious houses. It had previously belonged to the priory of St. Andrew, Northampton. Laurence Washington, from whom George Washington was seventh in direct descent, and his children and grandchildren lived in the manor house until 1610.

Sulgrave Manor was purchased in Jan. 1914 for £8,400 by the British committee for the celebration of 100 years of peace between Great Britain and the United States (1814–1914), and is vested in three ex-officio trustees: the American ambassador in London, the British ambassador in Washington and the regent of the Mount Vernon Ladies' Association of the Union. It is partly restored to its original condition, and furnished with choice furniture of the period. A formal reopening took place, on June 21,

BY COURTESY OF THE NATIONAL BUILDINGS RECORD

SULGRAVE MANOR, NORTHAMPTONSHIRE

1921. In 1927 the restoration was completed by the rebuilding of the western wing, which had been demolished in the 18th century. The funds were collected by the American Society of Colonial Dames, which body had, two years previously, raised a permanent endowment of over £20,000.

On the gable of the central porch are the royal arms of the Tudors; and in the right spandrel of the arch of the main doorway, the Washington arms, three mullets and two bars. The furnishings are gifts from British and American donors, and include an original Gilbert Stuart portrait of George Washington and many other treasures. (L. OF F.)

SULIDAE: *see* GANNET.

SULIMAN (properly SULAIMAN) **RANGE,** a range of mountains in West Pakistan, stretching for about 200 mi. from the Gomal pass (which gives access from northern Baluchistan to Dera Ismail Khan and the upper Indus) to the Marri and Bugti tribal country north of Jacobabad and the lower Indus. The highest point of the range is the twin peak of Takht-i-Sulaiman or Solomon's throne (the highest one being 11,290 ft.) at the northern end, which legend connects with Solomon's visit to Hindustan. It has now a *ziarat* or shrine visited by both Hindu and Mohammedan pilgrims. On both sides of the Takht-i-Sulaiman dwell the Shirani Pathans, those on the eastern slope being known as the Largha, those on western as the Bargha. Between the Shirani country and the Bugti hills the range passes through the territories of the Zmarais, Isots, Usteranas, Lunis, Bozdars, Khetrans and Agharis. (C. C. D.)

SULINA, the second Black sea port of Rumania, in the region of Constanţa. Pop. (1960 est.) 3,713. Grain is brought down the Danube for ocean-going steamers. The port was formerly under the European Danube commission. The depth of water at the Sulina entrance to the Danube is 24 ft. The mouth is generally blocked with ice in the winter.

SULITJELMA, a mountain (6,279 ft.) on the frontier between Norway and Sweden. It is covered with a snowfield from which many glaciers descend. The mountain is famous as a source of pyrites, rich in copper, which occurs as lodes in schists. From Sjönstaa steamers on the Langvand and a light railway give communication between the sea and Furulund, the mining centre.

SULLA, LUCIUS CORNELIUS (138–78 B.C.), Roman general and dictator, wrought great constitutional reforms. He sprang from an old but undistinguished patrician family and had in youth something of a reputation as a fop. As quaestor, however, in 107 he joined the staff of C. Marius in Africa and in 105, by a piece of trickery, contrived the capture of Jugurtha. Though by this he attracted attention to himself and earned Marius' jealousy, he continued to serve on Marius' staff in the war against the Cimbri and Teutones in Gaul. After his praetorship in 93 he governed Cilicia in 92 and secured from Mithradates VI Eupator, king of Pontus, the restoration to Cappadocia of King Ariobarzanes, whom he had expelled. He returned to Italy to fight

(as Marius also fought) in the Social War and was elected consul for 88, in which year he contracted important political alliances by marrying, as his fourth wife, Caecilia Metella, widow of M. Aemilius Scaurus. As consul he received command of the war against Mithradates (*q.v.*).

The tribune P. Sulpicius Rufus, however, in the interest of the Populares, carried a bill at Rome transferring the command in this war from Sulla to Marius. Sulla went at once to his legions in Campania and marched on Rome. Sulpicius was killed, Marius escaped and measures were hurriedly improvised to prevent a recurrence of this type of tribunician interference in Roman politics, which had been recurrent in the 45 years since the tribunate of Tiberius Gracchus. Sulla then set out for the east. Athens, which had joined Mithradates, was captured after a long siege in 86 and Archelaus was defeated at Chaeroneia and Orchomenus. Negotiations followed; and at Dardanus in the Troad in August 85, Sulla, anxious about the situation in Italy, offered Mithradates terms, which he accepted. He had to abandon all captured territory, to surrender his fleet and to pay an indemnity. After imposing a fine of 20,000 talents on the province of Asia and introducing emergency measures for the reinstitution of Roman taxation in the province, Sulla in the summer of 83 returned to Italy with an army of about 40,000 men and with a wealth of plunder which included, from Athens, the works of Aristotle, most of them still at that date unpublished. During his absence the Popular party had had its own way under the leadership of L. Cornelius Cinna, Marius having died in 86. Sulla's political measures of 88 had been rescinded; Sulla himself had been declared a public enemy; and an ineffective successor, the consular L. Valerius Flaccus, had been sent out to replace him. Flaccus, however, was murdered at the instance of his lieutenant C. Flavius Fimbria, who in his turn came to no less ignominious an end. On his return to Italy Sulla was joined by those opponents of the Popular regime who had saved their lives by going underground during his absence, notably by that Cn. Pompeius whom three years later he was to salute as "Magnus"—Pompey the Great. In 83 Pompey was 23 years old, and he commanded an army which he had raised in his native Picenum. Cinna having been murdered in 84, Sulla's opponents were Cn. Papirius Carbo and the younger Marius, consuls for 82, with forces raised in the main from Samnium and Etruria. There followed the first full-scale civil war in Roman history, ended in 82 by Sulla's victory in the battle of the Colline gate (in the northern suburbs of Rome) and by the siege and capture of Praeneste (Palestrina) held by the young Marius, who committed suicide after the surrender. The infamous proscriptions followed.

The *lex Valeria* of 82, whereby Sulla was appointed dictator, was contrary to precedent in that it set no limit of time on his tenure of the office. He celebrated a spectacular triumph for his successes in the east; and as dictator in 81 and as both dictator and consul in 80, he carried through a vast program of constitutional reform, a series of *leges Corneliae,* mainly designed to re-establish the supremacy of the senate in the Roman constitution. Much of this reform was statesmanlike and survived to the end of the republic: in particular, a new treason law, the *lex Cornelia maiestatis,* designed to check insubordination on the part of provincial governors and army commanders; an increase in the number of standing courts for criminal trials, the *quaestiones perpetuae;* and an arrangement by which the 20 quaestors elected each year should pass automatically into the senate. Others of his reforms were partisan and inevitably short-lived: the return to the system which C. Gracchus had overthrown, by which only senators might serve as jurors in the *quaestiones;* and the virtual elimination of the tribunate as an effective force in Roman political life, since the tribunes were now completely or virtually debarred from introducing legislation, and former tribunes were not allowed to hold the higher magistracies of state. Worst of all was the ruthless manner in which Sulla's discharged troops and supporters were settled in the Italian countryside, whether in colonies (for instance, Pompeii) or in districts from which the existing inhabitants were evicted. These settlers were the much-hated "Sullani."

In 79 Sulla startled Rome by resigning the dictatorship and retiring into private life. He lived for a year longer, in Campania, writing the memoirs of which Plutarch was to make such extensive use and enjoying the company of the raffish bohemian society in which his youth had been spent.

This resignation—"elementary ignorance," Julius Caesar called it later—constitutes a puzzle that has never been solved. Was Sulla just bored? Was his health beginning to fail? Or did he think that, with senatorial authority re-established, his task was finished? The evidence does not fully support J. Carcopino's hypothesis that Sulla had all the time been aiming not at the restoration of republican government but at establishing something like the principate which Augustus was to establish later; and that, his scheme detected, his most powerful backers, led by the Caecilii Metelli, turned on him and forced him to resign, using as instruments Pompey, who was certainly on bad terms with Sulla at the end, and Cicero, whose first important speech, *Pro Roscio Amerino,* in 80 was a bold attack on the Sullan regime.

Sulla was the most superstitious of men, believing himself to be the particular object of divine favour. He expressed his belief in his own luck by assuming, late in 82, the name Felix (in Greek, Epaphroditus). In the same spirit he had given his children earlier the unusual first names of Faustus and Fausta. Faustus married Pompey's daughter and was killed in 46, having fought on Pompey's side against Julius Caesar; Fausta, even in the Roman society of the last years of the republic, was a notorious profligate.

BIBLIOGRAPHY.—Ancient sources: Appian, *Bellum Civile,* i; Plutarch, *Marius* and *Sulla;* Sallust, *Bellum Iugurthinum;* Cicero, *Pro Roscio Amerino.* Modern writers: *Cambridge Ancient History,* vol. ix, chapters iv–vi (by Hugh Last, R. Gardner and M. Rostovtzeff), with extensive bibliography down to 1932; J. Carcopino, *Sulla ou la monarchie manquée* (Paris, 1931); F. Fröhlich, "L. Cornelius Sulla Felix," Pauly-Wissowa, *Real-Encyclopaedie,* iv, pp. 1522–66; Carolina Lanzani, *Lucio Cornelio Silla dittatore* (Milan, 1936). On Sulla's memoirs *see* H. Peter, *Historicorum Romanorum reliquiae,* i, 2nd ed. (Leipzig, 1914), cclxx–cclxxx. On his religious views and his title "Felix" *see* J. P. V. D. Balsdon, "Sulla Felix," *Journal of Roman Studies,* xli (1951), 1–10; and H. Erkell, *Augustus, Felicitas, Fortuna* (Göteborg, 1952).

(J. P. V. D. B.)

SULLIVAN, SIR ARTHUR SEYMOUR (1842–1900), English composer, was born in Lambeth on May 13, 1842, the son of a cultivated Irish musician who was bandmaster at the Royal Military college, Sandhurst, from 1845 to 1856, and taught at the Military School of Music at Kneller Hall from 1857 till his death in 1866. His mother, nee Mary Coghlan (1811–1882), was of Italian descent. Arthur Sullivan was brought up to music from boyhood, and he had learned to play every wind instrument in his father's band by the age of eight. He also had a fine treble voice, and became a chorister of the Chapel Royal. In 1856 he won the Mendelssohn scholarship at the Royal academy, where he studied under Sterndale Bennett, Arthur O'Leary and John Goss. In 1858 he was sent to study at Leipzig, where he had for teachers Moscheles and Plaidy for pianoforte, Hauptmann for counterpoint, Rietz and Reinecke for composition and F. David for orchestral playing and conducting.

After two years' hard study he returned to London in April 1861 and at his instigation Schumann's first symphony was introduced at one of the Crystal Palace winter concerts. The following year his own *Tempest* music was performed at the same concerts. Then followed his *Kenilworth* cantata (remembered chiefly for the charming duet, "How Sweet the Moonlight"), the *Sapphire Necklace* overture, and the five beautiful songs from Shakespeare, which include "Orpheus with his Lute," "Oh Mistress Mine" and "The Willow Song." His attractive personality, combined with his undoubted genius and brilliant promise, brought him many friends. Costa, who was conductor at Covent Garden, gave him the post of organist, and in 1864 he produced there his *L'Île Enchantée* ballet. Other works followed, and in 1867 he first showed his genius for light operatic music in his collaboration with F. C. Burnand in *Cox and Box* and in *Contrabandista.*

In 1871 Sullivan became acquainted with W. S. Gilbert (*q.v.*), and in that year they collaborated in a piece for the Gaiety theatre, called *Thespis; or, The Gods Grown Old,* which was a great success in spite of the limited vocal resources of the performers. In 1875 R. D'Oyly Carte, then acting as manager for

Selina Dolaro at the Royalty, suggested to Gilbert collaboration with Sullivan in a piece for that theatre. Gilbert had already suggested to Sullivan an operetta with its scene in a law court, and within three weeks of his completing the libretto of *Trial by Jury* the music was written. The piece succeeded beyond all expectation; and on the strength of its promise of further successes D'Oyly Carte formed his Comedy Opera company and took the Opéra Comique theatre. There in 1877 *The Sorcerer* was produced, George Grossmith and Rutland Barrington being in the cast. In 1878 *H.M.S. Pinafore* was brought out at the Opéra Comique, and ran for 700 nights. In America it was enthusiastically received and pirated right and left and the two authors, with D'Oyly Carte, went over to the U.S. in 1879, with a company of their own, in order to produce it in New York. To secure the American rights for their next opera, they brought out *The Pirates of Penzance* first at New York in 1879. In 1880, in London, it ran for nearly 400 nights. In 1881 *Patience* was produced at the Opéra Comique, and was transferred later in the year to the Savoy theatre. There all the following operas came out: *Iolanthe* (1882), *Princess Ida* (1884), *The Mikado* (1885), *Ruddigore* (1887), *The Yeomen of the Guard* (1888), *The Gondoliers* (1889). The vogue of the new type of light opera owed something to such admirable performers as George Grossmith, Rutland Barrington, Miss Jessie Bond, Miss Brandram, and later W. H. Denny and Walter Passmore; but these artistes only took advantage of the opportunities provided by the two authors. In place of the old adaptations of French *opéra bouffe* they had substituted a genuinely English product, humorous and delightful, without a tinge of vulgarity or the commonplace. But disagreements arose between them which caused a dissolution of partnership. Sullivan's next Savoy opera, *Haddon Hall* (1892), had a libretto by Sydney Grundy; and the resumption of Gilbert's collaboration in 1893 in *Utopia, Limited*, and again in 1896 in *The Grand Duke*, was not as successful as before. Sullivan's music, however, still showed its characteristic qualities in *The Chieftain* (1894)—largely an adaptation of *Contrabandista; The Beauty Stone* (1898), with a libretto by A. W. Pinero and J. Comyns Carr; and particularly in *The Rose of Persia* (1900), with Basil Hood.

In 1872 Sullivan's *Te Deum* for the recovery of the prince of Wales was performed at the Crystal Palace. In 1873 he produced at the Birmingham Musical festival his oratorio *The Light of the World*, in 1877 he wrote his incidental music to *Henry VIII*, in 1880 his sacred cantata *The Martyr of Antioch*, and in 1886 *The Golden Legend* was brought out at the Leeds festival. In 1891, for the opening of D'Oyly Carte's new English opera house in Shaftesbury avenue, later the Palace theatre, he wrote his "grand opera" *Ivanhoe* to a libretto by Julian Sturgis. The attempt to put an English opera on the stage for a long run was doomed to failure, but *Ivanhoe* contains many admirable pages. In 1892 he composed incidental music to Tennyson's *Foresters*. In 1897 he wrote a ballet for the Alhambra, called *Victoria and Merrie England*. Among his numerous songs, the best known are "Orpheus With His Lute," "Thou'rt Passing Hence" and, most famous of all, "The Lost Chord." This last, hackneyed as it became, was probably the most successful English song of the 19th century. The hymn tune "Onward, Christian Soldiers!" (1872) shows Sullivan in another light. In 1876 he accepted the principalship of the National Training School of Music, which he held for six years; this was the germ of the subsequent Royal college. He was conductor of the Leeds festivals from 1879 to 1898, besides being conductor of the Philharmonic society in 1885. He died on Nov. 22, 1900, and was buried in St. Paul's cathedral. Among works which he left unpublished may be mentioned a *Te Deum*, written for performance at the end of the South African War, and an unfinished Savoy opera to a libretto by Basil Hood, which, completed by Edward German, was produced in 1901 as *The Emerald Isle*. Sullivan was the one really popular English composer of any artistic standing in his time. One of the most agreeable companions, broad-minded and free from all affectation, he was intensely admired and loved in all circles of society; and though he suffered during many years from a painful ailment, he enjoyed life without being spoiled by success.

BIBLIOGRAPHY.—A. Lawrence, *Sir Arthur Sullivan: Life Story, Letters and Reminiscences* (1899); H. Saxe-Wyndham, *Arthur Seymour Sullivan* (1926); Herbert Sullivan and Newman Flower, *Sir Arthur Sullivan, His Life, Letters and Diaries* (1928).

SULLIVAN, HARRY STACK (1892–1949), U.S. psychiatrist, who originated a significant conception of psychiatry as the study of interpersonal relations, was born at Norwich, N.Y., on Feb. 21, 1892. He was vigorous, original, sensitive and uncompromising, and his views quickly became controversial, stimulating much study of the relation between social and group processes and mental disorder.

Sullivan, who received his medical degree in 1917, began his psychiatric work under William A. White at St. Elizabeth's hospital, Washington, D.C., in 1919, and turned to clinical research at Sheppard and Enoch Pratt hospital, Towson, Md., from 1923 to 1930. There he developed his formulation of psychiatry as the study of interpersonal relations, a point of view he tested by organizing a strikingly successful special ward for the group treatment of schizophrenic patients. After 1930 he devoted himself to elaborating his theoretical analysis, teaching, taking part in establishing the Washington School of Psychiatry (1936) and writing. His later views differed sharply from either psychobiological or psychoanalytic formulations, although owing something to both. Conceiving of anxiety as arising in the infant as a disturbance in the relation between the child and the mothering one, he taught that through the years of maturation the child develops a self-dynamism whose function is to guide behaviour in such a way as to diminish anxiety; thus, from its nature, the self-dynamism is in large measure formed by the personality of the mothering one. The disruptive effect of anxiety acts in opposition to tendencies to integrate situations with others. Consequently, both personality and the "dynamisms of difficulty"—psychiatric disorders—are interpersonal phenomena. Sullivan explicitly thought of the unique individual personality as a myth.

Following World War II, Sullivan's opinions led him into energetic efforts to apply psychiatric understanding to international tensions. He died suddenly of apoplexy in Paris, Jan. 14, 1949. His theories are set forth in his *The Interpersonal Theory of Psychiatry* (1953), his clinical studies in *The Psychiatric Interview* (1954) and *Clinical Studies in Psychiatry* (1956).

See Patrick Mullahy (ed.), *Contributions of Harry Stack Sullivan: a Symposium of Interpersonal Theory in Psychiatry and Social Science* (1952); Edith Jacobson, "Sullivan's Interpersonal Theory of Psychiatry," *J. Am. Psychoanal. Ass.*, 3:149–156 (1955), a critical review.

(A. H. Sn.)

SULLIVAN, JOHN (1740–1795), American soldier, was born in Somersworth (now Rollinsford), N.H., on Feb. 17, 1740. He practised law at Berwick, Me., and at Durham, N.H., was a member of the New Hampshire provisional assembly (1774), and also a member of the first continental congress. In June 1775 he was appointed brigadier general in the Continental army and aided in the siege of Boston. In the same year he took command of the American army in Canada and exhibited great skill in effecting a retreat from that province. Soon afterward, having been superseded in command by General Gates, he rejoined Washington's army and took part in the battle of Long Island, where he was taken prisoner. In December he was exchanged and succeeded Gen. Charles Lee in command as major general of the right wing of Washington's army. In the battle of Trenton he led an attack on the Hessians, and on Aug. 22, 1777, he led a night attack against the British and loyalists on Staten Island. He commanded the American right in the battle of Brandywine and took part in the battle of Germantown. In March 1778 he was placed in command in Rhode Island, and in the following summer plans were made for his co-operation with the French fleet under Count d'Estaing in an attack on Newport.

The Indian raids in western New York, especially the atrocities in the Wyoming and Cherry valleys, caused a retaliatory expedition to be sent to "chastise and humble the Six Nations," and Sullivan was chosen to lead the expedition. With about 4,000 men, he defeated the Iroquois and their loyalist allies at Newtown (now Elmira, N.Y.), burned their villages and destroyed their orchards

and crops. Although severely criticized for his conduct of the expedition, he received the thanks of congress in Oct. 1779, and in November resigned from the army. Sullivan was again a delegate to the Continental Congress in 1780–81 and from 1782 to 1785 he was attorney general of New Hampshire. He was president of the state in 1786–87 and in 1789, and in 1786 suppressed an insurrection at Exeter immediately preceding Shays's rebellion in Massachusetts. He presided over the New Hampshire convention which ratified the federal constitution in June 1788. From 1789 until his death at Durham, on Jan. 28, 1795, he was United States district judge in New Hampshire.

SULLIVAN, JOHN L. (Lawrence) (1858–1918), U.S. professional boxer, known as the "Great John L." and the "Boston Strong Boy." Born on Oct. 15, 1858, at Boston, Mass., Sullivan attended school there and for a time studied at Boston college. At the age of 19 he launched his professional boxing career. Sullivan was supreme at the rough-and-ready style, and in 1882 knocked out Paddy Ryan in nine rounds to win the championship. He held the title until Sept. 7, 1892, when James J. Corbett knocked him out in 21 rounds at New Orleans, La. In the last bare-knuckle fight of professional boxing, Sullivan knocked out Jake Kilrain in 1889 in a 75-round battle that lasted 2 hr. 15 min. 25 sec. Sullivan, who had earned more than $1,000,000 as champion but dissipated it all, later became an advocate of prohibition and gave many lectures on the subject before his death in modest circumstances on Feb. 2, 1918. (J. P. D.; X.)

SULLIVAN, LOUIS HENRI (1856–1924), was one of the pioneers of the modern movement in architecture. Born in Boston, Mass., on Sept. 3, 1856, he studied at the Massachusetts Institute of Technology, Cambridge, and at the École des Beaux-Arts in Paris. He worked as a draftsman in various offices, including those of Frank Furness in Philadelphia, Pa., and William Le Baron Jenney in Chicago, Ill., in 1873–74. A decade later Jenney, an engineering genius, invented the principle of skyscraper construction. (*See* Structural Engineering.) Sullivan was much interested in the many structural innovations made in Chicago during the generation after the fire of 1871. Dissatisfied with the widespread imitation of past styles that had prevailed during the preceding century, he sought to create a new architectural style, expressive of the functional problems and engineering advances of his day.

Sullivan's greatest work was done during his partnership with Dankmar Adler (1881–95). During those years the firm designed more than 100 structures. Outstanding were the Auditorium building (acquired by Roosevelt university), Chicago; the Transportation building at the Chicago Columbian exposition of 1893; the Gage building, Chicago; and two great skyscrapers, the Wainwright building in St. Louis, Mo., and the Prudential building in Buffalo, N.Y. In his steel-frame office buildings and skyscrapers Sullivan created designs harmonious with the purpose of the structure and appropriate to the new materials being used. During the years 1887–93 Sullivan was assisted by Frank Lloyd Wright (*q.v.*), who entered the office at 18 and became an ardent disciple of the older man, to whom he always referred as "the Master."

After the dissolution of the partnership with Adler, Sullivan's practice fell off drastically, but the score or more buildings he produced before his death included his masterpiece, the Carson Pirie Scott department store building in Chicago (1899–1904), and a number of fine banks in small midwestern towns, notably National Farmers' bank (later the Security Bank and Trust Company of Owatonna), Owatonna, Minn.

More influential than his buildings was his philosophy of architectural design, which he set forth in books, magazine articles and lectures. In these he bitterly attacked the imitation of historic styles—the architectural eclecticism of the 19th and 20th centuries—and called for a modern style expressive of the facts and spirit of the 20th century. Sullivan's theory of "functionalism" is but baldly stated in his slogan "form follows function." He believed that the architectural style of a building should be derived from and expressive of its function and structure. He believed that "each problem contains and suggests its own solution," that each building should have its own identity of beauty and that this must necessarily reflect its origin in present-day conditions, social, technical and spiritual. Another mainspring of Sullivan's thinking was an ardent belief in United States democracy, eloquently expressed in his Whitmanesque prose poetry and basic to his conviction that architecture is a democratic social art. Sullivan's most important published books were *Kindergarten Chats* (1901–02) and *The Autobiography of an Idea* (1924).

Sullivan died in almost complete poverty on April 14, 1924, in Chicago. After his death he came to be recognized, in both the United States and Europe, as one of the great pioneers of modern architecture. In 1943 he was posthumously awarded the gold medal of the American Institute of Architects. In 1956 the centennial of his birth was widely celebrated and an exhibition of his work was prepared at the Art Institute of Chicago for national and international showings. *See* also Modern Architecture.

See Hugh Morrison, *Louis Sullivan, Prophet of Modern Architecture* (1935); John Szarkowski, *The Idea of Louis Sullivan* (1956).
(Hr. M.)

SULLY, MAXIMILIEN DE BÉTHUNE, Duc de (1560–1641), French statesman, was born at the château of Rosny near Mantes on Dec. 13, 1560, son of François de Béthune, baron de Rosny (1532–75). Brought up in the Reformed faith, Maximilien joined the court of Henry of Navarre in 1571. He was taken to Paris by his patron and was studying at the college of Bourgogne at the time of the St. Bartholomew massacre, from which he escaped by discreetly carrying a Book of Hours under his arm. On the outbreak of civil war in 1575 he enlisted in the Huguenot army. In 1576 he accompanied the duke of Anjou on an expedition into the Netherlands in order to regain the former Rosny estates, but being unsuccessful he attached himself for a time to the prince of Orange. Later rejoining Henry of Navarre in Guienne, he displayed particular ability as an engineer. In 1583 he was Henry's special agent in Paris. In 1584 he married Anne de Courtenay, a wealthy heiress, who died, however, in 1589. On the renewal of civil war Rosny again joined Henry of Navarre, and at the battle of Ivry (1590) was seriously wounded. He counseled Henry IV's conversion to Roman Catholicism, but steadfastly refused himself to become a Roman Catholic.

Rosny became a member of the king's council of finance in 1596 and appears to have been sole superintendent as early as 1598. He authorized free exportation of grain and wine, reduced legal interest from 8⅓% to 6¼%, established a special court for trial of cases of peculation, forbade provincial governors to raise money on their own authority and otherwise removed many abuses of tax collecting, abolished several offices and by his honest, rigorous conduct of the country's finances was able to save between 1600 and 1610 an average of 1,000,000 livres a year. His achievements were not solely financial. In 1597 he was appointed grand commissioner of highways and public works and in 1599, superintendent of fortifications and grand master of artillery; in 1602 governor of Mantes and of Jargeau, captain general of the queen's *gens d'armes* and governor of the Bastille; in 1604 governor of Poitou; and in 1606 duke and peer of Sully, ranking next to princes of the blood. He declined the office of constable because he would not become a Roman Catholic. Sully encouraged agriculture, urged the free circulation of produce, promoted stock raising, forbade the destruction of the forests, drained swamps, built roads and bridges, planned a vast system of canals and actually began the canal of Briare. He strengthened the French military establishment; under his direction Évrard began the construction of a great line of defenses on the frontiers. Sully opposed the king's colonial policy as inconsistent with the French genius, and likewise showed little favour to industrial pursuits, although on the urgent solicitation of the king he established a few silk factories. He fought in company with Henry IV in Savoy (1600–01) and negotiated the treaty of peace in 1602; in 1603 he represented Henry at the court of James I of England; and throughout the reign he helped the king to put down insurrections of the nobles, whether Roman Catholic or Protestant. It was Sully, too, who arranged the marriage between Henry IV and Marie de Médicis.

The political role of Sully practically ended with the assassination of Henry IV on May 14, 1610. Although a member of the council of regency, his colleagues were not disposed to brook his domineering leadership, and after a stormy debate he resigned

as superintendent of finances on Jan. 26, 1611, and retired to private life. The baton of marshal of France was conferred on him on Sept. 18, 1634. He died at Villebon, on Dec. 22, 1641.

Sully was hated by Catholics because he was a Protestant, by Protestants because he was faithful to the king. An excellent man of business, he was implicitly trusted by Henry IV and contributed greatly to the recovery of France.

Sully left a curious collection of memoirs bearing the title, *Mémoires des sages et royales oeconomies d'estat, domestiques, politiques, et militaires de Henry le Grand. . . .* The *Mémoires* are very valuable for the history of the time, though they contain fictitious matter such as a mission by Sully to England in 1601 and the "grand design" for a Christian republic.

SULLY, THOMAS (1783–1872), U.S. painter, one of the best early American portrait painters, was born at Horncastle, Eng., on June 8 (or 19), 1783, of actor parents. They moved to the United States when Sully was nine years old, settling at Charleston, S.C. He was a pupil of Gilbert Stuart in Boston, and in 1809 he entered the studio of Benjamin West in London. After 1810 he made Philadelphia his home, but in 1837 again visited London, where he painted a full-length portrait of Queen Victoria for the St. George's society of Philadelphia. He died in Philadelphia on Nov. 5, 1872. Though Sully left about 500 subject pictures, he is best known for portraits. Among his 2,000 portraits are those of Stephen Decatur (City hall, New York); the actor George Frederick Cooke as Richard III, and La Fayette (both in the Pennsylvania Academy of the Fine Arts, Philadelphia); and Thomas Jefferson (U.S. Military academy, West Point, N.Y.). Many others are in galleries throughout the U.S.

SULLY PRUDHOMME (pseudonym of René François Armand Prudhomme) (1839–1907), French poet and philosophical and critical writer who won the Nobel prize for literature in 1901, was born in Paris, March 16, 1839. He was educated in science but gave up that study because of an eye injury. His first volume of poems *Stances et poèmes* (1865; new ed., 1866) was favourably reviewed by Sainte-Beuve. He published with the Parnassians (*q.v.*) and in *Épreuves* (1866) sounded the note of melancholy that characterized the rest of his work. Family troubles and a serious paralytic illness incurred from service in the Garde Mobile in 1870 ruined his health. After 1888 he wrote only prose, philosophy and criticism. His *Oeuvres* (1883–1908, 8 vol.) include both poetry and prose.

Elected to the academy in 1881, he won the Nobel prize in 1901, using most of the prize money to establish another prize for poetry. He died in Paris on Sept. 7, 1907.

See Edmond Esteve, *Sully-Prudhomme, poète sentimental et poète philosophe* (1927); Pierre Flottes, *Sully-Prudhomme et sa pensée* (1930).

SULMONA (ancient Sulmo), a city of Abruzzi e Molise, Italy. Pop. (1951) 17,689. Its cathedral of San Panfilo has a 14th-century portal and an 11th-century crypt. S. Francesco della Scarpa occupies the site of an older and larger church, the Romanesque portal of which still stands. Opposite is a picturesque aqueduct of 1256 with pointed arches. The Porta Napoli is a gate of the 14th century. Innocent VII was a native of the town. In the vicinity of the town is Monte Morrone where Pietro da Morrone lived as a hermit and founded a monastery for his hermits, who after his elevation to the papacy as Celestine V (1294) took the name of Celestines; the monastery (S. Spirito) is a prison. Sulmo was devastated by Hannibal (211 B.C.). It was famous for its ironsmiths. Charles V bestowed it on the viceroy of Naples, Charles de Lannoy.

SULPHUR. For a description of this element and its commercially important acid, *see* Sulfur; Sulfuric Acid.

SULPICIA, the name of two Roman poets. The earlier, daughter of Servius Sulpicius Rufus and niece of M. Valerius Messalla Corvinus, lived in the reign of Augustus. Her verses, preserved with those of Tibullus and for long attributed to him, consist of six short elegiac poems (*Corpus Tibullianum,* book iv, 7–12, 40 lines in all) addressed to a lover called Cerinthus (seldom now identified with the Cornutus mentioned by Tibullus in book ii, 2 and 3) and are preceded by five elegiac poems (book iv, 2–6) by another poet who comments sympathetically on the love affair. The latter poet may be Tibullus, but Esther Bréguet argues, from a detailed study of the language and metre, that the poet of 2–6 was the young Ovid, imitating the style of Tibullus.

The younger Sulpicia, wife of Calenus, lived during the reign of Domitian. Martial (x, 35 and 38) praises her wifely devotion, which was the subject of her poems. An extant satire (70 hexameters), discovered in 1493 and attributed to her, is probably of much later date.

Bibliography.—For Sulpicia I *see* E. Bréguet, *Le Roman de Sulpicia* (1946); editions of Tibullus by K. F. Smith (1913) and J. P. Postgate (1929); also Schanz-Hosius, *Geschichte der römischen Literatur,* vol. ii (1935). For the satire attributed to Sulpicia II *see* O. Jahn's edition of Juvenal and Persius as revised by F. Bücheler (1893); E. Bährens, *De Sulpiciae quae vocatur satira commentatio philologica* (1873); W. S. Teuffel, *History of Roman Literature,* Eng. trans., rev. ed. (1900); J. W. Duff, *Roman Satire* (1937). (L. J. D. R.)

SULPICIUS RUFUS, PUBLIUS (*c.* 124–88 B.C.), Roman orator and politician, supported the reforms of his friend the younger M. Livius Drusus (91 B.C.) and inherited his policy of generous concessions to the Italian allies. Though he renounced his patrician status to become a tribune of the plebs, he had shown, by prosecuting the tribune Gaius Norbanus in 94, that his political attitude was moderate. In the Social War he served as legate to Gnaeus Pompeius Strabo (89). As tribune in 88 he introduced several laws: (1) to unseat all senators owing more than 2,000 denarii; (2) to recall all men exiled under the Varian law of 90 B.C. (a measure which he himself had opposed earlier); (3) to transfer the command against Mithradates from Sulla, the senate's nominee, to Marius; and (4) to distribute freedmen and the newly enfranchised Italians among the 35 tribes (so as to make their voting power effective). Sulpicius' chief aim was probably championship of the new Italian citizens, but to implement this he needed wide support: hence his other measures, which would secure for him the backing of the people, of Marius and the equites. This bold challenge to the senate led to violence, and Sulpicius is said to have organized a group of 600 young equites and a larger bodyguard of 3,000 armed men. The consuls, Sulla and Q. Pompeius Rufus, declared some form of cessation of public business (perhaps a *justitium*), which Sulpicius denounced as illegal. Disturbances followed in the Forum, and Sulla was forced to call off the *justitium;* he then joined his army in Campania. At Rome Sulpicius' measures were carried. Sulla then marched on Rome and won the city. Marius and Sulpicius fled and were declared outlaws, and Sulpicius was caught and killed at Laurentum. His laws were declared invalid, as passed "by force." Sulpicius' policy is obscure, but he may well have been a liberal optimate, whose championship of the Italians led his enemies to depict him as a revolutionary demagogue rather than as a moderate reformer forced to go to greater lengths than he probably wished. (H. H. Sd.)

SULPICIUS RUFUS, SERVIUS (*c.* 106–43 B.C.), surnamed Lemonia from the tribe to which he belonged, was a Roman orator and jurist. He studied rhetoric with Cicero and accompanied him to Rhodes in 78 B.C. Finding that he would never be able to rival his teacher he gave up rhetoric for law.

In 63 he was a candidate for the consulship, but was defeated by L. Licinius Murena, whom he subsequently accused of bribery; in 51 he was successful.

In the civil war, after considerable hesitation, he threw in his lot with Caesar, who made him proconsul of Achaea in 46. He died in 43 while on a mission from the senate to Antony at Mutina. He was accorded a public funeral, and a statue was erected to his memory in front of the Rostra.

Two excellent specimens of Sulpicius' style were preserved in Cicero. Quintilian speaks of three orations by Sulpicius as still in existence; one of these was the speech against Murena, another *Pro* or *Contra Aufidium,* of whom nothing is known. He is also said to have been a writer of erotic poems. It was as a jurist, however, that Sulpicius was chiefly distinguished. He left behind him a large number of treatises, and he is often quoted in the Digest, although direct extracts are not found.

SULTANPUR, a town and district in Uttar Pradesh, republic of India. The town is on the Gumti river, midway between

Varanasi (Banaras) and Lucknow. Pop. (1961) 26,081.

SULTANPUR DISTRICT has an area of 1,713 sq.mi.; pop. (1961) 1,412,984. There are two other Sultanpurs in Punjab.

SULU, a province and archipelago in the Philippines extending about 170 mi. S.W. from the southwestern tip of Mindanao. Its area of 1,087 sq.mi. is divided into hundreds of islands, some volcanic and some coral, of which the most important are Jolo, Tawi Tawi, Sanga Sanga, Sibutu, Siasi and Cagayan Sulu. The 1960 provincial population of 326,898 was about 95% Muslim. These Muslim tribes, known as Moros, historically were fiercely independent (*see* MORO). During the Spanish occupation of the Philippines a troop garrison was stationed at the town of Jolo but Spanish rule never extended far beyond the fort. During the early period of U.S. rule, which began in 1899, several battles were fought against the Moros. The archipelago became part of Moro province (1903) and civil government was established 1914. On March 11, 1915, the reigning sultan of Sulu agreed to abdicate his rights of sovereignty and to retain only his right to be head of the Islamic sect. Following this agreement sporadic fighting continued between government forces and outlaw bands. A treaty signed April 1940 abolished the sultanate and Sulu became part of the Commonwealth, later the Republic, of the Philippines.

Cassava, coconuts and rice are the main agricultural products, but rice is not produced in sufficient quantity and some must be imported. Fishing and gathering sea shells (for the manufacture of buttons) are important activities. The Turtle Islands are famous as a source of turtle shells and eggs. Considerable illegal trade is carried on between Sulu and Borneo. The capital and chief port and trading centre is the town of Jolo on Jolo Island.

(R. E. HE.)

SULU SEA. A large, very deep interisland sea, located between the southwestern Philippine Islands and northeastern Borneo. About 350 mi. from north to south, 425 mi. in east to west dimension, and 135,000 sq.mi. in area, the Sulu sea is enclosed by various islands of the Philippines: Palawan on the west and northwest, Dalanganem and Cuyo Islands on the north, Panay and Negros on the east, Mindanao and the Sulu archipelago on the southeast, and British North Borneo on the southwest. The sea occupies a downfaulted block, in some places almost 3,000 fathoms (18,000 ft.) deep. The edges of this enclosed basin are seen in the bordering islands, particularly in the northeast-southwest orientation of Palawan and the Sulu archipelago. The sea is bisected by another northeast-southwest fracture line which is evidenced by the coral atolls of the Cagayan Islands, the Tubbataha reefs, and the volcanic cones of the Cagayan Sulu group. The sheltered sea is extensively used for interisland trade. Fishing is an important occupation and marine products include pearls, pearl shells, *bêche-de-mer,* shark fins and turtle eggs. Copra, timber and forest products of the surrounding islands provide the other commercial activities of the area.

The Sulu sea was long the stronghold of the fierce Moro pirates of the Sulu archipelago, who raided shipping as far away as Malaya. Punitive campaigns, started by the Spanish in the mid 19th century and terminated by U.S. troops in 1906, broke the power of the Moslem Moros and brought peace to the area. (L. A. P. G.)

SUMAC, the popular name for shrubs and small trees of the genus *Rhus* (family Anacardiaceae) with perhaps 50 species in southern Europe, Asia and eastern North America. They have a resinous, milky sap which hardens in the air; few staghorn branches; alternate, pinnate leaves; terminal clusters of small flowers with parts in fives; and small, one-seeded, hairy, red fruits. Closely related to and often included in *Rhus* is the genus *Schmaltzia* with about 30 species of many-branched shrubs with spicate clusters of flowers and fruit and variously compound or simple, deciduous or evergreen leaves. The distribution of *Schmaltzia* centres in Mexico.

The name sumac is given also to the commercial preparation of the dried and ground leaves of the Sicilian or tanners' sumac (*R. coriaria*) of southern Europe, long used in making leather. Several of the eastern North American species also were used in tanning. Among the most attractive North American shrubs are the staghorn sumac (*R. typhina*), sometimes a tree 30 ft. high, and the smooth sumac (*R. glabra*), sometimes 20 ft. high, both of which in autumn display highly coloured fruit and foliage. The dwarf or shining sumac (*R. copallina*) is a small shrub in the north and a tree 30 ft. high in the south. Best known in *Schmaltzia* is the aromatic sumac (*S. crenata*), 3 to 8 ft. high, with pleasant-scented foliage, which occurs in the eastern U.S., and its western counterpart, the skunkbush or squawbush (*S. trilobata,* perhaps better treated as a subspecies of *S. crenata*), with ill-scented foliage, found from Illinois to Oregon and southward. The mahogany sumac or sourberry (*S. integrifolia*) and the sugarbush (*S. ovata*) are elegant shrubs of southern California.

ROCHE

SMOOTH SUMAC (RHUS GLABRA)

The laurel sumac of the same area is sometimes placed in another segregate genus, *Malosma* (*M. laurina*).

Also related to *Rhus* is the genus *Toxicodendron,* which includes about 30 species containing an irritant poison in the sap. Best known in the U.S. are the swamp-dwelling poison sumac (*T. vernix*), poison oak (*T. quercifolia* of southeastern barrens and *T. diversiloba* of western thickets), and the nearly ubiquitous poison ivy (*q.v.*). Also poisonous is the Japanese varnish tree (*T. vernicifera*), which yields Japanese lacquer.

In addition to their value in tanning, sumacs were used by the Amerinds as sources of orange dyes, medicines, fuel, twigs for basket-weaving (*S. trilobata*), and fruit for cool drinks in summer and hot drinks, with maple sugar, in winter. *R. chinensis* of eastern Asia and *R. typhina* and *R. glabra,* especially the laciniate-leaved varieties, are planted for their ornamental foliage. Sumacs are valuable as sand binders, game cover and pioneers in the revegetation of denuded areas. (R. F. TH.)

SUMATRA, lying between 5° 39′ N. and 5° 57′ S., so that the equator divides it into two nearly equal parts, is the westernmost and, after Borneo, the largest of the Great Sunda Islands of the Malay archipelago and forms part of Indonesia. Its axis runs from southeast to northwest, from Sunda strait to Malacca passage; it is 1,069 mi. in length, and its greatest width is 246 mi. With its dependencies, Sumatra has an area of 182,859 sq.mi. (alone, 167,567 sq.mi.) and is nearly four times as large as Java and almost as large as Spain. The economic development of Sumatra, though it is far behind that of Java, is much further advanced than that of any other of the Indonesian islands.

Physical Structure.—In relief, Sumatra consists of a high mountain chain which runs along the western coast, descending eastward to a huge tract of flat, alluvial land, seamed with many large rivers and their scores of tributaries. Its great mountain chain is a link between the mountain system of Lower Burma, which is continued by way of Preparis and Great Coco Islands, the Andaman Islands and the Nicobars to Sumatra and thence to Java and on through the Lesser Sunda Islands to the Moluccas. The Sumatran chain extends for a distance of more than 1,000 mi. and contains numerous volcanic peaks of heights from 5,000 ft. to more than 12,000 ft. The whole system, known as the Bukit Barisan ("Parade of Mountains"), consists in general of two or more folded chains running parallel to each other, with a valley between, which is broken up into separate sections by the intrusion of volcanic massifs. Along this valley lies a string of mountain lakes, from south to north—Ranau, Kerintji, Singkarak, Manindjau and Toba. By far the largest of these is Lake Toba, which is 56 mi. long and 19 mi. wide, of great depth, and has a large island, Samosir.

The line of volcanic peaks stretches across Sumatra from Abongabong (9,892 ft.) and the mountains of Achin to Krakatau

(*q.v.*) off the coast at the southern end of the island. It includes Leuser (Loser) (11,092 ft.), Sinabung (8,040 ft.), Sibayak (7,257 ft.), Pasaman (Talakmau [Mount Ophir], 9,844 ft.), Talang (8,521), Merapi (9,485 ft.), Singalang-Tandikat (9,440 ft.), Kerintji (Kerinchi or Indrapura, 12,464 ft.) and Dempo (10,365 ft.). In all, 90 volcanoes have been counted, 12 of them active.

Sumatra, like Java and Borneo, is formed largely of strata of the Tertiary period, although it contains also two schistose formations, one of which is anterior to the carboniferous period. The Tertiary series are more complete than in Java; the volcanoes, so characteristic of the whole archipelago, are due to the Quaternary period. The western mountain system is composed largely of Archaean rocks, with occasional longitudinal bands of Jurassic and Cretaceous rock. The whole ridge is metalliferous, and erosion of the ranges exposes their mineral contents. The crystalline rocks (granites, basalts and their allies) are of various geological ages; they are exposed by denudation, or pushed through by intrusion. Intrusion causes change in some sedimentary rocks: sandstone changes into quartzite; limestone into marble; and clay or shale, under pressure, into slates.

The river system of Sumatra is extensive and of great value to the country. On the west coast, where the mountains approach the sea, the rivers run in valleys with a steep slope, have a very short lower course and are unnavigable, except near the mouth in the case of a few. The eastern rivers run through alluvial plains, have extensive drainage areas (they are often impeded by silt) and form the principal and often the only means of communication. South of 3° N., there are extremely important rivers, the first being the Asahan, which drains Lake Toba and is navigable by steamers for a portion of its length; next the Panai (Bila and Barumun), with a very wide mouth; then the Rokan, more than 120 mi. long and flowing into the Malacca straits through a wide, muddy estuary, navigable for 70 mi. inland; the Siak, with the port of Pakanbaru 100 mi. upstream; the Kampar, a very long river, navigable for ocean steamers for 18 mi.; the Inderagiri, navigable for ocean steamers as far as Cherinti; then the Hari, which springs from the peak of Indrapura, the largest of all the rivers of Sumatra, having a maximum navigability of just less than 500 mi.; and then the Musi, on which ocean-going vessels can reach Palembang, 54 mi. upstream, while steam launches and motor boats can go 200 mi. inland and small boats 330 mi. (below Palembang the Musi splits up into a number of channels which spread out amid a vast unhealthy swamp and slowly deposit their mud among the mangroves, thus gradually creating new land).

On the coast, there are small semicircular bays, but few large indentations, immunity from the southwest monsoon being provided for harbours only between Tapanuli and Indrapura. The north coast has precipitous cliffs crowned with dense vegetation in some parts; in others sandy beaches and well-cultivated plains, while the bay of Uleelheue gives shelter for shipping. The south coast consists, mainly, of two large indentations, Lampung bay and Semangka bay. The eastern shore of Lampung bay is mountainous and little indented, the western is much cut up; the centre of the bay has a good harbour (Telukbetung). Semangka has a mountainous coast line, with little indentation. The whole of the east coast is formed of morasses and sandbanks, which break up into innumerable islands, large and small, particularly about the central portion; and the coast line is deeply indented in many places and very irregular. The indentations are not permanent, the coast is constantly advancing, and it is difficult to establish practicable ports or landing places.

Sumatra has a chain of islands off the west coast which rise from the edge of the submarine platform and serve as a kind of outer barrier. They have an aggregate area of 5,769 sq.mi. Several are thickly populated, and some are of considerable size; most have rocky, reef-bound coasts, dangerous of approach.

Climate.—The climate of Sumatra resembles that of Java and is hot and extremely moist, but in most eastern coastal districts and often for many miles inland the heat is tempered by cooling land and sea breezes, the wind generally being northeast by day and southwest by night all the year round. The wind system of northern Sumatra differs from that of the greater part of Indonesia, the northeast monsoon blowing from December to March, and the southwest monsoon from May to October. During the southwest monsoon strong squalls from the northwest, known as "Sumatras," which are accompanied by thunder, lightning and rain, blow in the Malacca strait; they have been known also to occur during the northeast monsoon. On the west coast the influence of the northeast and southwest monsoons is felt as far south as 4° N. Between 4° N. and 2° N. there is a region of calms and light variable winds, which have a westerly tendency from March to November, when the northwest monsoon is blowing along the remainder of the west coast (*i.e.*, from 2° N. to 5° 57′ S.); and an easterly tendency from May to September, the period of the southeast monsoon on the west coast. Southern Sumatra has the highest temperatures, and the mean annual temperature for the lowlands is about 80° F. At Toba, in the Batak country (3,772 ft.), this is 69.6° F., and higher altitudes show a corresponding reduction. On the whole, temperature in Sumatra is slightly higher than in Java. March, April and May are the hottest months; January and February the coolest, but the difference does not exceed 2°. The different wind distribution causes a variation of the seasons. Thus in northern Sumatra October is the wettest month and February and March are the driest, while elsewhere the wettest months are December, January and February, and from July to September is the dry period. Accordingly the rainfall varies considerably. The annual average rainfall for northern Sumatra is 95.71 in., for eastern Sumatra 106.27 in., and for western Sumatra 122.32 in. The west monsoon gives the heavier rainfall and the fall is accentuated in western Sumatra by the high mountains.

Vegetation.—The vegetation of Sumatra bears a strong resemblance to that of India and of the Malay peninsula in the northern part of the island and to that of Java in the southern, but it has also many varieties distinct from either, and not only different varieties, but different genera from those of Java. The pine, *Pinus merkusii*, in the mountains of the north is not found elsewhere in the East Indies but occurs as far south as the equator. One of the most striking flowers is the huge *Rafflesia hasseltii*, with blossoms 18 in. across, growing on big lianas on the ground in the forests. There are immense areas of savanna country, covered with coarse grasses which stifle other forms of vegetation and give shelter to hordes of wild animals. These tracts occur as low as within 700 ft. of sea level. Myrtles, bamboo, rhododendrons, moss-loving orchids and palms are common, and orders strongly represented are Dipterocarpaceae, Begoniaceae, Nepenthaceae, Oxalidaceae, Melastomaceae, Myristicaceae, Connaraceae, Gesneriaceae and Ternstroemiaceae. There are zones of vegetation differing with altitude, but the general tendency is for all forms to grow at a lower level than is the case in Java. The Sumatran forest begins at a height of from 300 to 400 ft. and extends to all but the highest mountain peaks, giving a wealth of vegetation that produces very beautiful mountain scenery. The forests contain the finest timber-producing woods in the archipelago; there is not only a great variety of trees, but they possess both stature and beauty. Deciduous trees such as the oak, chestnut, etc., grow on lower slopes than in Java; and apart from ebony, ironwood, camphorwood and sandalwood there are also many species of resin and wild rubber-producing trees.

Animal Life.—Notwithstanding the proximity of the island to Java the fauna of Sumatra shows a greater resemblance to that of Borneo than to that of Java, this being particularly noticeable as regards the fauna of the east coast of Sumatra (that of the west coast and of the islands adjacent is more allied to the Javanese fauna). This striking difference in species between the fauna of Sumatra and Java led A. R. Wallace to assert that the first severance of Sumatra from Java is very ancient; that since the epoch of the first (volcanic) disturbance several distinct elevations and depressions may have taken place, the islands having perhaps been more than once joined with each other or with the mainland (of Asia) and again separated; and that successive waves of immigration may thus have modified the fauna and led to those anomalies in distribution which are so difficult to account for by any single operation of elevation or submergence.

The orangutan, common to Sumatra (in the northeast) and

Borneo, is unknown in Java; the siamang (*Hylobates syndactylus*) is found in Sumatra and in the Malay peninsula but not in Borneo or Java; there are ape species common to Borneo and Sumatra, also the elephant (introduced into Borneo), tapir (indigenous also in the Malay peninsula) and Malayan bear, all of which are unknown in Java, though, on the other hand, Sumatra shares the tiger with Java, while it is unknown in Borneo. The one-horned rhinoceros of Java is known also in Sumatra, which shares a two-horned variety with Borneo. The Sumatran fauna also includes among primates the gibbon *Hylobates lar*, the banded leaf monkey (*Presbytis melalophus*) and the pig-tailed macaque (*Macaca nemestrina*), both common to Borneo also; the silvered leaf monkey (*P. cristatus*) and crab-eating macaque (*M. irus*) common to Borneo and Java; the slow loris (*Nycticebus*); the tree shrews *Tupaia glis* (in Borneo and Java) and *T. minor* (in Borneo only); and the flying lemur, flying fox, bats, flying squirrels, rats and mice; the sambar (*Cervus unicolor*) and the serow (*Capricornis sumatraensis*); wild dog; wild boar; and civet cat.

There is no genus of birds peculiar to Sumatra but there are a few species and many subspecies. The avifauna is very similar to that of the Malay peninsula; in the extreme south a few Javanese forms are found. Among birds found in Sumatra but not in Java are the great Argus pheasant, the fire-back and other pheasants, and various partridges, the little Malay parrot (*Psittinus cyanurus*), the helmeted hornbill (*Rhinoplax vigil*), and the red-breasted bee-eater (*Nyctiornis amicta*). Other Sumatran birds include woodpeckers, pigeons and trogons. Insects, which also show a marked difference from those of Java, comprise the beautiful *Papilio memnon* butterfly, the *Papilio cöon*, and other species of this section, the *Kallima paralekta* (which attaches itself to and is able to imitate the leaf of a certain plant so closely as to secure itself from the attacks of enemies), flocks of *Cyrestis, Melanitis suradeva*, a new species of *Amnosia, Eurhinia fulva*, one of the prettiest species of *Oecophoridae*, two specimens of the curious genus *Homopsyche*, and the moth *Pentacitrotus transversa*. Lizards, large and small, frogs, tortoises and turtles abound, crocodiles frequent the mouths of rivers and are also found upstream; there are many varieties of poisonous snakes; the rocks around the coast yield many kinds of shellfish and the sea, ponds and rivers numerous sorts of edible fishes. (E. E. L.; X.)

History.—Some authorities suggest that Sumatra, not Ceylon, may have been the Taprobane of Pliny. By the end of the 7th century the descendants of immigrants from India had made Palembang the seat of the powerful Hindu kingdom of Sri Vijaya. In the 13th century the Arabs invaded Achin and Mohammedanism took firm hold of some of the most important states. On ancient inscriptions discovered in the Padang highlands, the island was called "the first Java"; Marco Polo called it "Java Minor."

It became known to Europeans as Sumatra through Ludovico di Varthema, an Italian, in 1505. The Portuguese established a trading post about 1509 and were driven out by the Dutch at the end of the century. For the next three centuries the Dutch were engaged in costly efforts to establish supremacy in Achin (Atjeh); a 30 years' war, started in 1873, is estimated to have cost 250,000 lives and £50,000,000 sterling. In 1602, the English first visited Achin, when Sir John Lancaster was well received; and 11 years later, to the annoyance of Dutch and Portuguese, they were given permission to start a factory. The Dutch established their sovereignty over Indrapura in 1664 and over Padang in 1666. In Palembang, where the Dutch had secured a modest concession in 1620, the sultans repeatedly sought to throw off the yoke: twice—in 1654 and 1819—the Dutch were actually driven out; and not till after a serious revolt in 1851 had been suppressed was their tenure secure. In Djambi the efforts of the sultans to escape the consequences of concessions were equally persistent. Throughout the 18th century Dutch and British maintained a constant rivalry in Sumatran waters. Bengkulen, where the English acquired a footing in 1685, was among the sources of irritation to the Dutch and was handed over to them in 1824 in exchange for Malacca. In Deli, in Menangkabau—whose princes claimed descent from Sultan Iskander (Alexander the Great)—and elsewhere, the Dutch established their authority in the teeth of repeated insurgence and generally as the result of the quarrels of native rulers who invoked their help and had to pay the price.

Except for certain regions of the interior (including almost the whole of Achin), Sumatra was occupied by the Japanese in World War II. After the defeat of Japan, the Dutch tried to re-establish themselves, but left Achin alone. Eventually the whole island was merged in the new state of Indonesia (*q.v.*), but the central government's control of vast parts of the interior was for some time only nominal: the Achinese, the Batak and the Menangkabau all demanded autonomy, the Achinese going so far as to support their claims by armed revolt. (C. H.; E. S.; X.)

Population.—The population of Sumatra, which was returned as numbering 7,841,175 at the census of 1930 (when it included 26,354 Europeans and Eurasians and 289,592 foreign Asiatics, mainly Chinese), was estimated at the middle of the 20th century to have risen to 12,000,000. The Indonesian government in the 1950s sponsored the immigration of large numbers of people from Java (10,000 families to South Sumatra and 3,200 to the Djambi area of West Sumatra in 1954).

The whole of the indigenous population had, probably, a common Malayo-Polynesian origin, but today all the races may be classified as Malay, in a wide sense, the pre-Malay blood of the country having been absorbed thoroughly. There is no race of true Negrito type.

Of these Malayan divisions the Achinese (*q.v.*) inhabit the ancient kingdom of Achin in the northwestern promontory and the whole of the north of the Gayo country.

Gayos.—The Gayos live in the northern central portion of the country, between the Achinese and the Batak, with the Alas on their west, living mainly in the valley of the Simpangkiri. They are essentially a highland and insular race and form an ethnographical unit, all speaking one language. Akin to the Batak, they are agricultural and pastoral, a friendly and honest race, living in well-built houses of wood, in small scattered villages. Their chief crop is rice; maize, cotton and sugar cane are grown; horses and cattle are reared; and forest products are collected (there is an export trade to the coast in livestock and forest products). Moslem by religion (with an animistic substratum), they are clever craftsmen in wood carving and plaiting and make weapons. They are patriarchal.

Batak Groups.—The Batak inhabit the mountainous region about Lake Toba and a part of the east coast. They are divided into several groups, differing considerably in language and customs. Of Malay-Polynesian origin, the Batak are nearly one-third Moslem; more than one-third are Christian; the rest animist. Cannibalism (which was largely punitive and ritualistic) and slavery have disappeared. The Batak possess a written language of their own; build compact houses on piles, with horned roofs, many families inhabiting one house; and have rice barns and a communal hall. The houses are beautifully decorated, the Batak being clever craftsmen in wood, ivory and copper (they also understand iron smelting). The Batak are good agriculturists (rice cultivation, wet and dry, being their chief occupation, though maize, coffee, fruit and vegetables are grown) and rear horses, cattle, buffaloes and pigs. They weave and dye cotton garments, make their own jewelry and pottery and can become keen traders and shopkeepers, teachers and even doctors.

Society is patriarchal, marriage exogamic, women are in a subordinate position, and polygamy is practised. There are village councils, a nobility; land belongs to a family, or village, but tenure differs.

Menangkabau (*Minang Kabau*).—South of the Batak, in the Padang highlands and vicinity, are the Menangkabau. At one time their ancient and powerful kingdom (which legend says rose on the ruins of a Hindu empire) covered the greater part of West Sumatra. They sent out emigrants to the Malay peninsula and represented the highest form of Malayan civilization.

Although converted to Islam, the Menangkabau retained their ancient custom of the matriarchate (*i.e.*, descent and inheritance in the female line), which is said to be preserved nowhere in a purer form than among them: a wife remains after marriage with her kinsfolk on the mother's side, and the husband, having no home of

his own, only visits his wife and resides in his mother's house; names, privileges and property derive from the mother's side; the eldest man of the elder female line is termed *mamak* and is the keeper of all the possessions of the family. Land is unalienable and always comes back into the possession of the family or village. Marriage between members of the same tribe is not allowed. Houses, raised from the ground, are very ornate, with horned ridge poles and finely carved wooden fronts, sometimes painted. Several families bearing the same patronymic live in the same house; and several houses, their rice barns, a communal house, a mosque, school, an inn and market form a village. There are priests and nobles; the people live by trade, agriculture, hunting and fishing. Their language is one of the chief branches of Malay (with a few words of Sanskrit origin); the Javanese script was once used but has been superseded by the Arabic.

Other Peoples.—South of the Menangkabau, in the Indrapura region, live the Kerintjis, a small community, partly Moslem, partly animist; and east of these, in Djambi, Malays, using the term as denoting a seafaring race (Moslems), speaking Malay and having the usual Malay characteristics, not fond of work, prone to finery, devoted to sport and gaming, friendly and intelligent and living in tribes.

In the Redjang-Lebong area there are the Redjangs, a rather truculent people, who work in the gold and silver mines of their country, are Moslems, possess a script of Indian origin and have Hindu antiquities scattered among them.

In the south are the Lampongs, who were largely under Hindu influence and attained a high degree of civilization which they have retained. Pepper culture is the chief occupation. Even before World War II, Javanese colonies were developed there. Among the primitive tribes, there are the nomadic Kubus of the Djambi mountains, almost dwarfs, with frizzled hair denoting Negritic blood. All are extremely shy and isolated peoples, but quite peaceable, nomadic or seminomadic, primitively clothed, living by fishing and hunting and rude agriculture. (For "Island peoples" see NIAS; MENTAWEI; BATU.) Enggano people are animistic Malays, fishers and agriculturists. (E. E. L.; X.)

Provinces and Towns.—Sumatra is divided administratively in six provinces: Achin (or Atjeh), North Sumatra, West Sumatra, Riouw, Djambi and South Sumatra. At the head of each province is a governor appointed by the central Indonesian government. The governor is assisted by a provincial council.

The chief town of Achin is Kutaradja with Uleelheue (at the northwestern extremity of the island) as its principal port. Along the west coast of Achin stand Meulaboh, Tapaktuan and Sibolga. Towns on the north coast are Sigli, Bireuen and, at the other side of Diamond point, Lhokseumawe and Idi. The major towns of the interior of Achin are Takingeun, Blangkedjeren and Kutatjane. The islands of Simuloeë and Bandjak are included in the province of Achin.

The chief town of North Sumatra is Medan (pop., 1951 est., 250,000), the largest city of the island; it is served by the port of Belawan. Other towns on the eastern coast are Langsa to the northwest and Tandjungbalai to the southeast, at the mouth of the Asahan, and Bagan-Siapiapi on the estuary of the Rokan. The only North Sumatran town of importance on the western coast is Airbangis. Nias and Batu islands are included in the province. Major towns of the interior are Rantauparapat and Padangsidimpuan.

West Sumatran chief town is Bukittinggi (formerly Fort de Kock) between the lakes of Manindjau and Singkarak, near the magnificent Buffalo canyon (with walls 200–300 ft. high) in the Padang highlands behind the great port of Padang, on the west coast. Other towns on this coast are Pariaman to the north and Balaiselasa to the south. In the hinterland are Pajakumbuh, Solok and Sawahlunto. The Mentawai Islands off the western coast are also part of West Sumatra.

The chief town of the Riouw province is Tandjungpinang on the Bintan Island, the largest of the Riouw archipelago. The last-named, as well as the Lingga archipelago, belong to the Riouw province, which includes the whole Inderagiri and Siak basins. The inland port of Pakanbaru (100 mi. up the river Siak) and Rengat on the Inderagiri are the major towns of the hinterland.

The city of Djambi, on the Hari river, is the chief town of the province of the same name. The wide low levels between the Bukit Barisan mountains and the eastern coast have no towns of importance.

The province of South Sumatra has its capital at Palembang on the Musi (*see* PALEMBANG). Its other great ports are Bengkulen on the western coast and Telukbetung on the Sunda strait in the south; there are also the smaller ports of Bintuhan and Kru south of Bengkulen. Inland, Baturadja stands northwest of Telukbetung and southwest of Palembang; and Lahat lies at the foot of the mountains, on the opposite side from Bengkulen. The islands of Bangka and Billiton (Belitung) are included in this province.

Communications.—In Achin and North Sumatra there is railway connection between Kutaradja and Belawan via Medan; this line branches into the hinterland of Medan and is continued southeastward from Medan to Rantauprapat, with a westerly branch to Pematangsiantar and an easterly one to Tandjungbalai. In West Sumatra, Padang is connected by rail on the one hand with Sungailimau and on the other with Padangpandjang, from which lines run via Bukittinggi to Pajakumbuh and via Solok to the mining area of Sawahlunto. In South Sumatra the railway from Palembang has two lines, one running southward via Baturadja to Telukbetung, the other westward to Muaraenim and thence northward to Lubuklinggau. Communications from east to west across the Bukit Barisan are of necessity by road. The importance of rivers as a means of communication, particularly for Riouw, Djambi and South Sumatra, has already been explained.

Internal air communications were in process of development by the middle of the 20th century, when airports had been established at Medan, Padang, Pakanbaru, Tandjungpinang, Rengat, Muntok, Pangkalpinang, Palembang, Branti, Bengkulen and Sibolga.

Agriculture.—Volcanic effusion has enriched the soil over large areas of the island; and irrigation, sponsored by the government, brings water with a high mineral content to the less fortunate regions where its fertilizing properties are required. Agriculture in Sumatra is of two types; indigenous, in the hands of smallholders; and European, on large estates.

The smallholders' main crop is rice, which is the staple food of the island and is grown entirely for home consumption. Even so, large quantities of rice have to be imported, as Sumatra's own crop is not sufficient to feed the population. Both wet (*sawah*) and dry (*ladang*) methods of cultivation are practised. In the hope of increasing the yield, the Indonesian government established a pool of equipment from which smallholders could borrow tractors and other sorts of machinery. Maize is also grown by smallholders (by the Batak as a primary crop) for home consumption. Other smallholders' crops are copra, spices (pepper, cloves and nutmegs), coffee, tobacco, betel nuts, peanuts, coconuts, kapok, cotton and gambier. Ebony, sandalwood and ironwood are cut in the forests, which also yield wild rubber, rattan and resins.

European agriculture on large estates is concerned principally with rubber, tobacco, tea, palm oil and hard fibres (sisal and ramie); and to a lesser extent with cinchona bark (for quinine), coffee, gambier and coca.

Rubber is grown in all six provinces of the island, but the greater part of the yield comes from the large estates: of Indonesia's 719 rubber estates, Sumatra has 242. These large estates, planted with *Hevea* trees and concentrated predominantly in North Sumatra (between Lake Toba and the east coast), produced in 1956 about 445,000 metric tons of rubber (60% of the country's total). The most important tobacco-growing area is also in North Sumatra, around Medan and Belawan (the leaf is fine and silky-looking and is particularly suitable for the outer wrapping of cigars). The output of tobacco amounted to 8,700 metric tons in 1952. The intensive cultivation of trees for palm oil, again mainly in North Sumatra, began in the 20th century; by the 1950s production in a normal year was about 240,000 metric tons, but adverse circumstances might reduce it considerably (*e.g.*, to 145,565 metric tons of palm oil, with 38,508 of palm kernels, in 1952). Tea is grown throughout the island, but again most extensively on the estates of North Sumatra; the crop suffered gravely from the

onset of blister blight in 1949, which necessitated the widespread use of fungicides; the yield in 1952, amounting to 9,123 metric tons, was less than half of what it had been in 1939. Hard fibres constitute the remaining major product of the North Sumatran estates; because, however, of the difficulties arising from the long growth-cycle of the *Agave* plant (from 8 to 10 years), the cultivation of sisal was to some extent being superseded, in the 1950s, by that of ramie, which the government encouraged; total production of hard fibres amounted to 20,230 metric tons in 1952. Most of the island's cinchona bark comes from the west coast of West Sumatra (estate production 199 metric tons in 1952). Coffee is grown by smallholders in the Lampung area of South Sumatra and also along the west coast as far north as the Tapanuli region. Most of the island's pepper crop comes from the Lampung area (around Telukbetung) and from the neighbourhood of Palembang, lesser quantities from Achin (in these areas the average yield of a single plant is 0.3–0.8 kg. of black pepper); on the island of Bangka a more intensive method of cultivation is in use.

The minor smallholders' crops are found mainly along the west coast of the island: betel nuts in North Sumatra; copra and nutmegs in North and West Sumatra; resins in West Sumatra; coconuts in West and South Sumatra; and cloves in South Sumatra. Copra is also grown by smallholders on the east coast in the Inderagiri and Hari basins.

Fisheries.—The great fishing port of Sumatra is Bagan-Siapiapi, which is the largest in Indonesia outside Java; by the 1950s it was handling an annual catch of about 80,000 tons. There are also fisheries at Sungsang (on the coast north of Palembang), at Lampung (near Telukbetung), at Bengkulen and on Bangka Island. Dried fish are exported; shellfish, trepangs (sea-slugs) and prawns (from which the delicacy known as *krupuk* is made) are also caught.

Minerals.—The major mineral resource of Sumatra proper is petroleum, found predominantly around Palembang and in the Djambi region but occurring also in North Sumatra; it is refined at Palembang. Sumatra produced in 1956 about 10,100,000 tons of crude petroleum (80% of the Indonesian total). On the islands of Bangka, Billiton and Singkep, however, there are rich deposits of tin, which by the 1950s had made Indonesia the second largest tin producer in the world; output amounted to 30,613 long tons in 1948 and to 35,003 in 1952 (65% of this total coming from Bangka, where the mines are government controlled). Bauxite is mined on the islands of the Riouw archipelago opposite Singapore and on the mainland in the Asahan valley inland from Tandjungbalai.

The principal coal fields are at Bukitasam (southwest of Muaraenim on the railway inland from Palembang) and on the Umbilin river near Sawahlunto (in the Padang highlands): by 1952 the average annual production of coal amounted to 600,000 tons from Bukitasam and to 78,000 from the Umbilin mines (the deposits of coal around Bengkulen were hardly exploited). There are also deposits of gold, silver, lead, copper, iron, antimony, cobalt, naphtha, sulfur, alum and saltpetre in the mountains.

Exports.—The accompanying table indicates the export trade of the principal Sumatran ports.

Exports, 1952
(Quantity in metric tons)

Product	Port of exit		
	Belawan	Padang	Palembang
Rubber: estate	157,000	...	...
smallholders	9,834	16,671	87,413
Tea	4,884	1,468	180
Coffee*	212	690	6,594
Cinchona bark*	...	625	...
Palm oil	120,035	...	...
Palm kernels	36,243	...	310
Hard fibres	18,893	...	...
Tobacco	2,980	...	...
Petroleum and petroleum products	....	...	5,537,629

*Mainly from smallholdings. *Cf.* estate production (1952): coffee 96 metric tons, cinchona bark 199.

BIBLIOGRAPHY.—For general information on the geography, history and economy of Sumatra consult the books cited in the bibliographies to the articles MALAY ARCHIPELAGO; INDONESIA. *See* also: O. J. A. Collet, *Terres et peuples de Sumatra* (Amsterdam, 1925); and E. M. Loeb, *Sumatra: Its History and Peoples* (Vienna, 1935). W. Marsden, *History of Sumatra*, 3rd ed. (London, 1811), is still interesting. (X.)

SUMBA (Dutch SOEMBA), also known as Sandalwood, one of the Lesser Sunda Islands, Indonesia; area 4,306 sq.mi. The island is dominated by heights of from 2,200 to over 3,000 ft. Several bays on the north coast give good anchorage, the best being that of Waingapu. The rivers, mostly unnavigable, are of importance for agriculture. A fertile soil and a fairly good rainfall (average 63.7 in. annually) help to produce good forests, grazing ground and food crops. The forests yield sandalwood (giving rise to the old name for the island), dye-woods, wax and wild cinnamon. On the grazing grounds are bred some of the finest horses in the country—the sandalwood breed—and fine cattle (Ongole), both of which figured largely as exports, while rice is raised on *sawahs* and *ladangs*, and maize, coffee, tobacco, coconuts, vegetables and fruit are grown, copra being exported. The estimated pop. is 182,000, consisting, apart from the indigenous Sumbanese, of people from Flores, Savu and Rotti. The Sumbanese are a Malayo-Papuan people of good physique. Except in the interior, where clothing is far more primitive, the *sarong* and *baju* are generally worn by women, and sometimes by men. Both sexes are fond of ornaments. They cultivate the soil, fish, mostly in rivers, not being a sea-loving race, collect forest products, edible birds' nests and turtles, make fishing-nets and plait baskets. There are workers in copper and iron, while the women weave, spin and dye, and make pottery. The natives are largely pagan (there are some Mohammedans) and make offerings on stone altars. A particular breed of horses is held sacred.

The Sumbanese enjoy feasts, music and dancing. Traditional customs include marriage by dowry and polygamy, fairly common among the upper classes. There is no written form of the Sumbanese language, which is allied to that of Savu. Sumba has numerous megalithic monuments.

The sandalwood of Sumba attracted attention to the island in the 17th century, when Sumba appears to have been a tributary of the state of Bima, in Sumbawa. Later Sumba became independent of Sumbawa, and treaties were concluded in 1756 with the Dutch. Between 1856 and 1874, new treaties were made, but there were frequent difficulties with the natives, due to piracy and slave raids. The trouble continued and in 1901 the Dutch were compelled to land troops. Most of the chiefs of the tribes were bound by the "short declaration," but there were disturbances in Sumba (Lamboja), necessitating a Dutch armed force in 1914.

Japanese forces occupied the island during World War II.
(E. E. L.; X.)

SUMBAWA (Dutch SOEMBAWA), one of the Lesser Sunda Islands, Indonesia, between Lombok and Flores, and separated from the former by Alas strait and from the latter by Sapeh strait and the Island of Komodo. Sumbawa with neighbouring islands, has an area of 5,965 sq.mi. From 30 to 100 mi. north are the Tenga and Sabalana (Paternoster) Islands, small low banks, surrounded by reefs. Mountain chains traverse the centre of each of the peninsulas. There are only a few stretches of alluvial land along the coast: off the southern shore great depths of water are soon reached in the Indian ocean. The mountains, some volcanic, are highest in the north. Mt. Tamboro, 9,255 ft. high, was very severely damaged in the eruption of 1815. No other volcanoes appeared to be active in modern times. The average annual rainfall (Bima) in 93 in.

Agriculture is carried on, rice being raised on *sawahs* and *ladangs*, while maize, cotton, coffee, tobacco, coconuts, onions and other vegetables, and fruit are grown. Horse and cattle breeding are important industries. The forests yield dye-woods (chiefly sapanwood) and teak.

The pop. (1930) of Sumbawa was 314,843. The natives of the western peninsula (Sumbawa) are of Malayan stock and allied to the Sassaks of Lombok, and those of the eastern peninsula (Bima, Sanggar and Dompo), are Malayo-Papuan.

Vessels of the Royal Packet Navigation company ordinarily called at Bima, Sumbawa, and Taliwang, affording communication with Macassar, the other Sunda Islands and Java ports. Sumbawa consisted originally of six states, all owing allegiance to either

Gowa or Macassar in Celebes—Bima, Dompo, Sanggar, Tambora, Papikat and Sumbawa; after the eruption in Tambora in 1815 that state and Papikat ceased to exist. In 1701 the Dutch arbitrated between Sumbawa and other states in the island engaged in civil war, and in 1765 the Dutch made a separate agreement with Sumbawa, revised in 1875 and again in 1905, under which it has a certain measure of self-government. A contract made with Bima in 1857 and again in 1886, on the lines of that made with Sumbawa, was revised in 1905. It allowed the state some self-governing powers. Sanggar was subject to the sultan of Ternate originally, then to the sultan of Macassar, and after the Bongay contract of 1667 it became a fief of the Dutch. A treaty made in 1858, and revised in 1905, defined its right of self-government. Dompo gave up allegiance to Macassar after the Bongay contract, was a party to the Sumbawa treaty of 1765 with the Dutch, and had its self-governing powers fixed by subsequent treaties. Sumbawa was occupied by Japan during World War II.

(E. E. L.; X.)

SUMBUL or SUMBAL, also called Musk Root, a drug occasionally employed in European medical practice. It consists of the root of *Ferula sumbul,* Hook., a tall Umbelliferous plant found in the north of Bokhara, its range apparently extending beyond the Amur. The action and uses of the drug are the same as those of asafetida (*q.v.*).

SUMERIAN LANGUAGE. Sumerian is the oldest written language in existence. First attested around 3100 B.C. in southern Mesopotamia, in an area not much larger than a county, it flourished in the course of the third millennium B.C. It was replaced around 2000 B.C. as a spoken language by the Semitic Akkadian (Assyro-Babylonian), but it continued in written usage almost to the end of the life of the Akkadian language (*q.v.*), around the beginning of the Christian era. Sumerian never extended much beyond its original boundaries in southern Mesopotamia; the small number of its native speakers was entirely out of proportion to the tremendous importance and influence Sumerian exercised on the development of the Mesopotamian and other ancient civilizations in all their stages.

Four periods of Sumerian can be distinguished: Archaic Sumerian, Old or Classical Sumerian, New Sumerian and Post-Sumerian.

Archaic Sumerian embraced a period from the time around 3100 B.C., when the first Sumerian records make their appearance, down to about 2500 B.C. Owing its origin to the needs arising from public economy and administration, the earliest Sumerian writing is almost exclusively represented by texts of business and administrative character. The other class of texts includes school texts in the form of simple exercises in writing signs and words. The Archaic Sumerian language is still very poorly understood, partly because of the difficulties surrounding the reading and interpretation of the early Sumerian writing and partly because of the meagreness of sources.

The Old or Classical period of Sumerian lasted from about 2500 to 2300 B.C. This period is represented mainly by records of the early rulers of Lagash. The use of the Sumerian language is attested not only in business, legal and administrative texts, but also in royal and private inscriptions, mostly of votive character, letters, both private and official, and incantations. The sources of this period are much more numerous than those of the preceding period, and the writing is explicit enough to make possible an adequate reconstruction of Sumerian grammar and lexicon.

During the period of the Sargonic dynasty, the Semitic Akkadians took over the political hegemony of Babylonia, marking a definite setback in the progress of the Sumerian language. During this period, the Akkadian language was used extensively throughout the whole area of the Old Akkadian empire, while the use of Sumerian was gradually limited to a small area in Sumer proper. Toward the end of the dynasty the Sumerian elements came again to the fore, marking in turn a setback to the progress of Akkadian. The New Sumerian period came to an end around 2000 B.C., when new inroads of the Semitic peoples from the desert succeeded in destroying the supremacy of the 3rd dynasty of Ur and in establishing the Semitic dynasties of Isin, Larsa and Babylon. The New Sumerian period is characterized by an extraordinary number of administrative texts.

The period of the dynasties of Isin, Larsa and Babylon is called the Old Babylonian period, after Babylon, which became the capital and the most important city in the country. During this period the Sumerians entirely lost their political entity, and the Sumerian language gradually disappeared from spoken use. However, it continued in written usage to the very end of the use of cuneiform writing. This is the last stage of the Sumerian language, called "Post-Sumerian."

In the early stages of the post-Sumerian period the use of written Sumerian is extensively attested in legal and administrative texts, as well as in the royal inscriptions. The latter are often bilingual, that is, in Sumerian and Babylonian. Many Sumerian literary compositions, which came down from the older Sumerian periods by way of oral tradition, were put down in writing for the first time in the Old Babylonian period. Many more were copied by industrious scribes from originals now lost. The rich Sumerian literature is represented by texts of varied nature, such as myths and epics, hymns and lamentations, rituals and incantations, and proverbs and the so-called "wisdom" compositions. For many centuries after the Old Babylonian period, the study of Sumerian continued in the Babylonian schools. As late as the 7th century, B.C., Assurbanipal, one of the last rulers of Assyria, boasted of being able to read the difficult Sumerian language, and from an even later period, in the Hellenistic period, there are some cuneiform tablets which show Sumerian words transcribed with the letters of Greek writing.

At the time of Christ, all knowledge of Sumerian disappeared along with cuneiform writing, and in the succeeding centuries even the name Sumer vanished from memory.

Unlike Assyria, Babylonia and Egypt, the histories and traditions of which are amply documented in the biblical and classical sources, there was nothing to be found in non-Mesopotamian sources anywhere to make one even suspect the existence of the Sumerians in antiquity, let alone to appreciate fully the important role they played in the history of early civilizations.

When the decipherment of cuneiform writing was achieved in the early decades of the 19th century, three languages written in cuneiform were discovered: Semitic Babylonian, Indo-European Persian and Elamite, of unknown linguistic affiliation. Only after the texts written in Babylonian had become better understood did scholars become aware of the existence of texts written in a language different from Babylonian. When the new language was discovered it was variously designated as Scythian or even Akkadian, that is, by the very name now given to the Semitic language spoken in Babylonia and Assyria. As knowledge of the new language grew, it was given the correct name of Sumerian.

The linguistic affinity of Sumerian has not yet been successfully established. Ural-Altaic (which includes Turkish), Dravidian, Brahui, Bantu and many other families of languages have been brought into comparison with Sumerian, but no theory has gained common acceptance. Sumerian is clearly an agglutinative language in that it preserves the root intact while it expresses the various morphemic changes, through the concatenation principle, by means of prefixed, infixed and suffixed elements. The difference between nouns and verbs, as it exists, *e.g.*, in the Indo-European or Semitic languages, is unknown to Sumerian. The word *dug* alone means both "speech" and "to speak" in Sumerian, the difference between the noun and the verb being indicated by syntax and different affixes.

The phonemic system of Sumerian consisted of four vowels *a, i, e, u,* and sixteen consonants, *b, d, g, h, k, l, m, n, p, r, s, ś, š, t, z.* In the noun, gender was not expressed. Plural number was indicated either by the suffixes *-me* (or *-me + esh*), *-hia,* and *-ene* or by reduplication, as in *kur + kur,* "mountains." The relational morphemes of the noun, corresponding approximately to the cases of the Latin declension, are: *-e* for the subject (nominative), *-a(k)* "of" (genitive), *-ra* "to," "for" (dative), *-a* "in," *-ta* "from" (ablative), *-da* "with" (commitative), and *-sh(e)* "for (a period of time)," "for (the purpose of)," etc.

The Sumerian verb, with its concatenation of various prefixes,

infixes and suffixes, presents a very complicated picture. The elements connected with the verb follow a rigid order: modal elements, tempo elements, relational elements, causative elements, object elements, verbal root, subject elements and intransitive present-future elements. In the preterit transitive active the order of object and subject elements is reversed. The verb can distinguish, in addition to person and number, transitive and intransitive, active and passive, and two tenses, present-future and preterit.

Several Sumerian dialects are known. Of these the most important are *eme-KU*, the official dialect of Sumerian, and *eme-SAL*, the dialect used often in the composition of hymns and incantations. (I. J. G.)

SUMER IS ICUMEN IN ("Summer is come in,"), famous 13th century English round which occupies a unique place in the history of music as standing alone in respect of the advanced character of its writing, as well as of its musical beauty, among the compositions of its period. As Father Anselm Hughes has put the matter in Grove's *Dictionary of Music*, this astonishing composition is pre-eminent in six respects, for (1) it is the oldest known canon; (2) it is the oldest known harmonized music which is frequently performed and enjoyed by singers and listeners today; (3) it is the oldest known 6-part composition; (4) it is one of the oldest known specimens of the use of what is now the major mode; (5) it is the oldest known specimen of groundbass; (6) it is the oldest known manuscript in which both secular and sacred words are written to the music. The belief is that it was composed about 1240, and it is ascribed to John of Fornsete, who wrote the manuscript, which came from Reading abbey. *See* article Music: *The Rise of Music in the Western World.*

SUMMANUS, a Roman god whose origin and nature were obscure even to the ancients. Varro considered him a Sabine deity, and there has been some modern inclination to regard him as Etruscan. His Latin name is an adjective ("highest") which suggests that he began as an epithet of Jupiter. Later he develops into a Jupiter of the night, having jurisdiction over the nocturnal sky and thunderbolt. He had a temple near the Circus Maximus built during the war with Pyrrhus (*c.* 278 B.C.). An annual festival was held on the dedication day of the temple (June 20), one of the features of which was the eating of sacrificial cakes baked in the shape of a wheel (*summanalia*). In late times Summanus is erroneously identified with Pluto.

See G. Wissowa, *Religion und Kultus*, 2nd ed., p. 135 (1912); W. W. Fowler, *Roman Festivals*, p. 160 ff. (1899); Platner-Ashby, *Topographical Dictionary*, s.v. (R. B. Ld.)

SUMMARY JURISDICTION. (1) **Meaning:** The phrase is now understood in English law as referring to jurisdiction, exercised by justices of the peace, which may result in a summary conviction or order. Blackstone divides the proceedings in courts of criminal jurisdiction into two kinds: summary and regular: and by regular he means such as involve trial by jury, the only mode of trial known to the Common law except in the case of contempt of court. In these regular proceedings there is only a preliminary hearing before justices, followed by either a dismissal or committal for trial by jury at Quarter Sessions or Assizes.

(2) **History.**—Being unknown to the Common law, this jurisdiction is entirely the creation of statutes. A brief sketch is to be found in Stephen's *History of the Criminal Law*, vol. 1. ch. 4, of the growth of the criminal jurisdiction. Speaking generally this growth has shown itself in two distinct forms: (a) The conferring of exclusive jurisdiction in a vast number of minor offences, mainly involving the infliction of pecuniary penalties, under such headings as Army, Game, Highway, Licensing, Motor Cars, Merchant Shipping, Post Office, Public Health, Revenue and Vagrancy. These offences have been created and are being created by Acts of Parliament or by by-laws of county councils or other local authorities made under statutory authority, or by orders of Departments. (b) The providing in indictable cases of a mode of trial, alternative to trial by jury and dependent for its adoption on the consent of the tribunal, of the accused and in some instances of the prosecution. These cases are referred to as indictable cases triable summarily.

In addition to summary criminal jurisdiction, justices have been endowed by many statutes with a summary civil jurisdiction of increasing scope and importance.

(3) **General Legislation.**—A court of summary jurisdiction is defined by the Interpretation Act 1889 as "Any justice or justices of the peace or other magistrate by whatever name called, to whom jurisdiction is given by or who is authorised to act under the Summary Jurisdiction Acts and whether acting under the S. J. Acts or any of them, or under any other Act or by virtue of his commission or under the Common Law." This does not apply to justices sitting as the licensing authority nor to the preliminary hearing by justices of indictable offences.

In the Interpretation Act "The Summary Jurisdiction Acts" means the S. J. Act 1848 and the S. J. Act 1879 and any Act past or future amending those Acts or either of them: thereby including amongst others, the Criminal Justice Administration Act 1914 and the Criminal Justice Act 1925. (*See* S.J. Act 1848 s. 35; S.J. Act 1879 s. 54 and C.J.A. Act 1914 s. 32 [i.].) These Acts prescribe the constitution of the courts and the procedure to be followed up to, during and following the hearing, in such civil cases as are not otherwise specially provided for, in purely summary criminal cases and in indictable cases triable summarily, and further specify what indictable offences are triable summarily, and the maximum punishment which can be inflicted on conviction. Rules made under these Acts of 1915 and 1926 deal with such matters as the register, accounts, payments and bail.

(4) **Constitution of the Courts.**—A court consisting of not more than one justice has a jurisdiction limited to receiving an information or complaint, issuing a summons or warrant (S.J. Act 1848 s. 1), adjourning the hearing to the next practicable sitting of a Petty Sessional Court, and where authorised by the statute creating the offence, trying the case and imposing a punishment of not more than 14 days imprisonment or a fine of not more than 20/–(S.J. Act 1879 s. 20).

A Petty Sessional Court is defined by the Interpretation Act 1889 as "A court of summary jurisdiction consisting of two or more justices when sitting in a Petty Sessional Court, and shall include the Lord Mayor of the City of London and any Alderman of that city, and any metropolitan or borough police magistrate or other stipendiary magistrate, sitting in a court-house or place at which he is authorised by law to do alone any act authorised to be done by more than one Justice of the Peace." It is by this court that summary jurisdiction must ordinarily be exercised (S.J. Act 1879 s. 20 [9 and 10]).

The position in London (excluding the city) is peculiar. Under the Metropolitan Police Courts Act 1839, a number of stipendiary magistrates (commonly known as police magistrates) are appointed and sit at 14 police courts. The jurisdiction of the justices for the county of London was not expressly taken away, and in theory they possess a concurrent jurisdiction, but inasmuch as by S. 42 of the Act of 1839 court fees in respect of proceedings elsewhere than in the police courts were prohibited, a practical obstacle was interposed which has prevented the jurisdiction being exercised except to a very limited extent. It is not suggested that thereby any public mischief has ensued. Outside London, some 17 stipendiary magistrates are appointed under various statutes for various cities and boroughs.

Offences under the Conspiracy and Protection of Property Act 1875 are to be prosecuted before, and jurisdiction under the Law of Distress Amendment Act 1908 is to be exercised by, a stipendiary magistrate where there is one.

(5) **Subject Matter of Criminal Jurisdiction.**—(a) *Summary Cases:* It has been already stated that *inclusive* summary jurisdiction has been conferred in numerous criminal cases both by or under special statutory authority: this statement requires qualification.

(i.) When the offence is punishable by imprisonment for more than three months the person charged may claim to be tried by jury: to this there are at least two exceptions: a male person who knowingly lives wholly or in part on the earnings of prostitution or who in a public place persistently solicits or importunes for immoral purposes. Punishment: 6 months' hard labour, with-

out right to trial by jury (Crim. Law Amend. Act 1912 S. 7 [2]).

(ii.) A like claim to the trial by jury may be asserted under the Explosives Act 1875 S. 92, and the Conspiracy and Protection of Property Act 1875 S. 9.

(iii.) Some offences are triable either summarily or by indictment, *e.g.*, assault, some forms of public indecency, cruelty to children and the offences mentioned above under C.L.A. Act 1912.

(b) *Indictable Cases Triable Summarily:* Where an adult is charged with one of the offences specified in the second schedule to the Criminal Justice Act 1925 and the court thinks it is expedient and the accused consents, the court may try the case summarily, unless the case affects the property or affairs of His Majesty or of a public body, or unless the Public Prosecutor is prosecuting, in which cases the consent of the prosecutor is also necessary. (C.J. Act. 1925 S. 24.)

This statute greatly increases the number of indictable offences triable summarily, and confers wider, though more elastic, powers of punishment than any previous statute. The power to fine, which is particularly appropriate in some cases of dishonesty is not possessed by courts of Quarter Sessions or Assize when dealing with felonies.

(6) **Criminal Procedure.**—(a) *Procuring the Attendance of the Accused:* Where the offence is originally indictable, the accused is brought before the court either on arrest without warrant or upon warrant or summons under the Indictable Offences Act 1848, and the charge cannot proceed in his absence.

In summary cases, unless there has been an arrest by a constable under statutory powers, information is laid before a justice as to the offence, whereupon either a summons or a warrant may be issued. For a summons the information may be oral and unsworn; for a warrant it must be in writing and upon oath. The information must be laid within six months from the commission of the offence unless a different time is limited by some particular statute (S.J. Act 1848 S. 11) or unless the offence is a continuing one. A warrant is not ordinarily granted in the first instance, except for a serious offence; but if the summons is disobeyed, a warrant may be issued upon proof of service. The practice on this matter varies in different courts and according to circumstances. Personal appearance is not in general necessary. The summons is served by a constable upon the accused either personally or by leaving it at his last or most usual place of abode, and should take place a reasonable time previous to the hearing.

(b) *The Hearing:* The hearing must take place in open court, and parties may appear by counsel or solicitor. If the accused does not appear and the court determines to proceed, the case must be proved in the ordinary way as if he had appeared. If the informant does not appear the case is dismissed, unless the court thinks proper to adjourn the hearing and to remand the accused either in custody (which would be very unusual) or on recognizance with or without surety. If the accused appears, he is asked to plead; if he pleads guilty, the court proceeds to judgment.

A police officer of whatever rank has no right *ex officio*, and should not be permitted to act as advocate or to take any part in the proceedings at all; that is, unless he is the actual informant: even in the latter event, whatever his actual right under the S.J. Act 1848, his activities should be confined, so far as possible, to those of a witness. The attendance of a witness may be enforced either by a Crown office *subpoena* or by a summons or warrant of a justice.

If there is a variance between the evidence and the information, and the party charged is prejudiced thereby, an adjournment on terms should be granted. At the close of the evidence for the prosecution, if the court is of opinion that there is a case to answer, the defendant may open his case and call his witnesses, subject to this: that if the defendant be the only witness on his side as to fact, he must be called immediately after the close of the evidence for the prosecution.

(c) *The Judgment:* The judgment follows the opinion of the majority of justices present at the hearing, and the chairman has no second or casting-vote. In the case of an equality of votes, unless a justice withdraws his vote, the case should be adjourned for a rehearing before another court, when, if the same result ensues, the case should be dismissed. If the court decides to dismiss the information either upon the non-appearance of the informant or as being of opinion that the charge is not proved, it makes an order to that effect and may award costs to the defendant: but costs do not follow the event so frequently as in civil cases. If the charge is without foundation and the defendant is in no way to blame, there seems no reason why he should bear any part of the costs of the defence, whether the prosecution is by the police or by a private person, but a variety of considerations, generally unstated, influence the courts in this matter, and it cannot be pretended that there is any general or settled practice. If the defendant has pleaded guilty or the court is of opinion that the charge is proved, the court may either:

(i.) Postpone judgment (either for consideration or to avoid inflicting sentence of imprisonment) and remand the defendant either in custody up to 8 days, in an appropriate case, or for that or a longer period, on recognizance, with or without sureties, or

(ii.) Convict the defendant, and commit him to the Assizes or Quarter Sessions in order that a sentence of detention in a Borstal institution may be passed, or

(iii.) Convict and sentence the defendant. The maximum term of imprisonment and the maximum fine which can be inflicted are in general fixed by the statute or bye law creating the offence: an order for costs can be included in the conviction. Whatever the provisions of the statute, or bye law, the imprisonment may be either with or without hard labour, and the court may reduce the term as it thinks fit or substitute for imprisonment (where no power to fine existed) a fine not exceeding £25 and not involving, on default, a longer term of imprisonment than that prescribed by the Act or bye law: and where a fine is prescribed, the court may, if it is imposed as in respect of a first offence, reduce the prescribed amount, or

(iv.) Without proceeding to conviction, where the court is of opinion that having regard to a variety of circumstances (*see* Probation of Offenders Act 1907 S. 1) make an order dismissing the information or charge, or discharge the defendant conditionally on his entering into a recognizance, with or without sureties, to be of good behaviour and to appear for conviction and sentence at any time within not more than 3 years.

(d) *Execution and Enforcement of Payment of Fines and Costs:* Fines and costs, ascertained by conviction, are enforced by warrant of distress and, in default of distress, imprisonment on a scale set out in S. 5 of the S.J. Act 1879. The court may issue a warrant of commitment in the first instance if the defendant has sufficient means to pay or if he does not ask for time to pay or has no fixed estate within the jurisdiction or for other special reasons. (*See* further C.J. Act 1914 S. 25.) Where no fine has been inflicted but costs have been ordered to be paid either by the defendant or, upon dismissal, by the informant, recovery is by distress under S.J. Act 1848 S. 18 and C.J. Act 1914 S. 25. There are exempt from distress: the wearing apparel of the defendant and his family, and, to the extent of £5, the tools and implements of his trade. Provision is made in C.J. and S.J. Acts and rules for payment by instalments.

(e) *Appeal:* (i.) *To Quarter Sessions:* There is a general right of appeal against a conviction (including an order under S. 1 of the Probation of Offenders Act 1907) where the person aggrieved did not plead guilty (C.J. Act 1914 S. 37 [1]: C.J. [Amendment] Act 1926 S. 1), and against sentence after a plea of guilty (C.J. Act 1928 S. 25). There is no right of appeal against a dismissal, unless specially given by some statute. The procedure is regulated by the S.J. Act 1879 to which reference should be made to master the technicalities which are numerous.

(ii.) *To the High Court by a Case Stated on a Point of Law:* Such an application may be made by either party under the S.J. Acts 1857 and 1879. A recognizance, similar to that required on appeal to Quarter Sessions, is necessary. Generally speaking, the parties agree upon the form of the case: failing agreement the court settles the case. The court may refuse to state a case but only on the ground of frivolity: the court

of King's Bench may rescind this refusal or order a case to be stated.

(iii.) The court of King's Bench has further powers, on certain limited grounds, to restrain the exercise of or control a refusal to exercise jurisdiction by writ of *prohibition* and *mandamus*.

(7) **Civil Procedure.**—(a) As before stated, justices have been authorised by many statutes to make orders for the payment of money or otherwise: and some of the statutes contain special provisions as to the procedure to be observed. Mention is made of those most in use. Bastardy is specially dealt with under the Bastardy Acts 1872 and 1923 and the Affiliation Orders Act 1918.

Disputes between employers and workmen or apprentices are provided for by the Employers and Workmen Act 1875.

A special code, applicable to certain differences between husbands and wives, is contained in the S.J. (Married Women) Act 1895 as amended by the Married Women (Maintenance) Act 1920 and the S.J. (Separation and Maintenance) Act 1925 and S. 5 of the Licensing Act 1902.

A limited jurisdiction in ejectment is conferred on justices by the Small Tenements Recovery Act 1838, at present qualified by a variety of Rent Restriction Acts.

In the absence of provision to the contrary in the Act conferring jurisdiction, the procedure is governed by such parts of the S.J. Acts and rules as deal with orders and the recovery of civil debts. A summons is issued on complaint and not on information; and no warrant can be issued to compel attendance. On the hearing the civil laws of evidence apply. To enforce an order for the payment of a civil debt or costs, in default of distress, a judgment summons, of which personal or substituted service is necessary, must be obtained. If on the hearing of this summons it is proved that the defendant has or had, since the date of the order, means to pay, he may be committed for any term not exceeding 6 weeks. (S.J. Act 1879 S. 35). Default in complying with an order other than for payment of money may be visited by an order to pay a sum (to be enforced as a civil debt) up to £1 a day during default, not exceeding £20 in the aggregate, or for imprisonment, until compliance, for a period not exceeding two months.

(b) *Appeal:* There is no general right of appeal to Quarter Sessions: in Bastardy, however, either party may so appeal, and various other statutes conferring civil jurisdiction include this right. For instance the Pawnbrokers Act 1872 (on refusal of a certificate) and the Money-Lenders Act 1927. In matrimonial cases a special form of appeal lies to the High Court, Probate, Divorce and Admiralty Division.

The S.J. Acts 1857 and 1879 as to appeal by stated case apply to civil determinations in like manner as they do to criminal determinations, and the powers of the High Court as to writs of *certiorari, mandamus* and *prohibition* can also be invoked.

(8) **General Observations.**—The public advantage of the courts is not confined to the work done in open court. The stipendiary magistrates, in London especially, act as friendly advisers of the poor, who flock to the courts for counsel and assistance. It must be confessed, however, that occasionally the need is felt for some further provision for the defence of the needy and ignorant persons in these courts. Societies for the protection of women and children play their part, the magistrate does what he can and even has been known to instruct a solicitor for the defence at the expense of the poor box. Yet still there is a gap to be filled. (H. W. W. Wil.)

Bibliography.—G. C. Oke, *Magisterial Synopsis* (14th ed. 1893) and *Magisterial Formulist* (10th ed. 1922); A. C. F. Boulton, *Law and Practice of a Case Stated by a court of summary jurisdiction* (1902); T. S. Pritchard, *The Jurisdiction, Practice and Procedure of the Quarter Sessions* (2nd ed. 1904); J. F. Archbold, *The Practice of the Court of Quarter Sessions* (6th ed. 1908); C. J. Douglas, *Summary Jurisdiction Procedure* (10th ed. 1926); W. Paley, *Summary Convictions* (9th ed. 1926); W. K. Wigram, *The Justice's Note-Book* (11th ed. 1927). *See* also current ed. of Stone's *Justices' Manual*.

THE UNITED STATES

Summary jurisdiction in the United States, as in England, is the creation of statutes. The American colonists were familiar with the institution of summary jurisdiction as developed in Stuart England and adapted it to the exigencies of their new life. Reasons such as their distrusts of bureaucratic tyranny, the lack of trained lawyers, the limited need for criminal legislation, made for a reduction in the extent of summary jurisdiction. That a vigorous and extensive jurisdiction over minor offences was exercised is, however, clear. (*See* Frankfurter and Corcoran, Petty Federal Offenses and the Constitutional Guaranty of Trial by Jury, 39 Harvard L. Rev. 917, 1926.) Some of the colonies gave general summary jurisdiction of all offences where the maximum punishment was a moderate fine and a moderate term of imprisonment. Special legislation conferred summary jurisdiction for petty offences such as swearing, gambling, illegal selling of liquor, profiteering, etc. The extent of this jurisdiction as being "a veritable social control in the daily lives of the people" is demonstrated by the fact that prior to the American Revolution the number of offences punished in the first instance by a single magistrate in Massachusetts amounted to 170, a figure equally illustrative of the extent to which summary jurisdiction prevailed in the other colonies.

Prior to the Revolution much antagonism was aroused by the Crown's denial of the right to trial by jury in a number of cases where traditionally such trial was had. This grievance of the colonists loomed large in the Declaration of Independence of 1776 and was responsible for the inclusion of provisions in the United States and State constitutions guaranteeing the right to trial by jury. The relationship of these constitutional provisions to the recognized summary jurisdiction of the magistrates has proved a controversial issue in American law. It is quite clear, however, that their effect was not to introduce jury trial in all cases but that, whatever restrictions they placed upon the statutory extensions of summary jurisdiction, they left the existing ambit of that jurisdiction, in the main, untouched. The history of that jurisdiction has controlled the courts in determining the scope of the constitutional provisions for trial by jury. Thus, the constitutional provision has been interpreted by various courts as not to deny the power of the legislature to confer summary jurisdiction upon magistrates in a variety of petty offences such as assault and battery, disorderly conduct, petty larceny, violations of game, food inspection, liquor, vagrancy, Sabbath-breaking and highway laws. The test of what offences can be entrusted to the summary jurisdiction of single magistrates has been framed by the New Jersey court of errors and appeals in Katz *v.* Eldridge, 97 N.J.L. 123, 151. (1922), as follows: "The offence must be a petty and trivial violation of regulations established under the police power of the State in order that the offender may be summarily tried, convicted, and punished without indictment by a grand jury and without trial by a petit jury. It must, of course, be assumed that the punishment for such a petty and trivial offence will also be comparatively petty and trivial, otherwise it would violate another provision of the State constitution which prohibits cruel and unusual punishments."

The constitutional problem arising from the guarantee of jury trials is present to a less degree in the grant of summary jurisdiction in civil cases. In some States the constitutional provision is expressly limited to criminal cases. In those where it includes civil causes its interpretation has followed the historical manner given to the provision applicable to criminal jurisdiction. To increase summary jurisdiction without infringing upon the constitutional right to jury trial, statutes commonly provide that unless such trial is expressly demanded by the defendant at a certain stage of the proceedings the magistrate is empowered to proceed as if the case fell within his summary jurisdiction.

The extent to which magistrates exercise summary jurisdiction varies with the statutes of each State. In criminal matters it is exercised only in the broad class of petty offences. The furthest extent to which this classification has been enlarged concerned the breach of a State liquor law where the penalty prescribed amounted to a fine of $500 and six months' imprisonment. No summary jurisdiction has as yet been conferred upon the judicial officials of the Federal Government. Federal crimes are limited

in character and of such a grave nature as would traditionally call for the benefit of a jury trial. The prohibition amendment to the United States constitution and the legislation thereunder punishing the manufacture, sale and transportation of intoxicating liquor did, however, thrust upon federal courts a large amount of trivial criminal business and brought to the forefront the desirability of establishing some system for federal summary jurisdiction.

In civil matters magistrates exercise a large summary jurisdiction. The abatement of nuisances, the collection of taxes, the enforcement of liens, actions upon official bonds, *quo warranto* proceedings, the enforcement of awards, bastardy proceedings, are examples of classes of cases commonly entrusted to single magistrates. In ordinary civil actions upon contracts, promissory notes, torts and the like, summary jurisdiction is conferred provided that the claim does not exceed a certain pecuniary limit. Actions involving title to real property are generally excluded from grants of summary jurisdiction.

The procedure in summary jurisdiction matters is more informal than otherwise. Matters of form are largely dispensed with and great liberality with respect to the proceedings is exercised. Actions are commenced by summons which contains a short statement of the cause of action. No formal pleadings are necessary and in many cases the pleadings are oral, the magistrate's entry showing their substance. Defenses are also pleaded orally. Many states, however, require the plaintiff to set out his cause of action in a bill of particulars and require the defendant to file a formal answer.

The admission of evidence is governed by general common law principles though less exactness in their application is required. Judgments may be sued upon or executed in the same fashion as other judgments of superior courts. Dockets are required to be kept, upon which a record of the proceedings must be entered. Review of the proceedings is generally accorded by statute by appeal or writ of error within certain limitations as to the amount involved. The review is limited to those issues which were presented to the single magistrate and to the jurisdiction of that magistrate. Provisional remedies of arrest and attachment are also generally conceded to magistrates invested with summary jurisdiction. (J. M. La.)

SUMMERALL, CHARLES PELOT (1867–1955), U.S. army officer who revolutionized the tactics and technique of field artillery, was born on March 4, 1867, near Lake City, Fla., and graduated from the U.S. Military academy in 1892. During the allied assault upon Peking in the Boxer uprising of 1900, Summerall, then an artillery lieutenant, displayed great gallantry by walking to a gate of the imperial city under hot Chinese fire to scratch marks indicating parts of the gate as targets for his gunners. In World War I Summerall commanded first a division and then an army corps.

At Soissons in 1918 Summerall's courage and complete disregard for his own safety in making personal reconnaissance under heavy enemy fire gained him the distinguished service cross. He became chief of staff of the U.S. army in 1926, retiring in 1930 to become president of the Citadel, a military college in South Carolina.

He died in Washington, D.C., on May 14, 1955.

Summerall's artillery method, first put into effect at Soissons in the Aisne-Marne offensive of July 1918, consisted in utilizing the flexibility of the modern field gun by massing the fire of every piece within range, regardless of its position, on the most important target of the moment. It was the Napoleonic concept of the *grande batterie* in reverse. (R. E. D.)

SUMMER SCHOOL. In the United States after the 1870s, and more especially after about 1905, universities and colleges offered instruction during the summer months. The earliest programs of summer study retained many of the qualities of the customary summer vacation. In 1869 Dean Shaler arranged a field trip in the mountains of Colorado for students of geology at Harvard university. Louis Agassiz of Harvard offered field work in a zoological laboratory established on Penikese Island in Buzzards bay, Mass., in 1873. In 1880 C. O. Whitman founded the Marine Biological laboratory at Woods Hole, Mass., as a centre for researchers in biology.

An early form of summer study which greatly influenced the growth of schools was the Chautauqua summer assembly (*see* LYCEUMS AND CHAUTAUQUAS) which began with a ten-day session in 1874, enlisting the co-operation of many notable scholars and educators. One of these, William Rainey Harper, who conducted courses in Hebrew and in Old Testament literature, became president of The University of Chicago in 1891. Moved by the success of his Chautauqua experience he provided for a university calendar of four quarters, of which the summer one was an integral part of the university year. Shortly thereafter other universities provided for summer sessions, many of them offering courses which could be counted toward the baccalaureate degree.

One of the factors contributing to the growth of summer schools was the demand for trained teachers in a rapidly expanding system of public schools. During the first quarter of the 19th century teachers depended for help upon the county "normals," one-day conventions which were held at many points throughout the country. By 1850 more elaborate institutes, often of two weeks duration, were common, and were provided for by law in several states.

The first six-week summer course in education for public-school teachers was offered at Martha's Vineyard, Mass., in 1878. This school emphasized teaching methods and in 1890 enrolled 700 students in its elementary, high school and academic departments. Because of the seasonal nature of public-school teaching and the rapid increase in numbers of public-school teachers, U.S. summer schools until the middle 1920s offered chiefly undergraduate courses in educational methods. After the 1920s, standards of qualification for public-school teachers reached so high a level that the emphasis of this aspect of summer schools, especially at the large universities, shifted toward postbaccalaureate courses.

A later development in summer schools was the introduction of highly specialized conferences, workshops and seminars for selected groups of professional and industrial people. Courses of this sort make the results of research at educational institutions available to industrial technicians, teachers and other professional persons.

Summer sessions of varying lengths are often provided within a single institution. At universities employing a calendar of four quarters, many courses offered in the summer quarter are so organized that students may enroll and receive credit for either the first or the last five or six weeks of the quarter. Other institutions schedule sessions of different lengths. Summer sessions in high schools became fairly common after World War I as a means of offering acceleration to superior students, making up failures in the regular school year and for general enrichment. (H. W. Ss.)

GREAT BRITAIN AND EUROPE

The summer school idea, which developed especially after the end of World War II, was due to three main causes: (1) evening classes alone were not sufficient to cater to the educational needs of workers; (2) university and school teachers found that the extension of their work beyond the lecture and classroom was of immense value and that the long summer holidays were clearly a time when the most ambitious experiments could be carried out; (3) group recreational activities (particularly for young people brought up in industrial centres) carried out by the principal youth movements, such as the Boy Scouts and Girl Guides, showed the value of recreational education of this kind.

The work of the folk high schools, started in Denmark by Bishop N. F. Grundtvig (*q.v.*) and his followers in the middle of the 19th century, provided one of the chief inspirations for the early experiments in summer schools. Great impetus to the movement came, naturally, with the introduction of paid holidays; and in Britain many new organizations sprang up to continue the work begun by the Workers' Educational association. Government departments, local authorities, professional associations, trades unions, learned societies and most voluntary youth organizations followed the lead. The movement spread in many

countries and became international at an early stage. Universities took the lead in making their premises available for students from abroad and summer courses beginning with language study rapidly developed in European university towns.

After World War II the rate of growth became extremely rapid, and international summer schools were one of the first preoccupations of UNESCO during its early years, as they continued to be.

In the second half of the 20th century there were few holiday resorts or university towns in Europe where buildings were not taken over during the summer months. Many were organized for a limited number of persons who were invited to take part and these were usually not advertised. Some general idea of the numbers of schools may be obtained from three sources: (1) in Great Britain the National Institute for Adult Education publishes each year a list of residential short courses, held between April and September in the United Kingdom, their list for one year containing details of well over 600 courses; (2) in Great Britain again, the Central Bureau for Educational Visits and Exchanges published a list which mentioned over 200 courses in Europe open to students from the United Kingdom, these courses being held in Austria, Belgium, Bulgaria, Denmark, the Republic of Ireland, Finland, France, the German Federal Republic, Greece, the Netherlands, Italy, Norway, Portugal, Spain, Sweden, Switzerland, the U.S.S.R. and Yugoslavia; (3) the UNESCO supplement to *Study Abroad,* which includes details of many hundreds more throughout the world.

These figures apply only to courses open to all and devoted to a formal course of study. Summer schools, where children of secondary and even of primary school age are taken to standing camps, schools and other centres for a wide range of educational activities, are even more numerous.

The term "summer school," formerly applied only to adult and university studies and to refresher courses for teachers and other professional persons, has grown in meaning to include the extension of all branches of education into the summer holiday period. One great value of the summer school lies in the bringing together of students in pleasant surroundings where they may intensify the study of particular subjects (especially geography, geology, botany and archaeology at field centres).

See also ADULT EDUCATION. (F. C. A. C.)

SUMMER TIME: *see* DAYLIGHT SAVING.

SUMMIT, a city of New Jersey, U.S., in Union county, on an escarpment of Watchung mountain adjacent to Springfield, Millburn and Chatham, is 10 mi. S.W. of Newark and 19 mi. W. of New York city. Founded as Turkey Hill in 1710, it was renamed Summit in 1837 and incorporated as a city in 1899. The site of a lookout post and signal tower for General Washington's troops alerting them to any movement of the British in the Revolutionary War is commemorated by a marker showing the location of the alarm gun "Old Sow."

A suburban residential community, its chief industries are chemicals, pharmaceuticals, textiles and floristry. Among its educational facilities is Kent Place school for girls.

For comparative population figures *see* table in NEW JERSEY: *Population.* (D. N. A.; M. P. M.)

SUMMONS: *see* PRACTICE AND PROCEDURE.

SUMNER, CHARLES (1811–1874), U.S. statesman, was born in Boston, Mass., on Jan. 6, 1811. He graduated in 1830 from Harvard college, and in 1834 from the Harvard law school. At the age of 23 he was admitted to the bar. The three years (1837–40) spent in Europe were years of fruitful study and experience.

In his 30th year Sumner returned to Boston, to settle down to the practice of law. But gradually he devoted less of his time to practice and more to lecturing in the Harvard law school.

Sumner co-operated effectively with Horace Mann for the improvement of the system of public education in Massachusetts. Prison reform and peace were other causes to which he gave support. He took an active part in the organizing of the Free Soil party, in revolt at the Whigs' nomination of a slave-holding southerner for the presidency; and in 1848 was defeated as a candidate for the national house of representatives. In 1851 control of the Massachusetts legislature was secured by the Democrats in coalition with the Free Soilers, but the Democrats refused to vote for Sumner, the Free Soilers' choice for U.S. senator, and urged the selection of some less radical candidate. A deadlock of more than three months ensued, finally resulting in the election (April 24) of Sumner by a majority of a single vote.

In the closing hours of his first session, in spite of strenuous efforts to prevent it, Sumner delivered (Aug. 26, 1852) a speech, "Freedom national; Slavery sectional," which marked a new era in U.S. history. The conventions of both the great parties had just affirmed the finality of every provision of the compromise of 1850. In 1856, at the very time when "border ruffians" were drawing their lines closer about the town of Lawrence, Kan., Sumner in the senate (May 19–20) laid bare the "crime against Kansas." He denounced the Kansas-Nebraska bill as in every respect a swindle, and held its authors, Stephen A. Douglas and Andrew P. Butler, up to the scorn of the world. Two days later Preston S. Brooks (1819–57), a congressman from South Carolina, confronted Sumner in the senate chamber, denounced his speech as a libel upon his state and upon Butler, his relative, and struck Sumner until he fell unconscious to the floor. This assault cost Sumner three years of heroic struggle to regain his health—years during which Massachusetts re-elected him, in the belief that in the senate chamber his vacant chair was the most eloquent pleader for free speech and resistance to slavery.

After the withdrawal of the Southern senators, Sumner was made chairman of the committee on foreign relations (March 8, 1861), a position for which he was pre-eminently fitted by his years of intimate acquaintance with European politics and statesmen. While the Civil War was in progress his letters from Cobden and Bright, from Gladstone and the duke of Argyll, at Lincoln's request were read by Sumner to the cabinet, and formed a chief source of light as to political thought in England. In the turmoil over the "Trent affair," it was Sumner's word that convinced Lincoln that Mason and Slidell must be given up, and that reconciled the public to that inevitable step.

Throughout the war Sumner had constituted himself the special champion of the Negro, being the most vigorous advocate of emancipation, of enlisting Negroes in the Union army and of the establishment of the Freedmen's bureau. The credit or the blame for imposing equal suffrage rights for Negroes upon the Southern states as a condition of reconstruction must rest with him.

In the impeachment proceedings against Andrew Johnson, he was one of the president's most implacable assailants. Sumner's opposition to Grant's pet scheme for the annexation of San Domingo (1870), brought upon him the president's bitter resentment. Sumner had always prized highly his popularity in England, but unhesitatingly sacrificed it in taking his stand on the adjustment of claims against England for breaches of neutrality.

Under pressure from the president, on the ground that Sumner was no longer on speaking terms with the secretary of state, he was deposed (March 10, 1871) from the chairmanship of the committee on foreign relations. Whether the chief cause of this humiliation was Grant's vindictiveness at Sumner's opposition to his San Domingo project or a genuine fear that the impossible demands which he insisted should be made upon England, would wreck the prospect of an adjustment with that country, cannot be determined. Sumner's last years were further saddened by the misconstruction put upon one of his most magnanimous acts. In 1872 he introduced in the senate a resolution providing that the names of battles with fellow citizens should not be placed on the regimental colours of the United States. The Massachusetts legislature denounced this battle flag resolution as "an insult to the loyal soldiery of the nation" and as "meeting the unqualified condemnation of the people of the Commonwealth." For more than a year all efforts—headed by the poet Whittier—to rescind that censure were without avail, but early in 1874 it was annulled. On March 10 against the advice of his physician, Sumner went to the senate—it was the day on which his colleague was to present the rescinding resolution. That night he was stricken with an acute attack of angina pectoris, and on the following day he died.

"The slave of principles, I call no party master," was the proud avowal with which Sumner began his service in the senate. His

was the first clear program proposed in congress for the reform of the civil service. It was his dauntless courage in denouncing compromise, in demanding the repeal of the Fugitive Slave act, and in insisting upon emancipation, that made him a great propelling force in the struggle that put an end to slavery.

SUMNER, EDWIN VOSE (1797–1863), U.S. soldier, was born at Boston, Mass., and entered the U.S. army in 1819. He took part in the Black Hawk War and the Mexican War. At the outbreak of the Civil War (1861) Sumner, then a brigadier general, was sent to relieve Sidney Johnston in command of the department of the Pacific. The following year, he was recalled to the army of the Potomac. At the battle of Fredericksburg he commanded the right grand division under Burnside. Upon Hooker's appointment to command of the eastern army, Sumner was assigned to the department of Missouri, and died suddenly, on March 21, 1863, while on his way there.

SUMNER, JAMES BATCHELLER (1887–1955), U.S. biochemist who discovered that enzymes can be crystallized, was awarded the 1946 Nobel prize in chemistry with W. M. Stanley and J. H. Northrop. Born Nov. 19, 1887, in Canton, Mass., Sumner received his bachelor's degree in 1910 and his doctor's degree in 1914 from Harvard university. He taught at Mt. Allison college, Sackville, N.B., and at Worcester Polytechnic institute. In 1914 he became assistant professor and in 1929 professor of chemistry at the medical college of Cornell university, Ithaca, N.Y., where he remained until his retirement in July 1955. He had studied in Brussels, Stockholm and Uppsala. He was a member of the commission for relief in the Belgian Educational foundation 1921–22. He was awarded the Scheele medal at Stockholm in 1937. In 1947 he became the director of the Cornell laboratory of enzyme chemistry, established by the trustees of the university in recognition of his investigations.

Sumner's early work included the study of urea formation and the estimation of sugar in urine. In 1926, after chemists had hoped to isolate an enzyme, Sumner observed urease, an enzyme essential to the nitrogen cycle. The first enzyme to be prepared in a laboratory, it was a remarkable contribution. In 1937 Sumner crystallized the important enzyme catalase, which influences reactions involving hydrogen peroxide and protects the living cells from this compound, which they produce during their respiratory process. Sumner contributed to the purification of co-enzyme I, lipoxidase and saccharase. He published *Biological Chemistry* in 1927. With G. F. Somers he wrote three editions of the two books *Chemistry and Methods of Enzymes* and *Laboratory Experiments in Biological Chemistry*, and with Karl Myrbäck compiled four books on *The Enzymes: Chemistry and Mechanism of Action* (1950–52). He died Aug. 12, 1955. (V. Bw.)

SUMNER, WILLIAM GRAHAM (1840–1910), U.S. social scientist, an influential teacher and prolific publicist, was born in Paterson, N.J., on Oct. 30, 1840. He studied at Yale (A.B., 1863); and at Geneva, Göttingen and Oxford (1863–66); tutored at Yale (1866–69); served as a priest in the Episcopal Church (1869–72); and was professor of political and social science at Yale (1872–1909). He was president of the American Sociological society in 1909. He died in Englewood, N.J., April 12, 1910.

Like Herbert Spencer, Sumner expounded in many essays on current issues his firm belief in *laissez faire*, individual liberty and the innate inequalities among men. As a social Darwinist he viewed competition for property and social status as resulting in a beneficent elimination of the ill-adapted and the preservation of racial soundness and cultural vigour. For him the middle class Protestant ethic of hard work, thrift and sobriety exemplified those virtues conducive to wholesome family life and sound public morality. Foreseeing the drift toward the welfare state, but viewing poverty as the natural result of inherent inferiorities, he opposed all proposals that smacked of paternalism because they impose unequal burdens on that class, his "forgotten men."

In his famous *Folkways* (1906), he found the origin of customs and morals in instinctive responses to the stimuli of hunger, sex, vanity and fear, selectively guided by pain and pleasure. He emphasized their inherent irrationality and resistance to reform. His notes became the basis of *The Science of Society*, 4 vol. (1927–28), edited by A. G. Keller.

See also ETHICS, HISTORY OF: *Ethical Relativity*.

(F. H. Hs.)

SUMPTUARY LAWS, a term originally used to denote regulations designed to restrict extravagance in food, drink, dress, household equipment and the like, usually on religious or moral grounds. More recently, the term has been applied to laws which either directly prohibit the manufacture or consumption of certain commodities or which influence their consumption indirectly through excise taxes. In the case of this more recent usage, the regulations are ordinarily defended on religious or moral grounds or on the grounds of public welfare.

Sumptuary laws have an ancient origin; numerous instances are to be found in ancient Greece. All the inhabitants of Laconia, for example, were forbidden to attend drinking entertainments and the Lacedaemonians were forbidden to possess a house or furniture that was the work of more elaborate implements than the axe and saw. The possession of gold or silver was forbidden to the citizens of Sparta and the use of iron money alone was permitted by the Lycurgean legislation. A system of sumptuary edicts and enactments was extensively developed in ancient Rome. The code of the Twelve Tables (*q.v.*) included provisions limiting expenditures on funerals. A series of laws beginning in 215 B.C. governed the materials of which garments could be made, the number of guests at entertainments and the sums to be expended on entertainments, and forbade the consumption of certain foods.

In later times, sumptuary laws were enacted in many countries of Europe. For example, in France, Philip IV issued regulations governing the dress and the table expenditures of the several orders of men in his kingdom and Charles V forbade the use of long-pointed shoes. Under later French kings the use of gold and silver embroidery, silk stuffs and fine linen was restricted. In England during the reign of Edward II a proclamation was issued against the "outrageous and excessive multitude of meats and dishes which the great men of the Kingdom had used, and still used, in their castles." Edward III in 1336 attempted to legislate against luxurious living and at the same time when costumes were regulated it was held that servants of gentlemen, merchants and artificers should have only one meal of flesh or fish per day. Another illustration may be found in 1433 in an act of parliament meeting at Perth which prescribed the manner of living of all orders in Scotland, even going so far as to limit the use of pies and baked meats to those who held the rank of baron or higher. Such legislation was brought to some of the American colonies in the 17th century. It appears, however, that most such legislation was not strictly enforced in the colonies.

In Japan sumptuary laws were passed with a frequency and minuteness of scope that had no parallel in the history of the western world. At the beginning of the 11th century an imperial edict regulated the size of houses and imposed restrictions as to materials that could be used in their construction. During the Tokugawa period sumptuary laws were passed in bewildering profusion regulating the most minute details of personal life.

In more modern times sumptuary laws have generally not gone as far as did the ancient enactments. An exception is to be found in the 18th amendment to the U.S. constitution (in effect 1919–33) that prohibited the manufacture, transport and sale of alcoholic beverages. Many additional illustrations are to be found of laws having the effect of sumptuary laws indirectly although enacted primarily for raising revenue. Excise taxes on tobacco, alcoholic beverages, jewelry, watches and other articles that are viewed as luxuries are levied in many parts of the world and are frequently defended, in part, on the ground of their beneficial moral effect.

See F. Baldwin, *Sumptuary Legislation and Personal Regulation in England* (1926). (FK. L. K.)

SUMTER, THOMAS (1734–1832), U.S. soldier, was born in Hanover county, Va., on Aug. 14, 1734. He served in the French and Indian War, and was present at Braddock's defeat in 1755. Later he moved to South Carolina. After the fall of Charleston in 1780, when his state was ravaged by the British, he

escaped to North Carolina, where he took the field as brigadier general at the head of a body of light horse and soon became one of the most active and able leaders of the south. His vigilance and bravery earned for him the sobriquet of "Gamecock." After successes over the British at Catawba and at Hanging Rock (Lancaster county), he was surprised and defeated by the British general, Tarleton, at Fishing Creek (Chester county). Raising a new force, he defeated Mayor Wemys at Fishdam (Union county) and repulsed Tarleton's attack at Blackstock (Union county) in Nov. 1780. At the time of his death at South Mount, S.C., on June 1, 1832, he was the last surviving general officer of the War of Independence.

See Edward McCrady, *The History of South Carolina in the Revolution* (1901–02); H. A. M. Smith, "General Thomas Sumter" in *Magazine of History* (vol. 8, 1908, and vol. 9, 1909).

SUMTER, a city of South Carolina, U.S., county seat of Sumter county, is 40 mi. E. of Columbia. Founded as Sumterville in 1799, and named after Revolutionary general Thomas Sumter, it was long a typical cotton plantation village in the heart of the low-country sand hills. Furniture, lumbering and textile industries came slowly, and it was chartered as Sumter in 1887. It claims to be the first city to have successfully adopted, in 1912, the council-manager form of government.

Essentially a town of the new south rather than the old, it maintains a leisurely southern atmosphere. Wide streets are lined with oak trees covered with Spanish moss. The iris gardens in the city are known for the old cyprus trees, azaleas and camellias, as well as iris.

For comparative population figures *see* table in SOUTH CAROLINA: *Population.* (G. H. Ct.)

SUMY, a town and *oblast* of the Ukrainian Soviet Socialist Republic, U.S.S.R., in the northeast of the Ukraine. The *oblast* lies along the southwestern edge of the Central Russian upland from which the land slopes gently toward the Dnieper river. Area 9,382 sq.mi.; pop. (1959) 1,513,718.

Drainage of the *oblast* is by tributaries of the Dnieper, notably the Seim, Sula and Psel rivers. In the north part its boundary lies along the Desna river, the basin of which is in the mixed forest zone, with extensive areas of marsh. South of the Seim the *oblast* is in the forest-steppe zone, but most of the former areas of oak forest have been cleared for cultivation and occupy not more than 5% of the surface. In the Desna basin soils are sod-podzols and elsewhere chernozems of high fertility. Two-thirds of the population are rural dwellers and agriculture dominates the economy.

The main crops are grains, sugar beet, tobacco, hemp and potatoes. Sugar refining is a major industry and the beet waste helps to maintain the large number of cattle.

There are 12 towns and 20 urban districts in the *oblast.* All the towns are small, the main industry being food production. Romny, in the southwest, dates from the early 10th century. Konotop and Shostka are minor industrial centres. During World War II the area was occupied (1941–43) by German troops.

The town of Sumy, the administrative centre of the *oblast,* is on the upper Psel river, about 68 mi. S.W. of Kursk. Pop. (1959) 98,015. Industries include manufacture of machinery for the chemical, coal-mining and metallurgical industries, sugar refining, shoemaking and the production of superphosphates. There is a teacher training college. Railways connect Sumy with Kharkov and the main Kiev-Kursk line. The town was founded as a Russian settlement in 1652 but there had been an earlier Slav settlement there in the 8th–9th centuries. (R. A. F.)

SUN. The sun is apparently the largest and brightest of the naked-eye stars but it is actually among the smallest and faintest. The illusion arises from its comparative nearness—the next nearest star is, in fact, nearly 300,000 times as far away.

TELESCOPIC INVESTIGATION

General Characteristics.—The sun presents to the telescopic view a dazzling white circular disk, with a sharp edge, whose diameter is an angular distance in the sky of about $\frac{1}{2}°$. Slight systematic variations of the angular diameter reveal that the distance from earth to sun is not constant, and that the earth's orbit is an ellipse with the sun at one of the focuses. Several methods have been employed to measure the distance of the earth from the sun in miles or other terrestrial units. The mean value throughout the year (from which the actual distance departs by less than 2%) is found to be about 93,000,000 mi. This is expressed astronomically by saying that the solar parallax is 8.79 (*i.e.,* 8.79 sec. of arc), the solar parallax being the difference in direction of two ends of the earth's equatorial radius as viewed from the sun.

The sun's linear diameter, in miles, follows at once from its distance and the observed angular diameter. It is approximately 865,000 mi.—about 109 times the diameter of the earth, but much smaller than many stars, some having diameters 800 times that of the sun.

Its mass is calculated from the force of mutual attraction of the earth and sun, the mass of the earth having been independently determined (*see* EARTH). The result (2×10^{33} g. or 2×10^{27} tons) is 333,000 times the mass of the earth and about at the middle of the range of stellar masses. Mass and diameter are sufficient data for the estimation of the mean density of the sun, viz. 1.41 times that of water. The density of the sun actually varies from a value much less than that of air in its outer layers, to about 60 times the density of ordinary rocks at its centre.

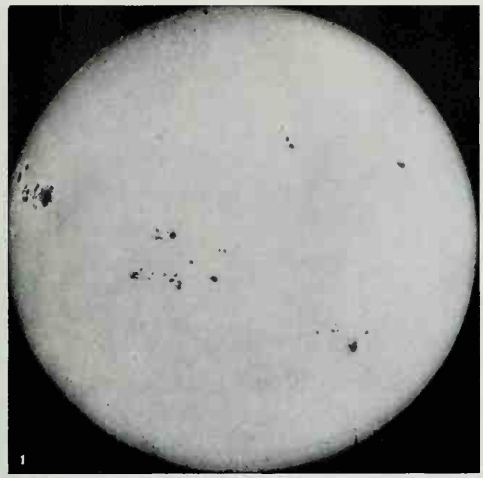
Comparison of measurements of the positions of the stars made over long periods shows that in the direction of the constellation Hercules the stars appear on the whole to be getting farther apart; in the opposite direction (the constellation Columba), they appear to be drawing together. This indicates that the sun and the members of the solar system are moving as a group with respect to the stars.

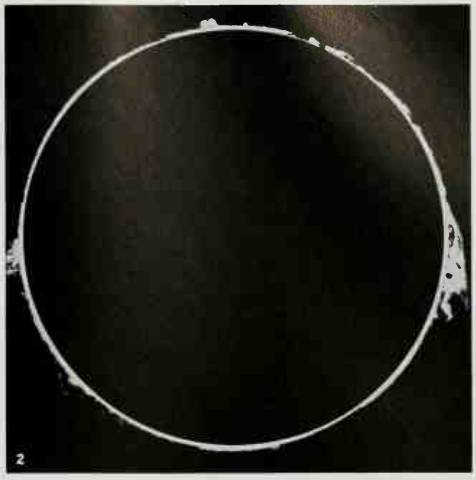
Surface Features.—The mensurational characteristics are supplemented by a voluminous body of physical data. When the surface of the sun is carefully observed through a telescope it is found to present a mottled, or grainy structure. The brightness of the disk is not uniform; the region near the circumference, or limb, is appreciably less bright than the centre of the disk—a fact which appears most obvious in photographs. Sunspots, which are frequently visible to the eye, when viewed through a densely smoked glass, are the most easily observed markings on the sun.

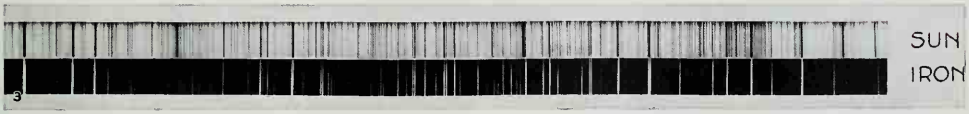
A sunspot consists of a central, apparently black, umbra, surrounded by a less dark penumbra. Each of these regions is in fact exceedingly bright, and appears dark only by contrast with the still brighter solar disk. Spots often are clustered together in groups, often in pairs or pairs of groups, and only rarely appear outside two zones on either side of the solar equator, extending about halfway toward the sun's north and south poles of rotation. The spots move steadily across the face of the sun so as to leave no doubt that the motion is caused by rotation of the sun about an axis. The sun, since it always appears circular in the sky, must be spherical, and its bright visible surface is called the photosphere. A remarkable feature of the sun's rotation is its variation with latitude; *i.e.,* the sun does not rotate as a rigid body; the nearer to its equator, the faster the rotation. Near the equator, the mean rotation period is 24.65 days; at solar latitude 20°, it is 25.19 days; at 35°, 26.63 days; and at 60°, 30.93 days.

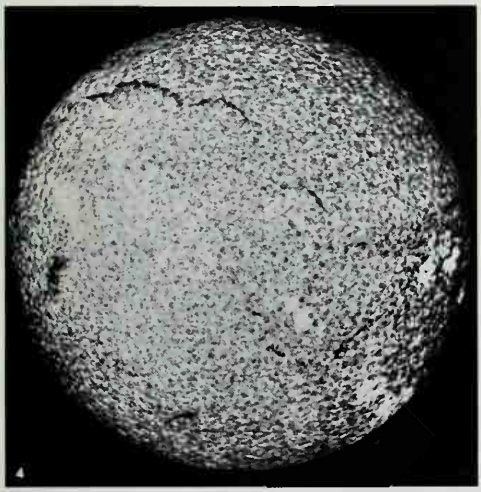
Individual sunspots appear spasmodically, remain visible for periods varying from a few hours to several months and then disappear. The apparently capricious behaviour becomes a striking regularity when large numbers of spots and great lengths of time are considered. If the number of spots appearing per year (or the total area covered by them) is plotted against time, as in the diagram, a nearly periodic relation is shown, the annual number reaching a maximum approximately every 11th year. A similar regularity characterizes the location of the spots. At a time of minimum those of a new cycle begin to appear in the higher latitudes, both north and south, of their appointed belts, and as the cycle progresses the place of outbreak gradually moves toward the equator. The thin lines in the graph illustrate this.

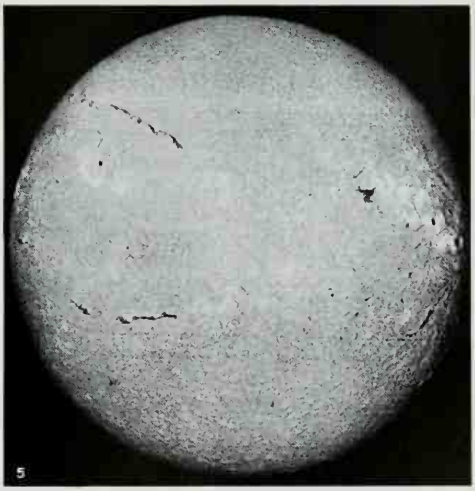
Sunspots are always accompanied by exceptionally bright areas on the photosphere, known as faculae. They are most easily seen near the limb, where the brightness of the photospheric background is diminished. Although less conspicuous than the spots,







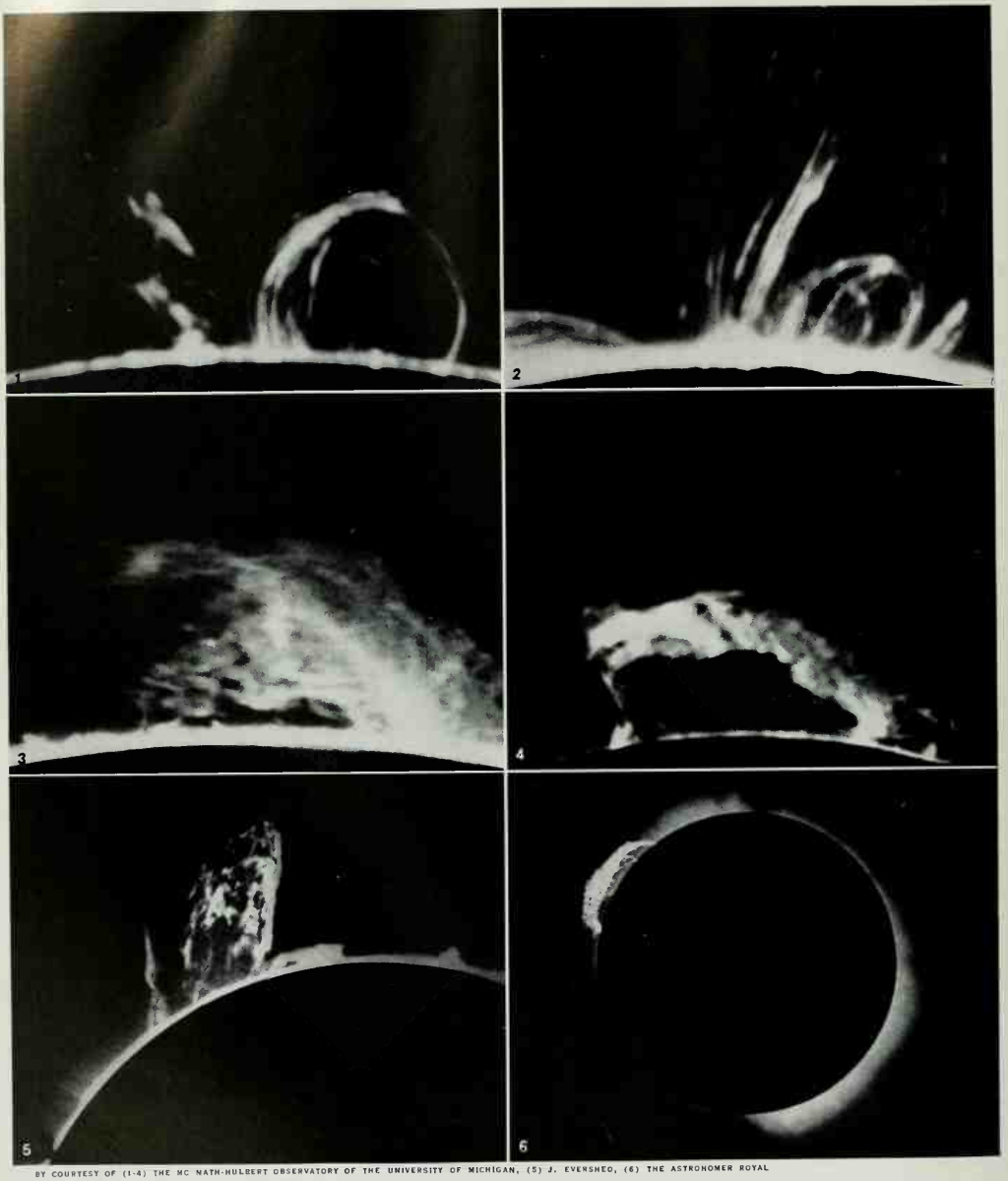




PHOTOGRAPHIC STUDIES OF THE SUN

1. The photosphere of the sun, showing an unusually large number of spots, faculae, and darkening at the limb (Rutherfurd)
2. Photograph taken in calcium light, the image of the photosphere having been artificially covered. Shows Chromosphere and Prominences (Evershed)
3. Comparison of solar spectrum with laboratory spectrum of iron (violet region). Presence of iron in the atmosphere of the sun is indicated by coincidence in positions of lines (Imperial College)
4. Spectroheliogram (photograph of sun made by monochromatic light) showing distribution of high-level calcium in the solar atmosphere (Deslandres)
5. Spectroheliogram showing distribution of hydrogen in the solar atmosphere (Deslandres)

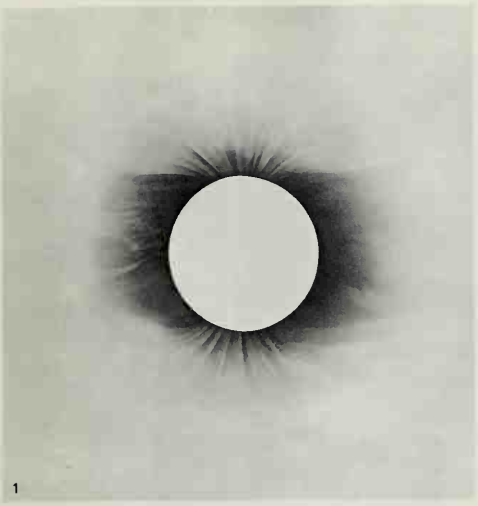
SUN

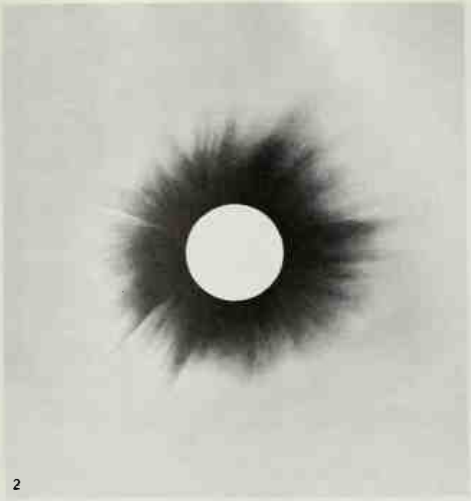


BY COURTESY OF (1-4) THE MC MATH-HULBERT OBSERVATORY OF THE UNIVERSITY OF MICHIGAN, (5) J. EVERSHED, (6) THE ASTRONOMER ROYAL

STUDIES OF SOLAR PROMINENCES

1. Prominence of Sept. 21, 1939, an example of the sunspot-type classification. All detectable motions were downward (*i.e.*, toward the surface of the sun). The photograph shows a loop formation that measured approximately 90,000 mi. in height and 105,000 mi. in width at the start of the record. The overall dimensions increased about 25% during the period of observation
2. Sunspot-type prominence of Sept. 7, 1939, recorded in a remarkable motion-picture scene of continuous and varied activity. As the scene progressed, the general pattern became intricate, developing a background of faint gigantic loops and many small condensations moving at high velocities toward the chromosphere. The longest streamer extended more than 150,000 mi. above the edge of the sun
3. Quasi-eruptive prominence of Aug. 24, 1939, photographed in the light of ionized calcium. The prominence was about 105,000 mi. high and extended about 155,000 mi. along the edge of the sun
4. The same prominence as in (3), photographed the next day
5. Prominences of the sun observed May 26, 1916, at Srinagar, Kashmir, India
6. Inner corona and prominence photographed during a total solar eclipse, May 29, 1919, observed at Sobral, Brazil. The red flame rose from a height of 130,000 mi. to more than 500,000 mi. above the surface of the sun in less than seven hours





BY COURTESY OF (1) MRS. E. W. MAUNDER. (2) THE ASTRONOMER ROYAL, ROYAL ASTRONOMICAL SOCIETY, LONDON. (3) J. E. WILLIS, U.S. NAVAL OBSERVATORY, NATIONAL GEOGRAPHIC SOCIETY. (4) THE LICK OBSERVATORY

CORONAE AT VARIOUS STAGES OF SUNSPOT ACTIVITY

1. Corona, May 18, 1901, near time of minimum sunspot activity. Characteristic sunspot minimum with short polar brushes and long equatorial streamers
2. Corona, Aug. 30, 1905, near time of maximum sunspot activity. The circular corona characteristic of a sunspot-maximum period and the long equatorial streamers which give it its typical form are here visible
3. Corona during total eclipse of June 8, 1937, characteristic of maximum sunspot activity. Photograph taken at Canton Island
4. Solar corona, Alhambra, Sp., Aug. 30, 1905. Photograph taken with 40-ft. telescope; exposure time 25 seconds

NOTE: THE RANGE OF BRIGHTNESS IN THE CORONA IS SO GREAT THAT ANY ONE PHOTOGRAPH CAN SHOW DETAIL EITHER OF THE INNER OR OUTER CORONA ONLY. FIGS. 1 AND 2 ARE FROM COMPOSITE DRAWINGS MADE FROM ORIGINAL NEGATIVES, WITH CONTRAST REDUCED

SUN

4:40 P.M. (E.D.T.); shortly before the beginning of the eclipse

5:15 P.M.; onset of the eclipse

5:35 P.M.

5:40 P.M.

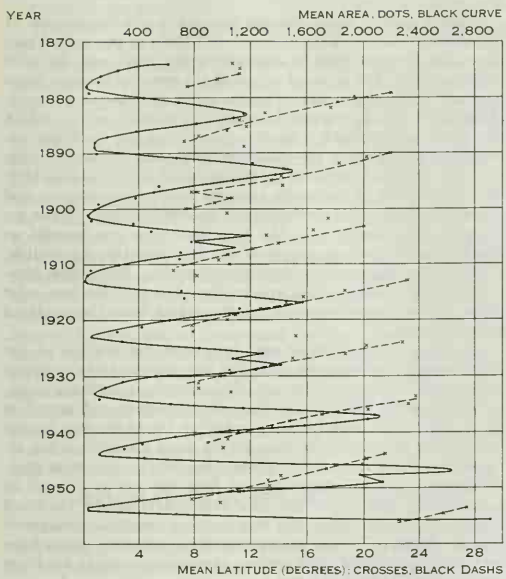
5:45 P.M.

6:00 P.M.

ECLIPSE OF THE SUN, JULY 20, 1963, AS PHOTOGRAPHED FROM NEW YORK CITY

PHOTOGRAPHS, (ALL) PAUL GREENWALD FROM SHOSTAL

the study of the faculae is equally important for an understanding of the sun.



MEAN AREAS AND HELIOGRAPHIC LATITUDES OF SUNSPOTS FOR EACH YEAR, 1874–1956

Areas in millionths of the visible hemisphere, heliographic latitudes in degrees

On the comparatively rare occasions on which the sun is observed in total eclipse, small red clouds are seen apparently floating above various points on the circumference of the dark moon. They belong in reality to the sun, and are known as prominences. Prominences are not seen through an ordinary telescope in full daylight because the intense photospheric light, diffused by earth's atmosphere and the optical parts of the telescope as by a ground glass, acts as a veil through which prominences and faint stars alike are invisible as individual objects. Prominences assume various shapes and sizes, sometimes reaching heights of hundreds of thousands of miles above the photosphere. The larger prominences can easily be seen during eclipses.

Prominences seem to be extensions of an outer solar atmosphere that is directly revealed at the beginning and ending of each total eclipse. When the moon is just covering the last remnant of the photosphere, a crescent of atmosphere remains exposed for a few seconds. It is easily seen in a small telescope as a brilliant red layer, and has been named the chromosphere.

The most striking solar eclipse phenomenon is the corona, a pearly white halo enveloping the sun and extending in more or less definite rays or streamers to distances of many solar radii. The brilliance of the corona diminishes very rapidly with distance from the sun's limb, and although its total brightness is about one-half that of the full moon, it is much fainter, area for area, than the solar prominences. No two aspects of the corona seen at different eclipses are identical; however, there is a relation between the shape of the corona and the sunspot period. At sunspot maximum the corona appears to extend equally in all directions from the sun's limb and appears round. At sunspot minimum, the poles of the sun are marked by comparatively small tufts of light, while from the equatorial regions long streamers shoot out to enormous distances.

A special type of telescope, called a coronagraph, can be constructed so that it diffuses almost no light and intercepts the image of the photosphere with a metal disk. On high mountain tops, where diffusion by earth's atmosphere is negligible on days of exceptional clarity, the innermost, brightest, portions of the corona can be observed directly with this instrument.

SPECTROSCOPIC INVESTIGATION

General Considerations.—The spectroscope is the most important of the auxiliary instruments used with the telescope for the investigation of the sun because it can select minute portions of the flood of solar radiation for intensive study. (*See* SPECTROSCOPY and SPECTROSCOPY, ASTRONOMICAL.)

Nearly all solar research depends at some point on a device for sorting radiation by its wave length. This results in the familiar sequence of colours seen in the rainbow. The word spectroscope in its most general meaning is used to indicate such a device. The human eye is sensitive to only a small band out of the vast range of wave lengths represented in the total radiation of the sun, but over this restricted range, the words "wave length" are analogous to a precise specification of colour. The spectrum, or ordered arrangement of radiation obtained from a spectroscope, of the sun's radiation consists mainly of a continuous bright background on which dark lines appear. This is called an absorption spectrum, and it confirms the direct eclipse observation that the photosphere is covered by a layer of glowing gases. The dark lines are evidence that the glowing gases are relatively cool. A qualitative spectroscopic chemical analysis of the sun's outer envelopes may be made by comparing the positions, or wave lengths, of the absorption lines with those of emission lines produced by known substances in the laboratory. In this manner the presence on the sun of 67 elements known on earth, and 18 chemical compounds, has been established. The spectroscopic chemical analysis of the sun can also be made quantitative by directing attention to the total amount of radiation extracted from the continuous background by the dark lines. The darkness of a line is related in a complicated way to the amount of material that produces it, but with the aid of laboratory calibrations of line strengths, the relative abundance of the chemical elements has been determined for the sun. The chemical constitution of the sun resembles that of the earth, with the exception that the lighter elements, especially hydrogen and helium account for almost all its mass.

The spectrum of individual features of the sun may be studied as well as the light of the entire solar disk. Just at the start of, or just before the end of the total phase of a solar eclipse, when the chromosphere still remains exposed for a few seconds, the spectrum of its light shows bright rather than dark lines, since this layer of the sun is semitransparent and there is no bright photospheric background to provide a source of radiation to be absorbed. Because of its evanescence, this spectrum is known as the flash spectrum, and it is essentially identical with the absorption spectrum in the position of its lines. The heights reached by the various substances have been determined from special observations at such times. The majority are confined to the lowest atmospheric layers, not more than 500 mi. high, near the top of the photosphere or the bottom of the chromosphere. Hydrogen, helium and ionized calcium can be traced to great heights, extending up to 8,000 or 9,000 mi. The radiation of hydrogen is the source of the intense red colour of this part of the sun's atmosphere—the chromosphere.

The chromosphere is so bright when viewed during a solar eclipse that it seems reasonable to expect success in the observation of its bright-line spectrum in full daylight. The observation is difficult because the photospheric light diffused by ordinary telescopes and earth's atmosphere spreads over and drowns the chromospheric radiations. Nevertheless, the hydrogen and helium emission lines are easily observable with very modest telescopic and spectroscopic equipment, but the fainter lines can only be seen with very powerful instruments fed by clean (nondiffusing) telescopes, such as the coronagraph, and located where the scattering of light by earth's atmosphere is as small as possible.

When the spectroscope is directed toward the corona, at eclipse time or with the aid of a coronagraph, it is found that a great deal of the coronal radiation yields the ordinary solar spectrum and seems to be only reflected sunlight. Actually, radiation from the corona contains four main components: a continuous spectrum

which is sunlight scattered by electrons; an absorption spectrum, which is sunlight diffracted by interplanetary dust particles; a series of bright emission lines which have not yet been produced in terrestrial laboratories; and radiation that can be detected using radio methods. The bright emission lines of the corona can be identified with radiations emitted by isolated and very highly ionized atoms of iron, nickel and calcium (*see* CORONA).

Spectra and Physical Conditions.—The spectrum of the solar corona is one of the best examples of the results of physical conditions in modifying the radiation emitted from various regions of the sun. After the effects on line spectra of temperatures, pressures, magnetic and electric fields, and motions of the sources of light had been established through terrestrial spectroscopic studies, these, and other physical parameters could be investigated on the sun by means of detailed investigation of the solar spectrum lines. Three of the more important applications to the sun have produced knowledge of temperatures, magnetic fields and motions.

An early spectroscopic laboratory result classified spectrum lines into high-temperature and low-temperature lines. This classification had an important application when spectra of the photosphere and a sunspot were compared. The spectrum of a spot contains evidence of compounds, of which only the separated elements contribute to the photospheric spectrum, and the spot spectrum shows a strengthening of low-temperature and a weakening of high-temperature lines when compared with the spectrum of the photosphere. The conclusion is that the sunspot region has a temperature 2,000° C. lower than the rest of the solar atmosphere.

A source of light placed in a magnetic field (*i.e.*, a region such as the neighbourhood of the poles of a magnet) has many of its spectrum lines split into a number of component lines, lying closely side by side. Unless the field is very strong, the components are too close together to be seen as distinct individual lines, but the original line may appear to be broadened. By means of special appliances very slight broadenings produced by weak magnetic fields can be detected. The lines in the spectrum of a sunspot show just the composite character that would result from the existence of a magnetic field in the spot in a direction perpendicular to the sun's surface.

Many years of observation show that all sunspots are magnets, some presenting a north, others a south pole to the surrounding space. In each pair of spots, or spot groups, the leaders and followers in the journey around the sun's axis have opposite polarities. Many apparently single spots have invisible companions which can be found by the magnetic splitting of the spectrum lines.

There is a reversal of polarity in the southern solar hemisphere with respect to the northern; thus, if the leader of a pair were a north pole in the northern hemisphere, it would be a south pole in the southern hemisphere, and vice versa; and, with few exceptions, this order would characterize all spots from one minimum to the next of a sunspot cycle. During the next cycle this order would be reversed, so that the approximate period from north pole preceding in the northern hemisphere to north pole preceding in the northern hemisphere is 22 years, just twice the interval between sunspot maxima.

More refined investigations have shown that weak magnetic fields are widespread over the solar surface. At times, these fields, which are much weaker than those of the spots, and generally only a few thousandths of the field in extremely large spots, give the impression that the sun as a whole is a magnet like the earth. (*See* GEOMAGNETISM.) Continuous observation of the sun's weak fields has shown that these vary in a complicated, apparently disorganized fashion, and do not define a magnetic field for the sun similar to earth's.

The effect on a spectrum line of relative motion of the source and observed along the line of sight (*see* DOPPLER, CHRISTIAN JOHANN; SPECTROSCOPY, ASTRONOMICAL), has had many and varied applications. The amount and direction of displacements of the spectrum lines relative to their normal positions indicate velocities of approach or recession.

Spectroscopic measurement of line shifts in stellar spectra has been used to find the speed of the sun's motion among the stars. Analysis of the line displacements shows that the solar system is moving at a speed of 12 mi. per second with respect to the local stellar system.

Another application to the sun consists in the observation of line shifts in spectra of the east and west limbs of the sun. These show that the east limb is approaching the earth and the west limb receding. The deduced values of the speed of rotation agree very well with those indicated by the spots, and the spectroscopic observations can be extended beyond the narrow belts to which the spots are confined. Similar spectra can be used for determining what lines in the solar spectrum properly belong to the sun, and which are produced by absorption in earth's atmosphere. The atmosphere of the earth contains much water vapour and other constituents that produce many absorption lines. These are not only intermingled with the solar lines, but are crowded so closely together in long ranges of the spectrum that no sunlight at all reaches the earth in these wave lengths. Such lines, however, occupy identical positions in the spectra of the two solar limbs and are therefore readily distinguished from the displaced lines of the sun.

One of the most difficult and important applications of the shifts arising from motions along the line of sight seeks to detect atmospheric currents in the sun. Solar storms sometimes occur, of incomparably greater fury than the hurricanes and tornadoes of the earth, but quite inconsequential as far as the sun is concerned, and the stormy motions of the solar atmosphere are revealed by distortions and displacements of the spectrum lines. Prominences are sometimes ejected from the sun as a result of such activity, and bits of the solar atmosphere can be projected from sun to earth, where they may produce auroras and magnetic storms. However, there are systematic movements which have been observed, and of these perhaps the most interesting are those occurring near sunspots. At the lower atmospheric levels, gases move upward and outward from a spot, while in the higher levels the movements are inward and downward, as if a spot were a sort of whirlpool into which the high-level gases are drawn. Indeed there is definite evidence of high-level circulatory movements around the axes of spots, which points strongly to the same conclusion.

Away from the spots, the solar atmosphere is in a constant, random, bubbling motion in rather strong contrast to the organized movement enforced by a sunspot region.

SPECTROHELIOGRAPHIC INVESTIGATION

The spectroheliograph is an instrument that uses a spectroscope as a light filter to reject all radiation but that in the centre of a dark absorption line or other small regions of the spectrum. The radiation transmitted by a spectroheliograph is used to form a monochromatic picture of the sun. (For construction and method of operation of this instrument *see* SPECTROHELIOGRAPH.) A variation of the spectroheliograph, the spectrohelioscope, can be employed for direct visual examination of the sun using light from any selected wave length in the visual part of the spectrum.

Somewhat similar photographs and views of the sun can be obtained through polarizing monochromatic filters, or multilayer interference filters. These devices are used to select small regions of the spectrum, and may be properly regarded as forms of the spectroscope in the general sense. The filters transmit a complete picture rather than dissecting it into lines and then reassembling it as is the case in the spectroheliograph, and they have found wide application in the production of monochromatic solar pictures.

The observation of the solar atmosphere with the instruments mentioned points the way to a fairly complete knowledge of the appearance of the sun in the radiation of various spectral lines and regions of the continuous spectrum. The observations verify and extend the evidence of the complicated structure of the sun which the ordinary telescope provides. The upper part of the sun's atmosphere is composed of minute prominences that form, disappear and re-form repeatedly, so that on the average the monochromatic view of the sun shows a rather permanent mottling. In the region of spot groups, bright-plage regions, the

monochromatic aspects of the faculae, are revealed, and the spectroheliographic studies have shown that sunspots always appear only in the plages.

The plages form before the sunspots and persist long after the spots have vanished.

Solar flares, the sudden, short-lived brightenings of small parts of the solar surface, also appear only in plages. The word "sudden" must be emphasized in the definition of a flare since the larger plage regions often change brightness at a slow rate. Among the markings on the sun's surface, revealed by monochromatic observation, flares are the most significant from the standpoint of the direct effects of the sun on earth. The outbreak of a flare, visible in hydrogen light, and accompanied exactly at the start of the visible disturbance by a sudden large increase in the solar radiation in the radio region of the spectrum, often are followed by a magnetic storm on earth after a delay of one to three days. Magnetic storms are accompanied by auroral displays, and it is assumed that the flares in some way eject nuclei of hydrogen atoms (or protons) and electrons and these, traveling at rather modest speeds of 400 to 600 mi. per second, impinge upon earth's atmosphere about 50 hours after the flare outbreak to produce the observed terrestrial disturbances. The connection between solar flares and the electrical equilibrium of the earth's ionosphere has been established so firmly that observations of the flares are important elements in the prediction of favourable or unfavourable conditions for the sending of long-distance radio messages. *See* also INTERNATIONAL GEOPHYSICAL YEAR.

INVESTIGATION OF THE SUN'S RADIATION

The Sun's Total Radiation.—Of the 3.9×10^{33} ergs emitted by the sun every second in the form of electromagnetic radiation, its attendant planets and their satellites receive about 1 part in 120,000,000 so that the significance of the process is primarily cosmic, even though of dominant importance for the thermal environment of the earth. The first step in seeking further knowledge of this fundamental process is the measurement of the rate of radiation with as great precision as possible. Various instruments must be used in this endeavour, since one must measure X-rays, ultraviolet radiation, light, infrared radiation and radio emissions.

Only the radiation received by a small area of the earth can be measured, but, knowing the dimensions of the solar system, the total amount can be deduced. A general method is to convert the sun's radiation into heat, which can be accurately measured making an allowance, estimated from separate investigation, for absorption by earth's atmosphere. The result is called the solar constant: the amount of energy which would fall perpendicularly per square centimetre per minute on a surface placed outside the earth's atmosphere at the earth's mean distance from the sun. The value of this constant is about 2.0 cal.

The results of measurement of the solar constant can be combined with similar laboratory measurements for obtaining an estimate of the sun's temperature. Extrapolations from experiment have shown that an ideal emitter of electromagnetic radiation—a black body—would radiate energy at a rate strictly proportional to the fourth power of its temperature. The assumption that the sun radiates like a black body is an oversimplification, yet it makes possible a temperature estimate which is an important clue to solar radiative processes. Since the sun is visible, energy is flowing out of it from hotter to cooler regions, and it cannot be uniform in temperature. It is necessary to consider what part of the sun has the temperature determined from the black-body assumption. The greatest part of the radiation is in the form of light, hence this region must be the photosphere: a relatively thin, superficial, layer of the sun, and the temperature estimated from the total radiation is that of this layer. The temperature is called the effective temperature, and has a value of 5,800° K. (absolute centigrade degrees).

The Sun's Analyzed Radiation.—The same instruments that are used in the measurement of the sun's total radiation can be employed in other ways to provide data for a slightly different method of estimating a temperature of the sun. The radiation from a black body is distributed among the various wave lengths in its continuous spectrum in a manner definitely related to the temperature. Therefore, if a photoelectric cell be passed along the solar spectrum through the visual region, and the amount of energy at each point in the continuous background be measured, the distribution of energy throughout the spectrum is obtained and the absolute temperature follows. (Allowance must be made for absorption by earth's atmosphere, which is not uniform for all wave lengths.) There are several methods of analyzing such observations to derive a temperature, and values are obtained in the range 4,500° to 6,300° K. (absolute). If the sun radiated exactly like a black body, the several temperatures would agree with one another and with the result obtained from the solar constant. The slight departures from agreement indicate that the sun is not quite in the ideal equilibrium demanded for a black body, but is sufficiently close to ensure that the physical phenomena near its surface are similar to those expected in an enclosure at a temperature of 5,000° to 6,000° K.

Although the assumption of black-body conditions provides a close representation of measurements of radiation made with ordinary and infrared light, there are two regions of the spectrum which are extraordinary in appearance and show large differences from expectations based on a solar model radiating perfectly at a temperature of 5,000° K. The two regions are: in the extreme ultraviolet, where observations must be made from rockets far above earth's atmosphere; and in the radio region, where special types of antennae and receivers make investigation possible from earth's surface.

The Sun's Radiation in the Extreme Ultraviolet.—In the extreme ultraviolet (wave length less than 0.00015 mm.), the continuous background of the sun's spectrum is undetectable and only bright lines appear. The principal line is the fundamental line of hydrogen. The character of this line indicates that it is formed in a region of the chromosphere where the temperature is increasing rapidly with height above the photosphere. Other bright, high-temperature lines are present to support this deduction. Thus the chromosphere, which is the source of the bright lines in the extreme ultraviolet part of the solar spectrum is a most unusual region, in which the temperature increases with distance away from the sun.

The Sun's Radiation in the Radio Region.—The result derived from the observations in the extreme ultraviolet is confirmed by measurements of the sun's radio radiation. In the ultraviolet part of the solar spectrum the wave lengths of solar radiation are less than 0.00015 mm., in the radio region the wave lengths are in the range 5 mm. to 10,000 mm. The shortest wave lengths originate near the base of the chromosphere, but the longest can only come from far out in the corona. Energy observed at the various wave lengths can be equated to the energy radiated from a black body of the same size as the sun's visible disk in order to derive an apparent temperature. Apparent temperatures, based on measurements of the radio energy received from the sun, range from about 6,400° K. for the shortest waves originating in the chromosphere near the sun's visible surface, to 1,200,000° K. for wave lengths of tens of metres originating entirely in the solar corona more than 1,000,000 mi. above.

On earth lightning flashes are the only natural sources of radio waves, but there are at least five different kinds of radio energy generated on sun: (1) The continuous, nearly invariable component from which the apparent temperatures are computed. (2) The continuous and steady contribution that is associated with the presence of plages on the sun's visible surface. (3) Erratic increases in energy of very short duration, but following closely upon one another, often so closely that they overlap. These may be accompanied by an increase in the steady component. (4) Sudden, very great, nonrecurrent increases beginning at the start of solar flares and associated with them. (5) Occasional isolated increases of very short duration. Sometimes the variable portion of the sun's radio energy quite overwhelms the continuous background. This is completely different from the situation in the visible spectrum where the variable features are difficult to detect. *See* also TELESCOPE, RADIO.

THEORIES OF SOLAR STRUCTURE AND EVOLUTION

Structure of the Sun.—A mathematical model of the structure of the sun can be based on data from observations similar to those described above, and data from terrestrial physics laboratories dealing with the behaviour of matter at temperatures, pressures and densities that range from the extremely high to the almost vanishingly small. According to these models, the temperature at the centre of the sun is about 20,000,000° K. Going out from the centre of the sun toward the surface about 4,200° K. is reached at the very top of the photosphere. In the chromosphere the temperature starts to rise, at first very slowly, and then more rapidly until at the beginning of the corona about 9,000 mi. above the photosphere, the temperature is 500,000° K. In the outer parts of the corona, 1,500,000 mi. from the centre of the sun, the radio observations are best fitted by a temperature of 1,500,000° K. Pressures start at 6,000,000,000,000 lb. per square inch (p.s.i.) at the centre; decrease to about atmospheric (15 p.s.i.) near the visible surface, and become inappreciable in the upper chromosphere and corona. The mathematical model indicates that 99% of the mass of the sun is concentrated within 0.6% of its radius, hence the density of the sun decreases rapidly away from the centre.

Only the most elemental forms of matter can exist at the temperatures encountered near the centre of the sun. Atomic nuclei must be stripped almost bare of their satellite electrons and the rates of motion of the elementary particles must be in the range 100 to 1,000 mi. per second. All of the bits of matter, electrons, neutrons and protons, that exist at the sun's centre participate in a process of continual interchange of energy between radiation and motion of particles. But there is an important difference in the behaviour of the atoms and electrons on one hand and the radiation on the other. The motions of the particles are kept within a limited range by the gravitational equilibrium of the whole sun, while electrostatic forces preserve a constant proportion between the number of nuclei and the number of electrons in each region. Radiation, however, is not so controlled. It works its way from the centre, where it is most intense, out to the surface and thence to space at the observed rate of 3.8×10^{33} ergs per second. The sun loses radiation at this prodigious rate, and has been doing so for many millions of years, while the electrons and atomic nuclei remain chained within its boundaries.

Source of the Sun's Radiation.—It has been realized for many years that the observed rate of energy loss by leaking of radiation from the solar surface must be balanced exactly by the rate of energy generation in the interior, since there is little evidence that the sun is heating or cooling. Because other sources of energy are insufficient, astronomers have sought a series of nuclear transformations which would liberate radiation at the required rate. The transformation of 1% of the sun's mass from hydrogen to helium supplies enough energy to keep it shining for fully 1,000,000,000 years. There are other conditions that must be satisfied in the selection of possible energy sources. The sun must not only be in balance as regards the flow of radiation; it must simultaneously be in mechanical equilibrium. The pressure, which is primarily determined by the temperature, must exactly balance the weight of the overlying layers at every point throughout the sun. Restrictions such as these on the possible choices have directly led to the conclusion: fusion between hydrogen nuclei, or protons, to produce helium is the source of the sun's energy.

The discovery of a satisfactory means of generating the required enormous energies and advances in knowledge of nuclear reactions have made possible the computation of a probable course of evolution for the sun.

Mathematical Models of Detailed Features of the Sun.—The general concepts involved in the computation of plausible models for the interior of the sun and the choice of a method of energy generation, if complete, would imply the existence of sunspots, the chromosphere and the prominences, the plages and the flares, the corona, the weird law of solar rotation, the magnetic fields—in short, all of the detailed and transient features that still require continuous observation. Considerable progress has been made in the study of solar details, but mutually consistent theories for the different structures do not exist. Perhaps a reason can be found in the realization that the observed deviations from equilibrium represent trifling phenomena on a solar scale, despite their remarkable effects on earth, and the idea of the sun as the perfectly calm, unchanging star of the mathematician and poet is an amazingly good approximation to the truth. *See* also STAR.

BIBLIOGRAPHY.—General accounts of the sun are contained in Donald H. Menzel, *Our Sun* (1949); Giorgio Abetti, *The Sun* (1957); M. A. Ellison, *The Sun and Its Influence* (1955); for more technical treatment *see* G. P. Kuiper (ed.), *The Sun* (1953); M. Waldmeier, *Ergebnisse und Probleme der Sonnenforschung* (1953); A. Unsöld, *Physik der Sternatmosphären* (1955). (H. D.; O. C. M.)

SUNBIRD, the name of a group of small birds of about 104 species forming the passerine family Nectariniidae. They live in the old world, chiefly the tropics; Africa is richest in species, then southern Asia and nearby islands, and one species reaches Australia. They are forest-, brush- and even garden-haunting birds. The males of many species have brilliant colours; iridescent blues, purples and greens with patches of yellow, red and black are common. However, some species are dull coloured, as the females usually are. Sunbirds flit about actively, perching among the blossoms and probing into the flowers for their food of small insects and nectar. For this they have long, slender, curved bills and extensible, more or less tubular, tongues. The usual type of nest is an oval, pensile structure attached to a twig. The entrance is at the side near the top, over which a porchlike roof projects. The eggs are usually spotted and two in number. (A. L. RD.)

SUN BITTERN (*Eurypyga helias*), a bird of Central and South America placed in a family by itself, Eurypygidae. It is distantly related to such birds as the kagua of New Caledonia, rails and cranes, and is grouped with them in the order Gruiformes. It is a slender bird about 17 in. long, somewhat raillike but with full wings, a long, ample tail and a plumage variegated with browns, yellows, black and white. It lives on the ground in wet forests and feeds on insects. Its display, involving spread wings and tail, is striking. (A. L. RD.)

BY COURTESY OF NEW YORK ZOOLOGICAL SOCIETY

SUN BITTERN (EURYPYGA HELIAS)

SUNBURY, the seat of Northumberland county, Pa., U.S., is an industrial and agricultural centre on the Susquehanna river, 40 mi. N. of Harrisburg, on the site of an old Indian village. Sunbury was surveyed in 1772, incorporated as a borough in 1797 and became a third-class city in 1922. Ft. Augusta, erected 1756, where the French were defeated during the French and Indian War, and the nearby home of Joseph Priestley, the discoverer of oxygen, are tourist attractions. In July 1883, Thomas Edison began operating in Sunbury the world's first three-wire central electric-lighting station. For comparative population figures *see* table in PENNSYLVANIA: *Population.* (A. J. FO.)

SUNBURY-ON-THAMES, an urban district of Middlesex, Eng., which under the London Government Act 1963 was to become part of Surrey on April 1, 1965 (*see* LONDON). It is a riverside resort and residential district 17 mi. S.W. of St. Paul's cathedral and it includes Shepperton, Littleton, Charlton and part of Ashford common. Pop. (1961) 33,403. Area 8.8 sq.mi. There are pumping works and filtration beds for London's water supply which belong to the Metropolitan Water board. The reservoirs at Littleton are very extensive. To the northeast is Kempton park, the manor house of which was a royal residence in the 14th century. The park is famous for its race meetings, the principal fixture being the Jubilee handicap, established in 1887. There are locks on the river at Sunbury and Shepperton, and at each place regattas are held during the summer.

SUN CHUANG-FANG (1885– ?), one of the group of men who rose to prominence and obtained a brief tenure of power in China during the years of strife which succeeded the death of Yuan Shih-kai. Born in Lingcheng in the province of Shantung, he

received military training at the Peiyang college and in Japan. In 1921 he obtained command of a division under Wu Pei-fu, and two years later he was appointed military governor of Chekiang. During 1924 and 1925 he extended his authority over Kiangsu and, after defeating his northern rivals, announced the consolidation of five provinces, Kiangsu, Chekiang, Anhwei, Kiangsi and Fukien, into an independent state under himself as governor general. He was the ablest and the most enlightened of the *tuchuns*, travelling continually through his territory to see for himself that the people were fairly treated.

He was on friendly terms with the foreign authorities of Shanghai, openly expressed shame at the backward state of Chinese cities and pleaded for foreign co-operation. In May 1927 he was swept aside in the advance of the Nationalist army from Canton to the Yangtze and fled to Peking. In 1928 he fought bravely but unsuccessfully at Hsuchowfu against Chiang Kai-shek's march on Peking. In 1933 he became a Buddhist monk. Later it was widely reported that he had been assassinated by a woman whose father he had executed.

SUNDA ISLANDS AND STRAIT. The Sunda, or Soenda, Islands extend from the Malay peninsula to the Moluccas and include the Greater Sunda Islands (Sumatra, Java, Borneo, Celebes and adjacent smaller islands) and the Lesser Sunda Islands (Bali, Lombok, Sumbawa, Sumba, Flores, Timor, Alor, Wetar, etc.). With the exception of Borneo, eastern Sumatra and nearby areas, they belong to the geologically unstable zone of island arcs, undersea ridges and active volcanism lying between Asia and Australia. The Sunda Islands include most of the land area of Indonesia. Only northern and northwestern Borneo and eastern Timor are not under Indonesian political control. Malaysian cultures and languages predominate in the area.

Sunda strait separates Java and Sumatra. At its narrowest point the small, low-lying island of Sangiang divides it into two channels, each about four miles wide. Although overshadowed in commercial and strategic significance by Malacca strait, it remains one of the two important passages uniting the Indian ocean and eastern Asia.

The spectacular eruption of Krakatau (*q.v.*) took place in Sunda strait in 1883.

See also INDONESIA, REPUBLIC OF and articles on the individual islands. (J. M. BT.)

SUN DANCE, an annual ceremony of Plains Indians (*q.v.*) (named "sun-gazing dance" by Dakota, "thirsting dance" or "offerings lodge" by other tribes). The ceremony was common to a score of tribes from Saskatchewan to Oklahoma and Texas. Late in the 19th century it was adopted by others west of the Rockies. Although each performance was initiated by an individual who, in distress, had vowed to dance for supernatural aid, it was a tribal affair. Ordinarily held once a year in early summer, it might be repeated for later pledgers.

The core of the performance was a dance of the pledger before a forked pole, joined by other votaries, until in an exalted state they obtained visions. The fasting supplicants danced hour after hour, staring at the sun, the source of "power." In self-mortification bits of their flesh or finger joints were cut off as offerings; among tribes central in the region—Dakota, Ponca (Ponka)—thongs stretched from the fork were attached to skewers thrust through the breast muscles of the dancers, against which they strained until they tore loose. Nearly all tribes erected a large circular enclosure, its rafters radiating from the centre pole; opposite the lodge's entrance was an altar on which were placed sacred "medicine bundles." At intervals the pledger's warrior society might dance to "help."

These features were elaborated by accretion of other rites and by ritualization of essentials, particularly by Arapaho, Cheyenne and Dakota: the centre pole was raised by invoking magic aid; pledgers feinted at entering the lodge four times, etc. Organization and motivation of the rite varied. By their dance Ponca pledgers joined the fraternity that owned the ceremony; Dakota self-torturers were candidates for rank among shamans; the Crow pledger sought specifically a vision assuring revenge against an enemy. Although focused on the individual pledgers, the ceremony assured blessings to all tribespeople. Proscribed by the U.S. government in 1904 because of its barbarities, its performance was resumed in the 1930s in tempered form as part of local Fourth of July festivities.

See L. Spier, "The Sun Dance of the Plains Indians," American Museum of Natural History *Anthropological Papers*, vol. 16, part 7 (1921). (L. SP.)

SUNDANESE. This powerful group, of about 12,000,000 persons, resides primarily in the highlands of southwestern Java. They are Proto-Malays much modified by later comers to the island. In their highland homes they were less influenced than the Javanese by Indian and Arabian contacts. They came under the rule of the Hindu-Javanese states and accepted Buddhism but were not stimulated to erect great shrines to that faith as was done in mid-Java. Later they accepted Islam but again without the great enthusiasm of their neighbours. Beneath this overlay many ancient customs and beliefs still survive.

Their culture is closely related to a widespread complex of Malayan traits. Stripped of imposed political units, the important centre of Sundanese life is the village ruled by a headman and a council of elders. Single-family houses are made of wood or bamboo and are raised on piles. Close to the settlement are individually owned lands, much of them devoted to terraced rice fields. Rice culture, ironworking and customs relating to marriage, birth and death ceremonies conform closely to the Malayan pattern although often mixed with elements of the more elaborate Indian ceremonies.

Opening of roads into the highlands, development of government and foreign-owned plantations and the establishment of village schools have all tended to erase many of the differences between the Sundanese and their neighbours. In recent years they have spread into neighbouring districts of Java and into the Lampong area of southern Sumatra.

BIBLIOGRAPHY.—A. Cabaton, *Java, Sumatra and Other Islands of the Dutch East Indies* (1911); A. F. Kramer, *West Indonesien* (1927); P. S. Veth, *Java*, 4 vol. (1896–1907); H. G. Q. Wales, *The Making of Greater India* (1950). (F.-C. CE.)

SUNDANESE LANGUAGE, like most of the languages of Indonesia, belongs to the family of Malayo-Polynesian languages; it is most closely related to the languages of western Indonesia, such as Malay and Javanese. It is the native language of more than 12,000,000 Indonesians and has a larger number of speakers than any Malayo-Polynesian language except Javanese. Most of its speakers are in West Java, a section, also known as Pasundan, which occupies approximately the western third of the island. The dialect with the highest prestige is that of Priangan (or Preanger). Like Javanese, Sundanese has status styles called *kasar* (informal), *lemes* (deferential) and *panengah* (like the Javanese *madya*, or "middle" style).

The Javanese (or Javanese-Sundanese) alphabet, the earliest used to write Sundanese, continues in use. An adapted Latin alphabet is also used. A modified Arabic alphabet, common in religious writings, is also employed for other purposes. Publication in Sundanese has been very limited; at mid-20th century, for example, only one newspaper was being published in Bandung. Sundanese literature is not very extensive. It is strongly affected by its Javanese counterpart; nevertheless it has independent elements in its myths, legends and humorous stories. (I. DN.)

SUNDARBANS (SUNDERBUNDS), a tract of country in Bengal, forming the seaward fringe of the Gangetic delta. It extends for about 170 mi. along the coast of the Bay of Bengal, from the estuary of the Hooghly (India) on the west to that of the Meghna (East Pakistan) on the east. It is intersected from north to south by large tidal rivers or estuaries, which are connected by numerous interlacing channels. The whole tract is a network of estuaries, rivers and creeks, which enclose a large number of flat, marshy islands. Many are covered with forest and a dense undergrowth standing in soft mud, half in and half out of the water. Formerly it ran inland from 60 to 80 mi. Some land was reclaimed, however, for cultivation. The characteristic tree is the sundari (*Heritiera littoralis*), from which the name of the tract has probably been derived. Along the coast the forest is composed of mangrove swamp, which in some places is separated from the sea by a line of sand

dunes. The population is sparse. South of the northern area of cultivation, the Sundarbans are practically uninhabited. The chief wild animals are tigers, of which many are man-eaters, leopards, deer and wild pig; crocodiles infest the estuaries.

SUNDAY, WILLIAM ASHLEY (BILLY) (1862–1935), U.S. evangelist, whose sermons gave vivid expression to the emotional upheaval caused by the United States' transition from a rural to an industrial society, was born at Ames, Ia., on Nov. 19, 1862. The son of a brickmason who died a private in the Union army (1862), Sunday spent four years in an orphans' home, grew up on a farm, completed high school and worked as an undertaker's assistant before entering professional baseball in 1883. He gave up baseball in 1891 to become a Y.M.C.A. worker. From 1896 until his death he conducted religious revivals in major cities throughout the United States. He was ordained in the Presbyterian Church in 1903. His theology was Fundamentalist.

Following in the footsteps of Dwight L. Moody (*q.v.*), and with the aid of a choir leader, Homer A. Rodeheaver, and a score of revival "specialists," Sunday conducted more than 300 revivals with an estimated total attendance of 100,000,000. Sunday claimed that 1,000,000 "hit the sawdust trail" to profess their conversion; he received more than $1,000,000 in "free-will offerings." Despite his sensationalism, Sunday was enthusiastically supported by the evangelical churches and by influential laymen. He played a prominent part in the prohibition movement and reached the peak of his fame in 1917 in his New York city revival. He died at Chicago, Ill., on Nov. 6, 1935.

See also REVIVAL, RELIGIOUS.

BIBLIOGRAPHY.—Elijah P. Brown, *The Real Billy Sunday* (1914); William T. Ellis, *Billy Sunday, the Man and His Message* (1936), which contains Sunday's brief autobiography; Homer A. Rodeheaver, *Twenty Years With Billy Sunday* (1936). The only analytical study of Sunday's career is that of William G. McLoughlin, Jr., *Billy Sunday Was His Real Name* (1955). (W. G. MCL.)

SUNDAY, the first day in the week, is the day particularly dedicated to worship in Christendom, in memory of Christ's Resurrection. Many Christian states, after the time of Constantine, supplemented the teaching of the church by enjoining Sunday rest from labour (except agricultural). The code of Justinian contained provisions to this effect.

English Laws.—In England prohibition of certain forms of work and pastimes has been continuous since the 7th century A.D., though it was not until the reign of Edward VI that Sunday observance was positively ordained, by the Act of Uniformity of 1551. By the statute of 1558–59 everyone had to go to church on Sunday or else be liable not only to the censures of the church but to a fine of 12*d*. This penalty was not formally repealed until 1846. In a highly abstract sense, the law of England still requires members of the Church of England to attend divine service on Sunday, but this has not been enforced for generations and the only importance of the laws of Sunday observance now lies in the negative restrictions they impose.

By the Sunday Observance act, 1677, tradesmen, artificers, labourers "or other person whatsoever" are forbidden to carry on their ordinary businesses. Acts have been passed to obviate some of the inconveniences in modern life caused by the 1677 act, which, however, did permit the sale of milk before 9 A.M. and after 4 P.M., and of victuals in cookshops and fried fish shops (*Bullen* v. *Ward*, 1905, 14 L.J.K.B., 916). As a result of an act passed in 1871 (34 and 35 Vict. c. 87), which makes difficult the prosecution of offenders under the 1677 act, the Lord's Day acts have been virtually abolished insofar as they refer to Sunday trading. The Factory and Workshop act (1901) forbids the employment of women, young persons or children on Sunday in a factory or workshop (s. 34). The Shops acts, 1912–22, tacitly imply that the 1677 act is still in general operation with regard to shops. A person of Jewish religion may carry on business on Sunday provided that: he does not do so on Saturday; he gives previous notice to the local authority; he posts a notice in his shop. Under the Licensing acts the sale of intoxicating liquor on a Sunday is confined to five hours. Fishing is permissible on a Sunday, except that one may only fish for salmon with rod and line (Salmon Fishery act, 1861). By the Game act, 1831, it is a crime to take or kill game (but not rabbits) on a Sunday. The law with regard to Sunday entertainments and performances is based upon the Sunday Observance act, 1781 (21 Geo. III c. 49), which prohibited public entertainments for payment, until its modification by the Sunday Entertainments act of 1932. This measure excludes the application of the 1781 act to any place used on Sundays for the purpose of a cinematograph or musical entertainment for which it is allowed or licensed to be used, or to any museum, picture gallery, zoological or botanical garden or aquarium, and to any place at which a lecture or debate is held on Sunday. Thus in effect there is a large measure of local option on this question.

Sunday is not considered to be a suitable day for the sitting of the courts or meetings of public bodies, though parliament has sat on Sunday in national emergencies, and Saturday sittings have been extended into the early hours of Sunday. Process may not be served nor persons arrested except in cases of treason, felony or breach of the peace (1677 act), but a justice may issue an arrest or search warrant. Though at common law a contract made on a Sunday is valid, the technical restrictions imposed by statute on Sunday labour, especially those contained in the act of 1677, render most contracts made and completed on that day void or voidable (*Smith* v. *Sparrow*, 1827, 4 Bing. 84). But any contract in respect of labour, business or work not done in the course of a man's ordinary calling is binding. A check drawn on a Sunday is binding on the drawer; in an action on a bill of exchange drawn on a Sunday the drawer was given judgment against the acceptor (*Begbie* v. *Levy*, 1830, 1 C. and J. 180). The position in regard to these matters has been laid down clearly by the Bills of Exchange act. A writ may not be served on a Sunday. Where any limit or period of time fixed in the high court exceeds seven days, the Sundays included (except in divorce proceedings) are counted in computing the time, provided that if the last day fall on a Sunday, the period must be computed so as to cover the next following working day. But where the period fixed is less than six days Sunday is not counted.

In Scotland and the British Commonwealth the laws as to Sunday have followed more or less the same course, except that, in the latter, native customs have been interfered with as little as possible and particular acts have been passed to meet special needs.

(W. LA.; W. T. WS.)

U.S. Laws.—Sunday legislation was quite general in the colonial period, Virginia having enacted such a law as early as 1629. Substantially all of the states have enacted statutes prohibiting certain types of activity on Sunday. In general, they prohibit work, labour, the transaction of business or the pursuit of one's ordinary calling (works of "charity" and "necessity" excepted), together with certain forms of amusement or recreation otherwise innocent. The validity of such laws has often been questioned, generally without success, on the ground that they infringe constitutional guarantees of religious freedom. In 1858 the California supreme court held a Sunday observance law unconstitutional on the ground that it sought "to enforce, as a religious institution, the observance of a day held sacred by the followers of one faith" (Ex parte *Newman*, 9 Calif. 502, 505), but three years later it reversed its position (Ex parte *Andrews*, 9 Calif. 678, 1861). Since then, the courts have upheld such laws with substantial unanimity, at first on frankly religious grounds but later on the ground that to prescribe periodic days of rest from customary labour is a legitimate exercise of the legislative power to provide for the physical and moral welfare of the community.

Theoretically, violation of a Sunday observance law may have four effects: it may subject the offender to the criminal penalties prescribed by the statute; it may make unenforceable contracts, including negotiable instruments, made or to be performed on Sunday; it may subject the offender to a civil disability, as by depriving him of the right to recover damages for injuries negligently inflicted on him while engaged in the prohibited activity; and it may impose liability on him for injuries inflicted by him in the course of the prohibited activity, although he is otherwise free from legal fault. The criminal enforcement of such laws is, however, notoriously lax. The defense of illegality based on such statutes is not often invoked in practice, and when invoked it may

fail for technical reasons. At one time a few states (notably in New England) denied the right of recovery to persons injured while traveling or working in violation of Sunday laws, on the theory that the offense was a contributing cause of the injury; but the evident injustice of thus augmenting the penalty brought about legislation abrogating the anomaly. The Vermont supreme court held that one hunting on Sunday in violation of the law is liable in a civil action for damages to his companion, for having injured him accidentally by the discharge of his gun (*White* v. *Levarn,* 93 Vt. 218, 1918), but other courts indicated a contrary view. In some states contracts made or to be performed on Sunday are expressly declared unenforceable, so that invalidity may result even though there is no violation of a criminal law.

By mid-20th century, Sunday laws had been relaxed in many states and cities, and liberal interpretation of statutory provisions exempting certain classes of activity from their scope had combined with this to produce conditions quite different from those which might be presumed from the superficially puritanical tenor of the laws. (B. Ce.)

SUNDAY SCHOOLS. Late in the 18th century the educational problems arising from the Industrial Revolution led to a new religious effort in the form of the Sunday school movement. Robert Raikes (1736–1811) established his first Sunday school in Gloucester, Eng., in 1780. The idea, however, did not originate with Raikes. Among earlier pioneers in this field were Joseph Alleine, the Puritan Father, who founded Sunday schools in England in the 17th century, and John Wesley, who held Sunday classes in Savannah, Ga., in 1737. Yet, by reason of his achievements in organization, Raikes is justly regarded as the founder of the English Sunday school. In its early days the Sunday school was of especial educational value because of its combination of secular and religious instruction; sometimes sessions were held on Saturday as well as Sunday. But with improved economic conditions and the establishment of day schools the Sunday schools became more restricted to religious instruction. Early in the 19th century the Sunday school movement became firmly established in the U.S., and expanded rapidly following the organization in 1824 of the American Sunday-School union and again after the introduction in 1872 of the international uniform Sunday school lesson system. From Great Britain and the U.S., Sunday schools have spread widely in other parts of the world.

SUNDERLAND, CHARLES SPENCER, 3RD EARL OF (1674–1722), English statesman, promoted the Hanoverian succession. He was the second son of the 2nd earl, but on the death of his elder brother (1688) he became heir and in 1702 succeeded to the peerage. He married in 1700 Anne Churchill, daughter of the famous duke of Marlborough.

Sunderland met George I in 1706 and when the elector became king was made lord lieutenant of Ireland. In Aug. 1715, he joined the cabinet as lord keeper of the privy seal, and in March 1718, became first lord of the treasury, but his connection, though slight, with the South Sea Bubble led to his political ruin. In April 1721 he resigned. He died in London on April 19, 1722.

SUNDERLAND, ROBERT SPENCER, 2ND EARL OF (1641–1702), English politician, turned with the political tides under Charles II, James II and William III. He was the only son of Henry Spencer (1620–43), 1st earl of Sunderland. After various diplomatic posts, he became a privy councilor in 1674 and secretary of state in 1679. He voted for the exclusion of James, duke of York, from the throne, and made overtures to William, prince of Orange, and consequently in 1681 he lost both his secretaryship and his seat on the privy council. Early in 1683, however, through the influence of the duchess of Portsmouth, he returned to office. When in Feb. 1685 James became king, Sunderland retained his position of secretary, to which was soon added that of lord president of the council. But although in 1688 he openly embraced the Roman Catholic faith, he grew hesitant in his support of the king's policies and in Oct. 1688 he was dismissed. Sunderland took refuge in the Netherlands, but in 1691 he was permitted to return to England. He declared himself a Protestant and gained a considerable influence with William III. He died at Althorp, on Sept. 28, 1702.

SUNDERLAND, a seaport, municipal, county and parliamentary borough situated at the mouth of the Wear river in County Durham, Eng., 13 mi. N.E. of Durham by road. Pop. (1961) 189,629. Area 13.4 sq.mi. It is a great shipbuilding centre (shipbuilding was recorded as early as 1346) and has also big ship-repairing and marine-engineering works. Coal was first shipped from the port in 1396 and the tonnage exported per year from Wearmouth and other collieries had reached 5,000,000 before World War II. There are several foundries and the largest ships' anchors are made there. Other manufactures include telecommunication equipment, ropes, aircraft engineering, furniture, clothing, confectionery, pottery (the Sunderland lustreware was an early product), glass, beer and many more.

The Wear river commissioners administer the port of Sunderland which has a well-equipped deep-water quay and several wet and dry docks. The two curved piers at the entrance, on one of which stands a lighthouse, enclose about 130 ac. In addition to coal and coke Sunderland exports iron and steel, machinery, paper, coal tar, pitch, glassware, etc. The main imports are timber, grain, sisal and hemp, esparto, iron ore, wood pulp, fuel oil and petroleum. Within the borough are the seaside resorts of Roker and Seaburn, with fine sands.

In Sunderland there are a technical college, affiliated to Durham university and specializing in electrical engineering and other subjects, a teachers' residential training college and a college of art. There is a museum and art gallery.

History.—In 674 a monastery was established on the north side of the river by St. Benedict Biscop and a church built, dedicated to St. Peter. The church is still in use and retains parts of the original building. The Venerable Bede was born at Wearmouth about 673 and studied at the monastery. Biscop brought over skilled Frankish workmen who made glass for the windows of the church—the first recorded instance of glassmaking in England.

The area in which the monastery was located became known as Monkwearmouth. Sunderland was the part "sundered" from the monastery by the river. Bishopwearmouth, immediately west of Sunderland, had been granted in 930 to the bishops of Durham, by Aethelstan. In 1382 there was probably a dock there and from the 16th century Bishopwearmouth seems to have been completely identified with Sunderland. In 1567 Wearmouth was one of the three ports in Durham where precautions were to be taken against pirates, while no mention is made of Sunderland. Monkwearmouth remained purely agricultural until 1775, when a shipbuilding yard was established and prospered to such an extent that by 1795 five similar yards were at work. Prosperity continued until the depression of the 1930s caused severe unemployment—41% of the insured population in 1934.

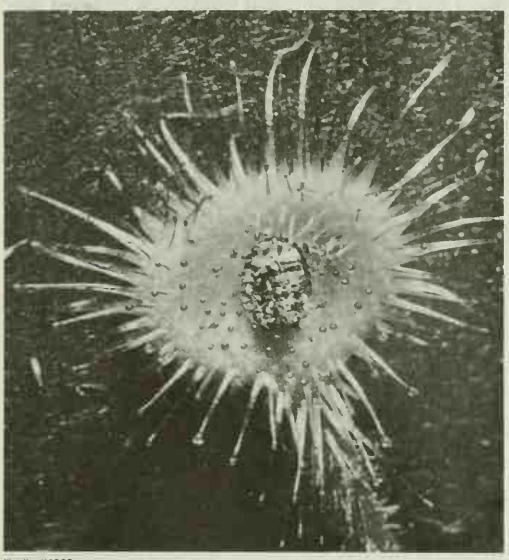
From the 9th to the 11th century the district suffered from successive raids from the Danes and the Scots. Hugh Pudsey, bishop of Durham, granted a charter toward the end of the 12th century, elevating Sunderland to borough status. In 1634 Bishop Thomas Morton granted a further charter. During the Great Rebellion Sunderland supported the parliamentary cause. Newcastle took the royalist side and an embargo on ships leaving the Tyne resulted in increased shipments from the Wear. In 1717 a board of harbour commissioners was appointed to develop the port.

In 1796 the river was spanned by what was then the longest single-span cast-iron bridge in the world (an aluminum alloy bridge opened in 1948 also created a record). Sunderland was incorporated in 1634 and became a parliamentary borough in 1832. It returns two members to parliament.

SUNDEW, an insectivorous plant of the genus *Drosera,* of which there are more than 90 species with world-wide distribution. There are four species in North America and three in Europe. The most extraordinary development of this genus is attained in Australia and south Africa, where, in addition to the low, soft, herbaceous plants found elsewhere, occur tall plants attaining a height of three feet with wiry stems which climb by the aid of adhesive leaf blades.

All species are carnivorous. They catch their prey by means of tentacles which spring in a regular pattern from the upper surface of the leaf blade. As many as 400 have been counted on a

leaf of *D. rotundifolia* 0.9 cm. in diameter. The tentacles stand distinctly separated. They vary in size, shorter in the middle of the blade and much longer (0.6 cm.) along and near the margins. In some species with linear leaves there is little or no difference in size. Each tentacle consists of a stalk, broader at the base and tapering to the top, which supports an oval or rounded gland of rather complex structure. The middle of this gland is occupied by a cylindrical mass of water vessels enclosed within a bell-



W. H. HODGE

SUNDEW (DROSERA) LEAF SHOWING AN INSECT TRAPPED BY THE DROPLETS OF MUCILAGE SECRETED BY THE GLANDS AT THE TOP OF THE TENTACLES

shaped layer of flat cells. The cells which form the edge of the bell extend to the outer surface of the gland. Covering this bell are two courses of cells normally filled with red or purplish sap. Each gland secretes and supports a droplet of mucilage, which, in multiple, accounts for the dewy appearance of the leaf. The red point of colour shining through the clear mucilage presents an appearance of rare beauty, which appears to attract insects. Connecting the gland with the vascular tissues of the leaf is a strand of vascular tissue extending through the stalk and base to the veins of the blade.

The tentacles are capable of motion. Darwin was the first to make extensive experimental studies on the matter. When a small insect alights on the leaf, its escape is prevented by the mucilage, which smears and smothers it during its struggles to escape. The presence of an insect provides a stimulus to neighbouring tentacles, which bend toward the prey and attach themselves to it by their mucilage, bringing a large contact surface to bear. A digestive ferment is secreted which attacks the nitrogenous content of the prey. The products of this digestion are then absorbed by the same glands. The movements of the tentacles are complex, but the net result is to bring the body of the prey toward the middle zone of the leaf blade, which itself can bend to surround the prey. There are also many minute sessile glands scattered over both leaf surfaces whose function is problematical. The leaf blade may be nearly orbicular, or linear and rushlike (it is then nonmotile), with various intermediate forms. Some are peltate (*i.e.*, having the stalk attached to the underside rather than to the base of the leaf) and circular, bowl shaped or lunate, others variously ovate to large strap shaped.

Reproduction takes place not only by seeds but also by leaf buds which spring normally from the bases of the tentacles, more abundantly on leaves which have been detached. In one or two species buds, in the form of gemmae, spring from the apex of the stem after the foliage leaves have been produced. Some species produce "droppers," special branches with negative geotropism, which penetrate the substrate. Their ends then enlarge into tubers. Closely related to *Drosera* are the genera *Drosophyllum* and *Byblis*, both carnivorous, and *Roridula*, once thought to be so.

See also CARNIVOROUS PLANTS.

See C. Darwin, *Insectivorous Plants* (1875); F. E. Lloyd, *The Carnivorous Plants* (1942). (F. E. L.)

SUNDIAL. A sundial is a device which indicates divisions of the day by the motion of the shadow of some object on which the sun's rays fall. It is one of the earliest examples of the application of man's knowledge of the apparent motion of the sun across the sky, and is constructed in various ways.

History.—The first device used was probably the gnomon. This consisted of a vertical stick or pillar, and the length of the shadow which it cast gave an indication of the time of day. By the 8th century B.C. it is clear that more precise devices were in use, and the earliest known sundial still preserved is an Egyptian shadow clock of green schist dating at least from this period. This clock consists of a straight base and a crosspiece set at 90° to it. The base was placed in an east–west direction and the shadow of the crosspiece fell across the base, upon which was inscribed a scale of six time divisions. Clocks of this kind are still in use in primitive parts of Egypt; they are positioned so that the crosspiece lies to the east before noon and to the west after noon.

Another early device was the hemispherical sundial or hemicycle. It is attributed to the Babylonian astronomer Berossus (Berosus, fl. *c.* 300 B.C.). Made of stone or wood, the instrument consisted of a cubical block into which was cut a hemispherical opening; to this was fixed a pointer or style, the end of the style lying at the centre of the hemispherical space. The path traced by the shadow of the style was, approximately, a circular arc. The length and position of the arc varied according to the seasons, so an appropriate number of arcs was inscribed on the internal surface of the hemisphere.

Each arc was divided into 12 equal divisions and each day, reckoned from sunrise to sunset, had therefore 12 equal intervals or hours. Because the length of the day varied according to the season, these hours likewise varied in length from season to season and even from day to day, and were consequently known as temporary hours. The dial of Berosus was widely used for many centuries, and, according to the Arab astronomer al-Battani (Albategnius, *c.* A.D. 850–930), was still in use in Moslem countries during the 10th century.

The Greeks, with their geometrical prowess, developed and constructed sundials of considerable complexity. Apollonius of Perga (*c.* 250 B.C.) developed the hemicyclium by using a surface of conic section upon which the hour lines were inscribed; this arrangement gave greater accuracy. Ptolemy used the analemma, a device which enabled shadows to be projected geometrically on to flat surfaces inclined at various angles to the horizontal. In general it appears that the Greeks constructed instruments with either vertical, horizontal or inclined dials, indicating time in temporary hours. The Tower of the Winds in Athens, which was octagonal in shape and dates from about 100 B.C., contains eight sundials. It is therefore clear that dials facing various cardinal points were in use for a long time.

The Romans also used sundials with temporary hours. The first dial set up in Rome in 290 B.C. had been captured from the Samnites, but it was not until almost 164 B.C. that a sundial was actually constructed for the city. In his great work *De Architectura*, M. Vitruvius (1st century B.C.) mentions many types of sundial, including portable ones.

The Arabs attached much importance to the study of sundials, the principle and design of which they derived from the Greeks. They increased the variety of designs available and, at the same time, simplified the processes of design and construction by using the principles of trigonometry. Abu'l Hasan, at the beginning of the 13th century A.D., wrote on the construction of hour lines on

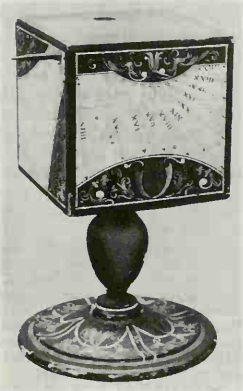
CROWN COPYRIGHT, SCIENCE MUSEUM, LONDON

FIG. 1.—PLASTER CAST OF A ROMAN HEMICYCLE SUNDIAL

cylindrical, conical and other surfaces and is credited also with the introduction of equal hours, at least for astronomical purposes. With the advent of mechanical clocks in the early 14th century it seems clear that sundials with equal hours gradually came into general use.

During the Renaissance the construction of sundials flourished; there was a great diversity of forms, and some dials of considerable artistic merit were produced. The making of sundials was a profitable business and required not only considerable manual skill but also some understanding of the theoretical principles of the apparent motion of the sun; in consequence, the methods of sundial design and construction were jealously guarded trade secrets. From the 16th century onward, however, the growing spirit of inquiry into natural phenomena caused this secrecy to give way to open discussion, and in Basel in 1531 the German cosmographer Sebastian Münster (1489–1552) published his *Compositio Horologiorum*, which contained details of sundials and also of a moondial.

By the 17th century the study of dialing, as it was then called, formed a branch of education. The German Jesuit astronomer Christopher Clavius (1537–1612) published in 1581 the *Gnomonices*, in which he virtually summed up all knowledge on dialing at that time. Universal sundials, adjustable for use in any latitude, were a feature of the 18th century, although instruments of such a kind had been constructed earlier. The increasing accuracy of clocks and watches made the sundial obsolete. By the 19th century sundials were, in general, used only as ornaments in gardens, although it is to be noted that certain instruments of a very precise kind, such as the heliochronometer, were designed and used for scientific purposes.



CROWN COPYRIGHT, SCIENCE MUSEUM, LONDON

FIG. 2.—FLORENTINE CUBICAL SUNDIAL SHOWING ITALIAN HOURS, *c.* 1560

General Principles.—Timekeeping by sundial depends on the changing position of the sun in the sky. As is well known, this apparent motion is caused by the diurnal rotation of the earth and by the motion of the earth in an elliptical orbit about the sun. The diurnal motion causes the sun to appear to rise in the east and to set in the west. The orbital motion gives rise to three other factors which the designer of a sundial has to take into account if correct time indications are to be obtained. The first of these is due to the earth's motion about the sun as a focus; because of this the apparent motion of the sun is retarded, so that it takes four minutes longer to move across the sky than do the stars.

The second factor is due to the elliptical nature of the earth's orbit: This causes the earth to move more slowly when far from the sun than when near to it, so that the velocity of the sun's apparent motion varies according to the time of year. Third, the inclination of the axis of diurnal rotation is inclined by 23° to the plane of the earth's orbit; therefore in summer the apparent path of the sun (known as the ecliptic) is higher in the sky than in winter and the duration of sunlight is longer. There is, of course, also the latitude of the observer to be considered; this affects the maximum height which the sun reaches above the horizon, and thus the inclination of the style or the position of the hour lines on a sundial.

Although it was not until the 17th century that the earth's orbit was correctly understood, sundials of considerable accuracy could nevertheless be constructed because dialing is concerned with the apparent motion of the sun. Temporary hours were sufficient until the advent of mechanical clocks, when reckoning of time in equal hours became necessary. Such equal hours give a mean solar time compared with the apparent solar time of temporary hours.

Mean solar time is based upon the behaviour of a fictitious body called the mean sun, which is assumed to travel along the celestial equator at a uniform speed. The mean sun coincides with the true sun at the equinoxes and solstices. The time difference between the two is known as the equation of time and is defined as mean solar time minus true solar time. The greatest numerical value of the equation of time, of the order of 16 minutes, occurs at a point between each equinox and solstice.

CROWN COPYRIGHT, SCIENCE MUSEUM, LONDON

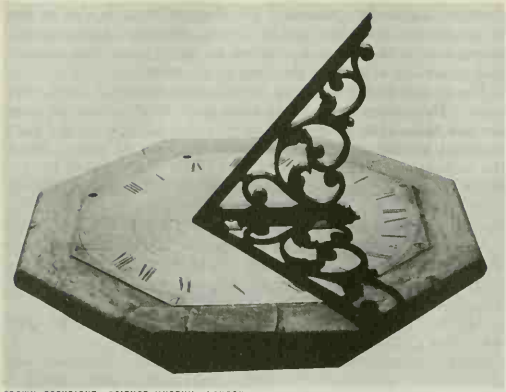
FIG. 3.—IVORY TABLET SUNDIAL BY HANS DUCHER, 1574

For indicating true solar time a sundial is oriented so that the style lies in the meridian and, in the cases of vertical, horizontal and inclined dials, with the style parallel to the earth's polar axis, *i.e.*, inclined to the horizontal by an angle equal to the latitude of the place at which the dial is to be used. At local noon as indicated by the true sun, the shadow cast by the style should fall on the noon mark engraved on the appropriate hour line. A sundial may be oriented by means of a compass or, if a compass is not available, by a plumb line fixed to the end of the style. The shadows cast by the style and the plumb line should coincide over the noon mark at local noon. Local noon as indicated by the true sun may be obtained by taking the zone time and making due allowance for the equation of time and for the longitude of the place at which the sundial is being set up.

Sundials for use in the southern hemisphere are like those for use in the northern, except that the hour lines are inscribed in the opposite direction. The noon mark and orientation are obtained in a similar manner.

Types of Sundial.—*T-Shaped Dials.*—These dials were used in Egypt in very ancient times. They usually consisted of either a simple inscribed bar with a perpendicular endpiece or a central block with two sloping sides cut in steps and with a perpendicular rectangular block at each end. In the latter form the steps did duty for hour lines and the instrument was not, of course, reversed at midday.

Hemispherical Dials.—In the original form of these dials the hemispherical depression was cut in a block of stone, but later only part of the hemisphere was cut and the depression was situated at one end of the block to allow for easier fabrication. As



CROWN COPYRIGHT, SCIENCE MUSEUM, LONDON

FIG. 4.—GARDEN SUNDIAL, 1718

noted, Apollonius of Perga used a conically shaped depression which gave greater accuracy. This type of sundial was revived in the late 16th century in Italy and Germany and took the shape of a cup or goblet. There the style was vertical and the hour lines were inscribed on the interior.

Sunken Dials.—In fashion during the Renaissance, sundials of this kind were usually made of stone. They were in essence a variation of the hemispherical dial, but the sunken depression took a great variety of forms. On one and the same dial cylindrical, triangular and heart-shaped depressions were sometimes cut, and it became popular to cut depressions of this kind into the buttresses of churches and into tombstones and other monumental stonework.

Cubical Dials.—Generally of painted wood and fixed to a stand, these sundials had a style set perpendicularly to the various faces. Different time scales were indicated on these faces; for example, a well-preserved cubical dial made in Florence about 1560 indicates "old Italian hours" (24 equal hours between sunrise and sunset) on its north and south faces, "Babylonian" hours (24 hours between sunrise and sunrise) on its top face and the old Jewish system (12 hours between sunrise and sunset and 12 hours of different duration between sunset and sunrise) on the east and west faces.

Column Dials.—Sundials of this type had a cylindrical column on the top of which was a horizontal style fixed to a cap which could be rotated to allow for the sun's position at different seasons. The instrument was oriented so that the shadow of the style fell vertically, and time was indicated by the shadow of the tip of the style on hour lines inscribed on the cylinder. In order to set up the sundial it was necessary to know whether the sun was east or west of the meridian and, in consequence, a compass was often fitted to the base of the mounting. Column dials date at least from the 16th century, and some 17th-century portable dials are known. In Tibet "time sticks," made since the mid-19th century, have been found, the gnomon being placed in a hole for the appropriate month. An example of a similar type, dating from the 10th century, was discovered during excavations in the cloister yard of Canterbury cathedral in 1939. This was of gold and silver and had a chain by which it could be suspended vertically. It is known as St. Dunstan's dial and may have been made by him. In both this and the Tibetan type of dial the shadow had to be vertical for a correct reading to be obtained.

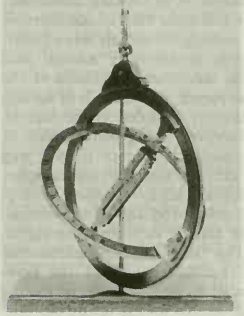
Ring Dials.—The basis of such a sundial was a metal ring, usually of about 3-in. diameter and 1-in. width. Sunlight passed through a pin hole in the rim on to the inside surface of the ring, on which were inscribed a series of vertical circles corresponding to the various seasons, time being indicated on inclined hour lines. Such dials were suspended from a shackle on the rim. Ring dials appear to have been constructed first in the late 17th century, and some were elaborately designed with various refinements permitting adjustment for any latitude.

Tablet Dials.—These were made from two rectangular pieces or tablets of ivory, metal or wood, hinged together at one end. The gnomon was usually a string which, when the two tablets were opened out at about 90°, became taut. The orientation of this kind of sundial was carried out by insuring that the shadow of the string fell at similar hour markings on both tablets. A compass was often included. Such dials were widely used in the 16th century.

Flat Dials.—Perhaps the best-known modern form of this type of sundial is that in which the dial is horizontal and the style is inclined from it by an angle equal to the latitude of the place where the dial is set up. Of such a kind is the garden sundial. Flat dials may, however, be so constructed that the dial is vertical or inclined by any amount to the horizontal. For any inclination the hour lines must be separately constructed.

Scratch Dials.—These, to be found on many early English church walls, consisted of a horizontal gnomon with hour lines scratched into the surrounding stonework. In the simplest type the day is divided into four "tides" of three hours each. A good example may be seen at Daglingworth. In the 7th-century dial on Bewcastle cross, hour lines have been intercalated between the early tide lines, which are marked with crosses.

Portable Sundials.—Many sundials of this kind have been made at different times. The requirements for such a dial are that it should not be too large, that it should be capable of adjustment in azimuth and also of adjustment for latitude if it is to be portable over considerable distances. Such was the "ham" dial (1st century A.D.), which consisted of a piece of metal shaped like a ham; the gnomon or style was a vertical arm, and the hour lines were engraved on the metal body. In use it was suspended vertically from a chain. Other examples are column, ring and tablet sundials.



CROWN COPYRIGHT, SCIENCE MUSEUM, LONDON

FIG. 5.—PORTABLE UNIVERSAL RING DIAL, *c.* 1800

Scientific Sundials. — During the 19th and early 20th centuries various sundials were developed to give results of an accuracy sufficient for scientific purposes. Of these perhaps the most notable were the solar chronometer made by the Scottish astronomer James Ferguson (1710–1776) and a universal heliochronometer of the early 20th century. The former consisted of two pointed rods parallel to the polar axis and close to each other, the shadow of these rods falling on a card mounted in a frame forming part of a cylinder of which the rods themselves were the axis. The hour lines on the card were so curved that an allowance was automatically made for the equation of time. Latitude and longitude corrections could also be made. The heliochronometer consisted of a disk set, by means of a quadrant, parallel to the celestial equator and mounted about a polar axis. Sunlight passed through a hole in one plate or on to a line on another plate, the time being indicated on the equatorial disk. The equation of time was allowed for by a cam which offset the first plate with respect to the second.

BIBLIOGRAPHY.—Ptolemy, *Almagest* bk. ii; *The Analemma;* M. Vitruvius, *De Architectura;* Johann Müller (Regiomontanus), *Kalendarium* (1489); A. Schöner, *Gnomonice . . . De descriptionibus horologiorum . . . omnis generis* (1562); S. Münster, *Compositio horologiorum* (1531); J. Delambre, *Astronomie ancienne* (1817). Works (20th century) in English: A. Gatty, *The Book of Sundials,* rev. ed. by H. K. F. Eden and E. Lloyd (1900); R. N. and M. L. Mayall, *Sundials* (1938); F. A. B. Ward, *Time Measurement,* pt. i, 2nd ed. (1958), pt. ii (1955). (CN. A. R.)

SUNDSVALL, a seaport of Sweden in the district (*län*) of Västernorrland, on a wide bay of the Baltic, at the north of the Selånger river, 360 mi. N. of Stockholm, the terminus of a branch from Ånge on the Northern railway.

Population in 1950 was 25,706. It was rebuilt in brick and stone after a fire in 1888.

SUNFISH, a name applied to several genera of North American perchlike fresh-water fishes of the family Centrarchidae and also applied to an unrelated group of marine species of the family Molidae, the ocean fishes which are often seen sunning or resting in the surface waters of temperate and tropical seas.

Fresh-Water Sunfishes.—Among the many kinds of fresh-water sunfishes, the bluegill (*L. macrochirus*) and the pumpkinseed (*L. gibbosus*) are the well-known plucky fighters prized by many fishermen. In general, the body, deep and rather compressed laterally, varies from a brilliant green to olive and, except in the bluegill, usually has orange or reddish spots on the sides; most species have a conspicuous dark membranous flap at the upper corner of the gill cover.

Sunfish inhabit ponds, lakes and quiet streams. They mature in spring and early summer, at which time the female excavates, using her body and tail, a shallow nest about two feet across. After the eggs are deposited and fertilized the male drives the female away and guards the nest till the eggs hatch and the young are able to shift for themselves. The bluegill sunfish, which may grow to be 12 to 14 in. and about 1½ lb., averages about 7–8 in. and less than ½ lb. It becomes quite stunted, however, when the food supply—crustaceans, snails, insects and larvae—is limited, as in an overpopulated body of water. The bluegill sunfish has been widely introduced where ecological conditions are favourable, often in farm ponds in combination with black bass.

Marine Sunfishes.—The Molidae are curious oval-shaped fishes that appear tailless, the body ending abruptly behind the high, single dorsal and anal fins. They have a cosmopolitan distribution in the warm seas. The common ocean sunfish, *Mola mola*, and the sharp-tailed sunfish, *Masturus lanceolatus*, are large, stout, deep-bodied fishes. They reach a length of nearly 11 ft. and a weight of perhaps two tons. They feed on a variety of pelagic and bottom-living invertebrates and on small fishes. The oblong sunfish (*Ranzania truncata*) is longer in proportion to its depth and more compressed. It grows to be about two feet long.

(L. A. Wd.)

SUNFLOWER. The common sunflower, *Helianthus annuus*, a member of the family Compositae (*q.v.*), is a native of the Great Plains region of the U.S. It is an annual herb with a rough hairy stem 3 to 15 ft. high, broad coarsely toothed rough leaves 3 to 12 in. long, and heads of flowers 3 to 6 in. wide in wild specimens and often 1 ft. or more in cultivated. Double forms are in cultivation, one (*globosus fistulosus*) having very large globular heads. The plant is valuable from an economic as well as from an ornamental point of view. The leaves are used as fodder, the flowers yield a yellow dye and the seeds contain oil and are used for food. It is cultivated in the U.S.S.R., England and other parts of Europe, in Egypt and on the subcontinent of India for the seeds, the yellow sweet oil obtained by compression being considered about equal to olive or almond oil for table use. Sunflower oilcake is used for stock and poultry feeding, and is ordinarily exported from the U.S.S.R. to Denmark, Sweden and elsewhere.

JOHN H. GERARD

PRAIRIE SUNFLOWERS (HELIANTHUS PETIOLARIS)

The genus *Helianthus* contains about 60 species, chiefly natives of North America, a few being found in Peru and Chile. They are tall, hardy annual or perennial herbs, several of which are of easy cultivation in gardens with moderately good soil. *H. decapetalus* is a perennial about five feet high with solitary heads about two inches across in slender twiggy branchlets; *multiflorus* is a beautiful species with several handsome double varieties; *H. orygalis* is a graceful perenial six to ten feet high, with drooping willowlike leaves and numerous comparatively small yellow flowerheads. *H. atrorubens* is a smaller plant, two to five feet high, the flower heads of which have a dark red or purple disk and yellow rays. There are many fine forms of this, some of which grow six to nine feet high and have much larger and finer flowers than the parent species. Other fine species are *H. giganteus*, 10 to 12 ft.; and *H. mollis*, 3 to 5 ft. *H. tuberosus* is the Jerusalem artichoke.

SUNG, a great Chinese dynasty and cultural period (960–1279) that marked one of China's most brilliant ages. The founder of the dynasty, Chao K'uang-yin (*q.v.*), named it for the ancient feudal state of Sung. It is sometimes called the Chao Sung to distinguish it from the Sung dynasty of the Liu house which ruled briefly in southern China (420–479). The era before 1127, when the capital was at Kai-feng near the junction of the Grand canal and the Yellow river, is often called the Northern Sung, while the following period, when the Yellow river valley was lost to the Jurchen empire and the capital was normally at Hangchow in Chekiang, is called the Southern Sung. The founder and his successors, scholarly in tastes and educated under strong Confucian influence, proved unusually conscientious in promoting the stability, prosperity and intellectual life of their empire. The Sung dynasty brought to full flower the trends of development that had begun in the T'ang (618–906). Its life, interests and accomplishments in many ways resembled those of Europe in the Renaissance.

Sung Civilization.—A characteristic of the age was the rise of commercial cities clustered along the principal canals and rivers and the southeast coast. Some were uniquely great for their times: Hangchow attained a population of 1,000,000 or more. These cities were walled and often symmetrically planned, with canals, broad avenues, market places, private gardens and public parks. Marco Polo described them soon after the dynasty's fall, when many of their features and customs were still preserved. Around the palaces and public buildings were the establishments of merchants, craftsmen, money-changers, dealers in intercity credit certificates, pawnshops, teahouses and wineshops. Citizens enjoyed the protection of state police and fire-fighting services, public baths, state relief, hospitals, orphanages and homes for the aged and destitute.

The canal system was extended to supplement the roads in linking the cities and to irrigate new farmland. Sea navigation served both internal and foreign trade—the size of the junks impressed even the Venetian Marco Polo. Long-distance exchange of common commodities made possible regionally specialized industries and agriculture. Improved strains of rice were introduced from Indochina and industrial entrepreneurs organized shops employing hundreds of workers. Iron, apparently smelted with the aid of coal, was cast in the form of pagodas and other objects of similar scale. Commerce was facilitated by a volume of copper coinage unparalleled in China before the 19th century and by the printing of paper money, whose circulation was successfully controlled until late in the 11th century.

With economic growth came social change. Merchants suffered less social stigma than before, and their progeny entered the bureaucracy more freely. Widespread printing of the classics and other writings from the 10th century, facilitated by the use of movable type in the 11th, opened the door to education for greater numbers. The private academies that flourished in more populous regions and state schools in several hundred prefectures and subprefectures prepared students for the national university at the capital. By increasing the numbers trained for governmental service, these developments broadened the distribution of influential offices, which during the earlier T'ang period had been largely monopolized by a few eminent families of north China.

The growing competition for bureaucratic positions encouraged a wider use of civil service examinations. Candidates were subjected to successive elimination through written tests on three levels; more than a hundred beginning the ordeal for each one who emerged successful; yet the numbers thus recruited rose from a

yearly average of under 20 to over 200. The effort to devise objective and meaningful tests for practical qualities led to long contentions over subject matter and testing methods. To preserve the anonymity of the candidate and ensure fairness in grading, examination papers were copied by clerks, examinees were identified by number only and three examiners read each paper. Higher officials were privileged to nominate junior relatives for admission to the bureaucracy, but the great stress on examination grades in promotion and the practice of recruiting many lower officials from the ranks of the clerical service ensured a considerable freedom of opportunity. Promotion of the more able and reliable was sought through legally guaranteed recommendation by higher officials and by annual merit ratings.

The bureaucracy was successful in drawing into its ranks nearly all of the ablest educated men. As a result, the outstanding achievements in fields of the greatest diversity were by men active in government and politics. The "universal man" was the typical figure of the time. A leading official might be creatively concerned with philosophy, history writing, poetry, science or painting, and most often with several of these: Ou-yang Hsiu, Fan Chung-yen, Ssu-ma Kuang, Wang An-shih, Shen Kua, Su Tung-p'o, Chu Hsi, Lu Hsiang-shan and Wen T'ien-hsiang were but a few of the most famous. Women were represented by such examples as the poet-archaeologist Li Ch'ing-chao. The emperor Hui Tsung —painter of the first rank, calligrapher, patron of art and archaeology—was the best-known instance of such versatility among the Sung emperors, but not the only one.

Political History of the Dynasty.—The character of officialdom and the tradition of the ruling house created an atmosphere favourable to spirited political debate. After the vigorous rule of Chao K'uang-yin (960–976) and his brother T'ai Tsung (976–997), both circumstances and the rulers' personalities tended to confine their role to that of selecting the chief officials and resolving divergences over policy. The emperor Chen Tsung (997–1022) was unsettled in mind during his later years; on the accession of the 11-year-old Jen Tsung (1022–63) the empress dowager ruled as regent; his successor, Ying Tsung (1063–67), was an invalid during his short reign. The authority deputed to the councilors of state was tempered by the influences of the censorate, the bureau of policy criticism and the bureau of academicians, and discussions at court on questions of policy often ranged officials of these and other bureaus in opposing groups. While more ardent followers of Confucius were never satisfied that his precepts of government through moral principle, character and reason were fully applied, the early Sung perhaps came closer than any other period to realizing the Confucian political ideal in practice.

But factional divisions grew acrid as economic and military problems became pressing. Rising farm population and equal land inheritance aggravated the burden of heavy taxes for the small farmer, who often lost his lands to larger landholders who could keep their lands in better condition and farm them more effectively. While the government depended increasingly on state monopolies or excises—salt, metals, wine, alum, aromatics and foreign imports —the total state income had begun to decline. Fan Chung-yen and others had proposed and sometimes implemented schemes to deal with some of these problems, but the difficulties increased. Wang An-shih offered a more radical and comprehensive program to deal with the mounting difficulties. Firm support by the emperor Shen Tsung (1067–85) enabled him to carry this out despite the antagonism of his leading contemporaries, which he invited by impatiently ousting from office all who questioned the prudence of any new measure. The magnitude of the problem and the bureaucrats' ineptness in applying the new laws limited their success. They were repealed under the regency of the empress dowager Hsuan-ho early in the reign of Che Tsung (1085–1100) and restored when the emperor assumed personal power. Despite the moderating efforts of the following ruler, Hui Tsung (*q.v.*; 1100–26), the struggle of the pro-Wang and anti-Wang factions continued for several generations, during which one or the other party was sometimes declared anathema and barred from all official rank or function.

In the meanwhile, the military situation had further deteriorated. After the initial achievement of recovering from the Khitan (Liao) empire all that it had seized of north China excepting a small strip within the Great Wall, and the subjection of the independent principalities in south China, the rulers curtailed the powers of their commanders to forestall a recurrence of the war lordism that had destroyed the T'ang. The peace of the 11th century was unbroken until midway in Jen Tsung's reign, when the new Tangut (Hsi-hsia) empire invaded the northwest frontier and the Khitans in the north were kept neutral only through appeasement payments. The regular troops were softened by ease, and the enemy was repelled only by the raising of militia corps. Wang An-shih included in his program a provision for regular militia reserves and for army horses to be kept by farmers. The inadequacy of this was shown when the Jurchen (Chin) empire, after overthrowing the Khitan in alliance with the Sung, invaded China and attacked Kai-feng. After a humiliating peace and the abdication of Hui Tsung to his son Ch'in Tsung (1126–27), the Jurchen renewed the fight, took both emperors captive and invaded the Yangtze valley. Under Kao Tsung (1127–62), who established the capital of the Sung at Hangchow, the Chinese responded to the emergency. Led by Yueh Fei and other generals they threw the invaders back. Perhaps for economic reasons, the councilor Ch'in Kuei nevertheless determinedly renounced the reconquest of the north, made peace with the Jurchen in 1141 and executed the recalcitrant Yueh Fei who was revered thereafter as China's national hero. Despite the fervid national spirit that could be aroused in times of crisis, the Chinese of the Sung had grown less martial and depended increasingly for military success on the technical development of weapons. Incendiary projectiles had long been used; explosive grenades appeared by the 13th century and by the end of the dynasty it appears that cannons were also employed. Such weapons also armed the large navy that was built up in the Southern Sung.

With such help the Sung Chinese enjoyed another century of relative prosperity and intellectual flowering, despite accumulating difficulties with monetary inflation and an unsolved farm problem, under the reigns of Hsiao Tsung (1162–89), Kuang Tsung (1189–94) and Ning Tsung (1194–1224). But China's security was precarious. The Jurchen, and then the Mongols who conquered them, in time discovered the secrets of China's new weapons, and technicians from occupied north China could copy them. Chinese technical skill was forced to surpass itself in creating still newer weapons to counter the earlier. Beginning in the second decade of the reign of Li Tsung (1224–74), the key fortresses guarding the northern approach by the Han river fell after a prolonged siege, and during the brief reigns of Kung Tsung (1274–76) and the child emperors Tuan Tsung (1276–78) and Ti-ping (1278–79) the Sung court sought successive refuges near the Canton estuary while Wen T'ien-hsiang and other loyal leaders fought defensive actions; it was believed that after a last unsuccessful naval engagement the Sung admiral perished in the sea with the infant ruler rather than surrender him to the Mongols.

Intellectual and Artistic Achievements.—In its cultural life the Sung dynasty was a time of beginning in some respects and of culmination in others. The rise of cities, new wealth and wider availability of education called into being new forms of entertainment to please simpler tastes and led a larger number to pursue more sophisticated forms of intellectual satisfaction. To the amusements of acrobatic troupes and displays of fireworks were added simple theatrical presentations that were to develop into the drama of later dynasties and itinerant storytellers whose prompting books held the plots of later historical and picaresque novels. The forms of verse developed by the great T'ang poets were employed subtly and elegantly as the habitual expression of educated men, reaching greatness at the hand of Su Tung-p'o, and a new form of poetry was brought to perfection—the writing of lyrics to well-known melodies. Great skill was devoted to artistic craftsmanship, and the manufacture of porcelain, while it did not reach the technical dexterity of the Ming and Manchu periods, attained perhaps its highest aesthetic perfection. Buddhist sculpture found its fullest grace; landscape painting, at the brushes of masters like Kuo Hsi, Hsia Kuei and Liang K'ai, attained suc-

cessive heights of expressiveness, subtle mood and mysticism that were scarcely to be touched again.

While Buddhism had passed its intellectual prime in China, its past attainments challenged reviving Confucianism to create a metaphysical system that could answer the problems the Buddhists had raised. Freedom of speculation prevailed, and private academies taught a diversity of interpretations; the two most influential formulations were those reached in the 12th century by Chu Hsi (*q.v.*), whose discipline stressed the extension of knowledge, and the more mystical Lu Hsiang-shan. While the system of Chu would become orthodox in later times, that of Lu was the forerunner of Wang Yang-ming's heterodoxy. The explanation of nature concerned systematic thinkers like Chu Hsi, and men like Shen Kua combined active interests in mathematics, theoretical speculation and the many advances that Chinese were making in practical technology, ranging from new medical practices to the development of the mariner's compass.

Ssu-ma Kuang and Li Tao typify the many historians who applied more critical methods, created new forms of historiography to present the causation and relationship of events and treated recent history in works impressive for size and detail. The thought of the time is expressed also in the popularity of the essay; and the individual personalities of the period are well preserved in the many collections of letters and other smaller pieces that were first published in large numbers. *See* also CHINA: *History*.

BIBLIOGRAPHY.—A. C. Moule, *Quinsai: With Other Notes on Marco Polo* (1957); J. P. Lo, "The Emergence of China as a Sea Power," *Far Eastern Quarterly*, xiv, 489–503 (1955); E. A. Kracke, Jr., *Civil Service in Early Sung China, 960–1067* (1953); J. T. C. Liu, *Reform in Sung China* (1959); Clara Candlin, *The Herald Wind, Translations of Sung Dynasty Poems, Lyrics, and Songs* (1955); Laurence Sickman and Alexander Soper, *The Art and Architecture of China*, pp. 94–145, 255–268 (1956). (E. A. KE.)

SUNIUM, a cape at the southern extremity of Attica, with a temple of Poseidon as landmark for ships, the modern Cape Colonna. The rocky promontory was fortified by wall and towers in 413 B.C. against the Spartans in Decelea, but was later seized by fugitive slaves from the Laurium mines. In the 4th century it was still a fortress. The temple, probably built in the time of Pericles in the place of an earlier one, similar but of tufa or "poros" stone, is shown by an inscription to be dedicated to Poseidon, not, as formerly supposed, to Athena, whose temple, of peculiar plan, with colonnade on two sides only, lies about ¼ mi. to the northeast. Of Poseidon's temple there are standing nine columns of the south side, two of the peristyle on the north, one of the *antae* and an inner column of the pronaos, of local white marble, which has suffered from the weather.

SUNN (SANN, SUNN HEMP or INDIAN HEMP) embraces at least two species of *Crotalaria*, of the legume family. *Crotalaria juncea* is best known as the source of an inner bark or bast fibre, which also is termed sunn. Sunn is a familiar plant in Asia, and has been naturalized in the West Indies. Other species of *Crotalaria* are employed as cover crops in the southern United States. Sunn fibre of commerce comes almost entirely from India, where *C. juncea* is grown and marketed much as is jute. The fibre is of better quality than jute, approaching hemp or flax in its characteristics. A good many thousands of tons are produced annually, much of it locally consumed for coarse cloth, cordage, sacking and calking. Sunn fibre imported into the United States has usually been employed in the making of special quality papers, such as that for cigarettes.

The plants are low shrubs with slender silky stems, undivided leaves and bright yellow flowers. At maturity seeds rattle loose in the pealike pod, giving rise to the name *Crotalaria*, from the Greek word for "rattle," source also for *Crotalus*, the rattlesnake. The fibres develop as bundles or strands in the inner bark next to the woody core of the stem, and are usually freed from the softer surrounding tissues by the process known as retting. Retting depends upon bacterial decomposition or "rotting" of the thin-walled weaker cells, and is completed in four or five days with sunn. The thick-walled resistant fibres can then be sloughed free of decomposing debris and core. The four- or five-foot-long fibre strands are dried and marketed as yellowish or grayish hands. Sunn fibres are essentially pure cellulose. Were they to contain any aromatic principles, certain to be evident as "off flavours" upon burning, sunn could not substitute for flax in the making of cigarette paper. (R. W. SY.)

SUNNIS (SUNNITES): *see* ISLAM.

SUNNYVALE, a city of Santa Clara county in northern California, U.S., 39 mi. S. of San Francisco, is noted primarily for its two major industries, electronics and guided missiles. Incorporated in 1912, the city adopted the council-manager form of government in 1949. Originally a part of the Mariano Castro land grant, it was known as Murphy's Station from 1850 to the turn of the century and then briefly as Encinal; the name Sunnyvale was officially adopted in 1912. Between World Wars I and II it was chiefly a cannery centre processing the fruit of the extensive orchards nearby. Industry came to the area during World War II and the Korean war; its population rose from 4,373 in 1940 to 52,898 in 1960. For comparative population figures *see* table in CALIFORNIA: *Population*. (J. J. HA.)

SUNSHINE, the sun's light; the direct rays neither scattered nor reflected; hence the warmth and light given by the sun's rays. Sunshine includes not only the visible portion of radiation from the sun, but also invisible ultraviolet and infrared rays (*see* also SUN). Radiation from the sun is the ultimate source of nearly all energy that is essential for the maintenance of plant and animal life on the earth and the operation of most natural phenomena on the surface of the earth. In particular, the amount and distribution in time and space of the solar radiation which is intercepted by the earth are the prime cause of weather and climate.

The energy output of the sun has its peak at 0.47μ (0.00047 mm.) and radiates 8 kw. per square centimetre into space. At the earth's surface the amount received is on an average 1 kw. per square metre. The energy measured at the outer boundary of the atmosphere, the so-called solar constant, is 2 cal. per square centimetre per minute on a surface normal to the rays at the mean distance between sun and earth. The amount of possible sunshine for each point on the earth's surface is determined by latitude and season, as shown in Table I.

TABLE I.—*Possible Duration of Sunshine in Various Northern Latitudes (Hours:Minutes)*

Date	10°	30°	50°	70°	90°
Feb. 15	11:49	11:10	10:07	7:20	0:00
April 15	12:21	12:53	13:44	16:06	24:00
June 15	12:42	14:04	16:21	24:00	24:00
Aug. 15	12:28	13:16	14:33	18:26	24:00
Oct. 15	11:55	11:28	10:47	9:03	0:00
Dec. 15	11:32	10:14	8:05	0:00	0:00
Longest day	12:35	13:56	16:09	24:00	24:00
Shortest day	11:25	10:04	7:51	0:00	0:00

Measurement of Sunshine.—Several devices are used to record sunshine.

The Campbell-Stokes Sunshine Recorder.—This instrument consists essentially of a segment of a spherical metal bowl, having a glass sphere placed concentrically within it and centred so that the sun's rays are focused on a card lining the bowl. Direct sunshine burns a spot or a line on the card. Differentiation between a record of a few seconds and a minute of intense sunlight is difficult because the two traces are similar. Also, when the sun is barely visible through smoke or thin clouds, its rays are too weak to make a record.

FIG. 1.—CAMPBELL-STOKES SUNSHINE RECORDER

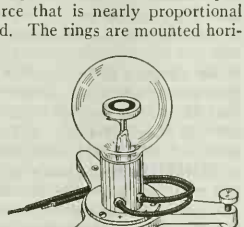
Modified forms of sunshine recorders use blueprint paper mounted on a cylinder. The Jordan type contains two semicylinders with narrow slits in their narrow sides and, like the Campbell-Stokes recorder, maintains its own time record.

The Marvin Sunshine Recorder.—This is an air thermometer in which an electrical circuit is closed through the expansion of mercury by thermal action of the sun.

The instruments so far mentioned merely indicate when the sun

shines.

The Eppley Pyrheliometer.—This unit accurately responds to nearly all the radiation which reaches the surface of the earth on a horizontal surface and gives quantitative results. The instrument consists of two concentric rings of equal area, one blackened and the other white coated. The hot junctions of a multiple thermopile of gold-palladium and platinum-rhodium alloys are attached to the lower side of the black ring, and the cold junctions are fastened to the lower side of the white ring. The differential in temperature between the two rings when radiation falls upon them creates an electromotive force that is nearly proportional to the amount of radiation received. The rings are mounted horizontally in the centre of a thin spherical glass bulb which is sealed to prevent deterioration of the receiving surfaces and also to prevent moisture from condensing within. The electromotive force generated by the receiver is recorded on a potentiometer and reduced to values in gram calories per minute per square centimetre.



THE EPPLEY LABORATORY, INC.

FIG. 2.—EPPLEY 180° SUNSHINE RECORDER

Quality of Sunshine.—The major portion of sunshine which penetrates through the atmosphere ranges in wave length from 0.29μ to 2.5μ. Ordinarily, sunshine is broken down into three major components: (1) the visible, or that part between 0.4μ and 0.8μ (one octave of wave length); (2) the ultraviolet, or that portion with wave lengths shorter than 0.4μ; and (3) the infrared with wave lengths longer than 0.8μ.

Because of a continual shift in the position of the maximum of the solar energy curve, no one factor can be used accurately to convert values in gram calories to values in foot-candles. However, the value of 6,500 foot-candles per gram calorie suffices for most practical purposes.

Visible Radiation.—Visible radiation is commonly termed light; however, the word light now has a broader meaning and includes the ultraviolet, or that invisible portion immediately adjoining the shortest wave length of visible radiation. The visible portion comprises nearly one-half of the total radiation received at the surface of the earth.

Ultraviolet Radiation.—While ultraviolet light comprises only a very small proportion of the total radiation, this component is extremely important. It produces vitamin D through the activation of ergosterol. It also is an important germicidal agent. Unfortunately, the contaminated atmosphere over large cities robs radiation of practically all the shorter wave lengths.

Ultraviolet and the near ultraviolet rays comprise the so-called actinic rays which are very potent in effecting chemical changes on photographic films.

Infrared radiation has its chief merit in its heat-producing quality. Close to one-half of total solar radiation received at the surface of the earth is infrared. The spectrobolometer is the most important instrument for measuring small components of sunlight.

Solar rays of different wave lengths are refracted on entering and emerging from a prism, the shorter wave lengths being refracted more than the infrared rays. It was through the use of the spectroscope and matched absorption bands that knowledge of the composition of the sun was determined.

The sunshine reaching the outer edge of the earth's atmosphere is closely equivalent to the radiation of a black body at a temperature of 5,800° to 6,000° C. On its path through the atmosphere this radiation is absorbed and weakened by various constituents of the atmosphere. It is also scattered by air molecules. This scattering, which occurs in the shorter wave lengths, causes the light-blue colour of the sky at the earth's surface. In the stratosphere, with fewer air molecules above the observer, the sky appears dark.

An effective absorber of solar radiation is ozone, which forms by a photochemical process at heights of 20 mi. and filters out most of the radiation below 0.3μ. Equally important as an absorber in the longer wave lengths is water vapour. A secondary absorber in the infrared range is carbon dioxide. These two filter out much of the solar energy below 1μ.

In the low layers of the atmosphere, dust scatters and absorbs effectively. At sunrise and sunset sunlight has a long path through the turbid air. The light is much attenuated and the short wave lengths are scattered. This causes the red colour of the sun at those times and permits one to look directly into the sun.

Duration of Sunshine.—At the surface of the earth cloudiness is the primary factor governing the amount of sunshine actually received, but even through dense clouds or fog some diffuse radiation still penetrates. The insolation received through various cloud types (in per cent of clear sky radiation) is: cirrus 83%, cirro-stratus 78%, alto-cumulus 49%, alto-stratus 41%, strato-cumulus 32%, stratus 25%, nimbo-stratus 20%, fog 18%.

The distribution of sunshine over the earth varies from over 4,000 hours per year; *i.e.*, more than 90% of the possible to less than 2,000 hours in the storm belts. The Sahara desert gets the largest amount, areas around Iceland and Scotland are among those getting least.

In the United States, Arizona shows the greatest amounts of sunshine, the Pacific northwest coast receives least. The variation throughout the year in terms of the possible duration is given for a selection of representative stations in Table II.

TABLE II.—*Monthly and Annual Sunshine* in Various Parts of the World*

Area	Station	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Annual
U.S. Atlantic	Boston, Mass.	47	56	57	56	58	63	64	63	61	58	49	49	57
	New York, N.Y.	51	58	59	60	60	64	65	63	64	63	56	53	60
	Washington, D.C.	46	54	56	57	61	63	64	62	62	61	53	46	58
	Miami, Fla.	65	72	73	72	68	61	65	67	62	61	64	65	66
U.S. central	Chicago, Ill.	44	59	53	56	63	69	73	70	65	61	47	41	59
	St. Louis, Mo.	48	49	56	59	64	68	72	68	67	65	54	44	61
	Little Rock, Ark.	45	53	57	62	67	72	74	74	68	66	58	46	63
	New Orleans, La.	49	50	57	63	68	65	58	59	64	69	60	46	59
U.S. mountains	Rapid City, S.D.	58	63	63	62	61	66	74	74	69	66	58	54	64
	Denver, Colo.	67	67	65	63	61	69	68	68	71	71	65	65	67
	Albuquerque, N.M.	71	72	73	76	79	84	77	76	80	81	80	71	77
	Phoenix, Ariz.	76	78	82	87	93	94	83	84	89	88	84	76	85
U.S. Pacific	Anchorage, Alaska	41	47	55	58	50	50	45	39	32	28	35	28	44
	Seattle, Wash.	25	35	43	50	53	53	62	56	51	35	25	22	40
	San Francisco, Calif.	53	57	63	69	70	75	68	63	70	70	62	54	66
	Los Angeles, Calif.	70	69	69	67	65	69	78	80	78	76	79	73	72
	Honolulu, Hawaii	62	64	60	62	64	66	67	70	70	68	63	60	65
Canada	Quebec, Que.	30	47	40	41	41	44	48	48	42	36	25	26	39
	Montreal, Que.	30	37	43	47	48	52	57	55	50	40	27	23	42
	Toronto, Ont.	37	37	41	47	50	57	61	59	55	45	29	24	45
	Vancouver, B.C.	18	28	37	43	49	46	60	59	47	33	20	15	38
Europe	Edinburgh, U.K.	21	29	31	33	35	35	32	32	31	31	25	20	31
	London-Kew., U.K.	17	22	29	38	42	40	40	41	38	28	20	17	33
	Paris, France	21	31	36	39	49	47	50	50	42	33	23	20	37
	Vienna, Aus.	23	25	34	42	47	50	55	55	41	36	24	17	40
	Rome, Italy	38	45	42	44	50	62	76	73	59	47	44	37	53
Near east	Cairo, Egy.	61	62	62	67	76	85	85	86	77	71	70	62	73
East Asia	Tokyo, Jap.	61	55	50	48	46	35	44	50	38	40	55	61	49
Caribbean	San Juan, P.R.	64	68	70	65	59	61	64	66	60	63	62	64	64

*In per cent of possible.

The sunshine shows in most places a rather regular variation throughout the day. A common situation, representative of many parts of the United States and other moderate latitudes, is cloudiness in the early morning, fewer clouds in the late morning and early afternoon, and an increase in cloudiness again in the late afternoon. This yields a sunshine distribution (in per cent of possible) as follows:

6–10 A.M.: 52%; 10 A.M.–2 P.M.: 73%; 2–6 P.M.: 60%.

Solar Energy.—For centuries man has dreamed of using free solar energy. The earliest exploitation was in evaporation pans for the production of salt from sea water. In 1878 a printing

press in Paris operated on solar energy. In 1913 Nile water for irrigation was pumped by a solar engine. An acre of land in an area where the daily radiation is 500 cal. per square centimetre per day receives 18,900,000 kilocalories. This is equivalent to 75,000,000 B.T.U. or nearly 5,000 kw.hr. of electricity. Solar engines of the late 1950s converted only about 1% to 5% into energy. Primary uses of solar energy were the demineralization of salt water, the heating of houses and as a fuel for small stoves. Solar furnaces for production of high temperatures for research have been constructed. Many curved mirrors focus the sunshine on a small area. Temperatures of 3,000° C. have been obtained in this way. Photogeneration of electricity in so-called solar batteries is obtained through semiconductors such as selenium (*q.v.*). A 5% efficiency has been obtained from pure silicon with a boron admixture, producing about 50 w. per square yard. Solar energy in conjunction with a heat-pump system has been used for refrigeration. Dissociation of water into hydrogen and oxygen has also been successfully accomplished by solar energy. Several systems have been devised to tide solar energy plants over sunless periods. These methods of storing energy include the use of hot-water tanks or hot-pebble beds through which water or gas percolates; salt hydration (using, *e.g.*, sodium sulfate) which can be reversed with liberation of heat; and pumping of water to higher reservoirs for later hydroelectric conversion. Photosynthesis in plants produces hydrocarbons from water and carbon dioxide by chlorophyll action. The theoretical requirement of photoenergy is 112 kilocalories for fixing 1 g. of carbon (or 8 quanta per carbon atom). The highest yield in natural agriculture has been obtained in sugar cane with a yield of 40 tons of dry matter per acre. This is an efficiency of 2% assuming 3,000,000 kilocalories of usable energy of the right wave lengths. The cultivation of algae shows the greatest promise of transforming water and carbon dioxide into hydrocarbons. Experiments have yielded 14.4 g. of dry *Chlorella pyrenoidosa* per square metre per day.

The energy available in various latitudes and geographic areas as derived from pyrheliometric observations is shown in Table III. (I. F. H.; H. E. Lg.)

TABLE III.—*Mean Radiation Received on Horizontal Surface*
(in gram calories per square centimetre per day)

Station	Lat.N.	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Annual
Yakarta, Indon.	5.7	368	390	408	404	382	368	396	435	460	431	405	378	402
Trivandrum, India	8.5	561	533	577	504	536	405	402	491	481	440	413	503	487
Miami, Fla.	25.8	368	434	538	557	590	548	539	522	471	380	357	330	470
Brownsville, Tex.	25.9	318	376	439	522	579	645	645	595	497	450	299	268	469
Delhi, India	28.5	361	441	509	599	610	571	491	486	478	524	442	353	489
Jerusalem, Israel	31.6	318	357	499	502	675	757	742	682	611	469	356	287	529
Phoenix, Ariz.	33.4	314	448	535	649	737	770	678	632	597	482	352	285	540
Atlanta, Ga.	33.6	220	295	400	524	571	601	562	525	439	389	272	212	418
Los Angeles, Calif.	33.9	264	384	400	517	616	605	650	594	502	385	295	244	462
Albuquerque, N.M.	35.1	314	412	513	634	686	763	693	642	589	473	349	287	523
Oak Ridge, Tenn.	36.0	162	242	334	445	537	568	559	488	450	350	214	161	376
Las Vegas, Nev.	36.1	281	407	523	620	713	755	673	631	565	441	317	254	515
Davis, Calif.	38.5	159	271	411	514	650	692	697	631	498	359	192	134	434
E. Wareham, Mass.	41.8	141	207	287	371	465	532	523	435	325	260	170	144	322
Ithaca, N.Y.	42.5	114	210	266	356	485	542	581	466	372	248	129	97	322
Davos, Switz.	46.8	138	223	350	478	516	538	526	436	339	235	163	121	339
Bochum, Ger.	51.5	44	91	182	294	339	399	363	320	225	122	51	31	205
Gdansk, Pol.	54.3	46	106	199	346	465	530	458	380	298	131	61	29	254
Stockholm, Swed.	59.3	31	78	176	290	403	443	400	302	215	101	38	21	208
Helsinki, Fin.	60.1	18	76	190	300	380	466	449	290	186	79	28	11	206
Fairbanks, Alaska	64.8	13	64	195	360	454	502	452	303	176	84	28	6	220

SUNSTONE, a variety of the mineral feldspar (*q.v.*) exhibiting a brilliant spangled appearance in certain directions, caused by minute scales of iron oxide arranged parallel to a cleavage. It is somewhat similar to aventurine. The best-known locality is in Tvedestrand, Nor. It is also found near Lake Baikal, Sib., and at several localities in the U.S., especially at Middletown, Pa., and Statesville, N.C. *See* also AVENTURINE.

SUNSTROKE: *see* HEATSTROKE.

SUN YAT-SEN (1866–1925), the leading figure in the Chinese revolution of 1911 and a national hero for all factions in modern China, was born in a farming village in Kwangtung province, near Canton. When he was 13 years old an elder brother who had emigrated to Hawaii sent for him; there he was educated and converted to Christianity in an Anglican school. After a brief return, to his village, where he showed himself impatient with traditional practices and beliefs, Sun left for Hong Kong in 1883. He completed training in medicine there in 1892, and practised briefly in Macao, but his real interests lay in politics, not in medicine. His experience in Hawaii and Hong Kong, his belief in Christianity, and his western education combined to make him eager for change in China. Because Sun Yat-sen was poorly grounded in the classical tradition of Chinese culture he drew his support from Chinese who were western-oriented as he was. He traveled constantly to recruit followers and solicit funds from Chinese living in southeast Asia, Japan and the western world. He also had many close friends in England and the United States. For a decade, during much of which he used the name Nakayama, Japan provided help for some of his many abortive revolts, and at the end of his life he looked to the Soviet Union for aid.

Revolutionary Activities.—In 1894 Sun Yat-sen made an unsuccessful attempt to present a memorial to the Viceroy Li Hung-chang urging reforms. The following year came the Japanese victory over China, and in its wake came Sun's first effort at rebellion in Canton. From then until the collapse of the Manchu (Ch'ing) dynasty he tried time after time to bring about a revolution. His tactics formed a link between the conspiratorial techniques of the old secret societies, which had always been strong in south China, and the more broadly based nationalist movements of modern times. Until 1911 Sun's program was dominated by a determination to rid China of the Manchus. To this end he allied himself with many other discontented groups. Japan, where thousands of Chinese students were to be found, was the scene of the organization of the T'ung-meng-hui (Combined League Society) in 1905. Although a broad program was set forth for the organization, its members were most firmly united on the nationalist, anti-Manchu goal.

Early Years of the Republic.—Sun Yat-sen was abroad when the Manchu dynasty fell, late in 1911; he returned to China a national hero. After a brief period as president of the Chinese republic, he resigned in favour of Yüan Shih-k'ai (*q.v.*), whose military strength forced the revolutionaries to compromise with him. It then became clear that Sun's republican associates could not agree on a political program. Sun trusted in the new constitutional machinery to control Yüan, but the latter, who enjoyed the financial support of international loan consortiums, was able to flout the will of the revolutionaries. In 1913 Sun staged an unsuccessful revolt against Yüan. Thereafter, once again a fugitive and conspirator, Sun entered upon a discouraging period during which he bargained desperately for foreign support. Henceforth he was less confident and trusting, and more concerned with practical politics and the strategy of power.

After Yüan Shih-k'ai's death in 1916 China entered the chaotic period of war-lord rule. By dealing with war lords in the Canton area Sun was able to set up a government there. But his real strength lay not with the war lords but with the students, merchants and workers of the modern cities, for whom he was becoming the best-known and most articulate representative of the new spirit of nationalism. When the Soviet Union turned its attention to China, Sun's movement seemed to offer the best prospects for a "united-front" drive against imperialism. By agreement with a communist agent in 1923 Sun received Soviet support for the reorganization of the Kuomintang (*q.v.*). The party was recast on communist lines, with leadership securely in the hands of the central executive committee. Soviet military advisers trained and equipped the new army headed by Chiang Kai-shek, and Chinese communists, who joined the Kuomintang as individuals, contributed their organizational skills and propaganda to undermine the war lords in preparation for the northern expedition that was to bring national unification.

Sun did not live to carry out this program. Shortly before the

northern expedition was to begin he made one last attempt to negotiate with the war lords. After his death in Peking on March 12, 1925, his last testament, a moving document that instructed his followers to carry on his work, became his legacy to modern China. Acclaimed as the leader of the Kuomintang and father of his country, he was for a time the centre of a state cult. In character and personality he was ideally suited for such a role, for his integrity, idealism, courage and sincerity had played a large part in bringing about the final victory. But the choice of Sun's writings as the centre of a new ideology was less fortunate.

Writings.—Sun had read widely and had drawn his ideas from many sources, but he was primarily an activist, intent upon the selection of materials suitable for the needs of the moment. His revolutionary program had begun to form in 1905 when he first enunciated the Three Principles of the People (*San Min Chu I*). "People's Rule" (nationalism), "People's Authority" (democracy) and "People's Livelihood" (sometimes given as socialism) became the slogans of the revolutionary movement. These principles received their fullest treatment in lectures Sun delivered at Canton shortly before his death. Under nationalism he denounced foreign controls on China and called for a rise in national spirit. His approach to livelihood owed much to the single-tax doctrines of Henry George (*q.v.*). Most significant was his treatment of democracy. He saw his party, the Kuomintang, acting as an elite, preparing the Chinese people for democracy against the day when they would be able to take over the task of governing themselves. There again Sun's ideas were garnered from many sources. He distinguished between the four powers of the people (election, recall, initiative and referendum) and the five powers of the government (legislative, executive, judicial, civil service examinations and censorate). The skilled elite would run this elaborate machinery just as an engineer would run a railroad train. Clearly this acceptance of the necessity of a period of one-party rule would defer indefinitely the establishment of democracy in China.

Perhaps because of the lack of a systematic treatment of these topics, Sun Yat-sen's writings were later made to serve many purposes. His mantle was claimed by Wang Ching-wei, who headed the puppet government at Nanking during the Japanese occupation, and also by Chiang Kai-shek, who carried the northern expedition to completion and headed the new Nationalist regime. Sun Yat-sen's widow, who was the sister of Mme Chiang Kai-shek, professed to find the program of the Communist government, which she served in high posts, in harmony with her husband's views. But it was above all the Nationalist government of Chiang Kai-shek that laid claim to Sun's memory. Sun Yat-sen's son by his first marriage, Sun Fo, held many important positions in the Nationalist administration. A magnificent mausoleum for Sun Yat-sen was erected on Purple mountain outside Nanking, and Sun's title as Kuomintang Leader, *Tsung-li*, remains his in perpetuity. *See* CHINA: *History; see* also references under "Sun Yat-sen" in the Index volume.

BIBLIOGRAPHY.—Sun Yat-sen, *Kidnapped in London* (1897), *The International Development of China* (1920), *Memoirs of a Chinese Revolutionary* (1927); *San Min Chu I*, tr. by Frank W. Price (1927); P. M. W. Linebarger, *Sun Yat Sen and the Chinese Republic* (1925); Lyon Sharman, *Sun Yat-sen: His Life and Its Meaning* (1934); Marius B. Jansen, *The Japanese and Sun Yat-sen* (1954); T. C. Woo, *The Kuomintang and the Future of the Chinese Revolution* (1928), *Tsung-li ch'üan-chi* ("Complete Collected Works of the Leader"), ed. by Hu Hanmin (1930). (M. B. J.)

SUPERANNUATION. The custom of providing pensions for aged employees who, after meritorious service, are no longer able to discharge their duties efficiently became increasingly common after the middle of the 19th century. The problem of mitigating the hardships of old age has been attacked from many angles and a great variety of solutions has been tried. In England superannuation schemes for discharged seamen and soldiers date back to the middle ages; these were followed by systems applicable to civil servants and other public officials, and the custom of making some kind of provision for old age gradually spread to the employers of labour on a large scale and even to businesses with relatively small staffs. The question has also received attention on national lines; this is treated under PENSIONS; SOCIAL SECURITY.

Between World Wars I and II the growing recognition of social problems led to the establishment of many pension funds. At the same time, the unemployment situation gave rise to a tendency to reduce the retirement age and to retire employees at the minimum age permissible to make way for the younger unemployed.

A further great expansion in the establishment of pension plans took place after World War II, but the situation otherwise was radically different. Shortage of labour, increased longevity and improved standards of fitness at older ages led to a reversal of the retirement policy, and older employees were encouraged to continue in service as long as they were capable of carrying out their duties. To this end, the maximum number of years that could be reckoned for pension purposes was raised in some cases; *e.g.*, from 40 to 45.

Further, in order to facilitate change of employment without the loss of accrued pension rights, a system was introduced in the public and semipublic services in Great Britain under which, by mutual agreement, a "transfer value" is paid out of the pension fund from which the employee is withdrawing into the fund of which he becomes a member, and he is then given credit in the latter scheme for an appropriate number of years of past service.

SUPERCONDUCTIVITY, a phenomenon which occurs in various pure metals and alloys when the material is cooled to within a few degrees of the absolute zero of temperature. Superconducting material is characterized by zero electrical resistance, *i.e.*, electric charge is apparently transported through the superconducting specimen without friction, and by the fact that no magnetic field can exist inside the material. The flux lines of a weak magnetic field are expelled from the specimen when it is cooled below its superconducting transition temperature; a moderately strong magnetic field, on the other hand, completely destroys the superconducting state of the material.

Superconductivity was discovered by Kamerlingh Onnes in 1911, but the perfect diamagnetism of superconductors was not discovered until 1933 (W. Meissner and R. Ochsenfeld). A microscopic interpretation of the superconducting state has been given by J. Bardeen, L. Cooper, J. Schrieffer and N. Bogoliubov (1957). *See* LOW-TEMPERATURE PHYSICS. (J. R. RE.)

SUPERFLUIDITY, a phenomenon which occurs in liquid helium when it is cooled to within 2.2° of the absolute zero of temperature, *i.e.*, when it is cooled below −271.0° C. This temperature, the so-called lambda point, marks the transition between two phases: liquid helium I and the low-temperature liquid helium II. The low-temperature phase appears to be a mixture of two fluids, the normal and the superfluid, with the fraction of superfluid increasing as the temperature is lowered. Certain properties of the superfluid give rise to its name—it can pass with ease through very narrow capillaries or small cracks, thereby showing no viscosity, and it conducts heat extremely well.

It is known that the helium molecule is a Bose-Einstein particle, and the phase transition at −271.0° C. is apparently closely related to the condensation mechanism of an ideal Bose-Einstein gas. *See* LOW-TEMPERATURE PHYSICS. (J. R. RE.)

SUPERINTENDENT, a term which, apart from its general use for an official in charge, has a distinct religious connotation, being applied, *e.g.*, to the head of a Sunday school and to the chief minister in a Methodist circuit. In its most important historical sense it refers to certain ecclesiastical officers of reformed churches of the Lutheran model.

In the confusion of the Reformation the name of superintendent was given to a class of men who discharged many of the functions of the older bishops, while bearing a character which in several respects was new. Only in Denmark was the name of "bishops" reserved for the new officers after the Lutheran model had been adopted and the older bishops had been deposed and imprisoned. It survived there, though no claim is made that it is the sign of formal apostolical succession. In Scotland the *First Book of Discipline* provided not only for ministers, teachers, elders and deacons, but also for superintendents and readers. The superintendents (who were appointed because of the scarcity of Protestant pastors) took charge of districts corresponding in some degree with the episcopal dioceses, and made annual reports to the general as-

sembly of the ecclesiastical and religious state of their provinces, in the churches of which they also preached.

Duties of Superintendents.—The distinctive character borne by the new officers was determined by the cardinal principles which Luther had laid down in his work regarding the religious functions of the state. He conceived of the secular government as an ordinance of God, and as being set to direct and control the external fortunes of the church.

He hoped that righteous magistrates would at all times form a sound court of appeal in times of ecclesiastical disorder, and that they would guard the interests of truth and justice more securely than had been done under papal jurisdiction. The superintendents, who now had to undertake large administrative responsibilities in the church, were therefore to be appointed by the civil power and to be answerable to it. They were to stand as intermediaries between the prince or magistrates on the one hand, and the ministers in their districts on the other.

In his earlier writings Luther had laid his main emphasis on the spiritual priesthood of all believers. Every sincere Christian was declared free not only to preach but also to administer the sacraments and to rebuke evil livers. The differences in office and function between the members implied no difference in rank, for the members of Christ's church were all members of His body, and Luther believed that they would all be ruled into true order and charity by the Head. But he was shaken by the Peasants' War, and his faith in the virtues of the average man never recovered itself.

The result was seen in his later writings, where he expresses his conviction that men need to be directed and restrained from without, and he looks to the state to undertake this duty. In the last resort the civil magistrates must take control of the church. His vindication for thus subordinating the ecclesiastical to the civil lay in his assumption that the rulers of a Christian land would themselves be Christian, and it was the Christian duty of the church to render obedience to those who had been ordained of God to bear rule. He was as firm a believer in a visible Catholic church as were any of those of whom he speaks as "the adherents of the old religion." Luther, always conservative in feeling, clung to an alliance with the state and denied that the repudiation by the reformers of papal authority had severed them from the visible church.

The character of the office and duties of the superintendent were not everywhere the same. Luther shrank from imposing any stereotyped forms and asked that the special circumstances of each separate district should be consulted. He hoped that as few changes as possible would be made, and trusted that the reformed doctrines would spread peacefully throughout the country. After the Diet of Speyer (1526) the civil authorities were invited to reorganize the Church in their respective dominions as they thought best. In the free towns superintendents were answerable to the city fathers for their good order. There were difficulties in the territories of the German princes, and in the case of Saxony Luther proposed to the elector that his first step should be to send out a commission of visitation which should report on the moral and spiritual condition of his principality, district by district. His proposal was carried out, and Luther himself became one of the visitors (1527–1528). He found the people in a state of such religious indifference and ignorance, and the clergy living often in such grossness, that his faith in their fitness to govern themselves ecclesiastically sank even lower than before, and he resisted all schemes for self-government such as had been proposed by Francis Lambert. The church organization which he devised for Saxony provided no place for democratic or representative elements: the grasp of the state must at all times be felt. The superintendent must speak at all times as a minister of the state, and the state must be represented in the synod to which he makes his first report, for upon the synod there must sit not only the pastors but also a delegate from every parish. If any appeal should be made from the decisions of the synod it must be heard in the court of the electoral prince, for he, as supreme civil ruler, possessed the *jus episcopale*, the right of oversight of the churches. Luther proposed that he should exercise this right by appointing a consistorial court composed in part of theologians and in part of canon lawyers, and it was thus that in 1542 the Wittenberg ecclesiastical consistory was formed. Other principalities adopted the model, so that the institution became common throughout the Lutheran churches.

In this scheme the superintendent (or superattendant) was charged with such part of the duty of the older bishops as had been purely administrative. He must concern himself with the discharge of their duties by the pastors of the churches, as well as with their character and demeanour. He must supervise their conduct of public worship, as well as give them licence to preach. He must take cognizance of their ministry to the indigent in their parishes, and of their management of the schools. He must further direct the studies of candidates for the pastoral office. He was answerable to the civil authorities to report all evil-living and false teaching.

The earliest occasion of the appointment of such a superintendent would seem to be found in the decisions of Prince John of Saxony about 1527. He assigns the duties of the office, and summons the newly appointed officer to give diligent heed to the conduct and teaching of the pastors under him, faithfully to warn them of all errors, and, in case they prove obstinate, to report them to the electoral court. He must further give close attention to the due observance of the marriage laws, for in this matter the previously appointed visitors to the principality had reported grave laxity.

Some of the smaller principalities appointed but a single superintendent for their territory, who, instead of being answerable to a consistory, sat as spiritual member on the territorial council, whilst in towns the superintendent was summoned to the town council whenever Church matters arose for discussion. In larger states there were various classes of superintendents with their respective duties severally assigned.

In modern times the functions of the superintendent have been somewhat confused in consequence of the introduction into Lutheran Church theory of inconsistent elements of Presbyterian and synodal type.

See T. M. Lindsay, *History of the Reformation* (1906), i. 400–416; and the articles "Kirchenordnung" and "Superintendent" in Herzog-Hauck's *Realencyklopädie für protestantische Theologie und Kirche*.

SUPERIOR, the most northwesterly city of Wisconsin, is situated between the bays of St. Louis and Allouez on Lake Superior; the seat of Douglas county. With its twin port, Duluth (*q.v.*), it forms the western terminus of the St. Lawrence seaway; through the natural harbour shared with Duluth are shipped iron ore, coal, grain and crude oil. The Superior harbour shipping facilities include four ore docks, averaging 2,250 ft. in length and 80.5 ft. in height, with a total capacity of 500,000 gross tons; eight grain elevators with a 30,500,000 ton capacity; and ten coal docks, capacity over 3,500,000 tons. Crude oil arrives at Superior from Canada through a 1,127-mi. pipeline to be refined. Other industries include shipbuilding and the manufacture of coal briquettes, flour, beer, dairy and wood products.

Pop. (1960) city 33,563; Duluth-Superior standard metropolitan statistical area (St. Louis county in Minnesota and Douglas county in Wisconsin) 276,596.

The Superior area was occupied originally by the Sioux and Chippewa Indians. In 1655 the Jesuit Father Claude Allouez built a mission near the bay which bears his name. Pierre Esprit Radisson and Médard Chouart des Groseilliers arrived in 1661. Trading posts were established about 1678 by Daniel Greysolon, sieur du Lhut (or Duluth) and about 1820 by the Hudson's Bay company. Permanent settlement did not take place until 1853, one year after the government survey made by George R. Stuntz. Lands were purchased and settlement was made at the eastern end of present Superior, the Allouez bay area, where names of avenues recall early speculators such as William Wilson Corcoran, former Sen. Robert J. Walker and Congressman John C. Breckinridge. A second settlement was made in 1882 by Gen. John H. Hammond; he organized the Land and River Improvement company which platted the townsite of West Superior on St. Louis bay. The third townsite was South Superior. Both West and South Superior were

joined with the original East End settlement and the united municipality was incorporated as the city of Superior in April 1889. Between 1941 and 1959 the city was administered by a city manager and seven councilmen; then the mayor-alderman type was adopted.

Superior is the site of a Wisconsin State college, established in 1893. The city has public and parochial high schools, a Carnegie library with several branches, and the Douglas County Historical museum.

Proximity to Lake Superior, numerous small lakes and Brule river, famous for trout fishing and canoeing, make the area a tourist haven. Along the St. Louis river is Billings park. Manitou falls, the largest in Wisconsin, is 14 mi. S. of Superior.

(E. F. Gr.)

SUPERIOR, LAKE, named from the French *Lac Supérieur,* meaning "Upper Lake," and the most northwesterly of the five Great Lakes of North America, is the largest body of fresh water in the world. It is bounded on the east and north by the province of Ontario, on the west by the state of Minnesota and on the south by Wisconsin and Michigan. The general form of Lake Superior is that of a wide crescent convex toward the north. It is 383 mi. long (on a straight line) in an east-west direction and its greatest width is 160 mi. The surface area of the lake is 31,820 sq.mi. and the area of its drainage basin, exclusive of lake area, is 48,180 sq.mi. The mean height of the lake above sea level is 602 ft., 22 ft. above that of Lake Huron, to which it is joined at its eastern extremity by the St. Marys river. Its maximum recorded depth is 1,333 ft., 731 ft. below sea level. (For comparison with the other Great Lakes, a discussion of their origins, connections, utilization, etc., *see* Great Lakes, The.)

History.—Pierre Radisson and Médard des Groseilliers traveled extensively on Lake Superior in 1659–60 and returned to the east with a valuable cargo of furs. Father Claude Allouez circumnavigated the lake by canoe in 1667, and prepared a map of the lake which was the most accurate drawn until recent times.

Daniel Greysolon, Sieur du Lhut, known as Duluth, opened Lake Superior to active trading in 1679. Fur trading by the French then flourished at intervals, but in 1763 the entire region was ceded to Great Britain. Lake Superior was under formal British control until the close of the American Revolution in 1783; after that, trade continued in the hands of the British until 1817 when John Jacob Astor's American Fur company took control of trade south of the international boundary line.

Physiography.—The lake receives the drainage of approximately 200 rivers. The largest tributary is the Nipigon river, entering from the north and draining Lake Nipigon, and the second largest is the St. Louis river, entering from the west at Duluth. Other principal rivers of the north shore are the Pigeon, which forms the international boundary northwestward from the lake, the Kaministikwia, which enters Thunder bay at Fort William, the Pic, the White and the Michipicoten. No large rivers empty into the lake from the south. The outlet, the St. Marys river, carries a mean annual discharge of 73,100 cu.ft. per second.

Going outward from the lake in any direction except the southeast, one soon comes to an escarpment above which is a distinct upland overlooking the lake basin. At many places the escarpment, 400 to 800 ft. high, descends into water 500 to 800 ft. deep. The coast line is picturesque, the north shore particularly being indented by deep bays surrounded by high cliffs. The principal islands in the lake are Isle Royal, 44 mi. long, near the northwestern shore, which is a U.S. national park; Michipicoten, 17 mi. long, in the eastern part; and the Apostle Islands, a group off the Wisconsin shore in the southwest. Abandoned shore lines indicate that the lake once stood about 200 ft. higher and drained southwestward to the Mississippi river while glacial ice occupied the eastern part of the basin, and that later the lake discharged into Green bay of Lake Michigan before its present outlet was uncovered by melting of the ice sheet.

The ancient (Pre-Cambrian) rocks which nearly surround Lake Superior contain valuable mineral deposits. For many years the Lake Superior district was the largest producer of copper in the United States. Iron ore was mined and smelted locally in the Negaunee and Marquette areas during 1848–54. The opening in 1855 of the ship canal along St. Marys river (*see* Sault Sainte Marie) was followed by regular shipment of iron ore from the Marquette district to the lower lakes. Subsequently iron was produced from many parts of the Lake Superior district, including the Mesabi range in Minnesota (*see* Iron and Steel). By the second half of the 20th century high-grade iron ores were becoming scarce but improved technology permitted the use of abundant taconite and other low-grade ores.

Harbours and Commerce.—Lake Superior is fairly well provided with natural harbours, and works of improvements have created additional harbours at various points.

Fort William at the mouth of the Kaministikwia river and Port Arthur, 4 mi. distant, with an artificial harbour, are the lake terminals of two Canadian transcontinental railway systems and they ship grain from the western prairies. Taconite Harbour, Minn., formed by joining two islands to the shore with breakwaters, and Two Harbours, Minn., on a natural bay improved with breakwaters, are important shipping points for iron ore. Duluth, Minn., and Superior, Wis., are located at the western end of the lake on a large natural harbour, formed by a sand bar across the mouth of the St. Louis river, with two entrances maintained by piers. This port ships a large tonnage of iron ore, grain and flour. The principal ports along the south shore of the lake are Ashland, Wis., with a harbour formed by a breakwater; Houghton and Hancock, Mich., located on a 25-mi.-long canal across the Keweenaw peninsula; and Marquette, Mich., with an artificial harbour. These ports ship large tonnages of ore. The traffic to and from Lake Superior via the United States and Canadian locks at Sault Ste. Marie amounts to approximately 100,000,000 tons annually.

A large part of the land surrounding Lake Superior is very sparsely settled and much of it is a wilderness. Extensive forests are held in federal, state, provincial and private timber lands. Hunting and sport fishing in season and tourism in the summer months form the basis for an increasingly important recreation industry in the region.

Bibliography.—Canadian Hydrographic Service, *Great Lakes Pilot,* vol. iii and charts (annually); U.S. Army Corps of Engineers, U.S. Lake Survey, *Great Lakes Pilot* and charts (annually); G. L. Nute, *Lake Superior* (1944). (J. L. Hh.)

SUPERNATURALISM, a belief in an otherworldly reality that, in one way or another, is found in all forms of religion.

The idea of nature is unknown to primitive man, who inhabits a wonder world charged with the sacred power of *mana* (*q.v.*) and spirits. Primitive man associates the uncanny, the unusual or the powerful of any kind (including so-called natural phenomena such as the rising of the sun and the prowess of a warrior) with the presence of a sacred power.

In the higher religions, a gulf usually is created between the supernatural and natural realms. Thus, for example, Christianity distinguishes between this world and the Kingdom of God, or nature and grace, or earth and heaven. Hinduism in its various forms draws a polar distinction either between nature and spirit or between *samsara* (the world) and *Brahman-atman* (the deity); and in the Vedanta school of Hinduism the world becomes *maya* (illusion).

Early Buddhism drew a radical distinction between the world as pain and suffering (*samsara*) and the redemptive goal of the disciple (*nirvana*), while refusing to look upon *nirvana* as a supernatural realm of existence. Certain later schools of Mahayana Buddhism postulated a mystical and paradoxical identification of *nirvana* and *samsara.* This identification of the world and the Beyond, the profane and the sacred, is virtually unique in the history of religions, although it has a parallel in the Hasidic school of Jewish mysticism.

Far eastern religions have never abandoned their primitive roots; and in Confucianism, Taoism and various sects of Chinese and Japanese Buddhism no genuine distinction is made between the natural and the supernatural.

The secularization of modern western civilization has created a gulf between the natural and the supernatural resting upon the

idea of an autonomous nature existing apart from the influence or control of God. Here, the world becomes a profane reality wholly isolated from the sacred. This situation led to Nietzsche's proclamation of the "death" of God and his condemnation of Christianity as a flight from reality and an escape to an otherworldly nothingness.

So pervasive is the modern experience of the unreality of the supernatural realm that certain schools of Protestant theology have collapsed the distinction between the natural and the supernatural. This is fully apparent in process theology, as exemplified by H. N. Wieman, and partially so in the existentialist school of dialectical theology led by Paul Tillich and Rudolf Bultmann. In the one case is found an affirmation of a rational and scientific universe that will admit neither the existence of a supernatural realm nor of a God who exists wholly apart from the world; in the other, a dialectical conception of faith that conceives it as leading the believer in some sense beyond the world while at the same time insisting that true faith leads the Christian into the depths of his present existence in such a way as to exclude the possibility of a belief in an actual supernatural realm.

BIBLIOGRAPHY.—R. R. Marett, *The Threshold of Religion* (1914); Rudolph Otto, *The Idea of the Holy* (1923); Mircea Eliade, *Patterns in Comparative Religion* (1958); H. N. Wieman, *The Source of Human Good* (1946); Paul Tillich, *The Protestant Era* (1948). (T. J. J. A.)

SUPERNORMAL CHILDREN, EDUCATION OF. L. M. Terman calls children "superior" if they have intelligence quotients of more than 110—*i.e.*, if they mature, mentally, more than 10% faster than the average. He designates as "gifted" those with intelligence quotients of more than 135. The education of "superior" children does not present a major problem. They do not diverge so widely from the average as to require any basically different educational approach.

"Gifted" children are almost a distinctive type, and require, for their adequate development, a distinctive kind of education. Not only are they from two to four or more years further advanced in mental capacity than other children of their age, but their superiority tends to be general. Contrary to popular notions, gifted children are, on the average, large for their age and well developed; and they excel in many personality traits. It is in this group that there reside the greatest potentialities for distinguished achievement in life. For this reason it is important to adapt their schooling to their needs.

The commonest, and least satisfactory, way of trying to adjust the school to the needs of gifted children, is to have them skip grades. This does almost nothing toward developing their special abilities, though better than retaining them in lower classes.

The next commonest practice consists of "ability grouping." Children are classified as X, Y, or Z according to their general ability as shown by intelligence tests, teachers' judgment, etc. The brightest, or X, group is, ideally, given an "enriched curriculum," going into each subject more intensively and extensively. The large majority of the children in the X group, however, are merely "superior" rather than "gifted," and the ability grouping is, accordingly, adjusted to their needs. It is not an adequate solution to the problem of the really gifted child.

A third method of caring for gifted children is through classes organized especially for them. Germany was the pioneer in this field, but in the United States by 1930 there were 135 such classes. They depart widely, in the best instances, from usual school practice. Routine drills and reviews give way to independent investigations and reports, the children using libraries, laboratories, shops, museums, and excursions. There is considerable freedom. Creative expression and originality are stressed.

The most common objection to such classes for gifted children is that they may tend to make children conceited. Proponents of special classes, however, claim that direct experience does not bear out this criticism, and argue that children competing with their peers are less prone to conceit than are those who are constantly associated with children of inferior ability.

A more practical objection lies in the fact that since less than one percent of school children are gifted, only the larger cities have a sufficient number, at the different grade levels, to make such classes feasible. In those cities, however, where special classes for gifted children have been organized, the consensus is that the gifted children get a real chance to develop their abilities; and as experience accumulates, such classes should be able to provide almost ideal conditions for them.

In smaller communities, the problem is more difficult. Hope lies in some of the newer general methods of education. So-called progressive schools, particularly, stress for all children just the types of education found most suitable for gifted ones—freedom, responsibility, and much more emphasis on original investigation.

Where individual progress is added to these "progressive" methods, as for example, in Winnetka, Ill., a partial solution appears. Gifted children are allowed to progress somewhat more rapidly; their individual programs are enriched with special assignments, electives, and responsibilities, and they are given considerable freedom, participation in group activities, and stimulus toward original investigation and creative work.

This kind of procedure in smaller communities, and, in large cities special classes for gifted children, are, up to the present, the best solution that has been found to the as yet inadequately solved problem of the education of gifted children. (C. WAS.)

SUPERNOVA: *see* NOVA AND SUPERNOVA.

SUPERSTITION. Like many words having a derogatory sense, "superstition" is often loosely and vaguely used, and is therefore not easy to define. To identify it, for example, with false belief or practice would be quite misleading. For in the first place, the field in which superstition is generally to be found, the magico-religious, is the very one in which the standard of truth and falsehood is most subjective and fluctuating; in the second, a custom or belief, religious or not, which is false may nevertheless be accepted at some times and places by men of enlightened intellect and conscience. To call a person superstitious, however, regularly implies more or less definite mental or moral deficiency. Thus, we should describe a man as superstitious who in a modern European or American community seriously believed that to break a mirror brought bad luck; but the original holders of that or a similar belief were perfectly reasonable in their views, according to their lights. Supposing that a shadow or reflection was in some sort a part of the soul, they naturally concluded that to break it, by breaking the substance upon which it was cast or reflected, was to injure the soul or life itself. They had merely been misled by a false inference. In like manner many savage customs and ideas, often described as superstitious by explorers or missionaries, turn out on investigation to be the product of quite sound reasoning, vitiated by the false premises on which it is based. To adhere irrationally to these premises after having their falsity clearly demonstrated might indeed be called superstition.

The word in question being of Latin origin, light may be thrown upon it by the usage of a Latin author. Vergil, in a well-known passage (*Aen.* viii. 187), characterizes *superstitio* as *uana* (empty, groundless) and also as *ueterum ignara deorum* (having no knowledge of ancient gods, *i.e.*, of well-tried and long-established religious ideas). If we turn to the Greek equivalent δεισιδαιμονία, literally "fear of superhuman powers," we find Theophrastos (*Characteres*, 28–[16]) ridiculing the δεισιδαίμων, not for holding polytheistic views nor for believing in omens, but for spending much time in the worship of obscure and foreign deities, and paying a ridiculous amount of attention to petty omens, such as a more sensible man would disregard.

We may now perhaps attempt to define superstition as *the acceptance of beliefs or practices groundless in themselves and inconsistent with the degree of enlightenment reached by the community to which one belongs.* It is clear that such a definition excludes, for instance, the mental attitude of one who, about the year A.D. 50, or in the middle ages, believed in astrology, which was accepted and defended by many, though not all, intelligent and well-educated persons of the time; it also excludes the action of a savage who adores what is to Europeans a ridiculous and nonexistent godling. But it includes, for example, those contemporaries of Plato who let themselves be deceived by the lower practitioners of Orphism, or by those moderns who appear seriously to believe in mascots. Even so, a certain element of vague-

ness is unavoidable, since agreement is far from being reached as to what ideas, and consequently what practices, especially those having reference to things not obviously material, are false and unenlightened. A long list might be drawn up of things which some would regard as false and outworn, others as plausible or even certainly true; it would range from theistic belief and religious observances of any sort to such matters as second sight and dowsing or water-finding. There still remain, however, a great number of ideas and actions which the consensus of educated modern opinion would regard as superstitious in the sense above defined. These remaining superstitions fall into two categories, *survivals* and *accretions*.

Survivals.—It is well known (*see* ANTHROPOLOGY; FOLKLORE) that a great many customs characteristic of a less advanced stage in social and intellectual evolution survive into higher stages, either as meaningless and fossilized customs or because a new significance and use has been found for them. Thus many performances originally magical survive, more or less modified, as games for children or adults; a new use, the satisfaction of the play-instinct, has been found for them, but in their details they show, on investigation and comparison, traces of their original purpose. Again, the very old Northern European custom, in origin probably a fertility-rite, of decorating the house with greens at Christmas, survives simply because it is a pretty and picturesque traditional usage. Some such customs have become matters of etiquette. The need being still felt for some kind of conventional gesture to express friendly feeling or desire for better acquaintance, the old, probably magical, gesture of "handshaking" is still in use, and from childhood, we are taught to use the right, or lucky, hand for the purpose. The reason now given is that it is polite to use the right hand rather than the left; here as in many other instances what is now polite was once magically good. Such things as these can hardly be called superstitions; certain customs being still found necessary in various spheres of modern life, and the old ones being in themselves inoffensive, it saves mental effort to retain them, although their original meaning was something quite foreign to our conceptions. But alongside of these some things survive which are useless or even obnoxious, on any theory of conduct, expressed or understood, save one involving some now exploded doctrine.

A good example is furnished by rest-days. The whole experience of mankind proves that it is advantageous to have occasional periods in which no avoidable work is done, since in the end the worker actually achieves more than if he had not rested, in addition to tiring himself much less; and for this there are sound physiological and psychological reasons. Moreover, it is the perfectly reasonable practice of members of the higher religions, such as Christianity and Islam, to set apart certain days (Sunday, Friday, etc.) for the intensified practice of their worship. But it is well known that, long before either an ethical religion appeared or the physiology of fatigue was studied, tabu-days, like the *gennas* of the Nagas, were observed by people in a savage or barbarous stage of culture, owing to the belief that, at regular or irregular intervals, the prevailing conditions were so magical that ordinary work would be unlucky or would prevent or hinder the beneficent effect to be expected from these conditions. Now the feeling that on certain days certain occupations, or all ordinary occupations, are particularly unlucky (or lucky) is quite common at far higher stages than these. Those who appear seriously to believe in the unluckiness of Friday for starting any work may rightly be called superstitious, for the ground of their conduct is ultimately the savage notion of the incompatibility of ordinary occupations with the spiritual atmosphere, so to call it, of a tabu-day, in this case a Church fast-day of comparatively modern origin. Not dissimilar considerations apply to the quite wide-spread avoidance of the number thirteen, ultimately based upon very ancient ideas of the peculiar character of this number.

Accretions.—Evidence of the survival, among a civilized population, of a savage type of mentality is furnished abundantly by the occurrence of quite new superstitious practices in our own day and in circles, not always uncultured, of Europeans and Americans. It is of course a commonplace that gamblers are superstitious, and many of their practices may fairly be deemed survivals, such as turning a chair around to "change the luck" (imitative or homoeopathic magic) or wearing a mascot (fetishism?). But examples are not wanting of entirely new beliefs, based upon modern conditions and having no near parallel in savage or ancient custom. A popular English newspaper finds it necessary, owing to its large circulation, to employ a number of presses. In order to check, and, where necessary, correct the work done by each of these, it has its title printed in a slightly different manner by each machine, the difference consisting of a varying number of small white dots in the large black lettering. A few years ago a firm belief grew up among its less instructed readers that these dots conveyed racing "tips"; thus, if there were three dots in the letter D, they signified that in the third race of that day a horse whose name began with D was the likely winner.

Here we have a definitely superstitious belief, although connected with no magical or religious idea, old or new. Clearly such a delusion would have been impossible but for two factors: (1) Instead of rationally crediting the paper with the possession of a good deal of information on various subjects, gathered through reporters and by other such means, these readers manifestly supposed it to have a quasi-omniscience, embracing such unknowable things as the result of a race not yet run. (2) Whereas it is in fact the custom of newspapers to state in plain language, in their sporting columns, which horse they think likeliest to win a given race, it seems to have been vaguely felt that it was more in keeping with the supposed oracular and infallible wisdom of the Press to adopt a symbolic style and set forth valuable information in a riddling form.

The above example throws some light on the question whether superstition is declining in modern civilizations. The statement that it is rapidly disappearing is often and confidently made; but on examination of the instances given in support of such assertions, it will be found that what are really disappearing are certain old and traditional forms of belief or custom, which have rather gone out of fashion than been overcome by reason. For example, in Britain and America very little is left of the old rites connected with the last sheaf. This is only partly due to the spread of education among country people; probably not many farm labourers could say either what the old rites were or why they were irrational; it is rather the result of the rapid drift of population towards industrial centres, and the consequent influence on the country of town ideas. Now the townsman, not being a producer of foodstuffs from the soil, is apt to laugh at the customs connected with them as rustic; compare the contemptuous tone of such words as *rustre, rusticus, boor,* ἄγροικος, etc. The irrational custom has been irrationally displaced. Here, therefore, is no evidence of the growth of a more logical and critical mentality, less prone than the old to superstition; and it has already been shown that superstitious beliefs are actually in process of creation in the towns themselves. To take another example; while the traditional belief in witches seems practically confined to a few country districts and small towns, and the old methods of divination hardly survive at all, save as amusements, fortune-telling by various methods, mostly pseudo-scientific, including a crude and degraded form of astrology, is by no means extinct, as may be seen by examining the advertisements of popular magazines and also the police-court reports. At best it may be said that superstition appears to be yielding, although very slowly, to the gradual rise of the average intelligence which there is some reason to suppose exists, and that it may in time disappear with other irrationalities.

BIBLIOGRAPHY.—For classical superstitions, *see* art. "Aberglaube" in Pauly-Wissowa, *Realencyklopädie*. For others, *see* ANTHROPOLOGY: *Bibliography;* SOCIAL ANTHROPOLOGY: *Bibliography*.

See also Lehmann, *Aberglaube und Zauberei* (1898); A. Wuttke, *Der deutsche Volksaberglaube der Gegenwart* (1900); Philip F. Waterman, *The Story of Superstition* (1929). (H. J. R.)

SUPER-TAX: *see* INCOME TAX.

SUPERTONIC, in music, the second degree of the diatonic tonal scale, as D in the scale of C, that is, the note next above the tonic, whence its name.

(*See* HARMONY.)

SUPILO, FRANO (1870–1917), Yugoslav statesman, was born of poor Croat parents at Cavtat (Ragusa Vecchia) in southern Dalmatia. In 1900 he became editor of *Novi List*, a Croat paper published at Fiume (Rijeka), which became a rallying ground for opposition to the intolerable regime of Count Khuen Héderváry.

In 1905, together with Anton Trumbić, Pero Čingrija and Smodlaka, he drew up the resolution of Fiume, which became the basis of political co-operation between Serbs and Croats in the critical period before World War I. When the Magyar coalition parties came into power in Hungary in April 1906, and placed the Serbo-Croat coalition in office at Zagreb, Supilo became a deputy at Budapest.

Within a year Magyarizing tendencies produced an acute conflict between Budapest and Zagreb, and for a time Supilo was the soul of Croat resistance. He thus became a marked man and was specially aimed at in the notorious forgeries by which the Austro-Hungarian foreign office attempted to prove Serbia's revolutionary intrigues in Croatia and thereby to justify the annexation of Bosnia in 1908.

At the Friedjung trial, however, in Dec. 1909, Supilo was triumphantly vindicated. He was the first to discover the secret negotiations with Italy which culminated in the treaty of London (April 26, 1915), and earnestly warned Sazonov of the disastrous complications to which it would give rise; but he was unable to prevent the promise of wide Slav territories on the eastern Adriatic to Italy in the event of victory. Supilo came into conflict with the Pan-Serb and reactionary tendencies of Pašić, and even withdrew from the Yugoslav committee, which in the early stages of the war he regarded as unduly subservient to Belgrade.

But before his premature death (in London, Sept. 23, 1917) he heartily endorsed the declaration of Corfu, issued in July 1917, which laid down the lines upon which the Yugoslav state was to be constructed.

SUPPÉ, FRANZ VON (Francesco Ezechiele Ermenegildo Cavaliere Suppé-Demelli) (1819–1895), Austrian composer of light opera, was of Belgian descent and was born at Spalato, Dalmatia. At 15 he produced a mass at the Franciscan church at Zara. Sent to study philosophy at Padua, Italy, he took lessons in music from Cigala and Ferrari, and on his father's death went to Vienna, where he became unpaid conductor of a small theatre, and later (1865) of the Leopoldstadt, where he remained until his death. His first success was *The Country Girl* (1847) and in 1858 he produced a singspiel, *Paragraph 3*, which became popular in Germany. Suppé wrote two grand operas, a symphony, quartets, a *Missa dalmatica* (1877), some church music and songs, as well as 31 operettas. Some of his works were produced in Paris, London and New York; among the most popular were *Zehn Mädchen und kein Mann* (1862), *Pique Dame* (1864), *Die schöne Galatea* (1865) and *Leichte Kavallerie* (1866).

He is chiefly remembered for his overtures, particularly *Dichter und Bauer* (*Poet and Peasant*), which has been arranged for 59 different combinations of instruments.

SUPPLEMENTARY ESTIMATES: *see* Budget, Governmental.

SUPPLY, provision; more particularly the money granted by a legislature to carry on the work of government. In the United Kingdom granting of supply is the exclusive right of the house of commons, and is carried out by two committees of the house, one of supply and the other of ways and means. In the United States supply originates in the house of representatives (*see* Budget, Governmental).

SUPPLY AND DEMAND. Economics, or political economy, is sometimes defined as the science of supply and demand. Although this is an inadequate definition it cannot be said to be altogether misleading. A very important part of economics, and the part which probably has the best title to the name of science, has to do with the operations of supply and demand and with the way in which variations of supply and demand are related to the movements of prices and to changes in the production and distribution of wealth. The "law of supply and demand" was not invented or discovered by the economists, however, nor do they lean very heavily upon it as a general explanatory formula. Long before there was any systematic analysis of economic processes men had observed that prices vary with supply and demand, and from the earliest days, traders have had to take account of that circumstance. The economist's task has been to scrutinize those characteristics of human behaviour and of the physical environment which determine the various forms or patterns in which supply and demand appear and to inquire into the complicated interactions of the demand for and the supply of different commodities and services.

Elementary Principles.—Consider the familiar theorem that the price of a commodity must be such as to make supply and demand equal. If supply is taken to mean the amount sold and demand the amount bought the theorem is mere tautology, for supply and demand become different names for the amount transferred from sellers to buyers at any price whatever. But if it be understood that demand means the amount which buyers would be willing to take at a specified price, that supply means similarly the amount which sellers would be willing to part with at a specified price, and that demand and supply vary in some systematic and continuous way and in opposite directions as the price is raised or lowered, the theorem has meaning and significance, for there will be one price, and only one price, at which supply and demand will be equal.

In another elementary theorem, namely that an increase of demand for a commodity will raise its price, that an increase of supply will lower it, and that a decrease of supply or of demand will have an opposite effect, other meanings are attached to changes of supply and demand. Here an increase of demand or supply means an increase of the amounts which will be taken at given prices, not an increase which is dependent upon a reduction of price. The general state of supply and demand, in the sense specified in the preceding paragraph, can be represented by lists or "schedules" of "supply prices" and "demand prices." In this other sense, however, supply and demand are regarded as independent variables, and a change of supply or demand means an alteration of the schedule of supply prices or demand prices, such as might come on the one hand from a change of consumer's preferences or an enlarging of the market or, on the other hand, from a change of costs of production.

It is proper to assume that at any given time the immediate general condition of supply might be represented by a schedule in which the progressively higher prices which are required to evoke a progressively larger supply are set forth. But if the commodity is one which can be produced more economically if produced in large quantities, the ultimate effect of an increase of demand, in the sense of an increase of the amounts which will be taken at specified prices, will ordinarily be to reduce the price per unit at which these larger amounts will be supplied. In a schedule of supply prices constructed on the assumption that sufficient time is allowed to permit the necessary economies to be effected, larger supply will be associated with lower prices. When the long-period schedule of supply prices is of this type, the commodity is said to be produced under conditions of decreasing costs or of increasing returns. When, on the other hand, because of the scarcity of some necessary productive factor, increased supply cannot be had, even in the long run, except at a higher price, the condition is described as one of increasing costs or of diminishing returns. The factors which give rise to increasing returns should not be confused with the circumstance that in many industries certain outlays (*e.g.*, for plant and equipment) have to be incurred in advance or with the further circumstance that in a growing industry such outlays are ordinarily considerably larger than the volume of output immediately in prospect would require. Under such circumstances the additional or "prime" costs incurred by reason of an increase of output may be relatively small. Furthermore, with a progressive increase of output there will be a progressive diminution of costs per unit of output, because the general, supplementary, or "overhead" costs will be spread over a large number of units. But although when the market is sluggish or when competition is especially keen, prices may be cut to a point where they barely suffice to cover the addi-

tional or "prime" costs, this condition, which cannot be lasting, should not be confused with a true condition of increasing returns, for this last condition is to be found only when a gradual increase of output is attended, in the long run, with genuine economies.

Interactions of Supply and Demand.—The results obtained by taking account only of the supply of and demand for a particular commodity in relation to its price are no more than a first approximation to the truth. In isolating, for reasons of practical convenience, the factors which determine the price of any one commodity, taken by itself, economists are accustomed to assume that the value of money, to both buyers and sellers, is constant. This means that no account is taken of the way in which changes in the amount of money which consumers expend for the one commodity will affect their ability to buy other commodities, or the way in which an increase of the production of the one commodity will affect the ability of producers to supply other commodities. There are many instances of joint or complementary demand, as for fruit and sugar or for automobiles and gasoline, and of joint supply, as of mutton and wool, of coal gas and coke, of cotton and cottonseed. The general rule is, however, that consumers' outlays for any one commodity can be increased only by reducing the amounts which they expend for other commodities, and that more of any one commodity can be produced only by displacing other possible uses of productive resources. This general rule is not inconsistent with the fact that, making abstraction of the use of money as a medium of exchange, the supply of any one commodity is an expression of the demand of its producers for other commodities and services.

There is a sense in which supply and demand, seen in the aggregate, are merely different aspects of a single situation. It is for this reason that some of the older economists held that general overproduction is impossible—a theorem which, though not really erroneous, has proved to be misleading. The *effective* demand of the producers of one commodity for other products depends not only upon how much they produce, but also upon the relative demand of other producers for that particular commodity as compared with other products. Only so far as the demand for a particular commodity is elastic is it true in any significant sense that an increase of its supply is an effective increase of demand for other commodities. There may be and often are maladjustments of supply and demand. Furthermore, production in general may at one time outrun and at another time fail to keep pace with the expansion of money incomes. In either event there will be general fluctuations of prices, attended, as experience shows, by changes in the relative levels of the prices of different classes of goods and services.

The general form of the relations of supply, demand and price which obtain when all products are taken into account can be depicted mathematically in systems of equations, and thus the general character of the whole interdependent structure of prices can be laid bare. But empirical (statistical) studies of the relations between the fluctuations of the production of various staple commodities and fluctuations of their prices have shown that the first approximation previously referred to is generally a useful and often a surprisingly accurate approximation. It is necessary, of course, to allow for the effects of contemporaneous changes of the general purchasing power of money, and it is sometimes necessary to allow also for the effects of other important disturbing circumstances. But it is not necessary to take account of complications of a secondary order of importance in order to obtain "empirical laws of demand" for such commodities as wheat, cotton, sugar, beef and potatoes which appear to be fairly reliable, at least over periods of some years.

Inelastic Supply.—The rule that supply and demand may be regarded as functions of or dependent upon price must be so interpreted, of course, as to allow for the circumstance that the supply of something is fixed and is in no way responsive to an increase of price. As the production of other goods increases, the prices of these nonreproducible forms of wealth must inevitably increase, unless the demand for them falls off. If these nonreproducible things are necessary instruments in the production of other goods, as land is, then other goods will be produced under conditions of diminishing returns, unless this disadvantage can be offset by improvements in productive processes or by cheaper supplies of other necessary productive instruments. For some purposes it is convenient to assume that the aggregate supply of reproducible goods, or of reproducible productive goods, is fixed for the time being. The problems of supply and demand then have to do merely with the apportioning, by exchange, of an existing stock of goods, or with the assigning of productive instruments to the most important of their various possible uses. Thus the increase of the supply of labour in a given industry or a given locality may be taken to depend largely upon a possible transfer of workers from other industries or other localities. Whether labour in the aggregate may be said to have a supply price (*i.e.,* to be responsive in the long run to an increase of wages) is a question to which the Malthusian theory of population gave a more nearly unqualified affirmative answer than would be supported by the present opinion of scholars. (*See* also DEMAND; ECONOMICS; and PRICE.) (A. Yo.)

SUPPLY AND TRANSPORT, MILITARY. An army may be divided into two component parts: the combat troops and the service forces. The combat troops cannot accomplish their mission without the support of the service forces which help them to move and furnish them with the supplies and services necessary to live and fight.

Militarily, the term "supplies" includes all things necessary to equip, maintain and operate armed forces. It includes such items as food, clothing, arms, ammunition, fuel, gasoline and lubricants, construction and maintenance materials, vehicles and varied types of machinery.

Supplies are obviously of little value to an army unless an adequate transportation system is available to move them to areas where they are needed. Lack of adequate transport in a wartime situation will frequently cause an army to abandon a tactical plan which is otherwise feasible. An army's transportation system includes all means of transportation—water, highway, rail and air—at its disposal both within its home country and within and en route to its theatres of operations. The establishment of such means creates lines of communication. The operation of the means of transport must be co-ordinated and controlled to eliminate bottlenecks, ensure that they are carrying the personnel and supplies they are capable of moving, and give the transportation system the flexibility necessary to function in changing tactical situations. Movements of troops and supplies are classified as *tactical* when surface contact with the enemy is considered possible, and as *administrative* when such contact is not expected.

The basic principle of supply envisages a steady flow of supplies from rear to front. Forces in rear areas are responsible for procuring and transporting supplies to combat troops. This relieves tactical commanders from such administrative burdens and permits them to devote their major attention to the defeat of the enemy. During the planning stage for a campaign or battle, an army commander informs the supply services supporting him of his operational plans. Thereupon the latter services estimate their capabilities to support the tactical plan logistically. Especially in a large-scale operation, the final tactical plan normally represents a balance between what the army commander wishes to accomplish and the ability of the supply services to furnish him with the matériel and transport necessary for success.

Until the early 17th century, armies lived off the country. Sometimes they bought supplies; more often they obtained what they needed by outright confiscation. This system of supply so ravished various countries in Europe during the Thirty Years' War that it came into popular disrepute. Thereafter, army commanders customarily paid for supplies and stored them in warehouses to be used as needed. From this beginning a system of supply and transport evolved which became increasingly effective with continued improvements in mechanical transport. The huge armies of the 20th century with their varied and specialized needs could not live off the land alone even if they wished to do so.

World War I.—This war taught the armies of the world two important lessons. The first was that, other things equal, the army that could best supply its troops in the field would be victorious.

The second was that the supply of a complex modern army involves the manpower and resources of an entire nation. Germany enjoyed an initial advantage in the war because its army had built up a tremendous reserve of supplies of all classes and had worked out a detailed plan of distribution to its forces in the field. The British, French and U.S. armies were not so well prepared. However, as the war progressed, the combined weight of Allied resources, which were speedily organized for war, began to tell. Production in the farms, mines and factories of the Central Powers decreased as more and more workers entered the fighting armies. Under the strain of providing for both their armies and their civilian populations, the Central Powers could not maintain the pace of production that the Allies had set. By the time the war had ended, it was clear that the ability of the Allies to produce more war matériel and move it faster was as much responsible for victory as were the efforts of their combat troops.

WORLD WAR II

This conflict was also essentially a war of supply. The brilliant strategy of high commanders, the bold tactics of battle leaders, the skill and courage of the fighting troops, all contributed to the final outcome; but victory went to the nations who had best organized their total economic resources to produce the matériel of war and the transport to move it, and troops, to the world's battlefields.

Logistical Preparedness for War.—Before war started on Sept. 1, 1939, with Germany's invasion of Poland, no nation even approached the state of economic and industrial preparedness for war that was to be required later. Under Hitler's leadership, Germany had stepped up production for war since 1933 and was much stronger militarily than its neighbours by 1939. Its industry and scientists had developed synthetic rubber and gasoline and had built plants to produce both these important war materials in which Germany was deficient. The extent of economic mobilization in 1939, however, was exaggerated; only on the eve of Germany's attack on Poland was there such a thing as a military procurement program worked out for a period as long as two years. Germany was not prepared for a long, sustained war.

Italy, unlike its ally, did not possess the economic resources necessary for even a short war involving a large-scale military effort—as the Italian general staff had informed Benito Mussolini prior to World War II. This became apparent early when Italy's inability to supply its army invading Greece in the autumn of 1940 contributed heavily to the defeat of that effort.

When Japan struck at Pearl Harbor, it was ostensibly well prepared for war. Its merchant fleet was the third largest in the world and it possessed the world's third largest navy. In Japan proper, in Manchuria and in Korea were highly developed systems of railroads. Japan had the industrial potential of Manchuria to supplement its home industries, which were harnessed to support the military effort. For years it had exported the products of a flourishing textile industry—often at the expense of the civilian population which needed these products—in order to import scrap iron, oil and machinery required for war production. There was clear evidence, however, that Japan's estimate of the quantity and type of equipment and supplies required to prosecute a major modern war was based largely on its experiences in the Russo-Japanese and Sino-Japanese wars, and influenced only to a small extent by the experiences of the major powers in World War I and in the early phases of World War II. Throughout World War II, the Japanese high command was to display a surprising disregard for the logistical aspects of war. In this connection, a statement made by Gen. Toshio Nishimura, assistant chief of staff to Gen. Tomoyuki Yamashita, was significant. "In our army," he said, "the officers do not like to study logistics—it does not interest them—they want to study only how to fight. Even in the military academy it is difficult to get the cadets to study logistics."

Although the axis nations had apparently not visualized the scope of the supply and transport requirements which World War II would entail, their state of military preparedness was clearly superior to that of the Allies who had done little to ready their economies for war. The U.S.S.R., however, was an exception. On the eve of World War II the Soviet Union was prepared both militarily and economically for the impending conflict. Following the advice of Nicolai Lenin, the soviet leaders had created a large army with modern equipment. Moreover, through a series of five-year plans, they had built a sturdy heavy industry for the acknowledged purpose of providing a broad economic base for war—in addition to manufacturing the implements of peace. The agricultural machinery industry, for instance, built tractors needed for Russian collectivized farms, but these factories were readily converted to manufacture tanks. Aided by foreign capital and technical assistance, the U.S.S.R. had also developed an aircraft and automobile industry necessary in providing transport for a modern army.

More important yet, Russian leaders in the early 1930s had started to build up huge industrial centres beyond the Ural mountains which could function undisturbed during an enemy attack on the western U.S.S.R. When Germany invaded the U.S.S.R., the *Wehrmacht* faced armies supplied by a civilian industry disposed in depth. The power of later soviet counteroffensives made by armies equipped with excellent tanks, planes and impressive numbers of artillery pieces was ample proof of the effectiveness of the U.S.S.R.'s war industry.

Meanwhile, the Red army was handicapped by deficiencies in the soviet transportation system. Roadnets in the U.S.S.R. were not well developed, and the railway system had not kept pace with the industrialization of the country. In addition to supplying the armed forces, the railways were also needed to move dismantled factories eastward before advancing Germans. This was a serious strain upon the soviet railroad system, especially during the early stages of the war. Another transportation deficiency was the lack of merchant ships. Unlike Britain, the U.S.S.R. could not dispatch fleets of supply vessels to Allied countries to pick up supplies which it lacked.

The other major Allied countries, Great Britain, France and the United States, had done little to prepare their economies for war prior to its outbreak. France had recognized the importance of total economic preparation for war in the Defense law of 1938 which provided for a minister of armaments. However, little or nothing had actually been accomplished in the way of economic mobilization when Germany struck. France depended more on the Maginot line and on manpower than on its military equipment, much of which was essentially of World War I type.

The general staff of the French army had prepared an elaborate supply plan to support field armies in a theatre of operations which was expected to be in northeastern France. This plan, however, was based on archaic World War I principles and was highly inflexible. It provided for supply concentrations in expected areas of battle operations and depended almost entirely on railroads to the exclusion of motor transport. It was to prove incapable of supporting field armies matched against highly mobile German forces.

Great Britain was similarly unprepared. Not until a month before Poland was invaded did the British create the ministry of supply, which was to be their major national co-ordinating agency for civilian and military supplies. The royal navy, of course, was superior to any that the axis nations could assemble, and the royal air force had new and modern planes. The British army, however, lacked adequate modern equipment. Transportation was a major factor in the British supply scheme since much of the needed war matériel had to come from member nations of the British empire and later, through lend-lease, from the United States.

Like France, the United States had long recognized the principle of industrial mobilization for war. As early as 1922, the navy and the war department had planned for such mobilization through the medium of the army and navy munitions board. Lack of funds, however, prevented adequate measures for preparedness. When World War II started in Europe in 1939, the U.S. found itself woefully short of equipment for the expanding military establishment. Meanwhile, the demands of France and Great Britain for military matériel were swamping a small munitions industry. As this industry expanded, it became apparent that a system of allocating the nation's materials to industry would have to be instituted; otherwise the country faced the

threat of uncontrolled competition among the various procuring agencies that had characterized 1917 procurement efforts. After several national agencies with limited authority had been created to supervise allocation, the War Production board was organized in Jan. 1942 as the major co-ordinator of industry for the national defense program. In March of the same year, the army service forces, at first called the services of supply, was organized to co-ordinate the administrative, supply (including procurement) and service activities for the war department as a whole.

Early Phases of the War in the West.—In May 1940 England, France, Belgium and the Netherlands had troops on the continent with total numbers exceeding those of Germany. Initial superiority in transport and supply organization, as well as the tactical aggressiveness of the German armies, however, enabled Germany to defeat them with ease. Less than a year after France had fallen, the *Wehrmacht* was driving deep into the U.S.S.R. and in early 1942, the *Afrika Korps* was threatening the Suez canal.

At this time, however, Allied resistance began to stiffen, and the German field armies began to feel the effects of overstretched supply lines and the weaknesses of a supply policy designed primarily for short campaigns. In an effort to pour as many new combat units into the field as possible and thus ensure quick victories, Adolf Hitler had insisted that 90% of newly produced equipment should be furnished to newly activated combat units. The shortsightedness of this policy in a sustained war became apparent as the lack of replacements for worn, damaged or captured equipment impaired the efficiency of fighting units which had been in combat for long periods.

Preparing for the long war which they knew lay ahead of them, Britain and the United States started to build up the armies and the tremendous amounts of supplies and transport needed for a major offensive. Meanwhile, they launched smaller-scale offensives in North Africa, Sicily and Italy. After the strength of the German eastward drives in the U.S.S.R. had been spent, Russian armies, supplied mostly by their factories beyond the Urals, mounted a large-scale counteroffensive.

Logistical Aspects of Campaigns in the West.—In 1942, the British and Americans were considering a cross-channel invasion of France from England to aid the hard-pressed Russians. The Allies, however, lacked the large numbers of invasion craft and the supplies and troops needed to make such an assault and thereafter to sustain an offensive. Accordingly, it was decided to invade North Africa.

North Africa.—The U.S. army service forces had built up enough matériel in the United Kingdom to equip the U.S. troops there who were to invade Africa. However, they had transported combat units to England at the expense of enough service units to store and catalogue this equipment adequately. As a result much of the matériel could not be readily located, and duplicate shipments from the United States became necessary. This experience provided an object lesson to the U.S. army, which thereafter attempted always to provide service troops commensurate with the amounts of supplies they were required to handle.

After North Africa had been invaded successfully, the transport factor immediately assumed importance in subsequent operations. Allied forces rushing eastward toward Tunisia soon outraced their supply support. Railroads along the North African coast line were single-track and had little usable rolling stock. Roads were poor and the Allies lacked adequate motor transportation. As a result, the defeat of axis forces was delayed.

Sicily.—One of the major factors influencing the Allied decision to attack Sicily was the desire to gain additional protection for the Allied supply life line through the Mediterranean to the Suez canal. The spectacular speed with which the Allies overran Sicily resulted in minimum losses of supplies. As a result, there was considerable excess equipment in North Africa and Sicily available for later assaults in Italy.

Italy.—It was political and operational factors, rather than logistical factors, that in the main determined the Allied decision to invade Italy. Italy was not self-sufficient and the Allies were obliged to ship coal, food and clothing for the Italian people in addition to supplies for their armies in Italy. The narrow front in the Italian peninsula made it difficult for the Allies to get behind the Germans to cut lines of communication. Mountainous terrain and an inadequate road net hindered service forces supplying combat troops.

Western Europe.—Logistical considerations had played a major role in the decision to attack *Festung Europa* from Great Britain. The United Kingdom provided a ready-made base of operations with modern transport and cargo facilities located only a few miles from the enemy. While conducting their operations in the Mediterranean, the Allies had continued to build up troops and equipment for Operation "Overlord," the cross-channel invasion. To keep British ports operating at full capacity, the United States preshipped equipment to the United Kingdom for the use of combat troops who followed later. In the year preceding the invasion, the U.S. preshipped a total of 5,530,000 measurement tons of such equipment and supplies. During the same period 1,600,000 men were moved into the United Kingdom.

The problem of moving supplies across the beaches to sustain the divisions employed in Operation "Overlord" was particularly difficult. Pending the seizure of French ports capable of handling large quantities of supplies, artificial harbours were built to be used at assault beaches.

In the Allied race across France following the Normandy invasion, supply lines stretched nearly to a breaking point. Lack of adequate supplies, especially gasoline, caused the combat units pursuing retreating German armies to halt. Not until the Allies captured Antwerp, Belgium, a port closer to their forward fighting elements, were the supply forces able to provide the masses of matériel needed to continue the plunge across the Rhine into the heart of Germany.

Two months after the Normandy assault, the Allies' landing in southern France gave them the ports of Marseille and Toulon, thus relieving some of the pressure on the congested channel ports of France. It also enabled the Allies to use the north-south railroad and road nets in France to supplement the congested east-west highways and railroads in northern France.

As this tremendous logistical build-up of the Allies increased, Germany could not keep pace. The basic reason for this was that Germany had not mobilized its economy as effectively as had the Allies for war. In 1944, for example, only about one-third of all German output went for direct war purposes, whereas 40% of U.S. production effort at the height of the war was used for implements of war. At one time, the percentage of British war production to total production may have gone as high as 45%. As Allied strategic bombing mounted in intensity, German transportation and industry suffered. Faced with the crushing weight of Allied matériel from east and west, its armies succumbed.

Campaigns in the East.—Japanese strategy from 1932 onward was aimed at denying Chinese armies the support of the industrial areas of Manchuria, and China itself. Concurrently, Japanese armies attempted to starve China into submission by seizing agricultural areas or by making periodic forays in strength to confiscate or destroy harvests.

In Dec. 1941 and Jan. 1942 the Japanese seized the ports of southern China, thus denying the shipment of Allied supplies to poorly equipped Chinese armies. In March 1942 the Japanese capture of Rangoon closed the Burma road, by which Allied supplies had been carried overland to the interior of China. Only by air shipment over the Himalaya (the "Hump") could supplies then be delivered. Planes, however, could not carry heavy construction machinery and fighting equipment, or adequate amounts of the supplies needed by Chinese armies. Accordingly, the Allies began the herculean task of constructing the Ledo road with parallel pipe lines through almost trackless wastes from Assam, India, to a point on the Burma-China border where the new road could link up with the portion of the old Burma road not controlled by the Japanese. Meanwhile, the Allies still had to get supplies from the supply base at Calcutta to the beginning of the Ledo road in Assam. This necessitated the improvement of railroads and barge lines between these two points.

The rough terrain and lack of existing means of ground transport forced British and U.S. troops operating in the China-Burma-

India theatre to rely on planes alone to move and supply them on many occasions.

Pacific Campaigns.—Early in the war, the United States had decided to contain the Japanese advance in the Pacific while making the major effort against the axis in the European theatre. Transportation was an important factor influencing the formulation of this strategic decision. The United States lacked enough shipping to move the troops and supplies required for powerful concurrent offensives in both the Pacific and Europe—and distances to Pacific fighting fronts were almost twice as far from U.S. supply bases.

While committed to a strategic defense in the Pacific, the United States launched limited offensives to seize advance bases for air and logistic support of later offensives in strength. The objective of the 1942 campaign in New Guinea was to shorten supply lines by establishing major bases along the New Guinea coast to support future operations. The primary reason for the assault on Guadalcanal was to seize a forward air base. The U.S. also built up supplies at such bases as New Caledonia, Espiritu Santo and Guadalcanal, so that by June 1943 there was sufficient strength for further attacks in the Solomons. The "island-hopping" technique employed in the South Pacific consisted of amphibious landings on shores that were trackless jungles. Such operations required a rapid build-up of supplies for fighting troops and the speedy construction of roads and airfields by specially trained and equipped engineer units.

Meanwhile, the by-passing of islands by amphibious U.S. forces made the enemy-held garrisons liabilities rather than assets to Japan. Cut off by U.S. submarine, surface and air blockade, Japanese troops could not be supplied and reinforced and rapidly lost their potentialities as combat forces.

By the summer of 1943 operations in the Mediterranean had stabilized sufficiently to release some shipping for other uses. These ships, along with large numbers of new ships produced by an accelerated construction program, were then available to carry large numbers of troops and supplies to the Pacific from the west coast of the United States or from the base facilities which had been developed in Hawaii. In accordance with long-range plans developed by the U.S. army and navy, supplies and troops were moved to take part in the long-awaited major offensive in the central Pacific.

By mid-1944 the joint chiefs of staff of the U.S. armed forces began to make detailed strategic plans for the seizure of the Philippine Islands to be used as bases for the eventual assault on Japan itself. There had been some question as to whether it was preferable to attack Formosa rather than Luzon in the Philippines for use as a major base. Here again the logistical factor decided the question. The army service forces, which had made logistical studies of the Philippine Islands as early as 1942, declared that Luzon, with its superior ports, road nets and airfield facilities, would provide the better base. Accordingly, operational plans in line with the recommendation of the army service forces were adopted.

In addition to the Allied offensive, there were two important logistical factors which contributed to the final Japanese capitulation in Aug. 1945. In the latter phases of the war, Japanese troops became deficient in almost all classes of supplies and equipment because of inadequate home production and inability to transport war supplies and reinforcements to the far-flung garrisons in the areas they had conquered, especially after the Allies had secured air and sea superiority in the Pacific.

KOREAN WAR

Military transport assumed a dominant role at the very outset of the Korean war when it became obvious that without immediate reinforcement the Republic of Korea army was doomed. Troop transport planes of the United States air force ferried an understrength infantry battalion to Pusan less than 12 hours after the decision had been made to commit United States troops in Korea. Seaborne units followed rapidly.

As the fighting progressed, superior transport enabled outnumbered United Nations ground forces to hold the enemy and then counterattack. Lacking adequate reserves during the defense of the Pusan bridgehead, the 8th army used trucks to shuttle ground units quickly from one threatened sector to another. Superior sea transport permitted U.N. troops to make an amphibious landing at Inchon on Sept. 15, 1950, and to help support logistically a subsequent offensive that virtually destroyed the Korean Communist army.

By contrast, lack of adequate transport hampered the Chinese communists when they entered the war. After an offensive, the Chinese were obliged to halt for long periods while they moved up ammunition, equipment and food by oxcart, porters and a limited number of motor vehicles and railroad cars. Attacks by United Nations aeroplanes further complicated Chinese resupply and reinforcement. Thus delayed, the Chinese could not exploit their initial successes against an overextended but more mobile 8th army.

The United States assumed the major responsibility for furnishing and transporting military supplies to United Nations forces in Korea. To fight a war 6,000 mi. from its shores, the United States expanded rapidly a military transport system organized after World War II. In 1949 the United States navy's Military Sea Transportation service (MSTS) had assumed responsibility for all ocean shipping. From June 25 to July 31, 1950, the MSTS moved 40,000 men, 500,000 measurement tons of cargo, and more than 2,000,000 barrels of petroleum to the far east. The United States Military Air Transport service (MATS), operated under control of the air force but manned jointly with air force and navy personnel, had been established in 1948. In the first week of the Korean war, MATS tripled its airlift tonnage from the United States to Japan. Five weeks later, the Pacific air lift had increased its pre-Korean tonnage by 3,000%.

Pusan provided a major port for United Nations forces. Despite shortages of experienced labourers, limited port and dock facilities and antiquated Korean equipment at Pusan, the United States army's transportation corps organized thousands of Korean stevedores and unloaded 300,000 tons of cargo at this port during July 1950 alone.

Korea's roadnet was limited and primitive. To supply United Nations troops in positions distant from usable roads, the Republic of Korea formed the civilian transportation corps. This organization, later integrated into the Korean services corps, furnished thousands of porters who carried ammunition and supplies to the most forward infantry units. The military helicopter also proved invaluable as a means of supplying combat units in inaccessible locations and evacuating wounded.

Japan, with its excellent ports, rail and roadnet, provided a superb supply base close to the Korean battleground. However, United States army divisions in Japan at the outbreak of the Korean war were organized for occupation and lacked much of the combat equipment they required, especially tanks, motor vehicles and anti-tank weapons. This equipment was rushed from the United States. At the same time the United States army instituted an extensive reclamation and rebuilding program to salvage thousands of motor vehicles and other equipment left on Pacific islands after World War II. The rapid redeployment of troops after World War II had left the United States army with insufficient men to maintain this equipment and most of it had fallen into disrepair.

When the Korean war brought sudden demands for additional shipping, the United States navy took about 2,000 ships of all types out of "moth balls." The United States air force, experiencing similar problems in obtaining additional equipment rapidly, employed some commercial transport planes to help overcome shortages.

The Korean war again emphasized the critical importance of military supply and transport in modern warfare. In the first four months of this war, the United States shipped more men and equipment to Korea than it did to North Africa 11 months after Pearl Harbor. Better supplied and better transported, outnumbered United Nations forces were able to defend the Republic of Korea with conspicuous success.

BIBLIOGRAPHY.—"How Does It Get to You?" *Armed Forces Talk*, no. 412 (June 27, 1952); *Army Information Digest, passim.*

MODERN MILITARY ORGANIZATION FOR SUPPLY

The supply services of a modern army are organized usually on a service, territorial and command basis.

Service Organization.—Specialized military supply functions are assigned to various vertical organizations within an army which have staff officer representatives at nearly all echelons of army command. In the United States army, for instance, the quartermaster corps has the special function of providing food and other supplies to the troops. There are quartermaster officers at levels from the division upward to perform this function. In addition to this staff organization, a service maintains troop units to perform the actual operations with which it is charged. Many of the services also maintain combat type units which handle special equipment of the service in battle.

Organization varies among armies, but the technical services of the United States army are typical. These services determine army requirements for the supplies in categories under their supervision and then procure, store and issue them. They also perform maintenance and other duties. Technical services, along with typical supplies which they handle and typical specialized duties which they perform, follow.

Chemical Corps.—Typical supplies which this service handles are smoke-producing equipment, bulk chemicals (toxic and nontoxic), flame-throwers, gas masks and decontaminating equipment. Typical services performed are decontamination of equipment, training of troops in gas warfare and maintenance of chemical depots.

Engineer Corps.—This service is responsible for such supplies and equipment as bridges, construction materials of all kinds, assault boats and maps. Among the services performed by engineer troops are the construction and maintenance of structures of various types, the clearing and operation of ports, surveying, mapping and reproduction of maps and the furnishing of electrical power.

Medical Service.—Medical supplies include instruments, drugs, plasma and therapeutic apparatus. The medical corps treats the sick and wounded and supervises measures taken by troops to maintain health.

Ordnance Corps.—This service handles most of the weapons, ammunition and motor vehicles used by the U.S. army. Ordnance troops help service this equipment.

Quartermaster Corps.—Food, clothing and gasoline are three of the major items which this service supplies. Among the services it performs are the operation of laundry and dry-cleaning facilities and graves registration.

Signal Corps.—Telephones, wire, radios, public address systems and photographic equipment are typical of the supplies for which the signal corps is responsible. Signal units install and repair communication facilities of various kinds, and take and develop photographs.

Transportation Corps.—This service handles railway and certain types of floating equipment. Its services include transportation of personnel and property of the U.S. army.

Territorial Organization.—A theatre of war includes all the sea, air and land masses which are actually involved, or which may become involved in a war. The actual fighting takes place in only a small portion of the theatre. The remainder is filled with combat troops preparing for or moving to battle and the numerous supply agencies which operate the lines of communication. For operations, these supply agencies are frequently placed under area commands which co-ordinate their efforts. There are such territorial supply commands in two of the major subdivisions of a theatre of war—the zone of the interior and the theatres of operations.

Zone of the Interior.—This area ordinarily includes all the territory of a nation exclusive of the theatres of operations. It may, under some circumstances, also include foreign territory, either allied, neutral or hostile. Within the zone of the interior is a major supply organization or group of organizations which co-ordinate the funnelling of troops and supplies to the theatres of operations.

Theatre of Operations.—This area includes that portion of the theatre of war necessary for tactical operations and for administration incident to these operations. Although there is no rigid organization of a theatre of operations, it normally contains a combat zone and a communications zone.

The combat zone comprises the land, sea and air masses required for combat operations and for the immediate administration of the combat units operating therein. The combat zone is divided normally into army areas which are in turn divided into corps and division areas. In a normal situation, an army within the combat zone will designate an army service area which extends from the corps rear boundaries to the combat zone rear boundary. In the army service area, supply and service units, organic or attached to the army, operate. These units may, however, be located anywhere throughout the combat zone, including territory under the control of subordinate commanders. In a sense, the army service area is the communications zone of the army, providing a logistic link between the army and the theatre communications zone.

The communications zone includes all the territory of the theatre of operations between the rear boundary of the theatre and the rear boundary of the combat zone. It provides space in which administrative units, including supply and transport agencies, function in support of the fighting units in the combat zone. The communications zone is a link between the zone of the interior and the combat zone. It is established primarily to free combat commanders of such responsibilities as supply and administration of territory not directly concerned with combat. For this reason, the communications zone is constantly pushed forward as the fighting troops advance. When practicable, the communications zone operates as a single unit; however, factors such as the territorial extent of the zone, enemy activity or the volume of support operations may necessitate decentralization of the zone into base sections and advance sections. In such a situation, the base section is normally closer to the zone of the interior while the advance section provides the link between the combat zone and the base section. The nature of combat operations and other special conditions determine the type of organization. After the establishment of a beach head on a hostile shore, for example, all service units supporting an army might be pooled into an army service command under direct control of the army commander. As more territory is gained, the army service command might be dissolved and an army service area and an army base established. A final step might occur when the communications zone or an advance section of the communications zone takes over this army base.

Command Organization.—In most military commands, whether large or small, commanders are responsible for the immediate supply and transport of their subordinate units. A theatre commander, for example, is responsible for the distribution of supplies and transportation to the army, navy and air theatre commands. An infantry regimental commander is likewise responsible for providing his battalions and companies with supplies and the vehicles needed to move them.

Each commander has a staff or its equivalent to assist him in discharging this responsibility. In the U.S. army, the assistant chief of staff G-4, or his counterpart in a larger or smaller unit, is responsible for supply planning and the supervision of supply operations. The G-4 also handles transportation for administrative moves of the command. The assistant chief of staff G-3 (plans and operations officer) supervises the transportation in combat operations.

SUPPLY OPERATIONS

On every level of command there must be a close liaison between the unit commander and his supply services in order that adequate supplies will be available at the time and place needed to support tactical operations. Because there is a time lag between requisitions for supplies and their arrival, all commanders must inform supply agencies of planned operations as early as possible.

Supply Requirements.—This is a statement of all the supplies needed by a military organization over a specified period of

time or to complete a particular project. Requirements for an organization may include:

Initial Equipment Requirements.—This is the equipment which a unit, such as one recently mobilized, needs in its initial issue.

Replacement and Consumption Requirements.—A unit needs issues of supplies to replace items worn out, destroyed or consumed.

Reserve Requirements.—Units are authorized to keep reserves of equipment and supplies which they need in case normal resupply channels are interrupted.

Project Requirements.—A unit has requirements for supplies not included in normal allowances when it is given a special operation to perform.

Levels of Supply.—Supplies are not pumped into a theatre at an absolutely uniform rate because of enemy interruptions of the supply lines and the imperfections of even the best transportation system. In order, therefore, to maintain a continuity of operations a theatre is authorized to maintain levels of supply. These are quantities of supplies, usually expressed in days, which it keeps on hand for future needs. In turn, the theatre authorizes its subordinate commands to maintain levels of supply which units procure by requisition on higher headquarters.

Procurement.—A theatre of operations obtains the supplies it needs by shipment from sources such as the zone of the interior lying outside the theatre, or by exploitation of resources within the theatre.

Supply from Zone of the Interior.—When a military force initiates a campaign that establishes a theatre of operations, it takes with it its initial equipment and supplies for replacement and maintenance for a specified period of time. As the campaign progresses, it receives an automatic issue of supplies from the zone of the interior on a prearranged schedule, based on anticipated needs. Later, as the theatre becomes organized, it can calculate more realistically its own supply needs; it then starts to requisition supplies from the zone of the interior based on its actual experiences and the automatic issue eventually ceases.

Supply from Local Resources.—Military forces habitually utilize local resources when possible to avoid the shipment of similar materials from distant military supply bases. When in allied areas, arrangements are made with the friendly government to obtain supplies and services needed. In hostile territory, troops may procure supplies by purchase, requisition, contributions or confiscation. In any case, the theatre commander prescribes the procedure by which subordinate commanders may utilize local resources in accordance with the law, rules of land warfare and regulations of higher authority. Any seizure of property not so authorized is punishable as looting.

Supply through Repair and Recovery.—In battle, much equipment and many supplies are abandoned, lost or damaged either through enemy action or through inadequate maintenance. Energetic efforts to collect the equipment which is strewn over any battlefield and to salvage damaged matériel greatly ease the strain on the supply system which must furnish replacements for lost or destroyed equipment. Every individual in an army is responsible for the recovery of salvageable supplies. When deemed appropriate, combat troops are often used to collect recoverable equipment after fighting is over. Commanders are responsible for the maintenance of supply discipline in their units to keep the loss and wastage of supplies to a minimum.

In divisions and higher units organic service troops salvage equipment as a normal function. When possible, they restore damaged equipment so that it can be used for the purpose originally intended. When this is not possible, they may convert the equipment for some other use. Within the theatre, some service units have the sole mission of retrieving abandoned and damaged supplies and equipment.

Marking and Packaging.—In order that supplies may go through the long supply line to their proper destination without confusion, it is essential that they be properly marked. Markings generally include the destination, type of supply, method of shipment, weight and cubage and the period in which shipment is to be made. For security reasons, some of this information is often in code.

Factors influencing the method used to package supplies include the need to conserve shipping space, protection against the elements and rough handling, and the fragility of the supplies. Often it is necessary to package supplies in such a way that they can be dropped from planes or broken down easily into one-man loads.

Storage.—At various points in a military transportation system, supplies are stored in depots where they are entered on stock records and held for further movement or distribution to the units which use them. Depots may be classified as branch depots (such as quartermaster or engineer services) according to the service which maintains them. Although general depots (which stock supplies used by two or more depots) are sometimes organized, branch depots are the type more often used. Depots may also be classified according to the type of supply they handle.

Within the communications zone is an elaborate depot system for storing supplies for further transportation. Within the combat zone are army depots from which are funnelled supplies to army railheads, truckheads or navigation heads, where distribution to troops is accomplished. A dominant factor in the selection of depot sites is proximity to means of transportation.

Distribution.—Distribution in a military sense includes all the operations incident to the movement of supplies within a command. A distribution system must be flexible to ensure the adequate supply of troops in spite of rapid changes in the tactical situation.

Within a theatre of operations, the first step in the distribution process is usually the movement of supplies from ports to depots within the communications zone. Such movement must be carefully controlled and expedited to avoid port congestion. After the supplies are classified and picked up on stock records in these initial depots, they start through the chain of depots that store them until they are distributed to user agencies.

There are two general methods employed to distribute supplies to user organizations distinguished by the point at which supplies are delivered to them:

Supply Point Distribution.—In this method, delivery of supplies is made by the supply agency of a higher echelon to a supply point established by subordinate units. The units which are to use the supplies employ their own transportation to pick up these supplies.

Unit Distribution.—Under this system, the higher echelon provides the transportation to haul the supplies directly to the bivouac areas of subordinate units or to some similar point of easy access.

In the U.S. army, field armies and divisions ordinarily employ the supply point distribution method, while regiments use the unit distribution method. (B. F. D.)

SUPRARENAL EXTRACT: *see* ADRENALINE AND NORADRENALINE.

SUPRARENAL GLANDS: *see* ADRENAL GLANDS.

SUPREME COURT OF JUDICATURE, in England, a court established by the Judicature act, 1873, consisting of the "court of appeal" and the "high court of justice." *See* COURT; PRACTICE AND PROCEDURE.

SUPREME COURT OF THE UNITED STATES, THE. The supreme court of the United States stands at the apex of a dual system of courts: state and federal. This position accounts for the distinctive and even paradoxical character of the court's function: it is a conventional tribunal whose authority can be exercised only in the decision of ordinary lawsuits, and at the same time, as the final expositor of the United States constitution, it is the voice and symbol of the aspirations of the nation. Within the framework of litigation, the supreme court marks the boundaries of authority between state and nation, state and state and government and citizen.

ORGANIZATION AND POWERS

Establishment and Membership.—The articles of confederation, which held together the American states from 1781 until adoption of the constitution, did not provide for a supreme court.

But the articles did authorize congress to appoint courts of appeal in cases of naval captures, and such a court functioned until 1787, deciding more than 100 appeals from state courts. In a limited sense this tribunal was a precursor of the present supreme court. The framers of the constitution in 1787 concluded that a supreme judicial tribunal was required in a political system that gave limited powers to the centre and that, even before the adoption of the Bill of Rights through amendments, imposed specific prohibitions on the state and national governments. Article iii, section 1, of the constitution provides:

> The judicial Power of the United States shall be vested in one supreme Court, and in such inferior Courts as the Congress may from time to time ordain and establish. The Judges, both of the supreme and inferior Courts, shall hold their Offices during good Behavior, and shall, at stated Times, receive for their Services, a Compensation, which shall not be diminished during their Continuance in Office.

The jurisdiction of the federal courts is defined in succeeding provisions of article iii. The major heads of jurisdiction reflect the special mission of these courts conceived as tribunals relatively detached from local pressures and biases. The authority of the federal courts extends to cases arising under the constitution, laws or treaties of the United States, controversies to which the United States is a party, controversies between states or between citizens of different states, cases of admiralty and maritime jurisdiction and cases affecting ambassadors, other public ministers and consuls. Of these courts, only the supreme court is given by article iii a status independent of the will of congress. Since the constitution is silent on such matters as the size of the court and the scope of its appellate jurisdiction in the judicial hierarchy, however, the supreme court itself is subject in important respects to congressional action.

The first Judiciary act, passed on Sept. 24, 1789, provided for a chief justice and five associate justices. In 1807 the membership was increased by law to seven, in 1837 to nine and in 1863 to ten. In 1866, as part of the effort by congress to curb the appointing power of Pres. Andrew Johnson, a statute reduced the membership to seven as vacancies should occur, and actually the court declined to eight members. But in 1869 an increase to nine was provided, at which the size of the court has remained. Tenure is during good behaviour, subject to expulsion by conviction on impeachment as provided in article ii, section 4, of the constitution. One justice of the court has been impeached—Samuel Chase, who was acquitted in 1805. By an act of 1937, a member of the court who has served for ten years or more may retire with full compensation upon reaching the age of 70.

Appointments to the supreme court, as to the lower federal courts, are made by the president with the advice and consent of the senate. Nine nominations have been rejected by the senate, the latest being that of John J. Parker in 1930. In fact all appointees have been lawyers, but beyond this there has been no pattern of selection. It can be said that geographic distribution has been a factor as well as the political outlook of the nominees: not so much their partisan affiliation as an outlook on problems of statecraft congenial to the chief executive and the senate. It must be added that a number of justices, enjoying the emancipation of the robe, have surprised and disappointed those who were responsible for their appointment.

The 14 members of the court who served as chief justice of the United States, with terms of service, follow:

John Jay (1789–95)	Melville W. Fuller (1888–1910)
John Rutledge (1795)	Edward D. White (1910–21)
Oliver Ellsworth (1796–1800)	William H. Taft (1921–30)
John Marshall (1801–35)	Charles E. Hughes (1930–41)
Roger B. Taney (1836–64)	Harlan F. Stone (1941–46)
Salmon P. Chase (1864–73)	Fred M. Vinson (1946–53)
Morrison R. Waite (1874–88)	Earl Warren (1953–)

The associate justices, with their terms of active service, are as follows:

John Rutledge (1789–91)	Samuel Chase (1796–1811)
William Cushing (1789–1810)	Bushrod Washington (1798–1829)
James Wilson (1789–98)	Alfred Moore (1800–1804)
John Blair (1789–96)	William Johnson (1804–34)
James Iredell (1790–99)	Brockholst Livingston (1806–23)
Thomas Johnson (1792–93)	Thomas Todd (1807–26)
William Paterson (1793–1806)	Gabriel Duval (1811–35)
Joseph Story (1811–45)	Oliver W. Holmes (1902–32)
Smith Thompson (1823–43)	William R. Day (1903–22)
Robert Trimble (1826–28)	William H. Moody (1906–10)
John McLean (1829–61)	Horace H. Lurton (1909–14)
Henry Baldwin (1830–44)	Charles E. Hughes (1910–16)
James M. Wayne (1835–67)	Willis Van Devanter (1910–37)
Philip P. Barbour (1836–41)	Joseph R. Lamar (1910–16)
John Catron (1837–65)	Mahlon Pitney (1912–22)
John McKinley (1837–52)	James C. McReynolds (1914–41)
Peter V. Daniel (1841–60)	Louis D. Brandeis (1916–39)
Samuel Nelson (1845–72)	John H. Clarke (1916–22)
Levi Woodbury (1845–51)	George Sutherland (1922–38)
Robert C. Grier (1846–70)	Pierce Butler (1922–39)
Benjamin R. Curtis (1851–57)	Edward T. Sanford (1923–30)
John A. Campbell (1853–61)	Harlan F. Stone (1925–41)
Nathan Clifford (1858–81)	Owen J. Roberts (1930–45)
Noah H. Swayne (1862–81)	Benjamin N. Cardozo (1932–38)
Samuel F. Miller (1862–90)	Hugo L. Black (1937–)
David Davis (1862–77)	Stanley F. Reed (1938–57)
Stephen J. Field (1863–97)	Felix Frankfurter (1939–62)
William Strong (1870–80)	William O. Douglas (1939–)
Joseph P. Bradley (1870–92)	Frank Murphy (1940–49)
Ward Hunt (1872–82)	James F. Byrnes (1941–42)
John M. Harlan (1877–1911)	Robert H. Jackson (1941–54)
William B. Woods (1880–87)	Wiley B. Rutledge (1943–49)
Stanley Matthews (1881–89)	Harold H. Burton (1945–58)
Horace Gray (1881–1902)	Tom C. Clark (1949–)
Samuel Blatchford (1882–93)	Sherman Minton (1949–56)
Lucius Q. C. Lamar (1888–93)	John M. Harlan (1955–)
David J. Brewer (1889–1910)	William J. Brennan (1956–)
Henry B. Brown (1890–1906)	Charles E. Whittaker (1957–62)
George Shiras, Jr. (1892–1903)	Potter Stewart (1958–)
Howell E. Jackson (1893–95)	Byron R. White (1962–)
Edward D. White (1894–1910)	Arthur J. Goldberg (1962–65)
Rufus W. Peckham (1895–1909)	Abe Fortas (1965–)
Joseph McKenna (1898–1925)	

Only two chief justices, Edward D. White and Harlan F. Stone, were advanced to that office while serving as associate justices. Charles Evans Hughes (chief justice 1930–41) served as associate justice before resigning in 1916 to run for president.

John Rutledge was an associate justice (1789–91) but resigned to become chief justice of the supreme court of South Carolina; he received a recess appointment as chief justice of the U.S. supreme court in 1795, but his nomination for an unlimited term was rejected by the senate. The longest tenure, 34 years, was that of Stephen J. Field.

Basic Functions.—In maintaining the constitutional order, the supreme court from an early date has exercised the power of declaring acts of congress or of the state legislatures to be unconstitutional. Neither branch of this power of judicial review is expressly conferred by the constitution, and neither was established without intense partisan and sectional resentment.

Judicial review of acts of congress, assumed to be proper in a case in 1796, was established as a function of the court in 1803 in the celebrated case of *Marbury* v. *Madison*. The opinion of Chief Justice Marshall, adopting in substance the argument of Alexander Hamilton in no. 78 of the *Federalist*, deduced the power as a logically inevitable element of the judicial process: if in the decision of a case it appears that a statute and the constitution are in conflict, the court, which is bound to decide according to law, must prefer the higher to the inferior law. This analysis, assuming as it does that the court rather than congress or the electorate has the controlling judgment on such issues of conflict, is not free of the question-begging fallacy, as is suggested by the practice in the Swiss federation, where the constitutional validity of acts of the federal legislature is not examinable by the supreme court. But this is by no means to countenance the charge of usurpation that has sometimes been leveled at Marshall. There were precedents for judicial review of legislation in the practice of the privy council in relation to the colonies and of U.S. state courts after the Revolution. Moreover, there were references in the Constitutional Convention, though oblique, to the function of judicial review in litigation as a preferable alternative to a council of revision exercising a kind of veto power over legislation deemed invalid.

The doctrine of judicial review was made more palatable to the anti-Federalist critics of Marshall by the actual consequence of the doctrine in the Marbury case itself: an act of congress which

the supreme court read as conferring original jurisdiction on the court beyond the warrant of article iii was declared invalid, the case was therefore dismissed for lack of jurisdiction, and a direct clash with the defendant, Pres. Thomas Jefferson's secretary of state, was averted. The irony of this act of abstention tended to blunt in popular consciousness the immediate effect of the court's assertion of power.

Judicial review of legislation of the states brought more severe immediate repercussions. Indeed the very power of the court to review decisions of state courts was vigorously challenged when it was laid down in an opinion of Justice Joseph Story in 1816, in the case of *Martin* v. *Hunter's Lessee*. For a number of years thereafter efforts were made in congress to repeal the section of the Judiciary act that conferred on the supreme court this appellate jurisdiction over state tribunals in cases turning on constitutional issues. No state or section had a monopoly on this spirit of resistance. Originally centred in the southern states, it was echoed on occasion in the North, as when in the period before the Civil War the state courts were held to the enforcement of the federal Fugitive-Slave law. A further ironic turn, in a subject studded with ironies, is that after the Civil War the power of judicial review served as an instrument of reconciliation, for the court declared unconstitutional, as bills of attainder, federal and state laws imposing test oaths intended to disqualify Confederate sympathizers from holding office or practising the professions.

Judicial review of legislation is not the only method whereby the court serves to maintain the constitutional order. Executive, administrative and judicial action also are subject to review in the supreme court. Moreover, some of the frictions and confusions of authority incident to a federal system are resolved in the court through the adjudication of controversies between states themselves, as for example those over boundaries or the allocation of water rights in interstate rivers. In this arbitral capacity the court serves as a substitute for diplomacy or war. Still another unifying function of the court is the interpretation of federal statutes, which would otherwise be subject to divergent applications in the lower courts.

For a time the supreme court essayed a further unifying role, that of developing a body of federal common (nonstatutory) law in cases arising in the diversity-of-citizenship jurisdiction of the federal district courts, but this enterprise introduced some anomalies of its own and it came to an end in 1938 with a decision holding that in such cases the common law to be applied is that of the state in which the federal district court is situated.

Jurisdiction and Procedure.—Since the functions of the court are performed within the framework of ordinary litigation, it is necessary to outline the procedures by which cases reach the court and are there decided. Relatively few cases are brought in the original jurisdiction of the court. These are controversies between states, between a state and the United States or between a state and citizens of another state, and cases affecting ambassadors, other public ministers and consuls. The great bulk of the court's business comes to it in its appellate jurisdiction. Depending on the nature of the decision in the state or lower federal court, the route to the supreme court is by appeal or certiorari. The difference between the two is that an appeal invokes the obligatory jurisdiction of the supreme court, certiorari its discretionary jurisdiction.

The development of this bifurcated jurisdiction reflects a response by congress to a long struggle by the court to cope with the volume of cases annually docketed. In 1891 a measure of relief was afforded by the Circuit Court of Appeals act, setting up intermediate courts (of which there are 11, in ten circuits and the District of Columbia) having final authority over appeals from federal district courts save in cases of exceptional public importance. The same act put a belated end to the circuit-riding duties of supreme court justices. The Judiciary act of 1925 carried the reforms further. Sponsored by the court itself, under the aegis of Chief Justice Taft, the act greatly limited the obligatory jurisdiction and gave the court a large measure of control of its business by subsuming most classes of cases under the head of review by certiorari; these can be disposed of on the moving papers, without full briefs or oral argument, where they do not qualify by their importance for an exercise of the court's discretionary power of review. Each member of the court votes on each petition for certiorari, as in other actions of the court. The affirmative vote of four justices suffices for the granting of certiorari, since at that stage the question is simply the appropriateness of the case for review. At mid-20th century the number of cases filed annually in the court was around 1,400. Of these about five-sixths are disposed of summarily, mainly by denial of petitions for certiorari.

The cases not thus disposed of are set down for oral argument after the filing of briefs. The time allowed for oral argument is limited, ordinarily to one hour for each side. At a conference of the justices following a week of oral arguments the cases heard during that week are decided, and the responsibility of writing the opinion of the court in each is assigned among the justices by the chief justice (or by the senior associate justice if the chief justice is in the minority in a case). The practice of writing dissenting opinions has been followed since the time of the Marshall court; the first resolute dissenter was Marshall's colleague William Johnson. In this tradition the supreme court stands midway between the British house of lords sitting as a court, where each participating member delivers an opinion, and the judicial committee of the privy council, where only a single opinion is given.

ROLE IN AMERICAN LIFE

Changing Contributions and Conflicts.—Any assessment of the unifying forces in United States society must ascribe an important role to the supreme court. Judicial review of state legislation takes its place in this regard with such physical factors as transportation and communication, and such political factors as national parties and presidential leadership. The chief technical instrument employed by the court in this aspect of its work has been the commerce clause. In terms conferring power on congress to regulate commerce among the several states, the clause has been applied from the time of Marshall to impose of its own force a negative on state laws of taxation or regulation that discriminate against or unduly burden interstate commerce. The common national market has thereby been maintained against the grosser demands of local commercial protectionism. The process has been one of accommodation, requiring in each case an appraisal of the state law in the light of local need compared with the inroad on the common market. Marshall's court tended to stress the national imperative, Taney's the concerns of local welfare, but at no time has it been possible for the court to avoid making close judgments on a pragmatic basis.

The assertion of national interest that spelled the doom of many state laws has served ultimately to vindicate national power when congress has legislated. Beginning with the Interstate Commerce act of 1887 and the Sherman Anti-Trust act of 1890, congress increasingly took over the regulation of sectors of the economy that previously were regulated ineffectively or not at all because of judicial restraints and practical inhibitions on the states. A series of decisions, of which the child labour case in 1918 was the most conspicuous, threatened for a time to create a no man's land in which neither the federal nor state governments would have constitutional power to legislate effectively—the federal government denied power because the objective of the law was employment practices, the state governments denied power because the interstate importation of articles of commerce is immune from state prohibition. Not until the overruling of the child labour case in 1941 was the federal power over commerce placed on a firm basis. Subsequently the court sustained far-reaching exertions of the commerce power by congress, of which perhaps the most striking was the setting of maximum acreage quotas for the production of wheat, including wheat consumed on the farm.

As the commerce clause has been the chief doctrinal source of national power over the economy, the due process and equal protection clauses have been the principal sources of protection of

persons and corporations against arbitrary or repressive acts of government. The guarantee of life, liberty and property against deprivation without due process of law, enforceable against the states by virtue of the post-Civil War 14th amendment, has been interpreted by the court to protect against the substance of legislation as well as against arbitrary procedures. During the first third of the 20th century the court invalidated on this score important social and economic measures, including laws of the states setting minimum wages and maximum prices. Liberty of contract was given a constitutional content that derived from a laissez-faire philosophy on the part of a dominant majority within the court. Against this paralyzing judicial treatment of experimental legislation Justices Oliver Wendell Holmes and Louis D. Brandeis voiced a current of powerful dissents. Beginning about 1940 the doctrine of these dissents became the law of the land, and thereafter no significant economic legislation was found repugnant to the constitutional guarantees.

The field of civil liberties has had a markedly different history. Not until the 1920s was it established that the rights safeguarded against federal incursion by the first amendment—the rights of freedom of speech, press, assembly and religious exercise—are safeguarded also against action by the states, as part of the liberty guaranteed by the 14th amendment. As the court has become more tolerant of economic controls, it has taken a more stringent view of laws that curtail the first amendment liberties and the procedural guarantees of the Bill of Rights. Insofar as these changing standards of judicial review have been rationalized, the warrant is that the special concern of the court is with process and structure; that political liberties are at the base of representative self-government; and that if normally the people must look to political processes for the correction of harsh or unwise legislation, the court for that very reason must be vigilant to strike down measures that repress the political process itself.

The guarantee of equal protection of the laws in the 14th amendment has had a development similar to the evolution of due process of law. Originally designed for the benefit of the emancipated Negroes, soon utilized as a shield against social legislation, it came at mid-20th century to serve its historic purpose as a barrier to racially discriminatory laws.

In a steady series of cases the court invalidated a variety of restrictions directed at Negroes: exclusion from voting in party primary elections, racial zoning, enforcement of restrictive covenants in real-estate deeds, exclusion from state-maintained universities, and enforced segregation in public schools and on public conveyances.

It is hardly surprising that a tribunal exercising authority of such gravity, delicacy and finality has frequently found itself the object of popular or partisan attack. Given the generality of the great constitutional provisions and the task of applying that document to an evolving social, economic and political order, the court cannot escape the responsibility of rendering judgments that provoke dissatisfaction in some quarters and for some time. Whether a judgment is viable will depend on whether it is the court or its critics who have more truly sensed the climate of the age rather than the weather of the day.

A number of the strongest presidents have been pitted at one time or another against the court. Thomas Jefferson and Andrew Jackson found certain of Marshall's opinions decidedly distasteful. Abraham Lincoln, faced with the Dred Scott decision, called for its re-examination, though he was careful to counsel against resistance to the decree itself. Theodore Roosevelt, irked by judicial invalidation of social legislation, urged in 1912 popular recall of judges and popular referenda on judicial decisions. Franklin D. Roosevelt, confronted with a supreme court that had upset a large part of his recovery program, generally by divided decisions, proposed a reorganization of the federal courts. His plan, presented to congress in 1937, provided for the appointment of an additional judge for each judge who remained on the bench after having served ten years and reached the age of 70; the maximum size of the supreme court was to be fixed at 15. The plan was defeated in congress. Meanwhile, however, the supreme court had given clear signs in its intervening decisions that it would take a more hospitable view of the scope of governmental power under the constitution.

Self-Limitation.—Cognizant of the risks of ill-advised decisions, the court has developed canons of self-limitation. It will not render advisory opinions, or decide a constitutional question when the case can be disposed of on another ground, or render a constitutional decision broader than required by the facts of the case before it or at the instance of a litigant whose interests are not actually in jeopardy.

A category of "political" questions has been evolved, moreover, which will be left for resolution by other departments of government without judicial review. Some of the most unfortunate decisions of the court have been marked by a disregard of the court's own canons.

Writing in 1928, Charles Evans Hughes characterized three ill-starred decisions as self-inflicted wounds: the Dred Scott case, the first legal tender cases and the income tax cases. (*See* Hughes, *The Supreme Court of the United States*, p. 50, 1928.) Each of these was in time uprooted: the first by the Civil War, the second by an overruling decision, the third by the 16th amendment to the constitution. Each was a needlessly precipitate decision: the first because it could have been rested on a nonconstitutional ground of state law; the second because one member of the court, holding a decisive vote, was on the verge of retirement for disability; the third because the interests of the complainants, corporate stockholders, did not require the relief that the court granted. It is noteworthy that judges who have been deeply sensitive to the demands on constitutional law made by a changing social order—Justice Brandeis most conspicuously—have also been most insistent on the court's observance of its professed canons of self-limitation.

Important Decisions.—The Marshall court laid the basis for a federal union capable of effective government. In *McCulloch* v. *Maryland* (1819), upholding the second Bank of the United States, the court gave a liberal interpretation to the necessary and proper clause of the constitution and thus equipped congress with ancillary powers of great potentiality. A counterpart was the decision in *Gibbons* v. *Ogden* (1824), holding that a state could not confer a monopoly of interstate steamboat traffic, thus establishing the court as a guardian of the free national market against local commercial self-interest. The Taney court, although on the whole more sympathetic to the concerns of the states, by no means reversed the current. *Bank of Augusta* v. *Earle* (1839) established the capacity of a corporation chartered in one state to do business in another; and the *Charles River Bridge Co.* v. *Warren Bridge Co.* case (1837), while sustaining the power of a state to charter a company in competition with one operating under a prior grant from the state, served to stimulate the development of new forms of interstate transportation. *The Genessee Chief* (1851) extended federal admiralty jurisdiction to inland navigable lakes and rivers, rejecting the English rule that such jurisdiction was confined to tidewaters. Pointing out the divergence between the English and American topography, Taney's opinion is at once a declaration of independence and an avowal that constitutional interpretation should conform to the facts of life.

Unhappily, Taney's reputation was derived too largely from the *Scott* v. *Sanford* (Dred Scott) case (1857), whose holding that congress lacked power to prohibit slavery in the territories foreclosed one avenue of compromise and helped to precipitate the Civil War. The war itself produced the *Ex parte Milligan* case (1866), a notable landmark in the field of civil liberties, holding that trial of civilians by military commission is unlawful outside the theatre of military operations while the civil courts are capable of functioning.

The post-Civil War amendments to the constitution were construed in the *Slaughterhouse* cases (1873) and the *Civil Rights* cases (1883) not to have worked a radical transformation in the basic structure of U.S. federalism. The first of these decisions ruled that the privileges and immunities of citizens of the United States, safeguarded by the 14th amendment, must be narrowly interpreted to include only those privileges peculiar to national citizenship and not such asserted rights as the opportunity to

engage in common callings. The second decision established the position that the amendments provide a federal guarantee against racial discrimination by state and not private action.

The due process clause of the 14th amendment was used in the late 19th and early 20th centuries to strike down economic legislation which did not commend itself to the social philosophy of a majority of the justices. Perhaps the high-water mark was reached in *Lochner* v. *New York* (1905), invalidating a law which limited the workday of bakers to ten hours: in time the case became less notable for the decision than for the stinging dissent of Justice Holmes. The dissenting position was definitely established as the law of the land in *Nebbia* v. *New York* (1934), a decision giving broad scope to economic legislation affecting prices, wages and hours of labour by renouncing the view that there is a closed category of businesses so affected with a public interest that for them alone is such regulation constitutionally permissible.

The New Deal program of Pres. Franklin D. Roosevelt at first foundered on restrictive constitutional decisions of the court. But the tide turned with the *National Labor Relations Board* v. *Jones and Laughlin Steel Corp.* case (1937), upholding federal authority over collective bargaining in industries in which work stoppages would affect interstate commerce. The opinion of Chief Justice Hughes was able to draw support from an earlier opinion of the same justice in the Shreveport case (1914), where federal authority over intrastate railroad rates was sustained on a showing of their effect upon competitive interstate traffic. Congressional power over industry and agriculture has been validated to an extent ample for legislative purposes. Restraints in the interest of the states have become a matter of congressional discretion rather than constitutional compulsion.

The decision in *Brown* v. *Board of Education* (1954), holding that separate but equal facilities for Negroes in public schools do not meet the constitutional standard of equal protection of the laws, created the greatest problem of compliance faced by the court since the Marshall period. But the court as an arbitral institution meanwhile won firm acceptance if not always equal understanding.

Symbolic and Pragmatic Role.—It has been argued with some force that the functions vested in the supreme court are tolerable only in a country that is agreed on its fundamental premises; that judicial review on the United States model would be impossible where there are deep philosophic cleavages; and that the U.S. practice is a corollary of the pragmatic American spirit. It can be said, however, that the relation is one of interaction. The practice of judicial review, by a reflexive effect, serves to soften the edges of ideological weapons and to promote accommodations on a working basis. The court operates in the context of specific facts; it does not so much choose between absolutes as endeavour to reach a mediating solution; and it holds itself open for reconsideration of doctrine in the light of changed circumstances and fuller experience.

By virtue of the reasoned elaboration of its opinions and the tradition of dissent it serves continually to clarify, refine and test the large philosophic ideals that are written into the constitution, and so to translate them into working principles for a federal union under a rule of law. Beyond its specific contributions, this symbolic and pragmatic function may be regarded as the most significant role of the supreme court in the life of the nation.

For the role of the United States supreme court in the development of the U.S. federal system and the history of judicial constitutional construction, *see* AMERICAN LAW and CONSTITUTION AND CONSTITUTIONAL LAW. *See* also UNITED STATES (OF AMERICA): *History;* COURT; and APPEAL.

Important decisions of the court are covered in detail in separate articles on various legal subjects of special interest, for example CIVIL LIBERTIES; CENSORSHIP; INTERSTATE COMMERCE; LABOUR LAW; and PUBLIC UTILITIES: *Regulation in the United States.*

See articles on famous justices, such as BRANDEIS, LOUIS DEMBITZ; CARDOZO, BENJAMIN NATHAN; HOLMES, OLIVER WENDELL; MARSHALL, JOHN; STONE, HARLAN FISKE; and STORY, JOSEPH. *See* also Index references under "Supreme Court of the United States, The" in the Index volume.

BIBLIOGRAPHY.—Charles Warren, *The Supreme Court in United States History,* 2 vol. (1926); Carl B. Swisher, *American Constitutional Development* (1943); Paul A. Freund, *On Understanding the Supreme Court* (1949); Robert H. Jackson, *The Struggle for Judicial Supremacy* (1941) and *Supreme Court in the American System of Government* (1955); Felix Frankfurter, *Of Law and Men* (1956); William O. Douglas, *We the Judges* (1956); Robert L. Stern and Eugene Gressman, *Supreme Court Practice,* 2nd ed. (1954).

(P. A. F.)

SURABAJA, the chief port of Java and the principal naval base of Indonesia, is situated on the northeastern coast of the island, 420 mi. (677 km.) E.S.E. of Djakarta. It lies at the western end of the Madura strait just above the Kali Mas river, one of the chief streams of the Brantas delta. The climate is humid. Pop. (1961) 666,002.

The second city in Java, Surabaja has modern houses and offices, hotels, cinemas and shops. In contrast, there are the street markets and the crowded downtown districts. The river Kali Mas is canalized, and there, as in Djakarta, the canal flows through the centre of the city. At the mouth of the river in a small park is a look-out station and from there railway and road run inland, in a southerly direction, past the old Fort Prins Hendrik (1835) to the old town, then to the upper town, and then beyond that to the suburbs of Gubeng and Wonokromo where a new residential and well-planned sector of the city has developed. In the suburb of Wonokromo, too, is Indonesia's biggest zoo.

Notable among the city's various churches is a large mosque (1868). Erlangga university has replaced the faculties of law and economics of the University of Indonesia which were formerly in Surabaja. There are also colleges for medicine and dentistry, and a naval college.

The area around the city is flat and, protected from extremes of climate by the island of Madura, is a rich agricultural region. Sugar cane is chiefly grown, but other crops include tobacco, maize and rice. Tapioca, coffee and cocoa are also produced locally. Cattle are bred and there is a large fishing fleet. The naval station is on the south side of the canal, while opposite are the commercial docks; these were destroyed by the Dutch during the Japanese invasion.

As well as being a major trading port, Surabaja is also an important industrial centre with several heavy industries. There are large railway locomotive workshops, and shipbuilding and ship-repair yards. There are textile mills and glass and cigarette factories. In the industrial suburb of Wonokromo there is an oil refinery. The bulk of Java's chief product, sugar, is shipped from Surabaja. Other goods exported include coffee, tobacco, rubber, spices and petroleum products.

Surabaja is linked with the railway system which connects the eastern and western coasts of Java and is thus in communication with Djakarta, Semarang, Jogjakarta and the other chief towns of the island. There is steamer communication with the chief ports of the world and with most of the ports of the Malay archipelago; Surabaja also lies on the main sea route from Singapore and Djakarta to Australia. From the airport at Tandjungperak, 5 mi. from the city, there are flights to various airports in the island including Djakarta and thence to the principal cities in Europe and Asia.

Surabaja was one of the chief centres of resistance to Dutch and British forces during the Indonesian war of independence which broke out in 1945 and terminated in 1949 with the transference of sovereignty to the republic of the united states of Indonesia.

SURABAJA, formerly a residency in Java, Netherlands Indies; area, 1,658 sq.mi.; bounded on the north by the residency of Bodjonegoro, on the west by that of Kediri, on the south by that of Malang, and on the east by the straits of Madura. Surabaja was one of the flattest residencies in Java, well watered by the Solo and the Brantas, and possessing a climate admirably adapted for the cultivation of sugar, the chief product, while tobacco, cassava (tapioca), maize, coffee, cocoa were grown, and the usual native crops—rice, pulses, fruit and vegetables. Cattle were bred extensively, forests in the western Rembang portion

yielding quantities of teak (there is a central lumber yard at Tjepu), and in this part too were extensive oil fields with a large refining installation of the Royal Dutch-Shell company at Tjepu. The population (1930) was 2,176,423, including 30,708 Europeans and 64,175 Chinese and other foreign Asians.

The capital was Surabaja (*q.v.*), pop. (1961) 1,007,945. Other towns were Grissee, 25 mi. N. of Surabaja (pop. [1958 est.] 35,434), a port of the old Dutch East India company and one of the first places of Dutch settlement in Java, where there was a trade in edible birds' nests; Modjokerto (pop. [1961] 51,732), a sugar industry centre, and the site of the Majapahit empire, with a museum of relics; Djombang, pop. (1958 est.) 36,855, a sugar centre; and Sidoardjo.

The main railway line from Batavia (Djakarta) crossed the residency, there were lines from Semarang to Surabaja city and thence to Pasuruan along the coast, and there was also excellent sea communication from the port of Surabaja. In 1618 the Dutch allied themselves with the Adipati of Surabaja, who favoured the Dutch against his overlord, the Sultan Ageng, and this helped to establish Dutch power in Surahaja. The residency was occupied by Japan in March 1942; it ceased to exist with the independence of Java in 1949. Surabaja regency has an area of 591 sq.mi., and a population (1961) of 666,002.

SURAJ-UD-DOWLAH (SIRAJ-UD-DAULA) (d. 1757), ruler of Bengal. The date of his birth is uncertain, but is generally placed between 1729 and 1736. His name was Mirza Mohammed, and he succeeded his grandfather Aliverdi Khan as nawab of Bengal on April 9, 1756. He was a cruel and profligate fanatic. Being offended with the English for giving protection to a native official who had escaped with treasure from Dacca, he attacked and took Calcutta on June 20, 1756. He then permitted the massacre known in history as "The Black Hole of Calcutta." (*See* CALCUTTA: *History;* and INDIA-PAKISTAN, SUBCONTINENT OF: *History.*) Calcutta was retaken by Clive and Admiral Watson on Jan. 2, 1757, and on June 23, Suraj-ud-Dowlah, routed at Plassey, fled to Rajmahal, where he was captured. He was put to death on July 4, 1757, at Murshidabad, by order of Miran, son of Mir Jafar, who had conspired against Suraj-ud-Dowlah.

SURAKARTA, a residency of the former Netherlands Indies in central Java, Indonesia; area 2,332 sq.mi., pop. (1961) 3,957,885. It is extremely hilly, except along the valley of the Solo, the principal river, which, with its tributaries, the Kaduan, etc., proved of great value to the area for agricultural purposes, watering the rich volcanic soil brought down from the mountains which flank either side of the Solo valley from north to south, and making Surakarta one of the most fertile parts of Java.

In the west the mountains are dominated by the Merbabu and Merapi groups; in the east by the Lawu group. The government of Surakarta was bounded on the north by Semarang, on the west by Kediri and Jokjakarta, on the east by Madiun, and on the south, a very small section of coast line, less than ten miles long, by the Indian ocean. Surakarta became a major tobacco-growing centre in Java; the production was concentrated mainly on estates, producing leaf tobacco.

Sugar also became an important export commodity. The sugar cane is grown extensively on the volcanic ground at the foot of the Merapi mountain. Coffee is a less important product. The government also developed the usual native cultures—rice, pulses, cassava, fruits and vegetables.

The susuhunan, or sultan of Surakarta, resided in the capital, Surakarta, pop. (1961) 367,626, also known as Solo where, too, the headquarters of the Dutch resident was established. Except in the oldest part of the town (where there is an old Dutch settlement, and Ft. Vastenburg, as well as the native quarter), there are fine, wide streets, well planned and with magnificent shade trees at their side, with many modern shops and residential buildings.

Surakarta has also a zoological garden, some very interesting houses of native nobles, and a large market.

The city is on the banks of the Solo, and also on the main railway line from Batavia (Djakarta) to Surabaja, which was built across the residency in the centre, from west to east, and served as the starting point of a line of railway to Semarang, via Gundih. Tram lines run from Surakarta south to Baturetno and west to Boyolali, the starting point for the ascent to the sanatorium of Selo, 4,500 ft., on Mt. Merapi.

The Dutch agreement of 1755 recognized two native rulers in Java, the sultan of Jogjakarta and the susuhunan of Surakarta. From about 1830 onward the native princes of Java ceased to be of any political importance.

The city and district were occupied by Japan during World War II.

See T. S. Raffles, *History of Java* (1817); M. L. van Deventer, *Daendals-Raffles* (1894). (E. E. L.; X.)

SURAT, a city and district of Gujarat state, India. There the English first established a factory on the mainland. In 1514 the Portuguese traveler Duarte Barbosa described it as an important seaport, and during the reigns of Akbar, Jahangir and Shah Jahan it rose to be the chief commercial city of India. At the end of the 16th century the Portuguese were undisputed masters of the Surat coast. But in 1612 Captain Thomas Best, and after him Captain Nicholas Downton, destroyed the Portuguese naval supremacy and made Surat the seat of a presidency under the English East India company. In 1664 Sir George Oxenden bravely defended the factory against Sivaji, but its prosperity received a fatal blow when Bombay was ceded to the company (1668) and shortly afterward made the capital of the company's possessions and the chief seat of their trade. At one time Surat's population was estimated at 800,000, but by the middle of the 19th century it had fallen to 80,000; by 1961 it had risen to 288,026. The city is on the left bank of the Tapti river, 14 mi. from its mouth, and on the Western railway, Bombay to Delhi, 157 mi. N. of Bombay. A moat and city wall divide the city proper, with its narrow streets and handsome houses, and the suburbs. The city is a centre of trade and there are cotton mills, factories for ginning and pressing cotton, rice-cleaning mills and paper, ice and soap works.

Fine cotton and silk goods are woven, and there are special manufactures of silk brocade, gold and silver wire, carpets, sandalwood and inlaid work. The ancient art of making superfine muslin, dormant since the early 19th century, was being revived in 1956. The city is the seat of three colleges of Gujarat university.

SURAT DISTRICT has an area of 4,856 sq.mi.; pop. (1961) 2,451,624. On June 10, 1948, it absorbed the former princely states of Dharampur, Bansda and Sachin, Navsari district of the former Baroda state, and other territory. It has a coast line of 80 mi., consisting of a barren stretch; behind this is a rich, highly cultivated plain, nearly 60 mi. in breadth at the mouth of the Tapti, but narrowing to only 15 mi. in the southern part, and on the northeast are the wild hills and jungle of the Dangs. Principal crops are cotton, millets, rice and pulses. The chief centres of trade are Bulsar ([1961] 37,586) and Surat. The well-built town of Navsari ([1961] 51,300), Ptolemy's *Nasaripa,* is an old Parsee settlement. Besides the main Western railway, a crossline runs eastward up the Tapti.

(The Surat Agency, a responsibility of the Bombay governor, consisted of the three states of Dharampur, Bansda and Sachin, together with the petty states of the Dangs and from 1933 was administered by the Baroda resident until absorption in Surat district in 1948.)

SURBITON, a municipal borough (1936) and parliamentary constituency of Surrey, Eng., which under the London Government Act 1963 was to be amalgamated on April 1, 1965, with Kingston-upon-Thames and Malden and Coombe to form the new London borough of Kingston upon Thames (*see* LONDON). Surbiton municipal borough is 12 mi. S.W. of Waterloo, London, with a frontage on the right bank of the Thames opposite Hampton Court park. Pop. (1961) 62,940. Area 7.4 sq.mi. It is largely residential, with some light engineering works on the Kingston bypass, and it includes Hook, Chessington (with its zoo) and Tolworth. Hillcroft college for women provides one-year residential courses in social studies and the arts and is open to foreign students speaking English. Subertone, as it was known in 1203, denotes South Barton as opposed to neighbouring Norbiton.

SURETY, in law, the party liable under a contract of guarantee (*q.v.*). In criminal practice sureties bound by recognizance are means of obtaining compliance with the order of a court of justice, to keep peace or otherwise.

See AGENCY.

SURFACE. There is no agreed definition of a surface. We may think of it as the boundary of a solid body, or as the division between two portions of space, or as the locus of an ∞^2 set of points, selected by some law from the ∞^3 points of space, and having a certain amount of continuity; in each case a plane is the simplest example; but none of these gives a precise or satisfactory definition. A very simple model illustrates the kind of difficulty that arises. If a surface partitions space, it must have two sides; but a long strip of paper, with the ends joined after one has received a half-twist, forms a surface with only one side.

At an ordinary point O of a surface, there exists a *tangent plane,* the unique limit of the plane OPQ, when the adjacent points $P,Q \to O$ by any paths lying on the surface, the angle POQ remaining finite. Any point where this limit does not exist is *singular.* The *normal* is the perpendicular to the tangent plane at its point of contact O.

A surface can also be regarded as the locus of a point satisfying a geometrical condition expressed by a single equation between its rectangular cartesian co-ordinates, say $f(x, y, z) = 0$. From this point of view, the definition of a surface depends on what types of function are admissible for f, and the question of continuity arises in another form. The surface is *analytic* if f is an analytic function, and in particular is *algebraic* if f is a polynomial. We shall use f as the name of the surface whose equation is $f = 0$.

We may also consider a surface as the locus of a curve varying according to some definite law. Thus a sphere is the locus of a meridian circle which rotates about the polar axis, or of a circle of latitude whose centre advances along the axis, while the radius first increases from zero and then decreases. Any surface can be thus regarded, for it is always the locus of its own sections by a fixed pencil of planes.

In particular, a *ruled surface* is the locus of a variable line, whose different positions are the *generators.* The tangent plane at any point on a generator contains the whole generator. Two alternative situations can arise: the plane may touch the surface all along the generator, or it may vary in a pencil (rotating about the generator) as the point of contact moves along the generator. If every generator is of the former kind, the surface is a *developable.* Such a surface has only ∞^1 tangent planes, one for each generator, whereas the general ruled (or nonruled) surface has ∞^2 tangent planes in accordance with the principle of duality. The simplest developables are cones and cylinders. Like them, any developable can be flattened out or "developed" into one plane, without tearing or stretching. Every developable that is neither a cone nor a cylinder is generated by the tangent lines of a certain twisted curve, the *edge of regression* of the surface. The tangent planes of the developable are the osculating planes of the curve. When the surface is developed into a plane, part of the plane is covered twice and part is uncovered, the edge of regression furnishing the boundary. The edge of regression may alternatively be described as the locus of the point on each generator that is closest to the "adjacent" generator. The corresponding locus on the general ruled surface is the *curve of striction* of the surface.

A surface f has ∞^3 tangent lines, each the limit of the join of two adjacent points. Except in the case of a developable, such a line is also the limiting intersection of adjacent tangent planes. To touch f imposes one condition on the ∞^4 lines of space, and f has in general ∞^2 *bitangents,* each touching it at two distinct points, and ∞^2 *inflexional tangents,* each meeting f in three adjacent points.

Order, Class, Rank.—The *order* of an algebraic surface is the degree of its equation; *i.e.*, of the polynomial f. It is the order of a plane section, and is also the number of intersections, real, coincident or imaginary, of f with any line. These are particular cases of more general properties of the order. Two surfaces of orders n, n_1 intersect in a curve of order nn_1, or in an aggregate of curves of this equivalent total order, when curves of contact and common multiple curves are counted a proper number of times. Three surfaces of orders n, n_1, n_2, have $n\,n_1\,n_2$ distinct points of intersection, or the equivalent of this when points of contact and common multiple points are properly counted.

The *class* of f is the degree of its equation regarded as an envelope, connecting the tangential co-ordinates of any one of its tangent planes. The properties of the class are dual to those of the order; it is the class of the tangent cone drawn to f from a general point, and the number of tangent planes through a general line.

The *rank* is an intermediate numerical characteristic, the number of the tangent lines that can be drawn through a given point and in a given plane through that point; it is the class of the plane section, and the order of the tangent cone.

If the polynomial f falls into rational factors $f \equiv f_1.f_2$, of degrees n_1, n_2 where $n_1 + n_2 = n$, then each factor by itself represents a distinct algebraic surface, and if a point lies on either, its co-ordinates satisfy the equation of f; regarded still as a single surface, f is *degenerate,* breaking up into the components f_1, f_2.

Polars.—The point P_λ dividing the segment PP_0 in the ratio $\lambda{:}1$ has four homogeneous co-ordinates of the form $x + \lambda x_0$. The condition that P_λ lies on f is that λ satisfies $F = 0$, where

$$F \equiv \lambda^n f_0 + \lambda^{n-1}\Delta f_0 + \cdots + \frac{\lambda^r}{(n-r)!}\Delta^{n-r} f_0 + \cdots + \frac{1}{n!}\Delta^n f_0$$

$$\equiv \frac{\lambda^n}{n!}\Delta_0{}^n f + \cdots + \frac{\lambda^r}{r!}\Delta_0{}^r f + \cdots + f,$$

and $f_0 \equiv f(x_0, y_0, z_0, w_0)$, $\Delta \equiv x\dfrac{\partial}{\partial x_0} + \text{etc.}$, $\Delta_0 \equiv x_0\dfrac{\partial}{\partial x} + \text{etc.}$

The surface $\Delta_0{}^r f$, for $r = 1, \ldots, n-1$, is the rth *polar surface* of P_0 with regard to f; the first polar passes through all the multiple points of f. In particular, $\Delta_0{}^{n-1} f$ or Δf_0 is its *polar plane,* and $\Delta^2 f_0$ its *polar quadric.* The locus of points whose polar quadrics are cones is the *Hessian* of f. Its equation of degree $4(n-2)$ expresses that the determinant of second derived functions of f vanishes.

If two roots of $F = 0$ are infinite, then P_0 lies on f, and P on the tangent plane at P_0; the two conditions are $f_0 = 0$, $\Delta f_0 = 0$. The latter represents the tangent plane at P_0, which is the polar plane of its point of contact. If P_0 is a multiple point, the tangent plane does not exist, and all the first derived functions of f vanish there.

If two roots of $F = 0$ are equal and finite, its discriminant vanishes. This equation of degree $n(n-1)^2$ in x, y, z, w is satisfied if PP_0 touches f at P; in general it represents the tangent cone from P_0 to f. Hence the class of f cannot exceed $n(n-1)^2$. It is less than this if f has multiple curves; for the condition is satisfied if P_λ is a multiple point of f, and the cone vertex P_0 standing on a multiple curve separates itself from the tangent cone.

If two roots of $F = 0$ are 0, then PP_0 touches f at P, and $f = 0$, $\Delta_0 f = 0$. The curve of contact of the tangent cone drawn from any point P_0 to f is the whole or part of the intersection of f with the first polar of P_0, any residual intersection consisting of the multiple curves of f. This curve of proper contact passes through all the isolated multiple points of f.

Indicatrix; Curvature.—In order to examine a surface in the neighbourhood of a particular point, we take this as origin O of rectangular cartesian co-ordinates, and arrange f in homogeneous functions of x, y, z, say, $f = u_0 + u_1 + \ldots + u_n$. If f consists of a single homogeneous function u_n it represents a cone vertex O of degree n. If f is general, the condition that O lies on the surface is $u_0 = 0$.

If $P\,(x, y, z)$ is a general point of space near O, at a distance small of first order, then the perpendicular distances of P from all general planes through O, viz.; x, y, z and all general linear functions of them, are small of first order. If P lies on f, the linear function u_1, being equal to $-(u_2 + \ldots + u_n)$, is small of second order; $u_1 = 0$ is the equation of the tangent plane, which in the neighbourhood of O lies infinitely closer to f than any other plane. Whenever we can safely neglect small quantities of second and higher order, the surface can be replaced by its tangent plane.

If we neglect terms of third and higher orders only, f can be

replaced by the quadric u_1+u_2, or more generally by any of the family $q\equiv u_1+u_2+u_1v_1$, where v_1 is a general linear function. These are the *osculating quadrics* of f at O, any one of which is a second approximation to the surface. The sections of f and q by any plane through O have the same curvature at O.

Let O be an ordinary point of f, with z as tangent plane, and therefore $x=y=0$ as normal. Then $f\equiv z+u_2+\ldots+u_n$, and the simplest osculating quadric is $q\equiv z+v_2$, where v_2 is the result of putting $z=0$ in u_2.

A plane $z=a$ small constant ζ, parallel and close to the tangent plane, cuts f approximately in the conic $v_2=-\zeta$, with centre on the normal, called the *indicatrix*.

If a plane through the normal meets the indicatrix in a diameter of length $2r$, the radius of curvature of the normal section is ρ, where $\frac{1}{\rho}=\frac{2\zeta}{r^2}$. An oblique plane meeting the tangent plane in the same line, and making a finite angle ϕ with the normal, cuts f in a section of radius of curvature $\rho\cos\phi$, the projection of ρ upon the plane (Meunier's theorem). This vanishes for the tangent plane, when the section is singular.

The major and minor axes of the indicatrix lie on the *principal* normal planes, making sections of maximum and minimum radii of curvature, say ρ_1, ρ_2. For a normal section inclined at an angle θ to the first principal section, $\frac{1}{\rho}=\frac{\cos^2\theta}{\rho_1}+\frac{\sin^2\theta}{\rho_2}$ (Euler's theorem).

According to the nature of the indicatrix, O is an *elliptic, hyperbolic* or *parabolic point* of f. If it is elliptic, all the normal curvatures are of the same sign, and the surface bends away from the tangent plane on one side only. The part near O of the section by $z=\zeta$ is a small real closed curve for ζ of one sign, and imaginary for ζ of the opposite sign. The section by the tangent plane $z=0$ has no real part in the neighbourhood except O itself, which is a double point with imaginary tangents. The surface is *synclastic*.

At a hyperbolic point, the surface is *anticlastic*. The principal curvatures are of opposite signs, and f bends away from the tangent plane on opposite sides in these directions. Two normal sections have inflexions, the curvatures vanishing; their planes pass through the inflexional tangents at O, which are parallel to the asymptotes of the indicatrix, and touch the section of f by the tangent plane, which has two real branches through O along which f crosses the plane. All sections near O have real branches there. If there is a line lying on f, it is one of the inflexional tangents at each of its points. *Conjugate* tangent lines are those parallel to conjugate diameters of the indicatrix.

At a parabolic point, the inflexional tangents coincide; the section by the tangent plane has a cusp. These points lie on the Hessian.

The normal at O does not in general meet the normal at an adjacent point, unless this lies on a principal section, in which case the normals intersect at one of the two *centres of curvature* of f at O. These coincide if the indicatrix is a circle; then O is an *umbilic*, all normal sections have the same curvature and all adjacent normals meet the normal at O.

Quadrics.—A surface of order 2 is a *quadric q;* in general its class and rank are each 2 also. The simplest example is the sphere (*q.v.*). Every plane section is a conic. Parallel sections are similar and similarly situated conics, whose centres lie on a line, a *diameter* of q. The diameters concur at the *centre* (which may lie at infinity), which bisects every chord of q through it. There are three mutually perpendicular planes of symmetry, the *principal planes,* intersecting by pairs in the *principal axes* which meet q in its *vertices*.

Through each point P of q there pass two generators, real or imaginary, lying wholly on q; their plane is the tangent plane to q at P, cutting q in this pair of lines. The generators fall into two systems, any one generator meeting all those of the other system and none of its own. A quadric is determined by any three generators of one system, and may be defined as the locus of a line meeting three fixed skew lines. There are six directions (only two real) of planes of circular section; the points of contact of the tangent planes parallel to these are the *umbilics*.

A quadric has three *focal conics* (two real), to each point S of which there corresponds a *directrix l,* such that the distance of any point of q from S is proportional to its distance from l measured parallel to one of the planes of circular section; S and l are focus and directrix of the section of q by the plane Sl, which is the normal plane at S to the focal conic.

A focus can also be defined as a point from which the enveloping cone is right-circular.

A quadric is fixed when its three principal sections are given; *i.e.,* it is determined by its trihedron of principal planes and the lengths of its principal axes. It depends on nine parameters. If the centre is finite, there are three main types of quadric, according to the types of the principal sections: *ellipsoid* (three ellipses), *hyperboloid of one sheet* (one ellipse, two hyperbolas), *hyperboloid of two sheets* (two hyperbolas, one section imaginary). If the centre is at infinity, one principal section is wholly at infinity; the other two are parabolas, meeting in the one finite vertex, and q is an *elliptic* or *hyperbolic paraboloid* according as these parabolas lie on the same or opposite sides of the tangent plane at the vertex. If the principal axes vanish, q is a quadric cone; if also the centre is at infinity, it is a cylinder. Finally, a degenerate quadric is a pair of intersecting, parallel or coincident planes.

The hyperboloid of one sheet and hyperbolic paraboloid have real generators, and models can be constructed by threads or wires; they are everywhere anticlastic. The ellipsoid, hyperboloid of two sheets and elliptic paraboloid have imaginary generators, and are everywhere synclastic. Referred to its principal axes, an ellipsoid of semi-axes a, b, c has the equation $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$. An elliptic paraboloid referred to its vertex as origin can have the equation $\frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{2z}{c}$; and the other types of quadric can be represented by equally simple equations.

Cubic Surfaces.—A surface f of order 3 is of class 12, or lower if it is singular. It possesses 27 lines lying on it, of which each meets 10 others, arranged in 5 intersecting pairs. There are 45 triangles lying on f, whose planes are tritangent, touching it at the vertices. We can also select, in 36 ways, a *double-six* from the 27 lines, consisting of two sets of six lines, each line meeting five of the other set and none of its own. Any plane through a line l cuts f in a residual conic meeting l in two points, at which the plane touches f. For two positions of the plane through l, the conic touches l, at a parabolic point of f, lying on the Hessian.

There can be as many as four double points on a cubic surface. If these lie at the vertices of the tetrahedron of reference, the equation can take the form $\frac{1}{x}+\frac{1}{y}+\frac{1}{z}+\frac{1}{w}=0$. Each of the six edges of the tetrahedron absorbs four of the 27 lines, and there is one triangle of distinct lines, each meeting two opposite edges of the tetrahedron. A cubic surface can have one double line; it is then ruled, each plane through the double line meeting it in a residual generator. Two of these pass through each point of the double line, and their plane meets f in a fixed line of the surface, the simple directrix.

Singularities.—A *singular point* of a locus is one at which there is no definite, single tangent plane. If the origin O is singular, and as above $f\equiv u_0+u_1+\ldots+u_n$, the four conditions are those involved by $u_0=0$, $u_1\equiv 0$. The simplest case is the *conic node*, an ordinary double point represented by $f\equiv u_2+u_3+\ldots+u_n$, where u_2 is a general quadratic function of x, y, z, giving the proper tangent cone. Any line through O meets f in two points coinciding at O, a generator of u_2 meets it in three points, and the six *lines of closest contact* given by $u_2=u_3=0$ meet it in four. The section of f by a general plane through O has a double point with two quite distinct branches, the tangents being generators of u_2; that by a plane through a line of closest contact has an inflexion on one branch. The section by a plane touching the tangent cone

has a cusp, the two branches merging into one cycle.

If a general line through O meets f in s points there, in the simplest case O is an ordinary s-fold point with a proper tangent cone u_s, where $f \equiv u_s + u_{s+1} + \ldots + u_n$; this involves $\frac{1}{6}s(s+1)(s+2)$ conditions. Any plane through O, or any surface having O as an ordinary point, meets f in a curve having an s-fold point there. If O is s_1-fold on the second surface, the two tangent cones having no common sheet, the curve of intersection has ss_1 branches through O. If O is also s_2-fold on a third surface, the three have ss_1s_2 points of intersection absorbed at O.

If the tangent cone breaks up, the simplest case is the *binode*, a double point whose quadric tangent cone consists of two *biplanes*, intersecting in the *edge*. The section by a general plane through O has a double point as for a conic node, that by a plane through the edge has a cusp, and that by one of the biplanes has a triple point. Taking the biplanes as planes of reference, we have $u_2 = xy$; then the equation says that either x or y is small of second order, or else both are small of orders between 1 and 2. A general point of f near O is close to a definite one of the two biplanes, and, in this sense, we can speak of two sheets of f, one touched by each biplane; they are not separate sheets, but become indistinguishable near the edge $x = y = 0$. There are three lines of closest contact in each biplane; these are in general different from the edge, which meets f in three points only.

The conic node and binode are *ordinary* singularities; *i.e.*, the tangent cone, proper or degenerate, has no repeated sheet. At an *extraordinary* singularity, the tangent cone u_s not only breaks up, but has a repeated sheet, u_r say, so that $u_s \equiv u_r^2 v_{s-2r}$. The simplest case is the *unode*, a double point with a repeated tangent plane, $f = u_1^2 + u_3 + \ldots + u_n$. The surface has two connected half-sheets, the general section having a cusp. There are three lines of closest contact; the section by a plane through one of these has a tacnode, that by the tangent plane has a triple point.

A general point O of a double curve is a binode of a special kind. The two sheets of f are quite distinct; every simple point of f near O belongs definitely to one sheet or the other, the points of the double curve belonging to both. There is no continuous passage at all from one sheet to the other near O; but on the double curve there lie a certain number of *pinch-points*, unodes of a special kind (which may be absorbed in higher singularities), at which the two tangent planes coincide, and the two sheets become connected.

A *tacnode* of the surface is a double point where the general plane section has a tacnode, which is the plane singularity equivalent to two adjacent double points. Thus O has an infinite set of double points of f adjacent to it, one in every direction in the tangent plane, which may be considered as an infinitesimal double curve adjacent to O. In a similar way, a tacnodal curve is equivalent to two adjacent ordinary double curves.

In general, a singular point, whether isolated or lying on a singular curve, is characterized, first, by its multiplicity; *i.e.*, the number of intersections of f with a general line through the point that are absorbed there; next, by the nature of the tangent cone and its sheets, and of the singularity of a general section; and further by the nature and arrangement of the higher singularities of special sections. By a series of suitable transformations, any singular point or curve of a surface can be analysed into an equivalent set of component singular points and curves of standard types, and any surface can be transformed into one having only elementary singularities, and in particular into one having no isolated singular points, but only ordinary multiple curves and such points of higher singularity as necessarily lie on them; or alternatively into a surface having only a double curve and a number of triple points which are also triple for the double curve. The complete theory is complicated. Isolated singular points may give rise on transformation to singular curves; and a singular curve may give rise, from the transformation of its special points of higher singularity, to other singular curves, whose general points may be of more complicated nature (though not of higher multiplicity) than the general points of the original curve.

Singular tangent planes are defined dually to the above. Any surface of order $\geqq 3$ must have singularities of one or other kind, or of both. A surface f of order n with no singular points is of class $n' = n(n-1)^2$; a surface of class n' with no singular planes is of order $n = n'(n'-1)^2$. These two equations are incompatible unless $n = n' = 2$, in which case f is a quadric.

If f has no point singularities, it has in general two series of double and stationary tangent planes, and its reciprocal has a double and a cuspidal curve. The characteristic numbers of these modify the equations relating order and class, and make them compatible. If f has higher singularities, there exists a complicated set of relations between the numerical characteristics; the simplest are the Plücker equations for the general plane section and tangent cone.

Curves on a Surface.—The interest of f depends partly on the simpler curves lying on it. There are always its plane sections, and its curves of total intersection with other surfaces; lower curves may arise as partial intersections, components of a total intersection that breaks up through acquiring additional singularities, the cutting surface either touching f or passing through one or more of its singularities. All curves on a plane are its total intersections with cones standing on the curves. On a quadric, the conic section by a tangent plane acquires a double point at the point of contact and breaks up into a pair of generators. A cubic surface contains 27 pencils of conics, cut out by planes through the 27 lines, and many other systems of curves of orders 3, 4, and 5, cut out by quadrics containing various sets of the lines.

Other important curves on f are defined by infinitesimal properties. An *asymptotic curve* is one whose tangent at any point is one of the inflexional tangents, parallel to an asymptote of the indicatrix. A *curve of curvature* is one whose tangent at any point is a principal tangent, parallel to an axis of the indicatrix. Two curves of each of these kinds can be drawn through any general point of f. A *geodesic* is a curve whose small arcs are each the shortest distance on the surface between their extremities. On a plane, these are lines; on a developable they are curves which become lines when the surface is flattened out. An infinite number of geodesics can be drawn through a general point of f, one starting in any direction.

Families of Surfaces.—If an equation contains a parameter λ it represents a singly infinite family of surfaces; the most important case is the linear family or *pencil*, when the coefficients are linear functions of λ and the equation has the form $f + \lambda\phi = 0$. Just one member of the family passes through each general point of space. The total intersection of any two of the family is the same, being that of f, ϕ. More generally, if f, ϕ are any two algebraic surfaces, the equation of any other containing their total intersection can be expressed as $\alpha f + \beta\phi = 0$, where α, β are polynomials (Noether's theorem). In the same way, $f + \lambda\phi + \mu\psi$ gives an ∞^2 linear family or *net* of surfaces; one and only one passes through any two given points; any three surfaces of the net have the same *base*, or set of common points and curves, as f, ϕ, ψ.

If f is of order n, it has $N \equiv \frac{1}{6}(n+1)(n+2)(n+3) - 1$ independent ratios of coefficients; all such surfaces form a linear family with N degrees of freedom, and one can be made to pass through N points in general position, each of which presents one independent condition to f. Thus a quadric can be drawn through nine points, and these determine it if they are general. However, seven points determine a net of quadrics, whose base consists of eight points, viz., the seven and another, associated with and determined by the seven, which offers no independent condition; but the passage of the quadric through it is a necessary consequence of its passage through the others. Thus nine points do not determine a quadric if any eight of them are an associated set of this nature.

To have a given point as a double point imposes four conditions on f; to contain a given line, $n+1$ conditions. The number of independent conditions presented to f by an assigned element of any sort is the *postulation* of that element. The conditions for a combination of elements may force f to contain other fixed elements, or to break up. If six out of nine points lie on a plane, the only quadric containing them all consists of this plane and the plane of the other three points.

Two surfaces are *orthogonal* if at every common point the angle between them, *i.e.*, between their normals, is a right-angle. If three surfaces are mutually orthogonal, their curves of intersection are curves of curvature on each surface. If three linear families are such that the three surfaces through any point are at right angles to each other, they form an *orthogonal system; f* can form part of such a system only if it satisfies a certain differential equation of third order. In the same way, a cubic family, with three members through any point, can be an orthogonal system, *e.g.*, the confocal quadrics

$$\frac{x^2}{a^2+\lambda}+\frac{y^2}{b^2+\lambda}+\frac{z^2}{c^2+\lambda}=1,$$

where the members through a point consist of an ellipsoid, a hyperboloid of one sheet, and a hyperboloid of two sheets.

Among the infinite number of surfaces having the same boundary, *i.e.*, a given closed curve through which each is to pass, the *minimal surface* is that whose area is least. This is the form assumed by a soap film bounded by a wire in the form of the given contour. The same property must attach to any small portion of the surface, and leads to a differential equation of second order, the boundary conditions determining the arbitrary elements of the solution. A minimal surface is anticlastic everywhere. If it were synclastic at a point P, a plane near the tangent plane at P on one side would cut it in a small plane closed curve, approximately the elliptic indicatrix, whose area would be less than that of the cap on which P lies; and the total area could be reduced without affecting the given boundary.

Parametric Representation; Genus.—The co-ordinates of any point on f can be represented as functions of two parameters. We may consider x, y as the parameters, and z given as an implicit function of these by the equation $f=0$. More often it is convenient to express each co-ordinate as an explicit function of two other parameters, say $x=\phi(u,v)$, $y=\psi(u,v)$, $z=\chi(u,v)$.

In general, ϕ, ψ, χ are many-valued functions; if there are r sets of values of x, y, z for each set of u,v, and we take u,v as co-ordinates of a point of a plane, then f is represented on this plane by an $(r, 1)$ point transformation. The nature of the representation depends largely on the form of the expression for the element ds of length of arc traced on the surface, given by

$$ds^2 = dx^2+dy^2+dz^2 = Edu^2+2Fdu\,dv+Gdv^2, \text{ say,}$$

where E, F, G are certain functions of u, v. Here $E^{\frac{1}{2}}du$, $G^{\frac{1}{2}}dv$ are the elements of length along the curves represented by $v=$ const., $u=$const., at the point whose parameters are u,v; while F depends also on the angle between these curves. If they are everywhere at right-angles, *e.g.*, if they are the curves of curvature, then $F=0$.

Rational surfaces are those which can be represented on a plane by a $(1, 1)$ point transformation. Then the co-ordinates of any point on f are rational functions of two parameters, the coordinates of the corresponding point of the plane. Of the different plane representations of f, the lowest is that for which the curves in the plane corresponding to the plane sections of f are of lowest order. Any two of these have n variable intersections, corresponding to the points of f on the common line of the two planes of section; the remaining intersections are fixed, at points of the plane fundamental for the representation. An important family of curves on f are those corresponding to the lines of the plane. Any two of these have one and only one variable intersection, the remainder falling at points of f fundamental for the representation. All quadrics and cubics are rational. Surfaces of higher degree are not, unless they have a sufficient number of singular points and curves.

There are several numerical characteristics of a surface that vanish if it is rational, and may be regarded as answering to the genus of a curve. The simplest is the *genus* as defined by Noether (*Flächengeschlecht*), the number of linearly independent *adjoint* surfaces of degree $n-4$, passing $s-1$ times through every ordinary s-fold curve of f, and $s-2$ times through every ordinary isolated s-fold point, with other conditions at extraordinary singularities. It is also the number of linearly independent double integrals of the form $\iint F(x,y,z)dxdy$, where F is a rational function and x,y,z are connected by $f=0$. If f has no singularities, its genus is $\frac{1}{6}(n-1)(n-2)(n-3)$.

BIBLIOGRAPHY.—The best-known textbooks are: G. Salmon, *A Treatise on the Analytic Geometry of Three Dimensions*, vol. i, 6th ed. (1914), vol. ii, 5th ed. (1915); J. G. Darboux, *Leçons sur la Théorie Générale des Surfaces* (1887–96). Full references are given in the *Encyklopädie der mathematischen Wissenschaften*, Band iii.

(H. P. Hu.; H. S. M. C.)

SURFACE HARDENING, a heat treatment of steel in which only the surface is hardened. Other surface-hardening treatments include facing, by fusing a hard metal onto a soft one, and shot peening (*see* BLAST CLEANING AND SHOT PEENING). The conventional quenching of hot steel in water or oil is literally surface hardening; however, it usually hardens not only the surface but some or all of the interior, too, depending on the size of the piece. The phrase "surface hardening" usually implies processes other than merely quenching such as carburizing, nitriding and induction hardening. The combination of a hard surface and soft interior is of inestimable value in modern engineering practices. Great strength and toughness in the core and extreme surface hardness result in a composite structure capable of withstanding certain kinds of stress to a high degree such as in gears or antifriction bearings. For less exacting requirements, low or moderate core properties and a high surface hardness can be obtained with cheaply fabricated, low-priced steels.

Carburizing.—One of the most important surface-hardening processes is carburizing in which steel is held at a high temperature (1,600° F. is common) for a long time (six hours, for instance) in a carbonaceous environment, usually a gas mixture of carbon monoxide and carbon dioxide, with or without hydrocarbons such as methane. This treatment causes carbon to diffuse into the steel to a depth of several thousandths or hundredths of an inch and leaves the carbon content of the surface much higher than that of the core; *e.g.*, 1.5% carbon compared with about .25%. When the carburized steel is quenched it becomes harder and more wear-resistant than similar steel of lower surface carbon content similarly quenched.

Carbonitriding and Cyaniding.—Carbonitriding, a variant of carburizing, is the same except that nitrogen as well as carbon diffuses into the steel. The nitrogen comes from ammonia introduced into the furnace atmosphere along with the carbonaceous gas. The cyaniding process gives a similar result; both carbon and nitrogen are supplied to the steel by diffusion from a liquid cyanide salt (usually potassium cyanide, KCN) in which the steel is immersed at about 1,600° F. Cyanides are violent poisons requiring rigid safety precautions when using the cyanide process of surface hardening.

Nitriding.—Another diffusion process producing hard surfaces is nitriding (*q.v.*) in which, by using ammonia without a carbonaceous gas, nitrogen goes into the steel and combines with other elements, principally chromium, already contained in the steel to form nitrides that are extremely hard and wear-resistant without the necessity for quenching the steel. Elimination of the quenching process is an advantage because dimensional changes that occur during quenching are avoided.

Induction Hardening.—From about 1935, another process of surface hardening has become important—induction hardening. In this process, steel is briefly heated by electrical induction so that only the surface portion has time to get hot. It is immediately cooled or quenched and the portion that was heated above the critical temperature (about 1,500° F.) becomes hard while the rest of the steel which was not heated by the brief flow of induced current remains unaffected. Induction hardening has become common in mass-production industries since World War II because it is a rapid process—heating time per piece may be five seconds or less—and because it often permits the use of a less costly steel than other hardening processes.

Flame Hardening.—Flame hardening, like induction hardening, involves no change in the chemical composition of the steel surface. In flame hardening, a high-temperature flame impinges directly on the work for a short time, heating the surface but not the interior of the piece. The depth of hardening is usually $\frac{1}{32}$

in. or greater. Bulky pieces are often hardened by this method. For instance, large gears 80 in. in diameter and weighing 2,500 lb. have been flame hardened to a depth of $\frac{3}{16}$ in. in production at a cost of about $15 each for the surface-hardening treatment. Small gears ($\frac{3}{4}$ in. face, 43 teeth) have been hardened at the rate of 150 per hour for eight cents apiece.

Choice of Methods.—The choice of a surface-hardening method depends on the design of the part, composition of the steel, type and depth of case needed, permissible distortion, production volume, initial investment and operating cost. Usually a given steel and treatment are compared with another steel-treatment combination.

The selection of case depth is also important because a deeper case usually costs appreciably more and may even be undesirable, especially on thin pieces, which become more brittle as the case depth is increased. (T. Ly.)

SURFACE TENSION. The title under which many surface phenomena—including Capillarity—are usually considered.

When the surface of water (or other liquid) in a tank is carefully examined it is found not to be perfectly level. The statement that water finds its own level is only approximate. The surface in fact becomes considerably curved near its edges where the liquid comes in contact with the wall of the tank.

If a vertical tube with a very narrow bore is placed with one end in the water, the liquid rises some distance above the level of the outer surface. The rise is greater the smaller the diameter of the bore. The rise of oil through a wick is a phenomenon of the same kind.

If a small quantity of mercury is poured on a horizontal plate it forms a drop. If the quantity is only a few cubic millimetres the drop is nearly spherical—a much larger quantity spreads out into a cake-shaped mass with a nearly flat top and with rounded edges.

A drop of oil placed on a clean water surface spreads almost instantaneously so as to form a very thin film. The thinness of the film can be judged from the colours that flash out or from the size of the area covered by even a small drop.

The phenomena of soap bubbles and the formation of froth in solutions of soap are also familiar ones. In these cases, also, thin films are formed.

The propagation of ripples of very short wave-length follows a different law from the propagation of deep sea waves.

When camphor is scraped so that the fragments fall upon a clean surface of water, they rush about on the surface with very rapid and rapidly changing motion. A minute quantity of oil placed on the surface brings them almost instantaneously to rest.

These and many allied phenomena, can all be studied under one heading. They arise from the existence of surfaces separating one medium from another.

The special forces which come into play are thence known as Surface forces. The rise in tubes of narrow bore—which are known as capillary tubes because the bore is as "fine as a hair" —is due to these forces. From the special way in which the effects arise in this case the action is called *capillary action*. This name, though often applied in other cases, is hardly applicable except to the case of narrow tubes; while the term *surface action* is applicable in all. The forces which are concerned in these phenomena are those which act between neighbouring parts of substances. These also produce the effects of cohesion.

Newton in the third edition of his *Opticks* refers to them in the following passage:—"The parts of all homogeneal hard Bodies, which fully touch one another, stick together very strongly . . . I . . . infer from their Cohesion that their Particles attract one another by some force which in immediate Contact is exceeding strong, at small distances performs the chymical Operations above mention'd, and reaches not far from the particles with any sensible Effect. . . . There are therefore Agents in Nature able to make the Particles of Bodies stick together by very strong Attractions. And it is the Business of experimental Philosophy to find them out."

These forces must be distinguished from those of gravity which also act between the particles of matter because these latter can act over very great distances. (The earth is controlled in its motion round the sun by gravitation.) Cohesion acts to any measurable extent only over minute distances. Sticks of chalk for drawing are made by highly compressing finely ground chalk and other materials. If a stick is broken and the two parts are brought into their original position as nearly as possible by hand they do not stick together because the neighbouring particles are still so far apart as to exert no sensible attraction. High pressure is needed for them to begin sensibly to attract each other. Before considering in detail how these attractions arise we give a historical summary of the development of the subject.

Historical.—(The following historical summary is taken from James Clerk Maxwell's classical article in the ninth edition of this Encyclopaedia, as modified by the 3rd Lord Rayleigh in the tenth edition.) According to J. C. Poggendorff (*Pogg. Ann.* ci. p. 551), Leonardo da Vinci must be considered as the discoverer of capillary phenomena, but the first accurate observations of the capillary action of tubes and glass plates were made by Francis Hawksbee (*Physico-Mechanical Experiments*, London, 1709, pp. 139–169; and *Phil. Trans.* 1711 and 1712), who ascribed the action to an attraction between the glass and the liquid. He observed that the effect was the same in thick tubes as in thin, and concluded that only those particles of the glass which are very near the surface have any influence on the phenomenon. Dr. James Jurin (*Phil. Trans.*, 1718, p. 739, and 1719, p. 1083) showed that the height at which the liquid is suspended depends on the section of the tube at the position of the meniscus, and is independent of the form of the lower part. Sir Isaac Newton devoted the 31st query in the last edition of his *Opticks* to molecular forces, and instanced several examples of the cohesion of liquids, such as the suspension of mercury in a barometer tube at more than double the height at which it usually stands. This arises from its adhesion to the tube, and the upper part of the mercury sustains a considerable tension, or negative pressure, without the separation of its parts. He considered the capillary phenomena to be of the same kind, but his explanation is not sufficiently explicit with respect to the nature and the limits of the action of the attractive force.

It is to be observed that, while these early speculators ascribe the phenomena to attraction, they do not distinctly assert that this attraction is sensible only at insensible distances, and that for all distances which we can directly measure the force is altogether insensible. The idea of such forces, however, had been distinctly formed by Newton, who gave the first example of the calculation of their effect in his theorem on the alteration of the path of a light-corpuscle when it enters or leaves a dense body.

Alexis Claude Clairault (*Théorie de la figure de la terre*, Paris, 1808, pp. 105, 128) appears to have been the first to show the necessity of taking account of the attraction between the parts of the fluid itself in order to explain the phenomena. He did not, however, recognize the fact that the distance at which the attraction is sensible is not only small but altogether insensible. J. A. von Segner (*Comment. Soc. Reg. Götting.* i. [1751] p. 301) introduced the very important idea of the surface-tension of liquids, which he ascribed to attractive forces, the sphere of whose action is so small "ut nullo adhuc sensu percipi potuerit."

In 1756 J. G. Leidenfrost (*De aquae communis nonnullis qualitatibus tractatus*, Duisburg) showed that a soap-bubble tends to contract, so that if the tube with which it was blown is left open the bubble will diminish in size and will expel through the tube the air which it contains.

In 1787 Gaspard Monge (*Mémoires de l'Acad. des Sciences*, 1787, p. 506) asserted that "by supposing the adherence of the particles of a fluid to have a sensible effect only at the surface itself and in the direction of the surface it would be easy to determine the curvature of the surfaces of fluids in the neighbourhood of the solid boundaries which contain them; that these surfaces would be *lintéariae* of which the tension, constant in all directions, would be everywhere equal to the adherence of two particles, and the phenomena of capillary tubes would then present nothing which could not be determined by analysis." He applied this principle of surface tension to the explanation of the apparent

attractions and repulsions between bodies floating on a liquid.

In 1802 John Leslie (*Phil. Mag.*, 1802, vol. xiv. p. 193) gave the first correct explanation of the rise of a liquid in a tube by considering the effect of the attraction of the solid on the very thin stratum of the liquid in contact with it. He did not, like the earlier speculators, suppose this attraction to act in an upward direction so as to support the fluid directly. He showed that the attraction is everywhere normal to the surface of the solid. The direct effect of the attraction is to increase the pressure of the stratum of the fluid in contact with the solid, so as to make it greater than the pressure within the fluid.

In 1804 Thomas Young (Essay on the "Cohesion of Fluids," *Phil. Trans.*, 1805, p. 65) founded the theory of capillary phenomena on the principle of surface tension. He also observed the constancy of the angle of contact of a liquid surface with a solid, and showed how from these two principles to deduce the phenomena of capillary action. His essay contains the solution of a great number of cases, including most of those afterwards solved by Laplace, but his methods of demonstration, though always correct, and often extremely elegant, are sometimes rendered obscure by his scrupulous avoidance of mathematical symbols. Having applied the secondary principle of surface tension to the various particular cases of capillary action, Young proceeded to deduce this surface tension from ulterior principles. He supposed the particles to act on one another with two different kinds of forces, one of which, the attractive forces of cohesion, extends to particles at a greater distance than those to which the repulsive force is confined. He further supposed that the attractive force is constant throughout the minute distance to which it extends, but that the repulsive force increases rapidly as the distance diminishes. He thus showed that at a curved part of the surface, a superficial particle would be urged towards the centre of curvature of the surface, and he gave reasons for concluding that this force is proportional to the sum of the curvatures of the surface in two normal planes at right angles to each other.

The subject was next taken up by Pierre Simon Laplace (*Mécanique céleste*, supplement to the tenth book, pub. in 1806). His results are in many respects identical with those of Young, but his methods of arriving at them are very different, being conducted entirely by mathematical calculations. For those who wish to study the molecular constitution of bodies it is necessary to study the effect of forces which are sensible only at insensible distances; and Laplace has furnished us with an example of the method of this study which has never been surpassed. He found for the pressure at a point in the interior of the fluid an expression of the form

$$p = K + \tfrac{1}{2}H(1/R + 1/R'),$$

where K is a constant pressure, probably very large, which, however, does not influence capillary phenomena, and therefore cannot be determined from observation of such phenomena; H is another constant on which all capillary phenomena depend; and R and R′ are the radii of curvature of any two normal sections of the surface at right angles to each other.

The next great step in the treatment of the subject was made by K. F. Gauss (*Principia generalia Theoriae Figurae Fluidorum in statu Aequilibrii*, Göttingen, 1830, or *Werke*, v. 29, Göttingen, 1867). The principle which he adopted is that of virtual velocities, a principle which under his hands was gradually transforming itself into what is now known as the principle of the conservation of energy. Instead of calculating the direction and magnitude of the resultant force on each particle arising from the action of neighbouring particles, he formed a single expression which is the aggregate of all the potentials arising from the mutual action between pairs of particles. This expression has been called the force-function. With its sign reversed it is now called the potential energy of the system. It consists of three parts, the first depending on the action of gravity, the second on the mutual action between the particles of the fluid, and the third on the action between the particles of the fluid and the particles of a solid or fluid in contact with it.

The condition of equilibrium is that this expression (which we may for the sake of distinctness call the potential energy) shall be a minimum. This condition when worked out gives not only the equation of the free surface in the form already established by Laplace, but the conditions of the angle of contact of this surface, with the surface of a solid.

In 1831 Siméon Denis Poisson published his *Nouvelle Théorie de l'action capillaire*. He maintained that there is a rapid variation of density near the surface of a liquid, and he gave very strong reasons, which have been only strengthened by subsequent discoveries, for believing that this is the case.

The result, however, of Poisson's investigation is practically equivalent to that already obtained by Laplace. In both theories the equation of the liquid surface is the same, involving a constant H, which can be determined only by experiment. The only difference is in the manner in which this quantity H depends on the law of the molecular forces and the law of density near the surface of the fluid, and as these laws are unknown to us we cannot obtain any test to discriminate between the two theories.

We have now described the principal forms of the theory of capillary action during its earlier development. In more recent times the method of Gauss has been modified so as to take account of the variation of density near the surface, and its language has been translated in terms of the modern doctrine of the conservation of energy. *See* Enrico Betti, *Teoria della Capillarità: Nuovo Cimento* (1867); a memoir by M. Stahl, "Ueber einige Punckte in der Theorie der Capillarerscheinungen," *Pogg. Ann.* cxxxix. p. 239 (1870); and J. D. Van der Waals' *Over de Continuiteit van den Gasen Vloeistoftoestand*. A good account of the subject from a mathematical point of view will be found in James Challis's "Report on the Theory of Capillary Attraction," *Brit. Assn. Report*, iv. p. 235 (1834).

J. A. F. Plateau (*Statique expérimentale et théorique des liquides*, 1873), who made elaborate study of the phenomena of surface tension, adopted the following method of getting rid of the effects of gravity. He formed a mixture of alcohol and water of the same density as olive oil, and then introduced a quantity of oil into the mixture. It assumes the form of a sphere under the action of surface tension alone. He then, by means of rings of iron-wire, discs and other contrivances, altered the form of certain parts of the surface of the oil. The free portions of the surface then assume new forms depending on the equilibrium of surface tension. In this way he produced a great many of the forms of equilibrium of a liquid under the action of surface tension alone, and compared them with the results of mathematical investigation. The debt which science owes to Plateau is not diminished by the fact that, while investigating these beautiful phenomena, he never himself saw them, having lost his sight in about 1840.

G. L. van der Mensbrugghe (*Mém. de l'Acad. Roy. de Belgique*, xxxvii., 1873) devised a great number of beautiful illustrations of the phenomena of surface tension, and showed their connection with the experiments of Charles Tomlinson on the figures formed by oils dropped on the clean surface of water.

Athanase Dupré in his 5th, 6th and 7th Memoirs on the Mechanical Theory of Heat (*Ann. de Chimie et de Physique*, 1866–68) applied the principles of thermodynamics to capillary phenomena, and the experiments of his son Paul were exceedingly ingenious and well devised, tracing the influence of surface tension in a great number of very different circumstances, and deducing from independent methods the numerical value of the surface tension. The experimental evidence which Dupré obtained bearing on the molecular structure of liquids must be very valuable, even if our present opinions on this subject should turn out to require modification.

GENERAL THEORY OF CAPILLARY ACTION

It is found by experiment that the forces between molecules to which cohesion is due only act perceptibly across very short distances. If we regard the forces between two particles as acting according to a law depending only upon their positions it must vary inversely as the distance according to a higher power than

that of the inverse square which is obeyed by gravitation. The experiments of Quincke and others show that the extreme range through which sensible effect is produced by them is certainly much less than a thousandth of a centimetre and it is probably less than one millionth.

In order to illustrate the important bearing of this limitation in range consider a molecule, A, well within a substance. A sphere

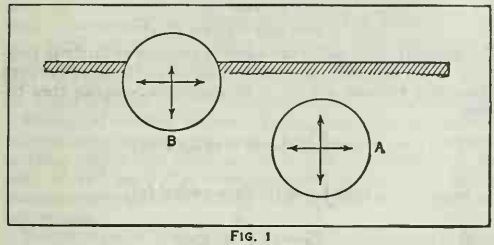


FIG. 1

of a thousandth of a centimetre radius may be drawn round it and the molecules outside this sphere will have no sensible influence on A. Those within the sphere attract A but in the case of a homogeneous isotropic body their pulls will be uniformly distributed and the resultant *force* on A will be zero. There will, however, be a pressure throughout the sphere due to the attraction: each spherical layer, being attracted, will compress the particles lying inside it. But if we consider a particle B very near the surface—nearer in fact than the range of action—the state of things will be found to be different. There will be more particles pulling B downwards than upwards so that all such particles as B will experience a resultant-*force* downwards. There will also be a *pressure* at B but its value is less the nearer B is to the surface. There is, therefore, a thin, indefinitely bounded, layer near the surface which is in a different condition from the main body of the substances. Many of the properties will be different in this surface layer from elsewhere. For example, corresponding to the different pressure, the density will be different. Since work is done in producing compression the potential energy will differ in the two regions. In the main body it can be written

$$\int \chi_0 \rho_0 dV = \chi_0 \rho_0 V_0$$

where V = volume, ρ density and χ the potential energy per unit mass; in the film, owing to its minute thickness, e, which we divide up into still thinner layers, de, and surface S we write

$$\text{Total mass} = V\rho_0 - S\int_0^e (\rho_0 - \rho)\, de$$

while the total potential energy (in terms of χ its value per unit mass) is

$$\int \chi \rho dV$$

or

$$V\chi_0\rho_0 - S\int_0^e (\chi_0\rho_0 - \chi\rho) de.$$

Multiplying the total mass by χ_0 and subtracting from the last expression

$$E - M\chi_0 = S\int_0^e (\chi - \chi_0)\rho de.$$

The right hand side is therefore an expression for that part of the energy which depends upon the existence of a surface S. The integral which multiplies it is the constant (or rather, factor) known as the *surface tension*. This integral is a measure of the work done in increasing the area of the surface. It must be carefully distinguished from the corresponding increase in the total energy because heat is drawn in simultaneously if the change takes place at constant temperature. It is in strictness a measure of the change in the "free-energy." (*See* THERMODYNAMICS: *Thermodynamics and Physical Chemistry.*)

By the principles of energy the potential energy (at constant temperature) tends to a minimum; in other words, any change that takes place spontaneously involves a diminution of this energy. Hence, whenever the surface tension is positive there will be a tendency for the surface to decrease, and the diminution will in fact take place unless it is resisted by other forces. This diminution takes place not by a contraction of the liquid but by a passage of the surface molecules into the body of the liquid. The properties of the body are not changed thereby. In this respect the phenomenon is quite different from the case of a stretched india-rubber film. In the case of the liquid the surface tension remains constant during the contraction of the surface area; in the case of the stretched india-rubber it diminishes along with the contraction.

We may express it otherwise by saying that if an imaginary straight line be drawn anywhere in the surface, when *equilibrium exists* there must be a force acting across the line in such a direction as to prevent further contraction of area. It is easy to show that this force per unit length of line is numerically the same as the surface-tension as defined above.

Dupré has described an arrangement by which the surface-tension of a liquid film may be illustrated. A piece of sheet metal is cut out in the form AA' (fig. 2). A very fine slip of metal is laid on it in the position BB' and the whole is dipped into a solution of soap. When it is taken out the rectangle $AA'CC'$ is filled by a liquid film. This film tends to contract on itself and the loose strip of metal will, if let go, be drawn towards AA'. If S is the area *of one face of the film* the potential energy is σS. If $AA' = b$ and $AC = a$ it is equal to σab. Hence if F is the force by which the slip is pulled towards AA', $F = \frac{d}{db}(\sigma ab) = \sigma a$; so that the force per unit length for one face is σ. Hence, σ is either the force per unit length or the potential energy per unit area.

It must be added that we have only considered one of the two faces and the force due to it. There is an equal force on the strip due to the second surface so that the total force is twice as great as the value taken. But the total area is twice as great also so that the conclusion remains unchanged.

There are other ways of illustrating the existence of the tension. Form a film as before but in a fixed frame. Tie a short length of spider line or fine unspun silk so as to form a flexible ring. If it is placed carefully on the film it usually takes an irregular shape. If the film inside the ring is destroyed by piercing it with a hot wire the spider line opens out into a circular form, this being the form which makes the surrounding surface least for a given length of line. Again, a small drop of mercury or a falling rain drop is practically spherical—the sphere being the form which has the smallest area for a given volume. A large drop fails to be spherical because gravitational forces are then strong enough to compete successfully with surface forces. Sometimes a difficulty is felt in connection with the assertion that the surface-tension is constant. Take a film such as we considered and place it vertically. The upward force on $AA' = 2\sigma a$ for the two faces, the downward force at the level CC' is also $2\sigma a$. The remaining force on the portion of the film between these lines is its weight. Hence the resultant force on the mass of the film is its weight and its acceleration should therefore be that of a freely falling body. Yet whatever may be found true for a pure substance, it cannot be true in general that the tension is the same at all parts of a large area. In the case of a soap film some adjustment must automatically take place (either by alteration of concentration or otherwise) which makes the upper tension a little greater than the lower. It is probably the power of such an adjustment being made in the case of a soap solution which is the leading requisite permitting a *durable* film to be made. Otherwise the film would literally fall to pieces.

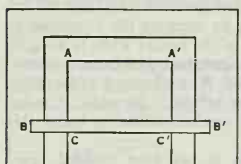


FIG. 2

Theory of Cohesion.—The mode in which cohesion is ex-

plained by assuming the existence of forces between elements (*i.e.*, infinitesimal volumes) of matter can be illustrated by supposing the substance of a body to be distributed continuously instead of partitioned into molecules. Such was the mode in which Laplace, more than a century ago, regarded it; and indeed it is the assumption that has always been made until recent times regarding the distribution of the substance of a body.

Let the force between two elementary volumes be supposed proportional to the product of the volumes and inversely as the nth power of their distances apart. The force required is that across unit area drawn anywhere well inside the body arising from the volumes on each side of it. The calculation is made in successive stages.

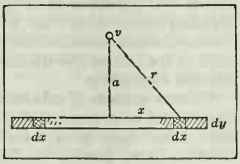


FIG. 3

(i.) Take a lamina of thickness dy. The force on an elementary volume v due to a zone of radius x and width dx is $v \cdot \frac{2\pi x\,dx}{r^n} \cdot \frac{a}{r}\,dy$ since the tangential components will cancel each other. Hence the whole lamina attracts the element with a force

$$2\pi va \int_a^\infty \frac{x\,dx}{r^{n+1}} \cdot dy = 2\pi va \int_a^\infty \frac{dr}{r^n}\,dy.$$

This becomes $\frac{2\pi va}{(n-1)a^{n-1}}\,dy$ or $\frac{2\pi v}{(n-1)a^{n-2}}\,dy$ since we assume the forces to vanish at infinity.

(ii.) Let v be an element in a parallel lamina of thickness da. Then the total force per unit area on such a lamina II. is obtained by putting $v = da$. Hence we find the force on a semi-infinite solid above lamina II. due to lamina I. by integrating from o to ∞ that is

$$\text{Force on semi-solid (per unit area)} = \frac{2\pi dy}{n-1} \int_0^\infty \frac{da}{a^{n-2}}$$

$$= \frac{2\pi}{(n-1)(n-3)} \left| \frac{1}{a^{n-3}} \right|_0^\infty$$

(iii.) The force (per unit area) on such a semi-infinite body due to a semi-infinite body in contact with it is obtained by integrating with respect to y from $y = 0$ to $y = \infty$. This force is the cohesion per unit area and is given by

$$K_0 = \frac{2\pi}{(n-1)(n-3)(n-4)} \left| \frac{1}{a^{n-4}} \right|_0^\infty$$

The unit of force used is that between two unit volumes separated by unit distance.

Lord Rayleigh remarks: "The pressure will therefore be infinite whatever n may be. If $n-4$ be negative the attraction of infinitely distant parts contributes to the result; while if $n-4$ be positive the parts in immediate contiguity act with infinite power. For the transition case, discussed by W. Sutherland (*Phil. Mag.* xxiv. p. 113, 1887) of $n=4$, K is also infinite." He adds "It seems therefore that nothing satisfactory can be arrived at under this head."

In more recent times, however, it has been realized (particularly by the Dutch school of physicists) that to treat the body as a *continuum* is bound to give erroneous results. Every substance is known to be built up of molecules. If we assume the existence of attractions depending on the distance between the centres of the molecules it is at the same time necessary to recognize that the two halves of the body can never approach to absolute contact. The lower limit in the last integration should be the least distance of separation between the lines of centres of contiguous layers. This will be a magnitude comparable with the molecular diameter, say s. The value we seek then becomes

$$K_0 = \frac{2\pi}{(n-1)(n-3)(n-4)} \cdot \frac{1}{s^{n-4}}.$$

There is now no objection on the above grounds to the power law provided that $n > 4$.

There can be no doubt that if the amount of matter compressed into unit volume (*i.e.*, the density ρ) be increased the value of the intrinsic pressure would be increased. Moreover, a factor Γ (akin to the gravitation constant) is also needed if K is to be expressed in ordinary dynamical units. Hence finally

$$K = \Gamma \cdot \frac{2\pi\rho^2}{(n-1)(n-3)(n-4)} \cdot \frac{1}{s^{n-4}}.$$

The quantity, K, is called the Laplace pressure or intrinsic pressure. More generally we could assume that the force between elementary volumes is $f(r)$. The successive integrals then become

i. $$2\pi a \int_a^\infty f(r)\,dr\,dy = 2\pi a\phi(a)\,dy;$$

ii. $$2\pi dy \int_y^\infty a\phi(a)\,da = 2\pi dy\,\psi(y);$$

iii. $$K_0 = 2\pi \int_s^\infty \psi(y)\,dy$$

where ϕa, ψy are written as the values of successive integrals. Finally K is obtained in ordinary units by multiplying by $\Gamma\rho^2$ For example if $f(r) = \epsilon^{-kr}$ where k is a constant we have

$$\psi(s) = \frac{ks+1}{k^3}\,\epsilon^{-ks}.$$

The difficulty in testing these formulae arises from the occurrence of three variables (k, s, Γ). The only one of these about which we know the value independently is s which for liquids must be of the same order as the molecular diameter. Caution is necessary in considering forces which vary so fast with the distance as cohesive forces undoubtedly do. A small error in the assumed value for s would lead to very considerable error in k and Γ. The intrinsic pressure *at the breaking point* of a solid in tension should be the breaking stress itself. The latter as actually measured is, however, only a lower limit of K since flaws in the material and irregularities in the application of the tension will lead to fracture occurring at a value lower than the ideal stress. The effect of cohesion on the characteristic equation of a gas was allowed for by van der Waals who, following Laplace, took the intrinsic pressure as varying with the density square alone. The equation becomes

$$(p + a\rho^2)(v - b) = RT$$

where at the same time b is introduced to represent the fact that the volume of the gas cannot be reduced to zero. Here $K = a\rho^2$ which for water at ordinary densities is between 10–20 Atmos. The fact that the equation turns out to be approximate only makes caution necessary. Moreover a is found to vary nearly inversely as the absolute temperature. We are obliged to assume, therefore, either that Γ is a function of the temperature or, what is more likely, that s is such a function. The latter assumption is indeed a very plausible one, for the least average distance of approach of the molecules must be affected by the motion of agitation amongst them and this increases as the square root of the absolute temperature.

Besides these considerations it must be pointed out that the modification worked out by the Dutch school only partly allows for the effects arising from the molecular character of matter. Knowing the forces between two molecules at a given distance apart, the actual value of the intrinsic pressure could only be correctly obtained by summing up the components (normal to the interface at which the pressure is being calculated) of all the forces between every pair of molecules each being calculated for the actual positions in which the molecules instantaneously are. It is true, that with forces that vary so fast with distance it is only the molecules which are fairly near to a point whose individual contributions are important: the contributions of the remainder may be estimated by an approximate method such as that followed by Laplace. Something has been done in this direction (Lennard-Jones, *Faraday Society*, "Discussion on co-

hesion," 1927) but the fact remains that though the law of force is probably greater than the inverse fifth it may vary up to the inverse seventh so far as present knowledge goes. If we take the higher of these two estimates the force between molecules which are neighbours and those that are next door to neighbours will be as 2^7 to 1, *i.e.*, in the ratio 128 to 1 or that between any layer and the next and next but one contiguous layer as 2^6 to 1, *i.e.*, in the ratio of 64 to 1. Hence all but a small percentage of the intrinsic pressure is due to the molecules in a small volume surrounding the point. This conclusion is important since it signifies that in any material (gas, liquid, solid) the intrinsic pressure will be uniform at all points except for a thin layer of usually negligible volume near its boundary.

Theory of Surface Forces.—This surface layer has theoretically no inner boundary but it may be taken, for most purposes, as being only a few molecules thick (mol. diameter is of the order of 10^{-8} cm.). The most prominent peculiarity is that it possesses more potential energy than the rest of the body per unit volume.

Now increase of free energy at constant temperature is equal to the external work done on a system. If we write $dW = \sigma dA$ σ is called the surface tension. To determine its value theoretically we estimate dW for the formation of a surface of unit area. Now if a body is split by a plane section and the two halves separated, a fresh surface is formed. The attraction per unit area between the halves at any distance x is

$$2\pi\int_x^\infty \psi(y)dy = K(x) \text{ say.}$$

The total work of separation is $\int_s^\infty K(x)\,dx$ and in the process a fresh area of 2 units is formed. Hence expressing the work done in ordinary units

$$2\sigma = 2\pi\Gamma\rho^2\int_s^\infty K(x)dx$$

(In previous editions many theorems were given [some of which date from Laplace] which depend, however, upon the lower limit being zero. It is sufficient to say that these theorems are not applicable to the formulae given here.)

If a power law is taken as the law of force between molecules it follows that

$$\sigma = \frac{\pi\rho^2\Gamma}{(n-1)(n-3)(n-4)(n-5)s^{n-5}}.$$

This law may be combined with that for K with the result that

$$\frac{K}{\sigma} = \frac{2(n-5)}{s}.$$

For water at ordinary temperature, approximately

$$K = 10{,}000 \text{ Atm.} = 10^{10} \text{ dynes/sq.cm.}$$
$$\sigma = 75 \text{ dynes/cm.}$$
$$s = 4\times10^{-8} \text{ cm.}$$

Whence, $$n-5 = \frac{10^{10}\times4\times10^{-8}}{2\times75} = \frac{8}{3}$$

and $$n = 7{\cdot}67.$$

It must not be forgotten, however, that K and s are obtained by very approximate methods only.

Interfacial Tensions.—When two liquids are in contact there is in general a surface-tension at the interface which is called interfacial tension. Denoting this by σ_{12} and the tensions of the liquids as σ_1 and σ_2 respectively, and carrying out a separation of the two liquids the value of σ_{12} can be determined. In the above process the interface between 1 and 2 disappears and two fresh surfaces appear at which the tensions are σ_1 and σ_2. Hence

$$\sigma_2+\sigma_1-\sigma_{12} = 2\pi\rho_1\rho_2\int_{\frac{s_1+s_2}{2}}^\infty K(x)dx$$

where the lower limit is the sum of the radii of the two kinds of molecules. Hence

$$\sigma_{12} = \pi\rho_2^2F(s_2)+\pi\rho_1^2F(s_1)-2\pi\rho_1\rho_2F\left(\frac{s_1+s_2}{2}\right).$$

Young took the functions F as being identical being all the same function of zero instead of the molecular diameters;

whence $$\sigma_{12} = \pi(\rho_2-\rho_1)^2F(0)$$

or $$\sqrt{\sigma_{12}} = \text{const.}\times(\rho_2-\rho_1).$$

Hence for three liquids taken pair by pair

$$\sqrt{\sigma_{12}}+\sqrt{\sigma_{23}}+\sqrt{\sigma_{31}} = 0.$$

These simplified relations are known not to be true. On the other hand if the power law is valid and is the same law for different substances the equations given here lead to

$$\sigma_{12} = \left(\frac{\rho_2^2}{\sigma_2^{n-5}}+\frac{\rho_1^2}{\sigma_1^{n-5}}-\frac{2\rho_1\rho_2}{\left(\frac{s_1+s_2}{2}\right)^{n-5}}\right)\frac{\pi}{(n-1)(n-3)(n-4)(n-5)}$$

and the square root law should not be expected to be true excepting in special cases in which the molecular diameters of the three substances were the same.

All surface-tensions with which we have to deal are in reality interfacial tensions since a liquid is always in contact with its own vapour or the gas in which it is immersed. The effect due to the gas is known, however, to be very small.

The normal pressure to be expected from the existence of a superficial tension is best obtained by considering in the first place a cylinder.

Consider an element of surface, $ds\times$unity. The tensions acting at its extreme edges are inclined to one another and have a resultant inwards. Their resultant is a normal force $2\sigma\sin\frac{d\theta}{2}$ or in the limit $\sigma d\theta$. But $\frac{ds}{d\theta} = R$ (the radius of curvature); hence the normal force is $\sigma\frac{ds}{R}$. Since this acts over an area equal to ds, the normal pressure is σ/R.

If we take a body of any shape and select a square-bounded element of the surface, the tensions in the plane of the diagram are equivalent to a normal force $\sigma ds_1\times\frac{ds_2}{R_1}$ and those acting on the other two edges to $\sigma ds_2\times\frac{ds_1}{R_2}$. Hence per unit area

$$p = \sigma\left(\frac{1}{R_1}+\frac{1}{R_2}\right).$$

Here R_1 and R_2 are the radii of curvature in two planes at right angles. Owing to Gauss's theorem of integral curvature the sum of two such curvatures at a point is a constant. The above theorem is therefore true whether the two rectangular planes considered are principal planes or otherwise. In the case of a sphere

$$p = \frac{2\sigma}{R}.$$

Effect of Temperature.—The variation of surface tension with temperature was not taken into account in Laplace's theory. The forces between molecules were taken, like gravitational forces, to be independent of temperature. The rise in a capillary tube was supposed to vary merely because the density varied. Experiment has shown, however, that the tension itself in all cases diminishes with rise in temperature. That it must ultimately diminish to zero can be inferred from the fact that according to the usual conceptions as to the critical state a liquid and its vapour will at that point become identical with one another; the tension at the interface must then be zero. Eotvös introduced the conception of molecular surface tension. By means of somewhat doubtful reasoning he concluded that $\sigma(Mv)^{\frac{2}{3}}$ (where Mv is the molecular volume), plotted against temperature should give

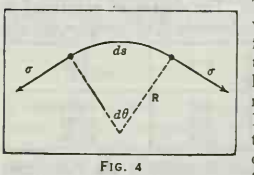


FIG. 4

the same kind of curve for different substances, and he found in fact that it was representable by $K(T_c-T)$ where T_c is the critical temperature and K a constant. According to Ramsay and Shields K should be the same constant for all non-associated substances. In reality K varies betyeen 1·5 and 2·6 and it is necessary also to change T_c-T to $T_c-T-\delta$ where δ is a small constant, so that the law is only a rough one. According to this equation the curve would be a straight line cutting the temperature axis a few degrees below the critical point.

In connection with this, it is important to remember that σ is only the free energy per unit area and not the total energy per unit area (u). Lord Kelvin was the first to prove that the variation of σ with temperature, involves that u must be greater than σ; or in other words, when an expansion of the surface takes place heat must be added to keep the temperature constant. The connection between these quantities he proved thermodynamically to be $u = \sigma - T\frac{d\sigma}{dT}$ (*vide* THERMODYNAMICS applied to chemistry). Now, at the critical point, u (which is only the *extra* energy due to the existence of unit surface and not the total energy of the whole body) must also vanish. But if both u and σ are then zero so also must $T\frac{d\sigma}{dT}$ be zero.

Hence, if our conceptions of the critical state are correct the curve of σ plotted against T instead of being a straight line must become horizontal at the critical point. In 1894, van der Waals showed that a formula of the type $\sigma = A(T_c-T)^n$ was to be expected and gave the average value of $n = 1·27$ as deducible from the experimental values. In 1916 A. Ferguson (*Phil. Mag.*, Jan. 1916), examining the data for 14 different organic substances showed that the best values of n ranged from 1·187 to 1·248. The advantageous character of this power equation is that it makes σ, u, and $\frac{d\sigma}{dT}$ all equal to zero at the critical point and it thus satisfies all that is required of it theoretically. Liquids, however, which are believed to be associated do not follow the above simple law.

The surface energy, u, also falls to zero at the critical point, but according to a different law. For taking the equation

$$u = \sigma - T\frac{d\sigma}{dT}$$

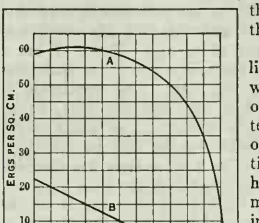
it follows that $u = A(T_c-T)^{n-1}(T_c+[n-1]T)$. The curves for σ and u are shown in fig. 5 for the case of benzene. The curve shows how nearly a straight line law satisfies the values of σ until the critical point is closely approached. An interesting relation has been given by Prof. E. T. Whittaker between u and the latent heat (internal) of *evaporation*. The internal latent heat (L_i) is obtained by subtracting the external work done from the ordinary latent heat. The relation in question is that $\frac{T_c}{T}\cdot\frac{u}{L_i}$ is nearly a constant, and that (in C.G.S. units) this constant is about unity for ethyl oxide, methyl formate, benzene and chlorobenzene and about 2 for carbon tetrachloride. The degree of constancy is indicated in the following table for benzene (Crit. temp. = 561.5A)

T°A	$\frac{T_c u}{TL_i}$	T°A	$\frac{T_c u}{TL_i}$
353	1·106	493	1·089
403	1·100	503	1·106
453	1·067	513	1·129

The relation is important since it indicates the close connection there is between the surface energy and the volume energy of the substances.

General Properties of Surface Tension.—The notion of surface tension, though it is only a derived one, enables us to deal with all the phenomena which were enumerated at the head of this article without making explicit reference to the more fundamental notion of molecular attractions upon which they in reality depend. It is appropriate to summarize here the leading characteristics of this property.

The tension of a liquid across any straight line drawn on the surface is normal to that line and is the same for all directions of the line and is measured by the force across an element of the line divided by the length of that element.



BY COURTESY OF THE FARADAY SOCIETY

FIG. 5

For any given homogeneous liquid surface, as the surface which separates water from air or oil from water, the surface tension is the same at every point of the surface and in every direction. In the case of mixtures, however, when circumstances demand it (*e.g.*, when the surface is inclined and the effects of a gravitation field require consideration) the concentration at the surface may adjust itself from point to point of the surface in such a way as to produce a corresponding variation in the value of the surface tension sufficient to satisfy the conditions of equilibrium. This effect is very minute and in nearly every problem may be ignored.

When the surface is curved the effect of the surface tension is to make the pressure on the concave side greater than that on the convex side by the amount $\sigma(C_1+C_2)$ where C_1 and C_2 are the curvatures in mutually perpendicular normal planes.

The tension of the surface separating two liquids cannot be deduced by any reliable method from the tensions of the liquids when separately in contact with air. The experiments of C. G. M. Marangoni, van der Mensbrugghe, and Quincke have led to results which show that the common surface between two liquids has a tension always less than the difference of the tensions of the separate liquids. This is usually referred to as Marangoni's rule.

If three liquids meet along a line the interfaces at the common edge must be parallel to the sides of a triangle proportional to the three tensions σ_{12}, σ_{23}, σ_{31} otherwise equilibrium cannot obtain (Neumann's rule). This triangle cannot exist unless two of the interfacial tensions are greater than the third. Marangoni's experimental rule shows that the triangle is always imaginary. Hence three pure fluids (of which one may be a gas) accordingly cannot remain in contact. If a drop of oil stands in a lenticular form upon a surface of water, it is because the water-surface is already contaminated with a greasy film.

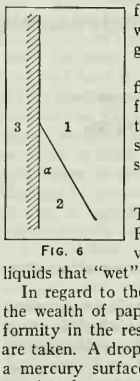


FIG. 6

When a solid body is in contact with two fluids the surface of the solid cannot alter its form but the angle at which the surface of contact of the two fluids meets the surface of the solid must depend on the values of the three surface tensions. For equilibrium

$$\sigma_{31}-\sigma_{32}=\sigma_{21}\cos\alpha.$$

The angle α is known as the angle of contact. For pure substances the angle is definite. It varies from zero (or very nearly so) for liquids that "wet" the solid to 130° or 140° for mercury on glass.

In regard to the experimental behaviour of water on mercury the wealth of papers is equalled by the remarkable lack of uniformity in the results obtained even when elaborate precautions are taken. A drop of water which fails to spread when placed on a mercury surface is often caused to spread by the action of pouring the mercury out of the dish—thus involving the creation of a new surface! Water shows a positive tendency to spread *in vacuo* on a drop *formed in vacuo!* (Burdon, *Faraday Society*, May, 1927).

Form of a Capillary Surface.—The form of the surface of a liquid acted on by gravity is easily determined if we assume that near the part considered the line of contact of the surface of the liquid with that of the solid bounding it is straight and horizontal.

This will be the case, for example, near a flat plate dipped into the liquid. Let A_2P_2 be the vertical plate supposed to be so wide that the edge effects do not count. We will consider the forces per unit depth (perpendicular to the paper) and will consider A_1A_2 as the standard level of the liquid far from the plate. The curvature of the surface at the point x, y, is

$$\frac{d\theta}{ds} = \sin\theta \frac{d\theta}{dy} = -\frac{d(\cos\theta)}{dy}$$

and consequently the pressure in the liquid at that point is less than the pressure outside it by $-\sigma \frac{d(\cos\theta)}{dy}$. But the pressure is less than that at a distant point where the surface is flat by $g\rho y$ or $g(\rho-\rho_0)y$ if allowance is made for the surrounding gas. Hence the differential equation for the surface is

$$-\frac{d(\cos\theta)}{dy} = \frac{g(\rho-\rho_0)}{\sigma} y$$

the integral of which is $1-\cos\theta = 2\sin^2\frac{\theta}{2} = \frac{g(\rho-\rho_0)}{\sigma} \cdot \frac{y^2}{2}$.

The shape of the surface is therefore the *elastica, i.e.*, the form taken by a uniform spring when equal and opposite forces are applied at its ends (*see* Thomson and Tait, *Natural Philosophy*, vol. i. p. 455). The equation might have been found by considering the elementary volume of a strip of liquid, dx, sustained above the normal level of the liquid by the surface forces at its outer and inner edges. The difference of the vertical component of these is $\sigma d(\sin\theta)$ and the weight of the strip is $g\rho y dx$ but since $\frac{d(\sin\theta)}{dx} = -\frac{d(\cos\theta)}{dy}$ the same result is obtained.

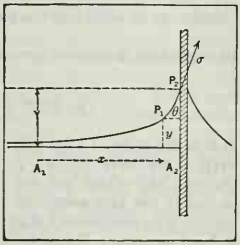
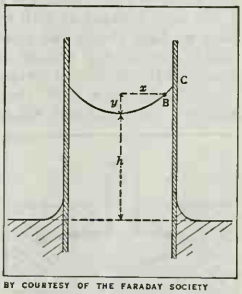


FIG. 7

If the angle of contact is α the value of θ at the plate is $\frac{\pi}{2} - \alpha$; hence

$$1-\sin\alpha = \frac{g(\rho-\rho_0)}{\sigma} \cdot \frac{Y^2}{2}$$

which gives the greatest height through which the liquid is raised. A thin sheet of glass suspended from the arm of a balance and just dipping into a liquid provides a simple way of determining the surface tension. The increase in the apparent weight of the sheet when the liquid has been raised by capillarity is σ multiplied by twice the horizontal length of the sheet. A sensitive form of this apparatus is known as Worthington's multiplier in which the strip is rolled into a vertical cylindrical spiral.



BY COURTESY OF THE FARADAY SOCIETY
FIG. 8

Ascent up a Capillary Tube.—A capillary tube dipped in a liquid provides a simple method of determining surface tension. To connect the surface tension with the rise (or descent) of the liquid in the tube measured from the level part of the outside surface we equate the effective weight of the liquid raised to the total force due to the tension. The effective weight (with the symbols shown on diagram) is

$$g(\rho-\rho_0)\int_0^z 2\pi x(h+y)dx$$

for an inner cylinder of radius x where ρ is the density of the liquid and ρ_0 that of the surrounding gas or vapour. The vertical component of the surface force acting all round the edge of the liquid surface of this cylinder is $2\pi x\sigma\sin\theta$ or $2\pi x\sigma\cos\alpha$ where θ is the angle between the surface and the horizontal and α is the "contact-angle" which is measured with respect to the solid surface (in this case the vertical). These two expressions are to be equated to one another.

If we can neglect y altogether compared with h and integrate between the limits 0 and r we obtain $2\pi r\sigma\sin\theta = g(\rho-\rho_0)\pi r^2 h$ whence $\sigma\sin\theta = \frac{1}{2}g(\rho-\rho_0)rh$. To seek for a better approximation we first differentiate the two sides of the equation whence

$$\frac{d(x\sin\theta)}{dx} = \frac{g(\rho-\rho_0)}{\sigma} x(h+y) = \frac{x(h+y)}{\beta^2}.$$

We may consider this as the standard exact equation. The left side is

$$\frac{x\,d\sin\theta}{dx} + \sin\theta \quad \text{or} \quad x\left(\frac{\cos\theta\, d\theta}{dx} + \frac{\sin\theta}{x}\right).$$

But $\cos\theta\frac{d\theta}{dx} = \frac{d\theta}{ds}$ (where ds is an element of the curve AC) and is $\therefore$ the curvature in the plane of the diagram at B; and $\frac{\sin\theta}{x}$ is the curvature of the surface at B in a plane at right angles to the diagram. Hence, calling these curvatures $\frac{1}{R_1}$ and $\frac{1}{R_2}$ we have

$$\sigma\left(\frac{1}{R_1}+\frac{1}{R_2}\right) = g(\rho-\rho_0)(h+y);$$

both of the terms here are expressions for the difference of pressure between the two sides of the surface at the point B.

Narrow Tubes.—The equation shows that the sum of these curvatures increases with the height; but when h is big and y small, as it is for a very narrow tube, the sum is nearly constant. Going back to the differential form of the equation and taking the case where the liquid wets the tube, we may consider the curve as practically circular, of radius r; the corresponding value of $y = r - \sqrt{r^2-x^2}$ and

$$\frac{d(x\sin\theta)}{dx} = \frac{1}{\beta^2}(h+y)x = \frac{x}{\beta^2}(h+r-\sqrt{r^2-x^2}) \text{ nearly whence, by integration}$$

$$x\sin\theta = \frac{1}{\beta^2}\left[\frac{hx^2}{2} + \frac{rx^2}{2} + \frac{1}{3}\{(r^2-x^2)^{\frac{3}{2}} - r^3\}\right].$$

or since $\sin\theta = 1$ when $x = r$

$$1 = \frac{r}{2\beta^2}\left(h + \frac{r}{3}\right).$$

By putting the radius of the circle as $R = \frac{r}{\cos\alpha}$ when the angle of contact is α instead of zero this becomes

$$\cos\alpha = \frac{r}{2\beta^2}\left[h + \frac{r}{3\cos^3\alpha}\{1 - 3\sin^2\alpha - 2\sin^3\alpha\}\right].$$

By calculating $\tan\theta$ from $\sin\theta$ and bearing in mind that

$$\tan\theta = dy/dx,$$

we can by subsequent integration obtain a closer value of y in terms of x which can, in turn, be introduced into the differential equation and so on. By such successive approximations Lord Rayleigh has obtained the equation for the case $\alpha = 0$ in the form

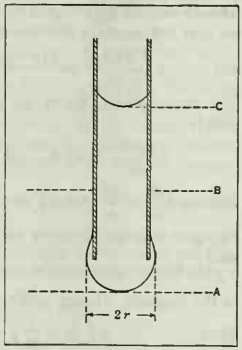
$$\sigma = \tfrac{1}{2}g(\rho-\rho_0)hr\left[1 + \frac{1}{3}\frac{r}{h} - 0{\cdot}1288\frac{r^2}{h^2} + 0{\cdot}1312\frac{r^3}{h^3}\right]$$

in which the coefficients are claimed to be correct to the approximation given. The equation can only be used in the case when r is small compared with h. It may be mentioned that a very near approach to the equation can be obtained by considering the surface as ellipsoidal with its minor axis vertical. This was first shown by Hagen and Desains. Various methods are employed for determining surface tension experimentally:

(*a*) From the rise in a capillary tube making use of the above equation.

(*b*) *Sentis's Method.*—A capillary tube is partly immersed in the liquid. It is then withdrawn vertically and a drop remains

clinging to the lower end; the position A of the vertex of the drop is noticed. The liquid in the vessel is then raised until it touches the vertex (when the column falls) and then raised further till the upper surface is at the same level C as at first. The liquid in the vessel is then at the level B. The vertical distance AB corresponds to h in the ordinary method but in the *correcting terms* h must be put negative. The width of the *drop* is in this case to be small compared with h.



BY COURTESY OF THE FARADAY SOCIETY
FIG. 9

(*c*) *Jaeger's Method.*—In this method an orifice (a "tip") is placed just under the surface of the liquid and the pressure of gas is increased until bubbles form. The *maximum* pressure of the gas (which is fairly sharply marked) is observed. The deduction of the applicable equation is a somewhat delicate matter because the problem is really a kinetic, not a static, one. The formula employed is

$$\sigma = \frac{p_{\max} r}{2}\left[1 - \frac{2}{3}\frac{r}{h}\right]$$

where $p_{\max}$ is the maximum difference of pressure between the level of the end of tube inside and outside, and r is the internal radius of the tube. The bubble is assumed to form within the internal circumference. In practice, however, it sometimes forms on the outside circumference. To obviate the uncertainty it is recommended to make the two circumferences as nearly equal as possible, but this is an experimental matter of great difficulty. The final accuracy depends chiefly on the measurement of the radius r. The subject of the formation of bubbles and drops requires much more study than it has received.

(*d*) *Drop Weight Method.*—This method is connected with the preceding one because the drops considered are those issuing from a narrow tube. It was employed by Tate (1864) who gave as the result of his observations:—Other things being the same, the weight of a drop of liquid is proportional to the diameter of the tube in which it is formed. Later, Quincke used it and gave the value $2\pi r\sigma$ as the weight of the suspended drop just before falling provided that the inflow of liquid is sufficiently slow. A portion of the drop is always left behind when the main part falls; but he considered that it might be neglected in the case of very small drops.

Rayleigh has discussed the question from the "dimensional" point of view. (*See* DIMENSIONAL ANALYSIS.) Assume that the mass M of the drop depends only upon the surface tension, σ, the value of acceleration due to gravity, g, the density of the liquid, ρ, and the inside radius, a, of the tube. Rayleigh shows that $\sigma a/g$ has the same dimensions as a *mass* and that $\sigma/g\rho a^2$ is a pure number. Since quantities that can be equated together must be quantities of the same kind, it follows that $M \propto (\sigma a/g)F(\sigma/g\rho a^2)$ where there is no restriction, imposed by this method of inquiry, upon the function F of the quantity in the second pair of brackets. Rayleigh finds by experiment that $gm/\sigma a$ is fairly constant for wide ranges in the diameter of the tube. For thin-walled tubes in the case of water the following values were obtained:

$\frac{\sigma}{g\rho a^2}$	$\frac{gM}{\sigma a}$	$\frac{\sigma}{g\rho a^2}$	$\frac{gM}{\sigma a}$
2·58	4·13	·277	3·78
1·16	3·97	·220	3·90
·708	3·80	·169	4·06
·441	3·73		

The mean value of the constant is thus about 3·8 instead of 2π which the imperfect theory gave. Further experiments by Harkins and Brown show that the constant approaches Quincke's value as the diameter becomes very small.

As a means for measuring surface tension this method is obviously not satisfactory; but for rough comparative values for liquids of like kind it is a very quick and easy method.

Wide Tubes.—When a is not small the approximations made above for narrow tubes are not suitable.

The question has been fully discussed by the late Lord Rayleigh to whose papers reference should be made (*v.* Rayleigh, *Proc. Roy. Soc.* A, 92, 184, 1915).

Bubbles and Drops.—These can be dealt with by similar methods. Air bubbles in a liquid can be formed of any size, from minute ones which are nearly spherical to large ones shaped like a flat cake. The excess pressure $p - p_0$, which is easily found by experiment, determines the total curvature at the lowest point N; being equal to $\frac{2}{R}$. If y is reckoned upwards as before, then

$$\frac{d(x\sin\theta)}{dx} = \left(\frac{2}{R} + \frac{y}{\beta^2}\right)x :$$

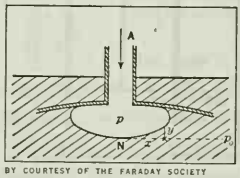
thus $\frac{2}{R}$ replaces h/β^2 in the previous problems. For very large bubbles, as for wide tubes, a good approximation can be obtained by neglecting the second curvature and also $\frac{2}{R}$, so that

$$(1 - \cos\theta) = \frac{y^2}{2\beta^2} = 2\sin^2\frac{\theta}{2}.$$

If H is the value of y for which the tangent to the curve becomes vertical we have $H = \sqrt{2}\cdot\beta$. This method has been used by Quincke and others but not always with bubbles large enough to justify the approximation that is made. The bubble is conveniently formed under a slightly concave surface so as to prevent it from escaping. Drops of mercury are easily formed above a concave surface. If y is measured *downwards* from the summit the same formulae hold as for bubbles.

Great attention has been paid to these methods owing to their use in determining surface-tension. The most thorough treatment from the practical point of view is given by Bashforth and Adams (*Capillary Action*, 1883). By means of infinite series calculated for each of the variables each term in the differential equation can be calculated to very high accuracy and tables are drawn up enabling the form to be determined for given weights of material and given surface tensions.

Thin Films.—In this section we consider thin films. The first group of cases consists of those in which the substance forms a thin sheet with gas on both sides, as for example, a soap bubble. Here we have always two surfaces to consider. When the thickness is considerably greater than the value of the range, ϵ of molecular forces the two surfaces are independent of one another; the surface-tension must then have the usual value for each. For very thin films this will not be so. If the thickness is less than the range, ϵ, the tension for each surface will be certainly less than the normal value. When the variation of density at different levels in the film is taken into account it is clear that the reduction in the surface-tension must begin when the thickness is twice ϵ.



BY COURTESY OF THE FARADAY SOCIETY
FIG. 10

A soap film is simply a small quantity of soapsuds spread out so as to present a large surface to the air. The soap solution may with great advantage be specially prepared. The addition of glycerine and resin enables more permanent films to be prepared. Bubbles may easily be blown 20 in. in diameter in the open air using a clay pipe or a small glass funnel. (The method needs no precise description.) Sir James Dewar has obtained bubbles that last almost indefinitely if evaporation be prevented by blowing them in a confined space saturated with water vapour.

The pressure inside a bubble is greater than that outside by

$2\sigma\left(\frac{1}{R}+\frac{1}{R}\right)$, *i.e.*, by $4\sigma/R$ where R is the radius of curvature.

The factor 2 arises because both surfaces give rise to normal components of forces. This result is easily obtained by considering the equilibrium of each *half* of the bubble. The force at the cut edge is $2\sigma \cdot 2\pi R$ and this is balanced by the excess pressure acting over the plane area πR^2. Hence the excess pressure is

$$\frac{2\sigma \cdot 2\pi R}{\pi R^2} = \frac{4\sigma}{R}.$$

If the end of the pipe or funnel is opened the bubble will contract because of the escape of high pressure air from inside. As it contracts the pressure rather paradoxically increases. It may be added that (neglecting gravity) the bubble is spherical because this *shape* makes the surface area (and therefore the potential energy) least for a given volume.

If bubbles are blown on the expanded ends of two funnels (the stems being stopped by the fingers) they can be made to coalesce by carefully bringing them into contact. (A slightly electrified rod brought not too near will assist the coalescence.) On drawing the funnels slowly apart a cylindrical bubble is obtained. The excess pressure is now $2\sigma/R$ where R is the radius of the opening of the funnel. Further separation causes the cylinder to contract in the middle; one of the two curvatures is now negative; the surface is an *anticlastic* one. With further extension of length the surface becomes unstable and the film collapses. It has been shown by Clerk Maxwell that a cylindrical film becomes unstable when its length is greater than its circumference.

Films on Liquids.—In the second group of cases a liquid or solid is on one side and a gas on the other. To this group belong all cases of the spreading of oil or other substance on water or other liquid or solid. This old familiar subject has recently acquired very great importance in its bearing upon molecular structures. The first to experiment in detail on such films was F. H. R. Lüdtge (*Pogg. Ann.* cxxxix., p. 620), who showed how a film of high surface tension is replaced by one of lower surface tension. Akin to these are experiments made on the erratic movements that are observed when fragments of camphor are sprinkled on a *clean* water surface. A trace of grease such as may be communicated to the surface by dipping a finger in the water may be sufficient practically to stop this motion. The thickness of oil which is required may be spoken of as the "camphor-point." The first to determine it was (the 3rd) Lord Rayleigh who in 1890 (*Roy. Soc. Proc.*) showed that the thickness of the film of olive oil, calculated as if continuous (*i.e.*, non-molecular) which corresponds to the camphor-point is about 2×10^{-6} mm.; *i.e.*, it is only a moderate multiple of the supposed diameter of a gaseous molecule and perhaps scarcely exceeds at all the diameter to be attributed to a molecule of oil; and he ultimately realized that this phenomenon was therefore entirely outside the scope of a theory of surface action such as Laplace's in which matter is regarded as continuous and that an explanation required a direct consideration of molecules.

In 1891 Miss Pockels (*Nature*, xliii., 437 [1891]) showed by means of a movable slider on water (the surface tension being measured by means of the attraction on a small disc) that the contaminating material on the surface could be squeezed up and concluded from her experiments that the water-surface can exist in two sharply-contrasted conditions; the normal condition, in which the displacement of the partition makes no impression on (the value of) the tension; and the anomalous condition in which every increase or decrease (of the surface) alters the tension.

The question was taken up again by Lord Rayleigh in 1899 (*Phil. Mag.* xlviii., 1899) with apparatus designed on the lines of that of Miss Pockels (but employing a different way of measuring the surface tension) and he concluded that the first drop in tension corresponded to a complete layer one molecule thick and that the diameter of a molecule of oil is about $1{\cdot}0\times10^{-6}$ mm.

Later investigations have been made by I. Langmuir (in America: *J. Amer. Chem. Soc.*, 1915 to 1918) and N. K. Adam (in England) and their coworkers. By measuring the surface area that can be completely covered by a weighed quantity of material Langmuir determined both the cross-sectional area and the length of a molecule and proved that the length for organic molecules like those of palmitic, stearic, cerotic acids, etc., was nearly five times the breadth, while in cetyl palmitate it is nearly ten times. Adam has developed the technique and has very much extended the theoretical interpretation (*Roy. Soc.*

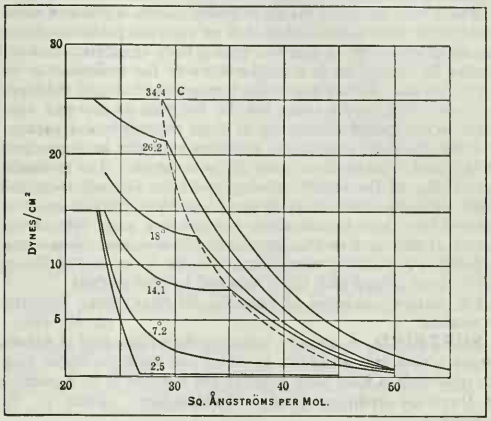


FIG. 11

Proc. A. 1921 to present time). Both Langmuir and he measured the amount of "oil" and measured directly the force required to compress the surface to any given area by means of a floating barrier. The actual force measured is the length of the barrier multiplied into the difference of the surface tension on the two sides of it. On the outside the surface is that of pure water—on the inside that of water on which a thin film has been formed. It is found that when the molecules in the film are so sparse as to be widely separated from one another the force due to the film is an expansive one.

This cannot be explained on Laplace's theory: it is necessary to modify it by allowing for the effects of thermal motion; the film in such a case is analogous to a two-dimensional gas. If the barrier is now moved so as gradually to compress the film various changes take place successively which are depicted in fig. 11 for the case of a film of myristic acid on weak HCl (N/100). The unit of area adopted is a square each side of which is an Ångström unit (*i.e.*, 10^{-10} metre), and the area specified is that occupied by one molecule of the acid. The ordinates of the curves are the difference of the tensions on the two sides of the barrier. The curves are isothermals extending from 2·5° to 34·4° C. They exhibit some of the characteristics of the p, v, curves for a condensible substance. At high temperatures the curves appear to be approximately rectangular hyperbolas. For small values of the area they approach the form for liquids. At intermediate stages, however, there is no constant pressure isothermal as in the analogous case of the vapour pressure of a liquid below the critical point (the form is more nearly that for the vapour pressure of a *mixture of two liquids;* this fact may indicate that the underlying water takes a part in the changes that occur).

It is not to be expected that there should be complete correspondence between the two classes of phenomena. There is, however, sufficient indication that the effects of thermal motions in the films cannot be neglected. The analogy with liquids and gases can be further illustrated by plotting FA against F, where F is the force applied per unit length; the resemblance to an Amagat diagram for a fluid is very striking (Schofield and Rideal, *Roy. Soc. Proc.* A, 109, p. 67; 110, p. 170). It is clear from the diagram that a film of myristic acid can be squeezed up until a molecule occupies less than 24 A^2. At the opposite end where the behaviour approaches that of a gas it is to be expected

that the equation would take the form $FA = RT$ where the value of R (allowing for only two degrees of freedom) should be 1.372 per molecule; at room temperature therefore FA should be about 400. This has been experimentally verified as a limiting value for very low surface pressures. With the long-chain fatty acids, esters and nitriles it is approached within 25%, the pressure being below 0.1 dyne per centimetre. With the dibasic ester $C_2H_5OOC.(CH_2)_{11}COOC_2H_5$ it has been verified within 10%.

The interpretation of the experimental results is still in a somewhat fluent state and it is clear that we must not press any *simple* gas analogy too far. A thin film upon a body of different material cannot be treated as in a similar state to the molecules in the body of a gas. Forces must exist between the film and the liquid beneath. This may account for the fact that at any rate some films do not spread indefinitely as a gas expands into a vacuum. Further, the most compressed state can only give an upper limit to the least cross sectional area of the molecule. The molecules are resting on the rapidly moving molecules beneath them and must share to some extent in their agitation. Still it must be granted that these investigations are throwing great light on the nature of films and on the dimensions of molecules. Where corroboration is possible, measurements made by means of X-rays are in good accord with those obtained by this method.

For certain properties of solutions *see* SOLUTIONS; THERMODYNAMICS. (A. W. Po.)

SURFBIRD, a stoutly built, sandpiperlike bird (*Aphriza virgata*), dark gray and white in colour and about ten inches long. Its relationships have been disputed but it seems to be a member of the snipe-sandpiper family (Scolopacidae), closest to the turnstones.

Its summer home is in the mountains of south central Alaska, where its nest, placed on rocky ground above timber line, was first discovered in 1926. In winter it lives along the Pacific coast of the Americas, from British Columbia to the Straits of Magellan. It feeds along the water line, preferring rocky or stormy shores where its favourite food, barnacles and mussels, is exposed at low tide. It is a rather quiet, inactive species, found in small flocks in winter, and is one of the lesser-known shore birds. (A. L. RD.)

SURF RIDING, a sport originating in Hawaii and now widely practised. For good surfing an open ocean exposure is required with a gently sloping bottom so that water depth at 100 to 1,000 yd. from shore, where the ride begins, is not over ten feet.

Where surf is light, flat boards four to six feet long by a foot wide, or air mattresses, ridden in a prone position, are used. For waves up to 20 ft. in height, and rides up to half a mile, a tapered hollow surf board 10 to 12 ft. long and 2 ft. wide, with a stabilizing fin near the back, is used. Body surfing without a board also is done. The rider first swims out to the point where the larger rollers peak up, but before the crest breaks. As the wave approaches him, he paddles his board toward shore to attain sufficient speed to start to coast down the face of the wave. Once he has caught the wave, he may rise to a standing position and ride it until it dies out near the beach. To increase speed and distance experts ride diagonally toward shore. (J. AN.)

SURGE, in general terms, is a sudden increase to an excessive or abnormal value and a fall from such a value subsequently. In meteorology, "surge" is the name applied to those long-period progressive changes of barometric pressure which occur simultaneously at many locations over a wide region and which are superimposed on short-period pressure fluctuations of a diurnal nature or variations caused by traveling atmospheric disturbances of a local character. A pressure rise or fall of two millibars in 24 hours that occurs simultaneously within an area of 3,000,000 sq.mi. corresponds to the order of magnitude of the phenomenon. Such pressure surges are commonly noted in regions close to the large semipermanent pressure centres of the earth such as the high-pressure area of the southern North Atlantic ocean. "Surge" is also a meteorological term often applied to a sudden increase in velocity of the large atmospheric wind streams, especially in the tropics. Such a wind surge occurs simultaneously over a large area and frequently spreads from place to place so that its progress can be followed on weather maps. In the trade-wind belts, meteorologists frequently refer to this event as the "surge of the trades" and, in the regions of the monsoon currents, as the "burst" or "surge of the monsoon." At a given location, an increase of wind speed from about 10 m.p.h. to 35 m.p.h. or more at all levels from the ground up to approximately 15,000 ft. can be expected in connection with a well-marked surge of the trades.

(D. C. McD.)

SURICATE (SLENDER-TAILED MEERKAT), a small South African mammal, *Suricata suricata,* belonging to the family that includes also the civet and mongoose. It ranges from Cape province to the Transvaal. The head and body are about 14 in. long and the tail half as much; the fur is long and soft, light grizzled gray in colour and banded with black on the lower part of the back. Suricates are sociable animals, living in holes in the rocks on the mountains and burrowing in the sandy soil of the plains. They feed chiefly on succulent bulbs, which they scratch up with long, curved, black claws on their forefeet. They are often domesticated as pets.

See CARNIVORE.

SURIGAO, a municipality and province of the Philippines. The province is on the northeast coast of Mindanao and includes the adjacent islands of Dinagat, Siargao and lesser isles. It lies within 10° of the equator so the climate is truly tropical. Surigao province with an area of 3,079 sq.mi., and a population (1960) of 359,997, is essentially hilly and mountainous with a narrow, often discontinuous coastal lowland. Surigao municipality, the provincial capital (pop. [1960] 37,439) has an average annual temperature of 80° F. and an annual rainfall of 144 in. with an average of 209 rainy days per year.

Agriculture is characterized principally as subsistance cropping by small landholders. In the mountains there is much shifting, rather than sedentary cultivation. The principal commercial products are copra, corn and abacá (Manila hemp). Both lumbering and mining are important industries with the largest sawmill at Bislig bay. Minerals include gold, iron ore and copper, but only gold was worked and its importance declined after 1950. Northeastern Surigao has one of the larger iron ore bodies of the world, but it is made up of discontinuous deposits and has both nickel and chromite impurities that have made it uneconomical to develop. (AN. C.)

SURINAM: *see* GUIANA: *Surinam.*

SURINAME, commercially the most important river in Dutch Guiana, South America, rises in the central forested area and flows 230 mi. to the Atlantic about 14 mi. below Paramaribo (*q.v.*), which stands on a shell reef on the west bank. Small ocean-going vessels may ascend about 37 mi. to Paranam and small motorboats about 100 mi. Beyond this are a number of rapids. There is an important source of bauxite at Paranam, 16 mi. upstream from Paramaribo. The most important tributaries are the Cottica and Commewijne which enter the estuary as a joint stream. A projected dam at Brokopondo will produce 150,000 kw. of electricity. (G. LN.)

SURINAM TOAD (*Pipa pipa*), a tongueless, tailless Amphibian (*q.v.*). It inhabits South America east of the Andes and north of the Amazon and is strictly aquatic, feeding largely on fish in quiet pools. It has an extremely flattened head; the snout and the angles of the jaws bear several loose skin flaps, the fingers of the forelimbs terminate in small star-shaped appendages and the eyes are minute and without lids. The hind limbs end in large fully webbed feet.

The eggs are carried on the back by the mother; the skin thickens and grows round the eggs until each is enclosed in a dermal cell, which is finally covered by a horny lid. The eggs may number about 100 and measure 5–7 mm. in diameter. They develop entirely within these pouches, and the young hop out as fully developed miniatures of the adult. Pairing takes place in the water, the male clasping the female round the waist. During oviposition the female extends from the cloaca a long bladderlike pouch, which the male carries forward between his breast and the back of the female. As each egg is released it is fertilized and attached by a natural adhesive to the female's back.

SURMA or BARAK, a river of Assam, India, and East Pakistan. It rises in the Barail range to the north of Manipur, its sources being among the southern spurs of Japvo. Thence its course is south with a slight westerly bearing, through the Manipur hills. The name Barak is given to the upper part of the river, in Manipur and Cachar. A short distance below Badarpur in Cachar it divides into two branches. The northern, which passes Sylhet, is called the Surma. The southern, which is called the Kusiara, subdivides into two branches, one called Bibiana or Kalni, and the other the Barak, both of which rejoin the Surma. At Bhairab Bazar in Mymensingh the Surma unites with the old channel of the Brahmaputra and becomes known as the Meghna until it joins the Ganges between Narayanganj and Chandpur.

The river is navigable by steamers as far as Silchar in the rains. The total length is about 560 mi.

SURPLICE, a liturgical vestment of the Christian church. It is a loose tunic of white linen or cotton material, with wide sleeves, worn as a choir vestment in the Roman Catholic Church (in which it is proper to all clerics) and as a eucharistic vestment in the Anglican Church and in some Lutheran and Free churches. The surplice appeared first in the middle ages, and it has no counterpart in the Eastern Churches. *See* also VESTMENTS, ECCLESIASTICAL.

SURPLUS WAR PROPERTY DISPOSAL (U.S.). Surplus war property disposal was first recognized in the U.S. as a special function in 1933 when the procurement division of the treasury department was assigned responsibility for disposing of the surplus property of World War I and even some property dating from previous wars. During World War II a virtually endless assortment of property of almost every description began to accumulate as surplus as the war developed and war requirements rapidly changed. The Surplus Property act of 1944, as amended in 1945, established a Surplus Property administration headed by a single administrator (previously a three-member board administered the act). An executive order effective Feb. 1, 1946, transferred the SPA's function to the War Assets corporation. Authority over surplus property abroad, however, became the responsibility of the state department. The War Assets administration, which succeeded the WAC on March 25, 1946, handled the entire disposal program for more than 90% of all surplus. Of the $27,100,000,000 worth of war surplus property declared to the War Assets administration, $25,600,000,000 had been disposed of by Oct. 1, 1949. In that year the WAA was abolished and its functions were transferred to the General Services administration for the express purpose of liquidation. (H. E. BY.; X.)

SURREALISM, a widespread revolutionary movement in the arts, promoted by the poet André Breton. The word itself was coined by the poet Guillaume Apollinaire, and the movement had its heyday, primarily in Paris, from 1918 to 1939, attacking tradition in all forms. It was an assertion of the belief that the world of the unconscious mind—as expressed by fantasies and dreams—has a reality superior to that of the phenomenal world. Surrealist expression aimed at a creative integration of both realities into a "super-reality." The theory owed much to the research of Sigmund Freud (*q.v.*), whose technique of probing the unconscious mind for the ultimate truth of human behaviour compares with the surrealist method of penetrating the depths of human personality to tap the inexhaustible sources of the imagination. In *Manifeste du surréalisme,* published in 1924 in Paris, Breton's own definition of surrealism was: "Pure psychic automatism, by which it is intended to express whether verbally or in writing, or in any other way, the real process of thought. It is the dictation of thought, free from any control by the reason and of any aesthetic or moral preoccupation."

Submission to mental promptings, appropriate as means of spontaneous writing, could apply only to the initial, conceptual phase of a painting. Since the movement was largely a literary expression, surrealist painting was, therefore, an illustrative art based on literary themes, using realistic and recognizable means and expressions. In abandoning rational, moral and aesthetic considerations, it echoed Dadaism (*q.v.*); but it was dissimilar in having a positive intention, namely to create a new, authentic vision of life by releasing dormant sensibilities. The release found infinitely varied expression in a fantastic art which, unprecedented as a movement, had affinities with the weird images of individual painters like Hieronymus Bosch (*q.v.*) or Giuseppe Arcimboldo. The emergence of surrealism in art was assisted directly by such notable pioneers as Kurt Schwitters, whose collages established strange relationships between disparate objects in juxtaposition; Marcel Duchamp (*q.v.*), whose "ready-mades" revealed the extraordinary significance of objects in abnormal contexts; and indirectly by Giorgio de Chirico, Paul Klee and Pablo Picasso (*qq.v.*). Exponents of surrealist art were Hans Arp, Salvador Dali (*q.v.*), Paul Delvaux, Jean Dubuffet, Matta Echaurren, Max Ernst, Morris Graves, René Magritte, André Masson, Joan Miró (*q.v.*), Francis Picabia (*q.v.*) and Yves Tanguy.

Surrealist expression also found outlet in literature (Jean Cocteau, Paul Éluard, Guillaume Apollinaire [*qq.v.*], and Louis Aragon), in sculpture (Alberto Giacometti) and in photography (Man Ray). The spirit of surrealism is also to be seen in advertising techniques where objects apparently unrelated are juxtapositioned.

BIBLIOGRAPHY.—A. Breton, *Manifeste du Surréalisme* (1924), *Poisson soluble* (1924) and *Le Surréalisme et la peinture* (1928); Alfred H. Barr, ed., *Fantastic Art, Dada, Surrealism* (1936); Alfred Schmeller, *Surrealism* (1956); M. Nadeau, *Histoire du Surréalisme* (1945); Y. Duplessis, *Le Surréalisme* (1950); D. Gascoyne, *A Short Survey of Surrealism* (1935). (F. W. W.-S.)

SURRENDER (IN LAW), a mode of alienation of real estate. It is defined by Lord Coke to be "the yielding up of an estate for life or years to him that hath an immediate estate in reversion or remainder" (Coke upon Littleton, 337 b). It is the converse of release, which is a conveyance by the reversioner or remainderman to the tenant of the particular estate. A surrender was the usual means of effecting the alienation of copyholds (*q.v.*). A surrender was required, after the Real Property act, 1845, to be by deed, except in the case of copyholds and of surrender by operation of law.

In Scots law surrender of a lease is represented by renunciation. The nearest approach to surrender of a copyhold is resignation *in remanentiam* (to the overlord) or resignation *in favorem* (to a purchaser). These modes of conveyance were practically superseded by the Conveyancing act, 1874.

In the U.S. surrender divides itself into express surrender and surrender by operation of law. The former is the surrender known to the early common law; the latter takes place when the tenant and the reversioner have performed acts inconsistent with the continued existence of the two former distinct estates, from which acts the law implies that a surrender has been made. Thus where there is an acceptance by the tenant of a new lease to begin at any time during the existence of the previous lease, or the relinquishment of possession by the tenant and the resumption of possession by the landlord, a surrender by operation of law is deemed to have taken place.

SURRENDER (IN WAR) is identical in international law with capitulation and refers to agreements between commanders of belligerent forces for yielding a body of troops, a fortress, a district or a whole country. Surrenders, as opposed to armistices, are necessarily unequal (*see* ARMISTICE). According to the American Rules of Land Warfare (1940, art. 247) "capitulations should contain in precise terms every condition to be observed on either side, except such conditions as are clearly implied by the laws of war." Capitulations usually specify the precise moment when hostilities shall cease, the fate of the persons surrendered, the disposition of arms, the evacuation of territory, the care of the sick and wounded and the taking over of civil government.

Unconditional surrenders, such as that after the American Civil War of the Confederate forces under Gen. Robert E. Lee at Appomattox Court House on April 9, 1865, and of the German and Japanese forces after World War II, specify the moment at which hostilities shall cease, but, as they leave the victor wide discretion, establish unilateral powers rather than contractual relations. If applicable to all enemy forces they terminate the war and entitle the victor to deal with the enemy's territory, not as occupied under the limitations of the law of war, but as subjugated and capable of

annexation. Following the unconditional surrender of the German high command on May 8, 1945, the four Allied governments by a declaration of June 5, 1945, "assumed supreme authority in respect to Germany, including all the powers possessed by the German Government, the High Command and any state, municipal or local government or authority," but it was added that this "does not effect the annexation of Germany." "The Unconditional Surrender of all Japanese forces" demanded by the Potsdam declaration of July 26, 1945, was effected by Japanese acceptance of this declaration on Aug. 14, 1945, and the signature of a surrender instrument on Sept. 2, 1945. The Japanese emperor was permitted to govern "subject to the Supreme Commander of the Allied Powers." (Q. W.)

SURREY, EARLDOM OF. This earldom is of uncertain but unquestionably early origin. A Norman count, William de Warenne (*c.* 1030–88), is generally regarded as its first holder and is thought to have been made an earl by William II about 1088. William and his successors were styled earls of Surrey or Earls Warenne indifferently, and the family became extinct when William, the 3rd earl, died in 1148. The second family to hold the earldom of Surrey was descended from Isabel de Warenne (d. 1199), daughter and heiress of Earl William, and her second husband, Hamelin Plantagenet (d. 1202), an illegitimate half-brother of King Henry II. Their descendants held the earldom until Earl John died without legitimate issue in 1347.

The earldom and estates of the Warennes now passed to John's nephew, Richard Fitzalan, earl of Arundel (*c.* 1313–76), being forfeited when Richard's son, Richard, was beheaded for treason in 1397. Then for about two years there was a duke of Surrey, the title being borne by Thomas Holand, earl of Kent (1374–1400), from 1397 until his degradation in 1399. In 1400 Richard Fitzalan's son, Sir Thomas Fitzalan (1381–1415), was restored to his father's honours and became earl of Arundel and earl of Surrey, but the latter earldom reverted to the Crown when he died. In 1451 John Mowbray (1444–1476), afterwards duke of Norfolk, was created earl of Surrey, but the title died with him.

The long connection of the Howards with the earldom of Surrey began in 1483 when Thomas Howard, afterwards duke of Norfolk, was created earl of Surrey. Since then, with brief exceptions, the title has been borne by the duke of Norfolk.

See the articles NORFOLK, EARLS AND DUKES OF; and ARUNDEL, EARLS OF; also G. E. C.(ockayne), *Complete Peerage*, vol. vii. (1896).

SURREY, HENRY HOWARD, EARL OF (1517?–1547), English poet, son of Lord Thomas Howard, afterwards 3rd duke of Norfolk, and his wife Elizabeth Stafford, daughter of the duke of Buckingham, was born probably in 1517. He succeeded to the courtesy title of earl of Surrey in 1524, when his father became duke of Norfolk. His early years were spent in the various houses of the Howards, chiefly at Kenninghall, Norfolk; he used also to stay at Windsor with young Henry Fitzroy, duke of Richmond, of whom his father was guardian. Anne Boleyn tried to arrange a marriage between the princess Mary and her kinsman, Surrey. The Spanish ambassador, in the hope of detaching the duke of Norfolk's interest from Anne Boleyn in favour of Catherine of Aragon, seems to have been inclined to favour the project, but Anne changed her mind, and as early as October 1530 arranged a marriage for Surrey with Lady Frances de Vere, daughter of the 15th earl of Oxford. This was concluded at the earliest possible date, in February 1532, but in consequence of the extreme youth of the contracting parties, Frances did not join her husband until 1535. In October Surrey accompanied Henry VIII. to Boulogne to meet Francis I., and, rejoining the duke of Richmond at Calais, he proceeded with him to the French court, where the two Englishmen were lodged with the French royal princes. Surrey created for himself a reputation for wisdom, soberness and good learning. Meanwhile in spite of his marriage with Frances de Vere, the project of a contract between him and the princess Mary was revived in a correspondence between the pope and the emperor, but rejected by the latter. Surrey returned to England in the autumn of 1533, when the duke of Richmond was recalled to marry his friend's sister, Mary Howard. Surrey made his home at his father's house of Kenninghall, and took his father's side in a dispute which led to a separation between the duke and duchess.

In May 1536 he filled his father's functions of earl marshal at the trial of his cousins Anne Boleyn and Lord Rochford. In the autumn of that year he took part with his father in the bloodless campaign against the rebels in Yorkshire and Lincolnshire, in the "Pilgrimage of Grace." Hasty in temper, and by no means friendly to the Seymour faction at court, Surrey struck a man who insinuated that he had secretly favoured the insurgents. For breaking the peace in the king's domain he was arrested (1537), but thanks to Cromwell, he was merely sent to reside for a time at Windsor. During this retirement he had leisure to devote himself to poetry. In 1539 he was again received into favour. In May 1540 he was one of the champions in the jousts celebrated at court.

The fall of Thomas Cromwell a month later increased the power of the Howards, and in August Henry VIII. married Surrey's cousin, Catherine Howard. Surrey was knighted early in 1541, and soon after he received the Garter and was made chancellor of the duchy of Lancaster. In 1542 he was imprisoned again for quarrelling, but was soon liberated. Shortly after his release he joined his father on the Scottish expedition. They laid waste the country, but retreated before the earl of Huntly, taking no part in the victorious operations that led up to Solway Moss.

Surrey's ties with Wyat, who was fifteen years his elder and of opposite politics, seem to have been rather literary than personal. He appears to have entered into closer relations with the younger Wyat, with whom he got into trouble for breaking the windows of the citizens of London on Feb. 2, 1543. In prison for this offence he probably wrote the satire on the city of London, in which he explains his escapade by a desire to rouse Londoners to a sense of their wickedness. In October he joined the English army co-operating with the imperial forces in Flanders, and in the campaign of the next year he served as field marshal under his father, and took part in the unsuccessful siege of Montreuil. In August 1545 he was again serving in France.

Surrey had always been an enemy to the Seymours, whom he regarded as upstarts, and he thwarted the proposed marriage of his sister, the duchess of Richmond, with Sir Thomas Seymour. He thus increased the enmity of the Seymours and added his sister to the already long list of the enemies which he had made by his haughty manner and brutal frankness. He was now accused of quartering with his own the arms of Edward the Confessor. The charge was a pretext covering graver suspicions. Surrey had declared that his father, the duke of Norfolk, as the premier duke in England, had the obvious right of acting as regent to Prince Edward. He also boasted of what he would do when his father had attained that position. This boast was magnified into a plot on the part of his father and himself to murder the king and the prince. The duke of Norfolk and his son were sent to the Tower on Dec. 12, 1546. The duchess of Richmond was one of the witnesses (*see* her depositions in Herbert of Cherbury, *Life and Reign of Henry VIII.*, 1649) against her brother, but her statements added nothing to the formal indictment. On Jan. 13, 1547 Surrey defended himself at the Guildhall on the charge of high treason for having illegally made use of the arms of Edward the Confessor, before hostile judges. He was condemned by a jury, packed for the occasion, to be hanged, drawn and quartered at Tyburn. This sentence was commuted to beheading. Surrey was executed on Tower Hill on the 19th of the month.

Surrey's name has been long connetted with the "Fair Geraldine," to whom his love poems were supposed to be addressed. The story is founded on the romantic fiction of Thomas Nashe, *The Unfortunate Traveller, or Life of Jack Wilton* (1594), according to which Surrey saw in a magic glass in the Netherlands the face of Geraldine, and then travelled throughout Europe challenging all comers to deny in full field the charms of the lady. At Florence he held a tournament in her honour, and was to do the same in other Italian cities when he was recalled by order of Henry VIII. The legend, deprived of its more glaring discrepancies with Surrey's life, was revived in Michael Drayton's *England's Heroicall Epistles* (1598). Geraldine was the daughter of the earl of Kildare, Lady Elizabeth Fitzgerald, who was brought up at the English court in company with the princess Elizabeth. (*See* James

Graves, a *Brief Memoir of Lady Elizabeth Fitzgerald*, 1874.) She was ten years old when in 1537 Surrey addressed to her the sonnet "From Tuskane came my ladies worthy race," and nothing more than a passing admiration of the child and an imaginative anticipation of her beauty can be attributed to Surrey.

His poems, which were the occupation of the leisure moments of his short and crowded life, were first printed in *Songs and Sonettes written by the ryght honorable Lorde Henry Howard late Earle of Surrey, and other* (apud Richardum Tottel, 1557). A second edition followed in July 1557 and others in 1559, 1565, 1567, 1574, 1585 and 1587. Although Surrey's name, probably because of his rank, stands first on the title-page, Wyat was the earlier in point of time of Henry's "courtly makers." Surrey, indeed, expressly acknowledges Wyat as his master in poetry. His sonnets, his elegy on Wyat and his lyrics served as models to generations of court poets.

As their poems appeared in one volume, long after the death of both, their names will always be closely associated. Surrey's contributions are distinguished by their impetuous eloquence and sweetness, and he introduced new smoothness and fluency into English verse. His chief innovation as a metrician lies outside the *Miscellany*. His translation of the second and fourth books of the *Aeneid* into blank verse—the first attempt at blank verse in English—was published separately by Tottel in the same year *Certain Bokes of Virgiles Aeneis turned into English meter*.

BIBLIOGRAPHY.—E. Arber's reprint of *Songs and Sonettes, English Reprints* (1870); the Roxburghe Club reprint of *Certain Bokes of Virgiles Aeneis* (1814); G. F. Nott, *The Works of Henry Howard, Earl of Surrey* (1815); and *The Poetical Works of Henry Howard, Earl of Surrey*, Aldine edition (1866). The best account of Surrey's life is in Edmond Bapst's *Deux Gentilhommes-poëtes de la cour de Henry VIII* (1891), which rectifies G. F. Nott's memoir in many points. *See* also Brewer and Gairdner, *Letters and State Papers of Henry VIII*; Lord Herbert of Cherbury, *Life and Raigne of Kinge Henry the Eighth* (1649); J. A. Froude, *History of England*, ch. xxi and xxii; W. J. Courthope, *History of English Poetry*, vol. ii, ch. iii (1897), where the extent and value of Surrey's innovations in English poetry are estimated; F. M. Padelford, *The MS. Poems of Henry Howard, Earl of Surrey* (1906); O. Fest, "Über Surreys Virgilübersetzung," in *Palästra*, vol. xxxiv (1903).

SURREY, a southeastern county of England, bounded north by the Thames (separating it from Buckinghamshire and Middlesex), east by Kent, south by Sussex, west by Hampshire and northwest by Berkshire. The administrative county of London bounds Surrey on the northeast. Under the London Government Act 1963, Middlesex ceases to exist and the administrative county of London, with additional areas, becomes Greater London on April 1, 1965. On the same date Surrey gains the urban districts of Staines and Sunbury-on-Thames from Middlesex and loses to Greater London the boroughs of Barnes, Beddington and Wallington, Kingston-upon-Thames, Malden and Coombe, Mitcham, Richmond (which includes the Royal Botanic gardens at Kew), Surbiton, Sutton and Cheam, and Wimbledon, and the urban districts of Carshalton, Coulsdon and Purley, and Merton and Morden. Croydon (*q.v.*), the only county borough in Surrey, also becomes part of Greater London. (*See* LONDON.)

This article is concerned with the county of Surrey before the implementation of the above act. The geographical area of the county (including Croydon) is 721.6 sq.mi.

The north of the county adjoining the Thames is low lying and the soil consists largely of brick earth overlying London Clay. To the west, on the edge of the Thames basin, Bagshot sands outcrop at Chobham common, and Bagshot heath and Chobham ridges (400 ft.) are capped by plateau gravels. Across the middle of the county, from east to west, run the chalk North Downs. In the east they form a wide band, rising to a height of 852 ft., but westward they narrow into the Hog's Back. The line of the Downs is broken by the river Mole between Dorking and Leatherhead, and by the Wey at Guildford. These are the county's two chief rivers, and they join the Thames at Molesey and Weybridge respectively. To the south of the chalk Downs is a wide band of Lower Greensand, which narrows eastward. On this greensand is found the highest land in the county, Leith hill (965 ft.) near the centre, and Hindhead (895 ft.) in the west. The rest of the county, except for a small triangle of Hastings beds in the extreme southeast, is composed of Weald Clay. The National Trust owned 11,443 ac. in Surrey and protected 556 ac. in 1962. In 1958 about 160 sq.mi. in the Surrey hills were established as "an area of outstanding natural beauty" under the National Parks and Access to the Countryside act, 1949. At Wisley are the gardens of the Royal Horticultural society.

History.—Paleolithic implements were found in the gravel beds of the Wey, the Mole, the Wandle and the Thames (for example, near Farnham, Guildford, Godalming, Mortlake and Wandsworth), and at older gravel plateaus such as Warlingham and Wimbledon common. Here too, as at Blackheath near Albury, there is evidence of Neolithic man. Hoards of bronze implements have been found at a number of places. Iron Age earthworks cover a wide area and include mounds at Weybridge, Farnham, Holmbury and one, the largest of all, near Lingfield.

It was an area of forest, swamp, heath and open downland through which the Romans passed at the time of the invasion of Julius Caesar (55 B.C.), using perhaps the ancient track along the Downs, later known as the Pilgrims' way. Romanization, however, did not come until the time of the emperor Claudius (A.D. 43). For the next four centuries Surrey was ruled by the Romans and though Roman remains are relatively few, the largest settlements appear to have been at Southwark, Kingston-upon-Thames, Farley Heath near Albury and Woodcote near Croydon. Remains of about two dozen villas lie, for the most part, along the northern and southern edges of the Downs, and include that of Ashtead to the north and Titsey to the south.

The Thames was bridged at Southwark as part of Watling street, leading from Dover and Canterbury. Stane street, from Chichester to London, was probably not constructed till the 3rd century; it ran via Ockley, Dorking and Merton. No Roman town survived the Saxon settlement of Surrey, and there are few place names of Roman origin. Surrey was never an independent Saxon kingdom. The name itself means southern district. Nearly all place names are of Saxon origin, a large proportion pagan, such as Friday Street (commemorating the goddess Frea [Frigg]). The scantiness of the population is indicated by the small number of hundreds (14) compared with neighbouring counties. The tithings into which they were divided generally became parishes after the establishment of Christianity. Coronations of West Saxon kings at Kingston during the 10th century are referred to by chroniclers. A coronation stone is preserved there.

During the reign of Edward the Confessor, Godwin, earl of Wessex, and members of his family, including Harold, held many manors in Surrey. After the battle of Hastings (1066) William I took these for himself or for his Norman followers. Held back in Southwark from entering London, he turned southward to the Pilgrims' way. Domesday Book shows the extent of the ravaging of the Normans through Limpsfield, Godstone and Gatton, and then westward through Albury, Shalford and Farnham, along the road to Winchester. The bishop of Winchester retained Farnham.

Chertsey abbey, according to the 13th-century cartulary, was founded in 666 by St. Erkenwald who became its first abbot. During the Danish wars (*c.* 871) the abbot and monks were killed and the abbey burned. Refounded for Benedictines in the 10th century it maintained a prominent position and held many manors, including Chertsey, Egham, Thorpe, Chobham, Cobham, Epsom, Sutton and Coulsdon, until it was dissolved in 1537. Later foundations include the Cluniac priory of Bermondsey (1082), the Cistercian abbey of Waverley (1128), the Augustine priories of Merton (founded 1117), Newark (*c.* 1190), Tandridge (*c.* 1200) and Reigate (1220).

Even in the 13th and 14th centuries there probably was a strong contrast between the outlying forests and the suburban parts of Surrey. Southwark, Lambeth and Bermondsey with their great suburban houses of ecclesiastical dignitaries, sharing in the life of London. Kingston, with its bridge over the Thames, was a market town, as was Guildford with its manufacture of cloth. As elsewhere the ravages of the Black Death intensified commercial depression and religious agitation, which led to the Peasants' Revolt of 1381.

Sheen already had royal importance when Richard II pulled down the palace after the death of his first queen. Henry V rebuilt it, founding at the same time a Carthusian house. Following the insurrections of the 15th century Surrey under the Tudors became a county of royal residences. Henry VII rebuilt the palace of Sheen, renaming it Richmond. Henry VIII and Elizabeth I resided there, and Henry VIII received Hampton Court, just across the Thames, from Cardinal Wolsey. This brought court officials to the villages on the Surrey side. After the dissolution of the monasteries (1536–39) Henry began to build a flamboyant hunting lodge in Surrey, Nonsuch (near Ewell). This was built partly of materials from the dissolved priory of Merton and on the site of the village of Cuddington which he destroyed. Stones from Waverley abbey were used in building Loseley house, and Oatlands palace (Weybridge), now in ruins, had stones from Chertsey abbey. The manors of the religious houses fell into the king's hands or were sold by him or his successors. Reigate priory was granted to William Howard (later Lord Howard of Effingham; *see* HOWARD); and his son, who was commander in chief against the Spanish Armada (1588), is buried in Reigate church. No reference to the Tudors, and in particular to the reign of Elizabeth I, could ignore the theatres of Southwark, more especially Shakespeare's Globe.

In the middle ages sheep were important in rural Surrey and by the 16th century a cloth trade was growing, for example, at Guildford, Godalming and Farnham. Fuller's earth was already known at Nutfield and until the 19th century was used mainly in the woolen industry. In the 16th century iron smelting with charcoal in the Weald areas was further developed because of the demand for cannons. Charcoal was also needed in the manufacture of gunpowder. Powder mills were established before 1570 at Rotherhithe, Wotton, Albury and Chilworth (the latter closed only during the war of 1914–18), and at Godstone in 1589. Glassmaking at Chiddingfold flourished until the beginning of the 20th century; it may have begun during the Roman occupation.

Surrey, like London, was overwhelmingly parliamentarian during the Civil War. It was not a major battle ground, though there was fighting in the county, for example during the royalist reversals leading to the stand at Kingston common (July 7, 1648).

By 1653 the canalization of the Wey from Guildford to the Thames, planned by Sir Richard Weston of Sutton Place, was completed. This facilitated the transport to London of timber, grain, iron, gunpowder and the produce of market gardens. Daniel Defoe in the early 18th century described Farnham as the greatest corn market in England, London excepted, grain being brought along the Wey to be ground and then sent onward to London by water.

Surrey quarter sessions records which survive from 1659–60 contain numerous references to Quaker meetings and religious dissent generally. At the July sessions, 1665, during the plague, it was ordered that the wells at Epsom be locked for the summer to prevent strangers coming from London on pretext of drinking the waters, thereby carrying disease along the roads.

Samuel Pepys and John Evelyn, both Surrey men, have many references to the county in their *Diaries*, including Epsom at its height as a spa. It was then that horse racing on the Downs became fashionable. It was already possible to go from London for the day. John Evelyn was also a great gardener, expert on trees and pioneer planter of conifers.

In Surrey much land was enclosed before the 18th century. Market gardens were a feature in the west and along the Thames, as at Mortlake, at Mitcham, with its lavender and physic gardens, in Bermondsey and Camberwell. These gardens gave place to housing developments. Godalming and Guildford declined as industries moved to the Midlands or to Yorkshire. Maps of the 18th and 19th centuries show copper and iron, oil, grain, paper, snuff, and calico printing mills, mainly along the Wandle. Elsewhere were chalk, gravel and stone quarries, brickworks and lime kilns.

The middle decades of the 18th century saw the beginning of improved roads in Surrey, under turnpike trusts. Navigations included the improved canalization of the Wey as far as Godalming (1762); the Basingstoke canal (1796) originally constructed to link the industrial town of Basingstoke with the port of London; and the Wey and Arun canal (1816). The development of railways brought to an end the turnpike road system and canal building. The Surrey Iron railway, from Wandsworth to Croydon and then to Merstham, worked by horses and intended to convey materials such as chalk and stone to the Thames, was the first public railway sanctioned by parliament (1801). Freight was carried on payment of a toll. The first railway for steam trains to pass through Surrey was the London and Southampton, opened as far as Woking in 1838. This was followed by the London and Brighton in 1841. The railways brought the county, particularly north of the Downs, continually nearer to London and resulted in intense house building, as at Wimbledon, or at Caterham after the building of the Caterham Valley railway, sanctioned in 1854.

William Cobbett (born at Farnham, 1763) in his *Rural Rides* observes political and social conditions in the 1820s. These must be viewed against the background of the wars with France and shortage of wheat. A wet summer and poor harvest helped to intensify the crisis of the autumn of 1830. In 1848 the Chartist rising ended at Kennington common.

Until 1889 Surrey stretched along the south bank of the Thames as far east as Rotherhithe and embraced ancient parishes now in London, including Wandsworth, Lambeth, Camberwell and Southwark. These formed part of the hundred of Brixton. The Reform act of 1832 disfranchised three boroughs and divided the county into two divisions, each returning two members. Lambeth became a borough and was extended. Rotherhithe and Bermondsey were included in the borough of Southwark. Further changes were made by the Reform acts of 1867 and 1885. By the second act the boroughs of the northeast, including Southwark, Lambeth and Croydon, were divided into 17 single-member districts. Apart from Croydon (which became a county borough) these boroughs became in 1889 part of the new county of London and a new phase in the history of the county, with its administrative headquarters at Kingston, began.

Population and Administration.—At the 1961 census the population of the administrative county was 1,478,541. The population of the 13 boroughs was as follows: Barnes, 39,057; Beddington and Wallington, 32,603; Epsom and Ewell, 71,159; Godalming, 15,780; Guildford, 53,976; Kingston-upon-Thames, 36,461; Malden and Coombe, 46,572; Mitcham, 63,690; Reigate, 53,751; Richmond, 41,024; Surbiton, 62,977; Sutton and Cheam, 79,008; Wimbledon, 57,312. Of the 15 urban districts five had a population exceeding 50,000: Carshalton, 57,484; Coulsdon and Purley, 74,926; Esher, 60,610; Merton and Morden, 68,011 and Woking, 67,519.

The headquarters of the Surrey county council are at County hall, Kingston-upon-Thames. There were prior to the formation of Greater London in April 1965 13 boroughs (named above); 15 urban districts; and 5 rural districts containing 69 civil parishes. Assizes and quarter sessions are both held at County hall. Guildford and Croydon have separate sessions. There are magistrates' courts for 12 petty sessional divisions and also for the boroughs of Guildford, Kingston, Reigate and Richmond. The northeast of the county (as far west as Esher) is in the metropolitan police district. The Surrey constabulary, covering the rest of the county, is divided into eight divisions. The parliamentary constituencies in the administrative county under the Representation of the People act (1948), as amended, consist of 10 county constituencies and 7 borough constituencies.

Until the 19th century the county lay in the diocese of Winchester, apart from those parishes belonging to Canterbury. In 1905 the diocese of Southwark was created. Parishes which had previously been transferred to the see of London and those which in 1877 had been transferred to Rochester, were transferred to Southwark. In 1927 the west of the county passed to the new diocese of Guildford. Its cathedral, begun in 1936 and consecrated in 1961, stands on Stag hill. It was designed by Sir Edward Maufe and contains sculpture by Eric Gill (d. 1940). Croydon remains in the diocese of Canterbury. The development of the county has been greatly influenced by its proximity to London,

and its good transport facilities have made it attractive both as a residential and recreational area.

There are racecourses in the county at Epsom Downs, where "the Derby" is run, Sandown Park (Esher) and Lingfield Park; the All England Lawn Tennis club has its headquarters at Wimbledon; Bisley is a centre for rifle shooting; the headquarters of the Surrey County Cricket club is the Oval in Kennington, London, but some matches are played at Guildford.

Industries and Communications.—The total acreage under crops and grass in the early 1960s exceeded 180,000 ac. of which more than 110,000 were arable land. This represents about 25% of the total area (excluding water) of the county, compared with about 9% in 1939. Wheat and oats are the chief grain crops. The cattle are mainly of the dairy type. There are extensive market and nursery gardens in the fertile brick-earth plain beside the Thames and in the Bracklesham beds at Chobham and Bagshot.

The level land adjoining the main roads and railways attracted groups of miscellaneous industries along the Kingston bypass near Raynes Park and Tolworth, and also at Addlestone and Salfords. More recently there has been industrial development at Camberley. Kingston-upon-Thames and Weybridge are important centres of general engineering and aircraft industry, while miscellaneous industries in the Wandle valley at Mitcham, Merton and Hackbridge are the modern development of the old watermill industries for which the Wandle was once well known.

The main roads from London to Penzance, Portsmouth, Brighton and the south coast pass through Surrey which is served also by the Southern region main and suburban lines of British railways. London Transport underground extends to Morden, Wimbledon and Richmond. Bus and coach services cross the county. The civil airport at Gatwick was opened in 1958.

BIBLIOGRAPHY.—J. Aubrey, *Natural History and Antiquities of the County of Surrey,* 5 vol. (1719); D. Lysons, *The Environs of London,* vol. i (1792–96); O. Manning and W. Bray, *The History and Antiquities of the County of Surrey,* 3 vol. (1804–14); also, in the British Museum, London, a 30 vol. edition, expanded by more than 6,000 illustrations, maps, etc., entitled *The History and Topography of the County of Surrey* (1847); Surrey Archaeological Society, *Surrey Archaeological Collections* (1858 *et seq.*); *The Domesday Book for the County of Surrey* (1876); E. W. Brayley, *A Topographical History of Surrey,* 4 vol., ed. and rev. by E. Walford (1878–81); *Victoria County History of Surrey,* 4 vol. (1902–14); J. E. B. Gover *et al., The Place-names of Surrey* (1934); F. J. C. Hearnshaw, *The Place of Surrey in the History of England* (1936); C. D. Hawley (ed.), *List of Antiquities in the Administrative County of Surrey* (1951); Land Utilisation Survey of Britain, *The Land of Britain,* ed. by L. D. Stamp, pt. 81 (1941); Surrey County Council, *Surrey Development Plan* (1953); *Publications* of the Surrey Record Society (1913 *et seq.*); *The Diary of John Evelyn,* ed. E. S. de Beer (1955); W. Cobbett, *Rural Rides in the Counties of Surrey . . . ,* ed. by E. W. Martin (1958); *Census 1961 England and Wales: County Report, Surrey.* (W. W. R.; M. Gz.)

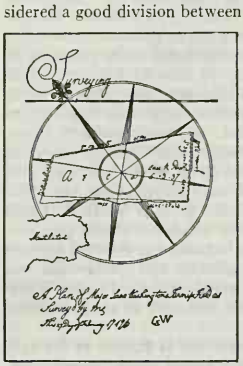
SURROGATE, a deputy of a bishop or an ecclesiastical judge, acting in the absence of his principal and strictly bound by the authority of the latter (Lat. *surrogare,* "to substitute for"). Canon 128 of the canons of 1603 lays down the qualifications necessary for the office of surrogate, and canon 123 the regulations for the appointment to the office. The office is of some importance in the United States as denoting the judge to whom the jurisdiction of the probate of wills, the grant of administration and of guardianship is confided.

SURTEES, ROBERT SMITH (1803–1864), English novelist and sports writer, was the second son of Anthony Surtees of Hamsterley Hall, Durham. Educated to be a solicitor, Surtees soon began to contribute to the *Sporting Magazine,* and in 1831 he published a treatise on the law relating to horses and particularly the law of warranty, entitled *The Horseman's Manual.* In the following year he helped to found the *New Sporting Magazine,* of which he was the editor for the next five years. To this periodical he contributed between 1832 and 1834 the papers which were afterward collected and published in 1838 as *Jorrocks's Jaunts and Jollities.* This humorous narrative of the sporting experiences of a cockney grocer, which suggested the more famous *Pickwick Papers* of Charles Dickens, is the work by which Surtees is chiefly remembered, though his novel *Handley Cross,* published in 1843, in which the character of "Jorrocks" is reintroduced as a master of foxhounds, also enjoyed a wide popularity. The former of these two books was illustrated by "Phiz" (H. K. Browne), and the latter, as well as most of Surtees's subsequent novels, by John Leech, whose pictures of "Jorrocks" are everywhere familiar and were the chief means of ensuring the lasting popularity of that humorous creation. Surtees wrote other novels, the last of which, *Mr. Facey Romford's Hounds* (1865), appeared after the author's death (March 16, 1864).

See R. S. Surtees, *Jorrocks's Jaunts and Jollities* (London, 1869), containing a biographical memoir of the author; W. P. Frith, *John Leech, His Life and Work* (2 vol., London, 1891); Samuel Halkett and J. Laing, *Dictionary of Anonymous and Pseudonymous Literature of Great Britain* (4 vol., Edinburgh, 1882–88).

SURVEYING is the science which has to do with the determination of the relative position or location of points on or near the earth's surface. This involves the art of measuring horizontal and vertical distances and of determining direction and angular relationships by the use of surveying instruments in the field. Extensive mathematical calculations are, quite often, a necessity along with the actual measurements. Information from the survey is generally portrayed graphically in the form of maps, charts, profiles, cross sections and other diagrams. Therefore, surveying in the field subsequently requires a considerable amount of computing and drafting.

Surveys may be classified in several different ways. Methods and instruments involved in surveying will vary considerably depending upon the magnitude and extent of the area and the purpose and use of the measurements. One method of differentiation involves a consideration of magnitude as to whether it is a plane or geodetic survey. The earth's surface which is very close to a sphere (actually a spheroid) has considerable curvature. In surveys of smaller magnitude involving limited areas or small parcels, the earth's surface is considered as a plane. In surveys of greater magnitude involving extended areas or surveys of whole states, countries or continents, account must be taken of this curvature. Special equipment and more precise equipment, procedures and calculations are involved. Sometimes an area of 100 sq.mi. is considered a good division between a plane survey and a geodetic survey. A second method of differentiation involves names which are descriptive of the particular function of the survey, such as land (or property) surveying, route surveying, topographic surveying, construction surveying, mine surveying, hydrographic surveying and geodetic (or control) surveying. These separate classifications will be discussed below. A third method of classification of surveys relates to the technical method or instruments used, such as transit-tape survey, stadia-azimuth survey, deflection angle survey, compass-chain survey, plane-table survey, trigonometric survey or photogrammetric survey. Precision denotes relative or apparent nearness to the truth, while accuracy denotes absolute nearness to the truth. The degree of accuracy is a ratio that expresses, in general terms, the accuracy attained or attainable in survey measurements. For example, a degree of accuracy of 1 in 5,000 indicates that there are errors in measurements amounting to one unit or less in each 5,000 units measured, or in the same proportion for a greater or lesser number of measured units. In horizontal distance measurement, the degree of accuracy is determined from the measured value and its probable error expressed as a ratio 1/distance. For a closed traverse it is usually expressed as the ratio of the linear error of closure to the total length of the traverse, $1/d$. For angular measurements in a traverse, it is usually expressed as the ratio of the angular closing error to the



BY COURTESY OF THE U.S. BUREAU OF RECLAMATION

FIG. 1.—A PLAN MADE BY GEORGE WASHINGTON IN 1746, INSCRIBED IN HIS OWN HANDWRITING

*Specifications for the Various Orders of Accuracy**

	1st Order	2nd Order	3rd Order
Triangulation:			
Base line length checks	1 in 25,000	1 in 10,000	1 in 5,000
Triangle closures, average	1 sec.	3 sec.	5 sec.
Triangle closures, maximum	3 sec.	5 sec.	10 sec.
Traverse:			
Closing error in position	1/25,000	1/10,000	1/5,000
Leveling:			
Closing checks	0.017 ft.$\sqrt{\text{mi.}}$	0.035 ft.$\sqrt{\text{mi.}}$	0.050 ft.$\sqrt{\text{mi.}}$
Closing checks	4 mm.$\sqrt{\text{km.}}$	8.4 mm.$\sqrt{\text{km.}}$	12 mm.$\sqrt{\text{km.}}$

*As approved by the U.S. Federal Board of Surveys and Maps, 1933.

square root of the number of angles, as $K\sqrt{n}$. For leveling it is expressed as the ratio of the error of closure in elevation to the square root of the horizontal length of the route in miles, as $K\sqrt{D}$. Prescribed standards have been established for various orders of accuracy as shown in the accompanying table. (K. S. Cs.)

History.—It is very probable that surveying had its origin in ancient Egypt. But long before the dynastic period in Egypt, the Neolithic man was probably able to make a rough kind of map based on his journeys, or a primitive plan to show tribal or property boundaries. Speculation is not needed. A plan of the villa of a great Egyptian noble was found in a Theban tomb of the 8th dynasty; in the tomb of one Menna at Thebes there is a representation on the walls of two chainmen surveying a field of corn; and in Ptolemaic and Roman papyri in the same country, measurements of plots of land are described. That the early Egyptians could carry out measurements with a considerable degree of accuracy is certain from a study of the dimensions of the Great Pyramid.

In Roman times the groma, which consists of two pairs of plumb lines suspended from the ends of two horizontal rods, at right angles to each other, was used to lay out lines at right angles. The metal parts of one of these gromas was found in 1912, in Pompeii. An early groma of the same type, but rougher construction, has been found in Egypt. The Romans also used 10-ft. rods, and bronze terminal pieces of such rods have been found at Enns, Aus., the foot in this case being 13.2 in. The Romans certainly made use of an instrument not unlike the plane table for determining the alignment of their roads. The Greeks used a form of log line for recording the distances run from point to point along the coast while making their slow voyage from the Indus to the Persian gulf about 325 B.C. Still earlier (as early as 1600 B.C.) it is said that the Chinese knew the value of the loadstone and possessed some form of magnetic compass. The earliest maps were based on inaccurate astronomical determination; not until medieval times, when the Arabs made use of the astrolabe, could nautical surveying really be said to begin. In 1450 the Arabs were acquainted with the use of the compass, and could make charts of the coast line of those countries which they visited. In 1498 Vasco da Gama saw a chart of the coast line of India, which was shown him by a Gujarati. Plane tables were in use in Europe in the 16th century and the principle of graphic triangulation and intersection was practiced by surveyors in England and elsewhere. In 1615 Willebrord Snell, the Dutch mathematician, measured an arc of meridian by instrumental triangulation.

The invention of printing and engraving in Europe in the 15th century and the age of the great discoveries which followed contributed much to the expansion of surveying and mapping activities. Modern developments in electronics, optics, photography, reproduction and aviation greatly influenced modern surveying. During World War II. the U.S. Army Map service and its affiliated organizations prepared more than 30,000 different maps and reproduced over 500,000,000 individual map sheets.

(C. F. A.-C.; H. St. J. W.; K. S. Cs.)

Horizontal Distance Measurement.—The distance between two points in surveying is understood to mean the horizontal distance. Frequently the slope distance is measured and is reduced mathematically to its horizontal equivalent before being used. Horizontal distances are reduced to sea level equivalent only in geodetic surveys. There are several methods of determining distances and the proper selection of the method to be used depends upon the cost, required precision and other considerations. The precision required on a reconnaissance would be quite different from that required on a horizontal control survey. The commoner methods of the pace, stadia, metallic tape, steel tape and invar tape are outlined below.

The method of pacing provides a rather rapid means of measuring, but yields results of a lower precision. The natural pace of an individual is usually about 27 to 30 in. and can be calibrated by pacing several times over a known distance. The natural pace is favoured over the artificial pace of three feet, especially over longer distances where the three-foot pace would be tiresome and hard to maintain. A passometer and pedometer are tally devices, similar to a watch in size, that can be strapped to the pacer and that register mechanically the number of paces taken or the distance directly in feet or miles. Distances obtained by pacing are not necessarily horizontal but are directly related to character of terrain. The usual precision of pacing is 1/100 to 1/200 and therefore is ideal for estimation, exploratory or reconnaissance work, rough mapping and checking other measurements.

The method of stadia is a rapid means of determining distance by an indirect method using the telescope of a surveying instrument and will be discussed under *Optical Distance Measuring Methods* below.

Tapes are made from several different materials, in various widths and lengths, and graduated in practically any manner desired. The most commonly used are metallic, steel and invar tapes. A metallic tape is made by weaving fine metal wires with cotton or linen threads into a tape which is waterproofed to prevent its stretching. It is usually $\frac{5}{8}$ in. wide, 50 to 100 ft. in length, and graduated in feet, tenths and half-tenths. Metallic tapes are principally used where a high degree of accuracy is not required and a light, flexible tape is desirable. Examples would be the cross sectioning for earthwork and the detailing for maps. The usual precision is 1/500 to 1/700. The steel tape is the most generally used method for direct linear measurement. It is made in a number of lengths, widths and thicknesses. The most commonly used length is 100 ft., but tapes are available in lengths of 50, 200, 300, 500 ft., in chain units and meter units. Most are $\frac{1}{4}$ or $\frac{5}{16}$ in. wide. Graduations in feet are generally feet and decimals. Steel tapes may be obtained with or without a reel and the usual precision of ordinary taping is 1/1,000 to 1/5,000; however, under proper conditions, precision taping can be accomplished at 1/10,000 to 1/30,000. Invar tapes are made of a nickel-steel alloy and therefore have a low coefficient of thermal expansion, approximately one-thirtieth that of steel. They are relatively soft and are easily kinked or broken. Invar tapes are used for the most accurate distance measurements made in surveying, such as base-line lengths for precise triangulation networks and city surveys. When properly used and adjusted, invar tape measurements can be made to a precision of 1/50,000 to 1/1,000,000.

Ordinarily taping on smooth level ground is accomplished by supporting the steel tape entirely on the ground, using taping pins as markers and counters, and using range poles to indicate the location of points and direction of lines. On uneven or sloping ground, the tape is held horizontal and a plumb bob is used to plumb vertically to the ground. Vertical positioning is accomplished by the use of a hand level. In precision taping, a number of precautions are followed and different techniques are employed. A number of factors affect the length of a tape and these factors can be either controlled or evaluated. These factors include standardization, temperature, slope, tension, sag, alignment, etc. A tape should be periodically compared with some permanent standard of length to determine its true length. The temperatures at which measured tape lengths are made are recorded and used in determining the temperature corrections. Spring balances are used to control the tension or pull which is applied and tension corrections are computed. Measurements are made on three-foot high stakes or bucks and vertical differences in elevations are obtained for computing slope corrections. The weight, length and tension of the tape are used for determining sag corrections. All of these corrections, when properly applied to the measured length of line, yield a measurement of high precision.

In years past, ordinary precision measurement was made with either an engineer's chain or a surveyor's (Gunter's) chain. The

engineer's chain was 100 ft. long and was composed of 100 links each 1 ft. long. The surveyor's chain was 66 ft. long and was composed of 100 links each 7.92 in. long. The surveyor's chain was a convenient unit of land measurement along with the rod, mile and acre units. Many modern land descriptions are still referenced with this unit. The terms chain and chaining have now mostly yielded to the terms tape and taping.

Other horizontal distance measuring devices include those that depend on observing the number of revolutions of a wheel. The mileage recorder on an automobile speedometer registers distances to 0.1 mi. with estimation possible to hundredths. A simple device for directly registering the number of revolutions of a wheel is an odometer and can be readily attached to any vehicle. When the circumference of the wheel is known, the relation between revolutions and distance is fixed.

Optical Distance Measuring Methods.—Distances measured optically depend upon the measurement of a length from a small base at one end and the angle subtended by that base at the other. The base may be fixed or variable and the subtended angle fixed or variable. The base may be at the observer's end or the far end and may be vertical or horizontal. Three examples of optical methods are stadia (tachymetry), subtense and range finder.

Stadia makes use of the telescope of the common surveying instrument. Two additional horizontal cross hairs are mounted on the reticule ring of a telescope. These are so spaced that the distance between the instrument and a distant point will be indicated by the intercept of the stadia hairs on a graduated rod held vertically on the point. Since the stadia hairs subtend a constant angle, the value of a stadia intercept is directly proportional to the distance between the observing instrument and the rod. The stadia interval factor is determined by the focal length of the telescope objective lens and the actual distance between stadia hairs and, for most modern instruments, is usually 100. On the external focusing-type telescope, an additional one-foot stadia constant must be applied. The usual precision of stadia is 1/300 to 1/1,000 and therefore useful in mapping, rough traverses and checking more precise measurements. On inclined sights, when the line of sight of the telescope is inclined upward or downward, the value of the vertical angle becomes a factor which is necessary to reduce stadia readings to horizontal and vertical distances. Stadia formulas, which have been derived, can be solved directly by special tables, by slide rules and by other special devices. A specially designed stadia arc is attached to some alidades and facilitates a rapid reduction of the stadia formula by eliminating the necessity of obtaining the vertical angle.

In the subtense method a bar or base of known length is arranged horizontally and the angle subtended is measured accurately by repetition procedures with a transit or theodolite. The subtense base may be of any convenient size. Sometimes signals are erected as ends of a subtense base as much as a quarter mile in length with the distance being taped. However, most subtense bars are specially designed instruments 6 to 20 ft. long. Subtense methods are mostly used in topographic and military surveying.

A range finder is an instrument consisting of a known base and arranged so that angles to a distant point are observed from the ends of the base and translated directly in distance units. The rangefinder principle has long been used in the military as a fire-control instrument.

Electronic Distance Measuring Methods.—With the advent of the electronic age, electronic methods were applied to the problems of distance measurement. The rapid development of radar during World War II as a navigational aid was quickly applied to peacetime uses in surveying long distances. Several of the radar systems which are of special note include shoran, loran, and Electronic Position Indicator (E.P.I.).

The geodimeter, a high-frequency light-signaling instrument, makes use of the speed of light in the determination of precise distances such as triangulation base lines longer than one mile. The distance is obtained indirectly by determining the time interval for a light beam to travel from the geodimeter to a distant plane mirror and return to the geodimeter.

The tellurometer is similar to the geodimeter but employs radio microwaves instead of light. Distances can be measured in daylight or darkness and in clear or foggy weather. The instrument is lightweight and portable.

Vertical Distance Measurement.—The difference in elevation between two points is the vertical distance between the two level surfaces in which the points lie. The elevation of a point is its vertical distance above or below an arbitrarily assumed level surface. The level surface used for reference is called the datum and the process of measuring vertical distances is called leveling. In considering the earth, a level surface is really a curved surface. Several techniques and methods are used including spirit leveling, trigonometric leveling and barometric leveling. Bench marks are fixed points of known elevation that are established at intervals throughout an area and provide permanent points of reference. Bench marks are usually referred to as mean sea level datum and are identified by metal caps set in concrete monuments, masonry walls, bridge abutments, rock outcrops or by spikes or large nails in poles or trees.

FIG. 2.—BENCH MARK TABLET USED IN THE UNITED STATES

In barometric leveling approximate differences in elevation can be determined by measuring the atmospheric pressure at the desired points using surveying altimeters. Atmospheric pressure varies inversely with the elevation of a point and thus these altimeters can be employed in making observations. However, barometric pressures do not remain constant and therefore precision measurements cannot be made. In order to detect these pressure changes, methods are used whereby observations are taken at a fixed station during the same period that other observations are made from point to point with a second altimeter. Barometric leveling is principally employed by reconnaissance or exploratory parties, especially in hilly or mountainous country.

Trigonometric leveling is an indirect method of determining differences in elevation by measuring vertical angles and slope or horizontal distances with the transit and using trigonometric calculations for reduction. Precision varies depending upon the instruments used, size of vertical angles and length of sights.

The most commonly used method for determining differences is a direct leveling procedure making use of a spirit level. A spirit level is used to fix the direction of a horizontal line of sight. The engineer's level, builder's level, hand level and precise level all make use of the spirit level principle. The engineer's level consists of a telescope which magnifies the image and provides a line of sight by means of intersecting cross hairs, a spirit level bubble vial which denotes the correct centred position, and a three or four foot-screw leveling head for adjusting the bubble, all of which are mounted on a tripod. The level vial and telescope are attached to a bar, which in turn is attached to a spindle. The telescope can be revolved only in a horizontal plane. When the instrument is properly adjusted and leveled, the line of sight in the telescope is horizontal. The telescope on the engineer's level is usually one which erects the image. The two types of engineer levels are the dumpy and the wye. In the dumpy level the telescope and bubble vial are rigidly attached to the instrument and in the wye level the telescope and vial are independent of the other parts of the instrument and may be lifted from the supports and turned end for end. The dumpy type is by far the more popular. The engineer level is used with a level rod for measuring differences in elevation. A level rod consists of a graduated 7-ft. wooden rod which can be extended to 13 ft. when needed. It is usually graduated with black markings on a white background in feet, tenths and hundredths. A sliding vernier target is used if the rod is to be read to thousands. Sometimes small spirit level vials are temporarily attached to the rod to ensure proper plumbing of the rod to vertical position. Differential leveling is the method of direct leveling used in determining differences in elevation. The level is

set up and leveled at a given position. A rodman holds the leveling rod alternately on the backsight and foresight turning points or bench marks and the levelman, operating the instrument, records the readings. The level and rod are moved alternately to new positions as the process is extended through an area. The difference between backsight and foresight rod readings yields a direct value for the elevation difference.

Spirit leveling methods are used extensively. The small builder's level is essential to setting construction grade stakes for building foundations, sewer lines and earthwork movement. The engineer's level is the most used level in all types of engineering work which require knowledge or control of the vertical dimension. The hand level is a hand-held metal sighting tube with plain glass covers at the ends with a level vial mounted on the tube which is used for rapidly measuring approximate differences. An Abney hand level is one mounted with a graduated arc for determining vertical angles. Precise leveling is accomplished with a precise level which usually has a three-screw leveling head, a telescope and bubble vial especially mounted so that they can be properly tilted with a tilting screw mounted under the eyepiece, a bull's eye level for approximate leveling, an optical-bubble viewer for observing the bubble from eyepiece position, and an optical system which gives improved sighting but with an inverted image. A precise level rod is about ten feet long and has a hardened-steel foot. It is graduated in metres, decimetres, centimetres or yards and decimals on a strip of invar metal fastened rigidly to the face of the rod and is inverted to correspond to the inverting of the telescope. The rod is equipped with a rod level and thermometer. Three cross-hair readings are taken on the front of the rod for each sighting and checked by reading other graduations on the back face.

A German-made self-leveling level was developed in 1950 which automatically levels by virtue of a pendulum-type arrangement of prisms, called a compensator, which always dampens down to the same position.

The "elevation meter," consisting of a three-wheeled trailer towed behind an automobile, was developed in the late 1940s. Distances are measured by revolutions of a wheel and inclinations with the vertical are measured by a pendulum. An electronic computer continuously multiplies and yields successive increments of elevation change. It has been successfully used in obtaining supplemental elevations.

Angular and Direction Measurement.—Angle and direction relationships are very important in surveying and can be expressed in a variety of ways and observed with several different instruments. Directions are expressed as bearings or azimuths which are referred to fixed reference lines. A reference line is called a meridian and can be either magnetic, true or assumed, depending upon its relationship to the earth's magnetic field, to the earth's north or south axis poles or to convenience. An azimuth is expressed as a horizontal angle measured clockwise from a reference meridian (north or south) and ranges in value from 0° to 360°. A bearing is expressed as a horizontal angle within a quadrant referred to a meridian. It is measured clockwise or counterclockwise, depending upon quadrant, from either north or south and ranges in value from 0° to 90°. A thorough knowledge of the intricacies of magnetic declination is desirable in understanding the relationships between the true and magnetic meridians.

Several methods are used in measuring angles depending on purpose, accuracy and convenience. Horizontal angles can be measured directly between two lines, can be measured by prolonging one line beyond point and observing deflection angle to the new line, can be observed directly on a compass box, or can be observed directly by carrying azimuths continuously on instrument. Multiple measurement of angles is desirable and these methods are referred to as repetition or direction. Instruments are available for making use of either basic method. The repetition method involves the use of a repeating instrument in adding successive observations of a horizontal angle by repetition and a mean taken to obtain an accurate measure of the angle. The direction method consists of measuring the direction to each station from one taken as the initial station. The angles are measured clockwise and the angle between any two observed stations is the difference of their directions.

Angles are commonly measured with the transit, but are also measured by means of the tape, compass, sextant, plane-table alidade and precise theodolite. The transit is referred to as a universal surveying instrument, because of the wide variety of uses for which it can be adapted. It may be used for measuring or laying out horizontal angles, directions, vertical angles, difference in elevation, distances and for prolonging straight lines. The essential features of a transit are a horizontal circle graduated in degrees which may be rotated and clamped in any position, a plate which rotates inside the graduated circle containing a vernier for accurate reading of the circle, and a telescope (line of sight) which is attached to and rotates with the vernier plate and also may be rotated vertically. The telescope is similar to that on an engineer's level. Verniers on most transits are so designed that the graduated circle can be read to the nearest minute, 30 sec., 20 sec. or 10 sec. The plates and telescope, mounted on four foot screws and tripod, are so manipulated that angles may be determined. A plumb bob attached to the centre of the instrument is used to assure centring over a station. In the 20th century the European-type transit has been increasingly used throughout the world. It essentially features a high degree of precision, an ease of operation, and compactness and lightweight. The horizontal and vertical graduated circles are read by means of an optical micrometer observed through an auxiliary telescope. All motions are enclosed and are dustproof and moistureproof. The instrument is mounted on a three-screw base and utilizes an optical plumbing device for centring over a point. Most models make use of a single eyepiece which means an inverted image. Measurements are generally made to the nearest minute of arc and the 0.1 min. by estimation.

The compass is an instrument which was formerly much used in land surveying, but, because of increasing accuracy requirements, has been replaced mostly by the transit. The compass, especially the hand compass, still has limited use in forest cruising, exploratory surveys and reconnaissance work. The modern transit has a compass attached and is used only as a check on transit work. The compass consists essentially of a circle graduated in quadrants from the north and south compass points to the east and west, with a magnetized steel needle mounted on a pivot at the centre, and counterbalanced for dip. Bearings can be read directly when the instrument is pointed on an unknown line.

The sextant, although principally used in navigation, is often employed in surveying to measure angles, especially in hydrographic surveying. The sextant is a hand-held instrument with a limb of one-sixth of a circle, but graduated and designed so that measurements can be made to 120°. Rays of light are reflected in order to span this large an angle. Angles may be measured while the observer is on a moving object, such as a launch from which soundings are also measured.

A tape is occasionally used in forming a triangle at the point to be measured and calculating the angle usually by trigonometry. Angles also can be determined graphically by means of the plane table and alidade. The essential features of a plane table are a drawing board mounted on a tripod and an alidade with a straight edge parallel to the line of sight. The alidade is essentially a telescope mounted on an upright standard and attached to a metal base which is used as a straight edge. The telescope can be rotated vertically and an arc permits reading the vertical angles.

The theodolite is a precision surveying instrument similar to the transit but larger and heavier in construction with a more finely graduated circle and more sensitive levels. Readings are made by means of micrometer microscopes to the nearest second of arc and by repetition and direction procedures to tenths. Because of the increased refinement needed in sighting the instrument, two closely spaced vertical cross hairs are employed. Theodolites are principally used in geodetic triangulation where precise angles are required and sightings are made on distant light signals at night.

Optical tooling is the application of surveying techniques and newly developed instruments to problems of shop measurement such as holding close tolerances in building large mock-ups and jigs, aligning shafts, setting surface plates, etc. With this new equipment measurements are made from a line of sight, vertical

plane and horizontal plane to accuracies of 0.001 in. by means of optical micrometers.

Triangulation.—Surveying procedures employ either the traverse method or the triangulation method. Both are employed extensively for setting up control for topographic surveys, land surveys and construction surveys. Triangulation consists essentially of a series of triangles in which many sides are common to adjacent triangles. All of the points or stations are occupied and the angles measured very carefully. The length of one base line in the system is also carefully measured. In trigonometry, if all the angles and one side of a triangle are known, the lengths of the remaining side can be computed. When one triangle is computed, the newly computed sides offer an entry into adjacent triangles and the process continues. Checks are made occasionally by comparing the computed distance with a new base line. Sometimes stations cannot be occupied, but angles from them can still be computed if the other two angles of the triangle in which they occur are measured. A triangulation arc system consists of a chain of figures, usually single triangles, polygons, quadrilaterals or some combination. A system over a broad area might take several of the above forms in a more or less irregular network scheme. After the angles are measured and the sides measured or computed, adjustments are made of the observations and computations are made of co-ordinates of the stations. The suitability of the figures depends to a great extent upon the shape of the figures selected. Angles near 0° or 180° are subject to large computational errors and are avoided when possible.

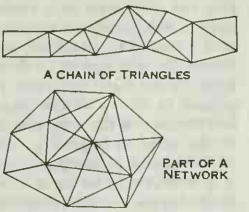


FIG. 3.—SYSTEMS OF TRIANGULATION

Traverse.—A traverse consists of a series of lengths and directions of straight lines between points, obtained by field measurements and used in determining positions of the points or stations. The lengths, as well as the directions, of all lines are field measured in contrast to the triangulation method. Traverses are classified and identified in a variety of ways. They are classified according to form, open or closed, depending upon whether the traverse is a long continuous one or one that closes on itself. Closed traverses provide checks on field work, where open traverses do not. Open traverses are used in highway or railroad preliminary lines and closed traverses are used in area or property surveys. Traverses are classified according to method used, as deflection-angle traverse, azimuth traverse, interior-angle traverse, etc. They are also classified according to instruments used, as transit-tape traverse, transit-stadia traverse, compass traverse, etc., and sometimes according to quality of results, as first-order traverse, second-order traverse, etc. Both triangulation and traverse methods afford convenient, precise and economical procedures for locating stations and the physical conditions involved usually dictate the method selected. Where the area is rough, hilly or swampy, obstructions probably forbid using a traverse. When stations have been located either by traversing or by triangulation, the details of a survey are referred to the traverse line or stations, and map plotting, area calculating, etc. can be continued.

Topographic Surveying.—A topographic survey is made for the purpose of determining the configuration or relief of the surface of the earth and of showing the location of natural and artificial features. A planimetric map presents only the horizontal positions of features and a topographic map shows the same features, but also presents the vertical position of features by showing relief in some measurable form. The form generally used is the contour which is an imaginary line on the ground, all points of which are at the same elevation above a specific datum surface, usually mean sea level. The contour or topographic map is an invaluable aid in most engineering construction projects, subdivision planning, military planning, industrial site selection and development of natural resources. Because of the wide range in the use of maps and the variation of the physical characteristics of different areas, there are several scales and methods used to develop the map. Large-scale mapping of small areas is usually rather simple in plan and execution, while small-scale mapping of extended areas is usually involved and difficult and is undertaken by the larger governmental or military agencies.

The aerial photograph is widely used in compiling maps and will be discussed separately later under *Photogrammetric Surveying*. Here only ground methods will be described. In compiling a topographic map, one should be thoroughly familiar with the characteristics of contour lines, imaginary lines of equal elevation. The proper choice of a contour interval, the vertical distance between contours, depends upon the scale, condition of terrain and intended purpose of the map. The construction of a topographic map involves setting up horizontal and vertical control over an area, locating detail and ground contour points and plotting the control, detail and contours on the map sheet. Horizontal control for an area usually consists of transit-tape or transit-stadia traverses or sometimes a triangulation system. Vertical control is established by differential leveling. The helicopter has added much to the ease of transportation in rough terrain.

In checkerboard mapping, the area is covered with a grid system of stakes set at the corner points and ground elevations determined at these grid corners and at essential intermediate locations with the level. The detail is also tied to the grid system and all of the data is located on the map sheet and the contours plotted. Regardless of the amount of control established and detail obtained, the compiler must do some freehand sketching to complete the contour lines. The checkerboard method is ideal for smaller areas where the terrain is gentle and a large-scale map is desired.

In transit-stadia mapping, points of a traverse or triangulation system are occupied and both the horizontal and vertical positions of the required detail and ground contour points are determined with a transit. The direction is usually recorded as an azimuth and the horizontal and vertical distances are measured by stadia. This data is transferred to the map sheet by paper protractor and scale and contours interpolated and sketched. Transit-stadia mapping can be accomplished over rougher terrain.

In plane-table mapping, the compilation is made directly in the field on the map sheet fixed on the plane table. The alidade takes the place of the transit and the directions and the horizontal and vertical distances are drafted directly on the sheet as they are observed. This sometimes eliminates the necessity of recording notes and also reduces the personnel needed on such an expedition. To use the plane table properly, one should understand the procedures of graphical traversing and triangulation and the principles of radiation, intersection and resection. Contours are sketched in the field where they can be viewed and not later in a drafting room. However, there are advantages and disadvantages for each method and there is no one better method, except as each individual job is considered. In order to test the accuracy of completed maps, test dimensions and test profiles are made and discrepancies analyzed. Contour maps are especially valuable in obtaining cross sections and profiles, determining earthwork quantities for grading operations, determining route locations, and areas and volumes of reservoirs. An excellent example of topographic mapping is the work accomplished by the United States geological survey in the task of preparing maps of the United States. The original plan was to map the country at a scale of 1:62,500 or 1:31,680. But with the increased use of maps and the increase in value of land and property, a scale of 1:24,000 has been adopted as standard for most regions. In most European countries, the governmental mapping has been completed; however, a continuing process of revision is necessary. Less than 2% of the land areas of the world are mapped adequately at a scale of 1:25,000 and less than 25% have reconnaissance maps at a scale of 1:250,000.

Route Surveying.—A route survey is made for the purpose of determining and laying out the lines and grades for linear construction of transportation and communication lines such as highways, railroads, pipelines, canals and transmission lines. Route surveying generally consists of determining the ground configuration and the location of objects, establishing the definite location of the route, and determining the necessary earthwork volumes.

Naturally, methods employed will vary somewhat between, say, a highway and a transmission line to be established between two end points. The typical sequence of operations involves a reconnaissance survey, preliminary survey, location survey and construction survey. The reconnaissance consists of an extensive general study of the area with a view toward selecting possible routes which might be feasible for further consideration. Maps and aerial photographs are useful to the locating engineer as he travels over the area by airplane, on the ground, or both. His cursory examination of the topography, physical features, distances, grades, vegetation and soils, will eliminate all but the better possibilities. From the economical standpoint, the reconnaissance is a highly important phase of the route survey. After one or two routes have been fixed for further study by the reconnaissance, a preliminary survey is begun. A narrow strip map is made of the country along the proposed route of sufficient width to make minor adjustments of location. The preliminary line is staked out on the ground with a transit and tape and the strip mapping completed by one of several methods. Elevations along the route are obtained with a level and are plotted to form a profile of the line. A profile is a vertical section along the ground surface. When these data are plotted, a centreline of the proposed route can be tentatively established on the map and is referred to as the paper location. The final line is calculated and the necessary curves computed. The final field location is then located on the ground by transit surveys, and cross sections and profiles are obtained with a level. These cross sections and profiles help to establish the final grade line and to indicate the amount of earthwork to be moved. Additional construction surveys are required to set grade stakes, slope stakes, alignment stakes and stakes for various structures such as bridges and culverts.

Mine Surveying.—Mine surveying includes underground surveying and mineral-land surveying. Underground surveying involves the survey operations performed beneath the earth's surface in connection with mining, tunneling, cave exploration and subterranean passageway construction. Underground surveys are essentially the same as surveys on the surface in that the purpose of the angle and distance measurements is to obtain the horizontal and vertical positions or co-ordinates of a point. However, underground surveys do differ in several ways. The station or point is usually in the roof instead of in the floor; the telescope, cross hairs and graduations must be artificially illuminated; distances are usually measured on the slope; and the transit tripod must have adjustable legs to work in low areas and on irregular or steep surfaces. The transit ordinarily used in underground work is the regular engineer's transit with an extension leg tripod and a prismatic eyepiece or an auxiliary telescope attached to one end of the horizontal axis or above the main telescope so that large vertical angles may be made for steep sightings. This side auxiliary telescope which is offset from the vertical axis causes an eccentricity and affects the observed values of horizontal angles. When the auxiliary telescope is mounted on top of the main telescope, the eccentricity affects the observed vertical angles. Mineral-land surveying involving location and patent surveys is essentially similar to land surveys on the surface involving the determination of land boundaries. Some differences arise because it is possible, by special laws, to obtain rights to subsurface mineral veins or lodes only.

Land Surveying.—Land surveying is one of the oldest branches of engineering known to man, the principles of which have been passed down through the centuries. Land surveys involve the original surveys to locate and monument the boundaries of property. Operations also include the preparation of a plat and a legal description of the limits of property and of the area included, the resurveys to recover and remonument property corners, and the surveys necessary when property is subdivided into smaller parcels. Property laws vary in different countries; however, the essentials of land surveying as practised in the various countries are similar in principle. The description of a parcel of land may take several forms such as describing by metes and bounds, by stating location and size in a rectangular system of land subdivision, by stating co-ordinates of property corners with reference to a plane co-ordinate system, or by reference to numbered lots on a subdivision plat. The primary function of the land surveyor is to carry out the field surveys; however, he must not only be familiar with the technical procedures but also with the legal aspects of property and boundaries. Land surveys are usually completed by the traverse method using a transit and tape and directions recorded as bearings. The precision required generally depends upon the value of the land. Formerly the surveyor's compass and chain were used extensively and many property descriptions in older deeds are recorded in those units. When a parcel of land is defined by giving the courses (directions) and lengths of all boundaries it is said to be described by metes and bounds. Such a description starts at a point of beginning, which is properly referenced by ties to witness objects, and, in turn, gives the bearing and distance of the sides around the parcel until it is closed back on the point of beginning. Many descriptions are of this type. In the opening of new public land for settlement and development, it is desirable to set up a better system of land subdivision. A good example is the development of the western part of the United States. Congress passed legislation establishing a rectangular system of surveys so that land could be systematically partitioned and transferred. The public lands were divided by true meridional lines and by parallels of latitude to form land townships approximately six miles square. Each township was subdivided into sections one mile square of approximately 640 ac. All 36 sections were numbered uniformly. Sections were further subdivided into half-sections, quarter-sections, etc. The principal line and standard parallel formed the initial point from which all land in the area was referred. Because meridians tend to converge when extended northward, the rectangular system was not the absolutely perfect system, as indeed no system would be on a sphere, and occasional correction lines were needed. A description of this type might be written: the northeast quarter of the southwest quarter of section 8, township 23 north, range 4 west, of the initial point of the second principal meridian in Indiana, containing 40 ac., more or less, according to the United States survey. Geodetic triangulation and traverse monuments of known geographic positions (latitude, longitude and azimuth) have been available, but computations were generally too involved to be used to any great extent as reference marks by land surveyors.

In the 20th century, plane co-ordinate systems have been developed for many states and countries. These grid systems involve relatively simple calculations and their use in describing tracts of land is increasing. Once the co-ordinates of the property corners have been determined, the boundaries may be relocated even though all corner monuments have been destroyed. The permanency of the co-ordinate system is obvious. Co-ordinates are generally not used by themselves but in conjunction with one of the other methods of description. In urban areas, land is subdivided in lots and numbered and may be legally described by reference to the recorded subdivision plat. The land surveyor is confronted by many legal problems such as the relocation of lost and obliterated corners, excess and deficiency, riparian rights, adverse possession, etc. Most states and countries require land surveyors to be registered, either directly or as a civil engineer.

Construction Surveying.—Construction surveys consist of the measurements necessary in the various stages of conceiving, planning, locating and erecting the many varied projects of construction. Such surveys are conducted to obtain data essential to the planning and cost estimating of a proposed project and to establish control points, lines and grades for construction operations. These surveys generally involve making a topographic survey of the site to be used in planning the structure, establishing on the ground a system of both vertical and horizontal control stakes to be used in making measurements of earthwork, giving line and grade as needed for construction and making final measurements to verify completed construction as a basis of payment to the contractor. On many projects, line and grade are conveniently given by the use of batter boards. Construction surveys are an integral part of the construction procedure in the building of highways, railways, airports, streets, bridges, culverts, tunnels, sewers, pipelines, canals, aqueducts, buildings, homes and dams, and generally involve the application of plane surveying.

Hydrographic Surveying.—In hydrographic surveying, surveys are made to determine the depth of water and configuration and character of bottom, the location of shore lines, rocks, shoals, reefs, lights and buoys, the direction and force of currents and the height and times of tides and water stages. The surveys are intended primarily for the preparation of nautical navigational charts, but are also used in planning and controlling engineering projects such as dams, reservoirs, tunnels, bridges, docks and other river and harbour improvements. In a sense, hydrographic surveying is quite similar to topographic surveying, the principal difference being that of locating points on the water surface, sometimes at great distance, and determining the depth of water at these points instead of elevations of land topography. The same horizontal control systems of triangulation or traverse are needed along the shore lines, and the same for vertical control in the form of bench marks. Shore line details of topography and prominent cultural features such as buoys and lighthouses are located by stadia or plane table so that these details might later be used as points of reference for range line and sounding work. The determination of depth of water at a point is made by soundings. Depths are referred to water level at the time of observation, but are later corrected to a desired datum which has been determined by tide gauge. Soundings are made by sounding rod, sounding line or fathometer. A sounding rod is used in areas where there are small amounts of current and depths up to three fathoms. A sounding line with attached lead and markers at intervals along the line has been popular. Increasing use is being made of the fathometer, an electrical echo-sounding device, which automatically records the depth by measuring the amount of time needed for an impulse of the sound from the ship to travel to the bottom and return as an echo. Some fathometers indicate depths on a lighted dial and others plot depths directly on graph paper, giving a continuing record as the ship moves along a predetermined course. Before the corrected soundings can be plotted on a map, their location with respect to the shore must be determined. Several methods are employed in determining this location, using range lines and angle measurement. Modern methods utilizing electronics are being increasingly used for position finding from two or more previously located shore stations. In shoran, loran, Decca and E.P.I., a radio impulse is sent out by the ship to the shore station and returned to the ship. The time elapsed between the two impulses can be measured to 0.00000001 sec. and, knowing the velocity of radio waves, the distance can be determined accurately. Several methods of plotting soundings on a map are used, depending on which of the foregoing methods was used. A three-armed protractor is useful in graphically solving the three-point problem.

A hydrographic map (or nautical chart) shows depths in fathoms (or feet for shoal areas) referred to mean low-water level and contours on land in feet above high water. A wire-drag procedure is used to find a near surface obstruction such as a reef or pinnacle rock. Current velocities are normally determined by the use of floats or one of a number of types of current meters. Current meters consist essentially of a propeller or a cupped wheel which revolves with the impact of flowing water and a counting device to indicate the number of revolutions of the wheel per time interval. Tide gauges, for measuring the height of the tide, are generally of the nonrecording type, such as the staff gauge and various types of float and pressure gauges requiring the presence of an attendant, and the automatic or self-registering gauge usually activated by a float which provides a continuous record of tidal variations with a passage of time.

Geodetic Surveying.—Geodetic surveys are necessary where the area is of such magnitude and the desired accuracy and precision is of such a high order as to require taking into account the curvature, figure and size of the earth. Surveys of this nature are of the highest calibre and generally form the basic framework of control, both horizontal and vertical, to which surveys of a lower order are tied. Therefore, many refer to them as geodetic control surveys. Basic horizontal control is established by precise first- and second-order geodetic triangulation and traverse. Basic vertical control is established by precise geodetic leveling. Both necessitate extremely precise instrumental equipment and observational techniques, capable and conscientious personnel and detailed computations. The resulting horizontal data are expressed in the form of geodetic latitudes and longitudes of definite established and monumented points and include distances and azimuths of all lines. Descriptions and geographic positions of these stations are available to the public and serve many useful purposes. This information provides the rigid framework for all types of mapping and charting, serves in locating national and other governmental boundaries, perpetuates the surveys of properly connected private boundaries, enables local surveys and maps to be tied to the national plan, provides a check in accuracy for local surveys and serves as a basis for many scientific studies such as earth movement, figure of the earth and geophysics.

An excellent example of a national geodetic triangulation net is that of the United States which is being performed by the U.S. coast and geodetic survey. The first surveys were along the Atlantic coast and then long transcontinental arcs were established across the country and tied together. Before computations could be finalized a datum was necessarily adopted and it was based on a point in Kansas known as station Meades Ranch. Geographic positions were adopted for this station and along with Canada and Mexico, the United States thus established the North American Datum. This network of first-order triangulation arcs is substantially completed and the areas between are to be covered by improved and standard classes of second-order work. Eventually such monuments will exist five or six miles apart throughout the country. In triangulation of such order, base lines must be measured with an invar tape and angles with precise theodolites. Base lines are measured as described previously with all the necessary precautions and corrections applied. Because of the distances involved the points occupied must be high to be intervisible. Stations are generally chosen on mountain peaks, hill tops and high ground, and towers must be built to elevate the instrument and to provide working space for the observing party. Many wooden stands are built; however, in flat plain areas, tall towers, of more than 100 ft., must be constructed. Steel towers, which can be relatively easily assembled, disassembled and transported, have revolutionized the speed of triangulation. Aluminum towers have been considered. The involved adjustments and computations of field data are the job of skilled mathematicians and computers. Precise traversing is essentially a series of connected base lines and is a part of a triangulation system augmenting the widely scattered triangulation stations. Precise first-order leveling forms the basic net for elevations or vertical control and is analogous to precise first-order triangulation, which forms the basic net for horizontal control. Precise leveling differs from ordinary leveling only in the sensitiveness of the instruments used, the care with which observations are made and the refinement of the computations. The result of first-order leveling is a network of leveling throughout a country with permanently set bench marks of known elevations above mean sea level. This allows the many surveying and civil engineering operations to be based on a common reference datum for purposes of co-ordination and comparison. Instruments used are the precise level and precise level rod which were previously described. Corrections for curvature of the earth and refraction of the air are necessarily applied. There are several methods of triangulation which are an outgrowth of scientific advances during World War II. One is shoran, an electronic method by which distance measurements are made by determination of the travel time of a radio signal. Observations are made by flying an airplane, equipped with the shoran transmitter-receiver-indicator units, across the line about midway between two ground stations equipped with shoran transponders. Several corrections must be applied to convert the shoran distance to geodetic ground distance. Accurate knowledge of the altitude, temperature, pressure and humidity are essential. Shoran is best suited for longer distances, 500 to 600 mi., and is particularly applicable in triangulation connection of continents and off-lying islands or where widely spaced area coverage is desirable. Another development is flare triangulation, a method of extending triangulation between points which exceed the sight limitations of ground means. The method depends upon height and intensity of the flare and clear weather visibility. It involves

several simultaneous radio-controlled theodolite observations on parachute flares dropped from an airplane about midway between stations. A third development is the Bergstrand geodimeter which is based upon the measurement of the travel time of a light wave to determine distances.

Photogrammetric Surveying.—Photogrammetry is the science or art of obtaining reliable measurements by means of photographs. ' Photogrammetric surveying is the application of photogrammetry to the determination of contours, dimensions, positions, etc., of parts of the earth's surface. The principles of photogrammetry are employed in many other scientific fields, but the greatest development has been in the field of surveying and mapping. It is not a new science, but only in the first half of the 20th century has the knowledge become general because of rapid developments in photography, optics and aviation. The type where photographs are taken from points on the ground surface is called ground or terrestrial photogrammetry. Aerial photogrammetry involves vertical or oblique photographs taken from an air-borne vehicle. Stereophotogrammetry means the use of overlapping pairs of photographs in a stereoscopic viewing or plotting device which gives a three-dimensional model for study. Terrestrial photogrammetry is quite similar to plane-table surveying with the camera replacing the plane-table. Modern phototheodolites of utmost precision can quickly record the terrain to be mapped. However, the process of conversion of a terrestrial photograph to the map plane is a cumbersome process and the view commanded by an air station is much preferred. A phototheodolite is especially advantageous in mountainous terrain where elevations are desired. The desired scale of a vertical photograph can be controlled or computed by the relationship of the flight altitude to the focal length of the aerial camera or by the relationship of a measured distance on the photograph to the corresponding distance on the ground. For surveying purposes, photography must give complete coverage of the area together with generous overlapping. Parallel flight strips are flown with usually 60% forward overlap and 30% side lap between adjacent strips. Most photography is made with a general purpose single-lens aerial camera, but there are some multiple-lens cameras that are used to considerable advantage in some types of aerial mapping. A vertical photograph of an area of perfectly flat terrain without any tip or tilt of the camera axis will yield an excellent perspective image of the area. Image points on photographs of irregular terrain are displaced caused by this relief and it is impossible to avoid this displacement. Very seldom is a photograph taken without some camera tilt or tip as the airplane attempts to fly its course, and therefore displacements of all points result. It is through measurements of these differences in scale and displacements in position that photogrammetrists are able to function. The displacement caused by relief and displacement caused by tilt afford the two major difficulties to be surmounted because these act from different points on the plane of the photograph. In addition, there are distortions of the image caused by photographic imperfections such as lens aberrations, film curvature, film shrinkage and camera movement during exposure. All of these factors—tilt, relief and photographic imperfections —make the task of preparing accurate topographic maps an involved and interesting problem.

One solution to the problems involved in aerial mapping is by method of radial-line plotting, which is accomplished by mechanical-armed templets, acetate templets or slotted templets. The positions of points in the radial-line method are plotted in their correct relative position to one another by a system of intersection and resection, using the radial centre of each photograph as the origin of the radials through the images of the points. This method is often referred to as graphical triangulation and there are some who prefer analytical methods of solution. The methods most commonly used in mapping make use of stereoscopic plotting instruments and the principle of stereoscopy. The problem of measuring vertical and horizontal dimensions in a stereoscopic model is overcome by using the principle of the floating mark. Many stereoscopic instruments have been developed and most serve well for some specific purpose. Some are simple, inexpensive and of low accuracy, and others are intricate, expensive instruments of the highest precision. In solving the problems of tilt and relief displacements, each instrument uses the same general method although there is wide difference in mechanical details. A pair of overlapping photographs or diapositives is mounted in the instrument to recover the relative orientation of the negatives at the instant of exposure. A miniature model of the terrain is created and is viewed by a stereoscopic viewing system. The model is brought to the desired scale and oriented to the proper datum as provided by ground-survey control points. Map detail and contours are transferred to the manuscript sheet by use of the floating mark. The model re-created accurately, as it would appear in nature, automatically solves the problem of tilt and relief displacements. Some of the instruments used are the sketchmaster (nonstereoscopic), KEK plotter, contour finder, stereocomparagraph, multiplex, Kelsh plotter, Balplex, stereoplanigraph, aerocartograph, stereotopograph and autograph. Problems of photographic imperfections have not been altogether solved and much research remains in this area.

Even though photogrammetry is over 100 years old, it was only after World War II that it had profound influence on the economics of surveying and mapping. The plane table, as the basic procedure for making topographic maps, has been replaced by photogrammetric methods. Ground surveys are still needed in considerable number, but in major mapping operation the field operations are supplemental to the photogrammetric operations.

BIBLIOGRAPHY.—American Society of Civil Engineers, *Definitions of Surveying, Mapping, and Related Terms,* Benjamin E. Beavin *et al.,* Manual No. 34 (1954). *History:* L. A. Brown, *The Story of Maps* (1949); E. R. Kiely, *Surveying Instruments, Their History and Classroom Use,* National Council of Teachers of Mathematics, Nineteenth Yearbook (1947). *Hydrography:* K. T. Adams, *Hydrographic Manual,* Spec. Publ. 143, U.S. Coast and Geodetic Survey (1942). *Geodesy:* Howard S. Rappleye, *Manual of Geodetic Leveling,* Spec. Publ. 239, U.S. Coast and Geodetic Survey (1948); F. R. Gossett, *Manual of Geodetic Triangulation, Spec. Publ. 247,* U.S. Coast and Geodetic Survey (1950); Guy Bomford, *Geodesy* (1952). *Photogrammetry:* L. W. Swanson, *Topographic Manual,* pt. ii, Spec. Publ. 249, U.S. Coast and Geodetic Survey (1950); American Society of Photogrammetry, *Manual of Photogrammetry* (1952). (K. S. Cs.)

"SURVILLE, CLOTILDE DE," the supposed author of the *Poésies de Clotilde.* The generally accepted legend gave the following account of her. Marguerite Éléonore Clotilde de Vallon Challis, dame de Surville, was born in the early years of the 15th century at Vallon. In 1421 she married Bérenger de Surville, who was killed at the siege of Orleans in 1428. Her husband's absence at the war inspired her heroic verses and his death her elegiac poems. The last of her poems is a *chant royal* addressed to Charles VIII. In 1803 Charles Vanderbourg published as the *Poésies de Clotilde* some 40 poems dealing with love and war. Research showed that the documents, communicated by the marquis de Surville to Vanderbourg, were not genuine, and the account of Clotilde herself was proved to be inaccurate.

SUSA (SHUSHAN, SHUSH), capital of Susiana (Elam) and chief residence of Darius I and his successors. It lay under the Zagros range at about lat. 32°, near the bank of the Karkheh (Choaspes) river and close to the Karun. (For early history *see* ELAM.)

The site, identified (1850) by W. K. Loftus, has on it four mounds. One, rising about 38 m., holds the citadel. A second to the east represents the palace of Darius I and was excavated by M. Dieulafoy. The enameled bricks taken from its walls are in the Louvre, Paris. A third mound to the south contains the royal Elamite city, while the fourth mound consists of the poorer houses. Excavation of the citadel was begun by J. de Morgan in 1897. It yielded the obelisk of Manistusu, the stele of Naram-Sin and the code of Hammurabi (the latter in the winter of 1901–02). (*See* BABYLONIA AND ASSYRIA; IRAN: *Archaeology.*)

The finest pottery came in the lowest strata, 25 m. below the surface, and belongs to two different civilizations, both in the Neolithic period. The earlier is characterized by vases of fine red clay, wheel-made, in a few well-defined shapes, and all with thin, polished sides. The decorations applied in black paint or red-brown ferruginous earth consist of bold geometrical patterns, often combined with spirited studies from nature. The pottery of the second period shows a retrogression, being coarser and porous.

Above the early strata come remains of Elamite and early Babylonian civilizations, inscribed objects from the latter bearing pictorial characters from which the cuneiform was evolved. The upper portions of the mounds disclosed inscribed Achaemenian

monuments, Greek pottery and inscriptions of the 4th century B.C., coins of the kings of Elam and Parthian and Sassanian relics. Moslem tradition says that the tomb of the prophet Daniel lay in the bed of the Karkheh river and a mosque was built on the bank opposite the supposed spot. Until after the 14th century the city was a flourishing centre of a district famous for silk, sugar cane and oranges.

See also Index references under "Susa" in the Index volume.

BIBLIOGRAPHY.—W. K. Loftus, *Travels and Researches in Chaldaea and Susiana* (1857); M. Dieulafoy, *L'Art antique de la Perse* (1884–85), *L'Acropole de Suse* (1890); A. Billerbeck, *Susa* (1893); J. de Morgan, *Mémoires de la delegation en Perse*, vol. i–xxxii (1899); H. Frankfort, *Studies in Early Pottery of the Near East* (1924); E. E. Herzfeld, *Archaeological History of Iran* (1935), *Iran in the Ancient East* (1941); G. G. Cameron, *History of Early Iran* (1936); A. E. Christensen, *L'Iran sous les Sassanides* (1936); A. T. Olmstead, *History of the Persian Empire (Achaemenid Period)* (1948; 1949); R. Ghirshman, *Iran: From the Earliest Times to the Islamic Conquest*, Eng. trans. (1954). (R. LEV.; T. C. Y.)

SUSA, Tunisia: *see* SOUSSE.

SUSANNAH: *see* DANIEL, BOOK OF.

SU SHIH: *see* SU TUNG-P'O.

SUSLIK, the Russian name for the European ground squirrel, *Citellus citellus*. It is found from central Europe east to central Asia. *See* GROUND SQUIRREL; SPERMOPHILE; GOPHER.

SUSO (SEUSE), **HEINRICH** (*c.* 1296–1366), Swabian mystic, was born at Überlingen on Lake Constance, Ger., or, possibly, at Constance itself. At the age of 14 he joined the Dominican order in Constance where, five years later, he underwent a religious experience that gave a lasting and distinctive character to his spiritual life: all marked by his self-dedication, made at that time, to the Eternal Wisdom. (Testimony of the so-called *Life of the Servant* concerning the early decades of his religious life is open to serious reserves.) Sometime before 1327 he went to Cologne for further theological studies and, together with Johann Tauler, was a pupil of Meister Eckhart (*qq.v.*). His first work, the *Little Book of Truth* (*c.* 1329), was occasioned by the controversies stirred up by the trial of his teacher. About ten years later he composed in Latin the *Horologe of Wisdom*, the most durably popular and influential of his writings. The *Little Book of Eternal Wisdom* (of which, at one time, the *Horologe* was thought to be a free translation) seems to come from the period of his maturity about 1348. The most attractive of the Rhineland mystics and the easiest to comprehend, Suso presents a doctrine, Neoplatonic and Thomist, which in orientation is Eckhartian but in content quite his own.

He died at Ulm on Jan. 25, 1366. He was beatified in 1831, and his feast is celebrated on March 2.

The most reliable text of the German writings is that established by K. Bihlmeyer, *Heinrich Seuse: Deutsche Schriften* (1907); of the *Horologe* it is *Colloquia Dominicana . . . a Beato Henrico Susone dictata* (1923). The best translations are those of J. M. Clark: *The Life of the Servant* (1952) and *Little Book of Eternal Wisdom* and *Little Book of Truth* (1953).

See C. Gröber, *Der Mystiker Heinrich Seuse* (1941); J. A. Bizet, *Henri Suso et le déclin de la scholastique* (1946); J. M. Clark, *The Great German Mystics*, pp. 54–74, with bibliography, pp. 114–117 (1949). (E. O'B.)

SUSQUEHANNA, the longest river in the eastern United States, rises in Otsego lake, central New York, and flows southwestward into Chenango county, N.Y., where it heads southward into Susquehanna county of northeastern Pennsylvania. The river then changes its direction northwestward and enters south-central New York for a short distance before again turning southward into Bradford county, Pa. After flowing southeastward through the Allegheny plateau it enters the Ridge and Valley province north of Wilkes-Barre and changes its course southwestward, following one of the major longitudinal valleys. At Sunbury this part of the river, called the North Fork, is joined by the West Branch. The stream then turns southeastward, crossing five ridges and forming water gaps, after which it traverses the Piedmont plateau and empties into Chesapeake bay at Havre de Grace, Md.

Prior to the Pleistocene ice age the Susquehanna drained most of central New York state, but due to morainal deposits across the state the northern tributaries were blocked. At the mouth of the Susquehanna a submarine channel extending eastward across the Continental Shelf indicates the presence of the former valley of the river.

The West Branch of the Susquehanna, 160 mi. in length, rises in the Allegheny plateau of Indiana county, west-central Pennsylvania. The river flows northeastward to Williamsport and then turns southward to Sunbury. The North Fork falls 1,923 ft. and the West Branch falls 688 ft. The major tributaries are the Chemung and Juniata rivers.

Important cities on the Susquehanna river are Binghamton, N.Y., and Wilkes-Barre, Berwick, Sunbury and Harrisburg, Pa. Williamsport is the major city of the West Branch. Due to rock obstructions and rapids the Susquehanna is not navigable; an abandoned canal at its mouth once made the river navigable for a short distance. The Susquehanna presents the greatest waterpower potential of rivers in the eastern United States. There are hydroelectric power plants at Holtwood and Safe Harbor, Pa., and Conowingo, Md. (E. W. MI.)

SUSSEX, EARLS OF. The early history of the earldom of Sussex, an English title, is involved in some obscurity, owing to the fact that under the Norman kings titles of earls were often indifferently derived from a county, from its chief town or from the earl's principal residence, although the distinctive mark of an earl was deemed to be his right to "the third penny" of the pleas of a county (*see* EARL).

Thus in the 12th century the same person is sometimes found described as earl of Sussex, sometimes as earl of Chichester and sometimes as earl of Arundel, while the inclusion of the counties of Sussex and Surrey under the jurisdiction of a single sheriff led at one time to even further confusion. The difficulty was increased by the crown's admission in 1433 that the possession of the castle of Arundel carried with it the right to the title of earl of Arundel, though a later investigation proved the invalidity of the claim.

The Complete Peerage I (ed. 1910), 231–2, holds that Roger de Montgomery, who received grants from William the Conqueror of a large part of the county of Sussex, including the city of Chichester and the castle and honour of Arundel, besides lands in Shropshire with the castles of Shrewsbury and Montgomery, "may be considered" the first earl of Sussex. Whatever his titles were, they were forfeited to the crown when his son Robert was attainted in 1102, and the forfeited estates were conferred by Henry I on his second wife Adelicia, who after Henry's death married William de Albini, or d'Aubigny. The latter was created earl of Sussex by King Stephen in 1141, and "the third penny" of that county was confirmed to him by an instrument of the reign of Henry II, in which, however, he is styled earl of Arundel, a designation by which he was more generally known. His grandson William, 3rd earl of Sussex, was one of King John's sureties for the observance of Magna Carta; and in 1243, the earldom reverted to the crown on the death of Hugh de Albini, 5th earl of the line. His wife, the daughter of the earl of Surrey, held the estates of the earldom in her dower and after her death her brother, and his successor, John de Warenne (1286–1347), earls of Surrey, were sometimes incorrectly called earl of Sussex. (*See* ARUNDEL, EARLS OF.)

Ratclyffe Earls.—For nearly 200 years, from 1347 to 1529, the title of earl of Sussex did not exist in the English peerage. In 1529, however, it was conferred on Robert Ratclyffe (Radcliffe, Radclyffe: *c.* 1483–1542), who had been made Viscount Fitzwalter in 1525. In 1540, Ratclyffe was appointed great chamberlain of England. With Edward, the 6th earl of the Ratclyffe creation (*c.* 1552–1643), the title became extinct.

Savile Earls.—In 1644 Thomas Savile (1590–*c.* 1659), son of John Savile, 1st Baron Savile of Pontefract (1566–1630), was created earl of Sussex. Savile opposed Thomas Wentworth, afterward earl of Strafford, the rivalry between the Saviles and the Wentworths being of long standing in the history of Yorkshire. During the Great Rebellion Savile played a double game, and was imprisoned at different times by both parties. His later years were

spent in retirement at Howley Hall, where he died about 1659. He was succeeded in the earldom of Sussex by his son James, on whose death without issue in 1671 the title became extinct.

It was revived in 1674 in favour of Thomas Lennard, 15th Baron Dacre, whose wife Anne (d. 1722) was a daughter of the famous duchess of Cleveland by King Charles II, and the title again became extinct at this nobleman's death in 1715. The title was next conferred in 1717 on Talbot Yelverton, 2nd Viscount de Longueville and 16th Baron Grey de Ruthyn (1690–1731), from whom it descended to his two sons successively, becoming once more extinct on the death of the younger of these, Henry (b. 1729) in 1799.

Royal Earls.—In 1801 Prince Augustus Frederick (1773–1843), the sixth son of George III, was created duke of Sussex. Spending his early years abroad, the prince was married in Rome in 1793 to Lady Augusta (d. 1830) daughter of John Murray, 4th earl of Dunmore. The ceremony was repeated in London and two children were born, but under the Royal Marriage act of 1772 the Court of Arches declared the union illegal. The children took the name of d'Este. Sussex was a man of liberal ideas; he favoured the abolition of the slave trade, the repeal of the corn laws and the removal of the civil disabilities of Roman Catholics, Dissenters and Jews. His second wife, Cecilia, widow of Sir George Buggin, was created duchess of Inverness in 1840. He died at Kensington palace on April 21, 1843. The older title of earl of Sussex was conferred in 1874 upon Prince Arthur, the third son of Queen Victoria, who at the same time was created duke of Connaught and Strathearn. At the death of his grandson, who had succeeded to the titles in 1942, in 1943, it became extinct.

SUSSEX, THOMAS RADCLIFFE or RATCLYFFE, 3RD EARL OF (*c.* 1525–1583), was lord lieutenant of Ireland and lord lieutenant of the north under Queen Elizabeth I. The eldest son of Henry, 2nd earl of Sussex, by his first wife, Elizabeth, daughter of Thomas Howard, 2nd duke of Norfolk, he was born about 1525, and after his father's succession to the earldom in 1542 was styled Viscount Fitzwalter. After serving in the army abroad, he was employed in 1551 in negotiating a marriage between Edward VI and a daughter of Henry II, king of France. His prominence in the kingdom was shown by his inclusion among the signatories to the letters patent of June 16, 1553, settling the crown on Lady Jane Grey; but he nevertheless won favour with Queen Mary, who employed him in arranging her marriage with Philip of Spain, and who raised him to the peerage as Baron Fitzwalter in Aug. 1553. In April 1556, Fitzwalter was appointed lord deputy of Ireland. The measures enjoined upon Fitzwalter by the government in London comprised the reversal of the partial attempts that had been made during the short reign of Edward VI to promote Protestantism in Ireland, and the "plantation" by English settlers of that part of the country then known as Offaly and Leix. But Fitzwalter first of all found it necessary to make an expedition into Ulster. Having defeated Shane O'Neill and his allies the MacDonnells, the lord deputy, who by the death of his father in Feb. 1557 became earl of Sussex, returned to Dublin, where he summoned a parliament in June of that year. Sussex then took the field against Donough O'Conor, whom he failed to capture, and afterward against O'Neill, whose lands in Tyrone he ravaged, restoring to their nominal rights the earl of Tyrone and his reputed son Matthew O'Neill, baron of Dungannon. (*See* O'NEILL.) In June of the following year Sussex turned his attention to the west, where the head of the O'Briens had ousted his nephew Conor O'Brien, earl of Thomond, from his possessions, and refused to pay allegiance to the crown; he forced Limerick to open its gates to him, restored Thomond, and proclaimed Donnell O'Brien a traitor. He took part in the ceremonial of Queen Elizabeth I's coronation in Jan. 1559; and in the following July he returned to Ireland with a fresh commission, now as lord lieutenant, from the new queen, whose policy required him to come to terms if possible with the troublesome leaders of the O'Neills and the MacDonnells. Sussex was recalled, at his own request, in Oct. 1565. His government of Ireland had not, however, been without fruit. Sussex was the first representative of the English crown who enforced authority to any considerable extent beyond the limits of the Pale.

On his return to England, Sussex immediately threw himself into opposition to the earl of Leicester. In 1566 and the following year Elizabeth employed him in negotiations for a marriage with the archduke Charles. When this project failed Sussex returned from Vienna in March 1568. In July he was appointed lord president of the council of the north and in 1569 lord lieutenant of the north in which office he had to deal with the rebellion of the earls of Northumberland and Westmorland. In 1570 he laid waste the border, invaded Scotland, and raided the country round Dumfries, reducing the rebel leaders to complete submission. In December he became a privy councillor and in July 1572 lord chamberlain. He was henceforth in frequent attendance on Queen Elizabeth, both in her progresses through the country and at court. He favoured her proposed marriage to Francis, duke of Anjou, and was an advocate of a conciliatory policy toward Spain. He died in London on June 9, 1583.

The earl of Sussex was a patron of literature and of the drama. He was twice married: first to Elizabeth, daughter of Thomas Wriothesley, earl of Southampton; and secondly to Frances, daughter of Sir William Sidney. His second wife was the foundress of Sidney Sussex college at Cambridge, which she endowed by her will. The earl left no children, and at his death his titles passed to his brother Henry. (R. B. WM.)

SUSSEX, a county of southern England bounded on the south by the English Channel, on the west by Hampshire, and on the north by Surrey and in part by Kent, which bounds it on the east. The geographical area is 1,457 sq.mi. (3,774 sq.km.).

Physical Geography.—From the chalk hills of the western border, the South Downs (average height 500 ft. [150 m.]; Ditchling Beacon, 813 ft. [248 m.]) lead east-southeast to meet the sea between Brighton and Eastbourne as bold cliffs, notably Beachy Head. Farther east, between the rich alluvial pastures of Pevensey Levels and Walland Marsh (part of Romney Marsh), the Hastings Sands reach the sea, stretching back in a broken range northwest to the Forest Ridge of Ashdown Forest (837 ft. [255 m.] at Crowborough) and St. Leonard's Forest. The South Downs slope gently to the south, with zones of greensand and gault leading to a belt of fertile brick-earth loam extending from near Shoreham-by-Sea to Hampshire and broadening out south of Chichester into the flat headland of Selsey Bill. The northern slopes of the Downs are abrupt, and there the greensand and gault belts lead to the heavy clay of the Weald (*q.v.*). The coastline has changed within historic times, the land being constantly eroded by the sea, particularly around Selsey, where the site of the Saxon cathedral is now probably a mile out to sea. The lagoon of Pagham Harbour was partly reclaimed at an early date but was inundated by the sea about 1300, recovered in 1876, and again lost to the sea in 1911. The drift of shingle tends to force eastward the mouths of the rivers that rise in the Forest Ridge and flow south through gaps in the Downs, but new outlets have in some instances been made by storms, as at Newhaven in *c.* 1590, or by dredging, as at Shoreham-by-Sea in 1815. Wide deposits of shingle now separate the sea from the medieval ports of Winchelsea and Pevensey. (L. F. S.)

Archaeology.—The influence of geology on the distribution of human activities at different stages of culture is striking, and Sussex has an important record of prehistoric sites and discoveries. The Paleolithic period is poorly represented, especially since the Piltdown discoveries, world famous for 40 years, were proved fraudulent in 1953. A few implements have been found in the raised beaches, especially at Slindon, and in the river gravels near Pulborough. The Mesolithic food-gatherers left a scatter of remains along the sand ridges of the Weald and on Tertiary deposits near Peacehaven. Primitive agricultural communities from the Neolithic to the Roman period preferred the chalk, though they hunted in the Weald. Four of the rare Neolithic "causewayed camps" lie on the Downs, the two larger ones being at Whitehawk (Brighton racecourse) and the Trundle (inner ring) at Goodwood. There is also a group of flint-mining sites near Worthing, including the well-known Cissbury and Harrow Hill flint mines. The Early and Middle Bronze Ages are poorly represented but include bell barrows on Bow Hill and on Monkton Down, Treyford,

as well as the vanished Hove Barrow, which yielded a magnificent amber cup and other objects. Remains of the Late Bronze Age are abundant, including several examples of the earliest identifiable farm sites with their field systems, notably at Plumpton Plain near Brighton, New Barn Down near Worthing, and Itford Hill near Lewes. Many bronze implements have been found, singly and in hoards, as well as gold ornaments, including the lost Mountfield treasure near Battle. The Iron Age is notable for the hill forts on the Downs, especially the Trundle (outer ring) at Goodwood, Cissbury near Worthing, and the Caburn near Lewes. The first two were constructed perhaps in the 3rd century B.C. and abandoned before the Roman occupation; the Caburn was rather later and appears to have been captured by the Romans soon after A.D. 43. Farm sites of the Iron Age have been found on Park Brow near Cissbury and elsewhere, and many gold coins of the Belgic kings of the Commius dynasty have been found at Selsey.

(Et. C. C.)

History.—About the beginning of the Christian era a dynasty of British chieftains was established in the peninsula of Selsey. When the Romans invaded Britain in A.D. 43, the last of these chiefs, Cogidumnus, proved so useful an ally that he was made *legatus,* or official administrator, and given an extended kingdom (*regnum*). His centre, Noviomagus (Chichester), was developed as a tribal capital by the Romans, who made a great road (Stane Street) to London and others to Calleva (Silchester) and Venta Belgarum (Winchester), and minor roads connecting the downland corn districts and the Wealden iron mines with London. The fertile coast and downland area west of the Pevensey lagoons was developed with many villas or large farms. During the 3rd and 4th centuries fortresses were built for defense against pirates, one being at Anderida (Pevensey, where half the 20-ft. [6 m.] high walls still stand). After the Roman troops left Britain, about 410, civilization decayed, and in 477 a band of Saxons under Aella (Ella) and his three sons landed near Selsey and occupied the country, slowly fighting their way eastward until in 491 they stormed the last stronghold of Andredesceaster—usually assumed to be Pevensey—and massacred its defenders. The South Saxon conquest seems to have been ruthless and complete: no Celtic place-names survived, except those of three rivers—Tarrant (later Arun), Lavant, and Limene (later Rother). Aella's son Cissa probably occupied the walled Roman city called Chichester after him; but Sussex, the land of the South Saxons, soon fell under the control of Wessex, except for a (probably later) settlement in the east near Hastings, which seems to have retained some independence until 1011, when the Danes are said to have overrun it and the rest of Sussex.

In 681 Aethelwalh, one of the semi-independent kings, received the exiled Bishop Wilfrid, gave him large estates, and encouraged or ordered his subjects to accept Christianity. Wilfrid established a monastic bishopric at Selsey, and during the next two centuries churches were built and endowed throughout the county. Estates on the coast of Sussex were granted to Norman monasteries, notably to the abbey of Fécamp. In 1066, when Harold II was elected king, William I landed at Pevensey and marched to Hastings, near which town the decisive Battle of Hastings was fought on Oct. 14. On the site William built his votive Battle Abbey. Probably in order to secure his communications between Normandy and London, William divided Sussex into five blocks, each running from the coast, with a port, to the northern bounds of the county. Each of these districts, known as rapes, *i.e.,* districts that were "roped off," was in the hands of a trusted Norman lord, whose castles were built at Hastings, Pevensey, Lewes, Bramber, and Arundel. Later, about 1250, a sixth rape, Chichester, was separated for administrative purposes from Arundel. In the Domesday Book of 1086 the liberty of the abbot of Battle is also called a rape, but in all other records it is called the lowy. The boundaries of the rapes ran roughly parallel from north to south, and each rape contained a river, a castle, and a forest. From Domesday it is clear that the northern boundary of Sussex was still uncertain and that scarcely any villages existed in the wooded districts of the Weald, though many areas there were attached to manors in the settled south, serving at first as pastures for swine but being gradually cleared and settled. In 1086 the chief towns were Chichester, to which city the cathedral see had been moved from Selsey in 1075, and Lewes, which as early as 930 had been one of the few boroughs in England to be allowed two mints. Of Hastings, then an important harbour and one of the Cinque Ports (*q.v.*), no details are given. The Norman Conquest brought much building of churches and the foundation of the great Cluniac priory of Lewes and of "alien priories" (cells of Norman abbeys) at Boxgrove, Sele, and Wilmington.

During the reign of Henry III the see of Chichester was occupied by Bishop Richard de Wyche, whose saintly character led to his canonization in 1262, the shrine of St. Richard in the cathedral becoming a centre of pilgrimage. Although the castles of Sussex were held by supporters of King Henry III, it was on the Downs above Lewes that the royal forces were decisively defeated by Simon de Montfort in 1264; but Pevensey Castle held out until the king was restored to power. About this time a series of storms damaged the Cinque Port of Hastings and threatened to destroy the "ancient town" of Winchelsea. Accordingly, Edward I removed the inhabitants of Winchelsea to a safer site, where he laid out a town on a plan of rectangular blocks. The coast, however, soon suffered other assaults, the Hundred Years' War bringing constant raids by the French, who burned Hastings, Rye, and Winchelsea. The plague known as the Black Death hit the county heavily in 1349, killing about a third of the population, and there were further outbreaks in 1361 and 1366. Distress and misgovernment led the men of Sussex to support the Peasants' Revolt in 1381 and to take a prominent part in Jack Cade's Rebellion in 1450.

The Tudor period brought increased prosperity, but its importance in Sussex lies mainly in the realm of religion. On the whole, Sussex accepted the Protestant reform, so much so that during the short reign of Mary I, 30 men and women were burned, at Lewes and elsewhere, as heretics. The memory of these martyrs gave a strongly antipapal, and later puritan, tinge to popular feeling, though there were pockets of Roman Catholicism in the county. The county was therefore predominantly on the side of the Parliament during the Civil War, though Arundel Castle and the city of Chichester were for a time held for the king, and it was from Brighton that Charles II escaped to France after his defeat at Worcester. Puritan tendencies in the diocese are shown by the resignation or expulsion of more than 60 incumbents under the Act of Uniformity in 1662 and by the growth of the Quakers in spite of persecution.

During the last quarter of the 18th century, troubles with France led to the formation of local volunteers, and the movement was speeded up by the French Revolution, many of whose victims sought shelter in the Sussex coastal towns. A great camp was established in 1793 on the Downs near Brighton, and others near Eastbourne and Bexhill; when Napoleon I threatened invasion the coast was fortified by a chain of Martello towers, of which the westernmost was at Seaford. Brighton had since 1750 been developing into a fashionable watering place, especially from the first visit (1783) of the Prince Regent (afterward George IV) and the building of the Royal Pavilion. With the cult of sea bathing other towns developed at Bognor (*see* Bognor Regis), Hastings (*q.v.*), and St. Leonards, Eastbourne (*q.v.*), and Worthing (*q.v.*).

The growth of coastal resorts, the urbanization of almost the whole coast, and the conversion of much of it into "dormitory towns" for London were the main features of the history of Sussex after the beginning of the 19th century.

Architecture.—Of monastic remains the chief are those of Battle Abbey, with its splendid gatehouse, Bayham Abbey, and Lewes Priory. Extensive remains of Robertsbridge Abbey and of the Augustinian priories of Michelham and Shulbrede and the alien priory of Wilmington are incorporated in modern houses. The friary chapels of Chichester, Winchelsea, and Rye are important, as is the medieval hospital of St. Mary at Chichester. Sussex is rich in churches built before or shortly after the Norman Conquest, such as Bosham, Sompting, and Worth. There are many mural paintings; an early group arose under the influence of Lewes Priory.

Pevensey Castle, built within the impressive walls of a Roman

fortress, has an early Norman keep but was mostly built about 1250. At Lewes the early keep and later barbican are noteworthy; at Arundel only the keep is ancient, and Hastings and Bramber are mere fragments. Amberley Castle dates from 1377 and the lovely moated castle of Bodiam from 1386. Herstmonceux Castle (now occupied by the Royal Observatory), a brick building begun in 1444, was rather a defensible great house than a true castle, in this resembling the ruined early Tudor house of Cowdray. Of later great houses, Parham, Petworth, and Up Park (the last two belonging to the National Trust) are outstanding. But it is in its wealth of minor houses that Sussex is particularly strong. The medieval examples are mostly timber framed; later building is mainly in brick, *e.g.*, at Chichester and Lewes; and the good tradition, come down through the charming Regency buildings of Brighton, was revived in the planning of many council estates. (*See* also ARUNDEL; BATTLE; BEXHILL; BRIGHTON; CHICHESTER; HASTINGS; HERSTMONCEUX; LEWES; PETWORTH; PEVENSEY; RYE; SEAFORD; SHOREHAM-BY-SEA; WINCHELSEA.)

Population and Administration.—At the first census in 1801 the population was, in round figures, 159,500; in 1851, 337,000; in 1901, 605,000; in 1951, 937,339; and in 1961, 1,077,517 (East Sussex 665,904, West Sussex 411,613). The respective figures for Brighton were: 7,300; 65,500; 102,300; 156,486; 163,159; and for Hove: 100; 4,100; 29,700; 69,535; and 72,973, showing the progressive urbanization of the county. In East Sussex are three county boroughs: Brighton, Eastbourne (1961, 60,918), Hastings (66,478); four municipal boroughs: Bexhill (28,941), Hove, Lewes (13,645), Rye (4,438); and six urban and five rural districts. West Sussex has three municipal boroughs: Arundel (2,617), Chichester (20,124), Worthing (80,329), and five urban and six rural districts. Crawley (*q.v.*) New Town, designated in 1946, had a population of 54,047 in 1961.

The formation of the two administrative counties of East and West Sussex in 1889 regularized a division that is traceable in Saxon times, corresponding to the two archdeaconries of Lewes and Chichester. The creation of five rapes, each originally with its own sheriff in addition to the king's sheriff of the whole county, obscured this division, but a long struggle over whether the County Court should be held at Chichester or Lewes was compromised in 1504 by an order for alternate sessions. From the reign of Elizabeth I, Quarter Sessions records were kept separately for East and West Sussex, with one joint session yearly. After the grant of the honour of Pevensey to John of Gaunt in 1372, this district became part of the duchy of Lancaster; the privileges of the manor of Bosham and of the Cinque Ports, the ecclesiastical lowy of Battle Abbey, and the peculiars of the archbishop of Canterbury complicated the medieval jurisdiction of the county.

From 1295, members of Parliament were returned more or less regularly by the city of Chichester and the boroughs of Arundel, Bramber, East Grinstead, Horsham, Lewes, Midhurst, Shoreham-by-Sea, and Steyning, and the Cinque Ports of Hastings, Rye, Seaford, and Winchelsea. Many of these constituencies were small and very corrupt, and under the Reform Act of 1832 five lost both members and four lost one, but Brighton was assigned two members. There are four county and four borough constituencies in East Sussex. West Sussex returns four members to Parliament. Brighton returns two members, Eastbourne one, and Hastings one. The University of Sussex was founded at Brighton in 1961, with its headquarters in the 18th-century Stanmer House.

The Economy.—Sussex is still heavily wooded; in early times the Weald was tree covered, the clay producing oaks and the lighter soils favouring beech. The small fields there became good corn-growing areas, after improved drainage was provided. Most arable land, however, has always lain on chalk and greensand and along the coastal plain. This last, especially near Worthing, became a centre of market gardening. Hops are grown in the east, though less widely than formerly. Under John Ellman of Glynde between 1780 and 1830 the breed of Southdown sheep became famous, but since 1870 sheep have almost vanished from the Downs. Cattle have increased, particularly for dairy herds, and the rearing and fattening of poultry is also an active industry. Nor is the harvest of the sea negligible, though less important than in medieval times, when Winchelsea was famous for plaice, Arundel for mullet, and Selsey for cockles, and the Cinque Ports were as much concerned with herring as with naval transport and piracy.

The sea was once the source of the salt industry; a quarter of all salt pans recorded in Domesday Book were in Sussex. Salt boiling depended on fuel, and the Sussex woods supplied fuel also for glassworks round Kirdford and Wisborough Green from the 13th to the 17th century, as well as for pottery, tile, and, later, brick kilns. Sussex oak was in demand for naval and other shipbuilding. The plentiful supply of fuel and iron ore made the county a centre of the iron industry from Roman times until the end of the 18th century. From Tudor times Sussex cast-iron ordnance, as well as domestic objects, were in demand; but when it became possible to use pit coal for smelting, the industry moved north to the coalfields.

Although the Weald was not impenetrable, Sussex roads were notoriously bad, and in winter often impassable, until the coming of turnpike roads toward the end of the 18th century. A railway from London to Brighton was opened in 1841, and the county is now served by the Southern Region of British Railways. A regular cross-channel service to Dieppe is maintained from Newhaven.

Kingley Vale and Lullington Heath are nature reserves; 10,618 ac. (4,297 ha.) of Sussex were owned by the National Trust in 1963 and 1,414 ac. (572 ha.) protected; 15 areas belonged to the Forestry Commission.

BIBLIOGRAPHY.—J. Dallaway and E. Cartwright, *A History of the Western Division of Sussex* (1815–32); T. W. Horsfield, *The History, Antiquities, and Topography . . . of Sussex*, 2 vol. (1835); Sussex Archaeological Society, *Sussex Archaeological Collections* (1846 *et seq.*); M. A. Lower, *The Worthies of Sussex . . .* (1865); W. D. Parish, *A Dictionary of the Sussex Dialect . . .* (rev. 1958); Sussex Record Society Publications (1901 *et seq.*); *Victoria County History of Sussex*, 6 vol. (1905 *et seq.*); A. Mawer, F. M. Stenton, and J. E. B. Gover, *Place-names of Sussex*, 2 vol. (1930); E. Straker, *Wealden Iron . . .* (1931); H. C. K. Henderson and W. H. Briault, *Sussex*, "Land of Britain Series," no. 83 and 84 (1942); I. D. Margary, *Roman Ways in the Weald* (1948); E. C. Curwen, *Archaeology of Sussex*, 2nd ed. rev. (1954). (L. F. S.)

SUSSEX, KINGDOM OF (*Súð Seaxe; i.e.*, the South Saxons), one of the kingdoms of Anglo-Saxon Britain, the boundaries of which coincided in general with those of the modern county of Sussex. According to the tradition given in the Anglo-Saxon Chronicle, it was in 477 that a certain Ella (Aella) led the invaders ashore at a place called Cymenes ora and defeated the inhabitants. A further battle at a place called Mearcredesburne is recorded for 485, and in the annal for 491 we read that Ella and Cissa his son sacked Anderida (Pevensey) and slew the inhabitants. Ella is the first king of the invading race whom Bede describes as exercising supremacy over the southern English.

The history of Sussex now becomes a blank until 607 when Ceolwulf of Wessex is found fighting against the South Saxons. In 681 Bishop Wilfrid of York, on his expulsion from Northumbria by Ecgfrith, retired into Sussex, where he remained until 686 converting its pagan inhabitants. According to Bede, Aethelwalh, king of Sussex, had been previously baptized in Mercia at the suggestion of Wulfhere, who presented him with the Isle of Wight and the district about the river Meon in Hampshire. After Wilfrid's exertions in relieving a famine which occurred in Sussex the king granted to him 87 hides in and near the peninsula of Selsey which, with a lapse until 709 after Wilfrid's retirement, remained the seat of the South Saxon bishopric until the Norman Conquest. Shortly afterward, however, Aethelwalh was slain and his kingdom ravaged by the exiled West Saxon prince Ceadwalla. The latter was eventually expelled by two princes Berhthun and Andhun, who assumed the government of the kingdom. In 686 South Saxons attacked Hlothhere, king of Kent, in support of his nephew Eadric, but soon afterward Berhthun was killed and the kingdom subjugated for a time by Ceadwalla, who had now become king of Wessex.

Of the later South Saxon kings we have little knowledge. In 692 a grant is made by a king called Nothelm to his sister, which is witnessed by two other kings called Nunna and "Uuattus." Nunna is probably to be identified with Nun, described in the Chronicle as the kinsman of Ine of Wessex who fought with him against Gerent, king of the West Welsh, in 710. According to Bede, Sussex was subject to Ine for many years. A grant, dated by Birch about 725, is made by Nunna to Eadberht, bishop of Selsey, and to this too "Uuattus" appears as a witness. In 722 we find Ine of Wessex at war with the South Saxons, apparently because they were supporting a certain Aldbryht, probably an exile from Wessex. An undated grant is made by Nunna about this time, which is witnessed by a King Aethelberht. After this we hear nothing more until shortly before 765, when a grant of land is made by a king named Aldwulf with two other kings, Aelfwald

and Oslac, as witnesses. In 765 and 770 grants are made by a King Osmund, the latter of which is witnessed by Offa of Mercia. Offa also appears as witness to two charters of an Aethelberht, king of the South Saxons, and in 772 he grants land himself in Sussex, with Oswald, *dux* of the South Saxons, as a witness. It is probable that about this time Offa definitely annexed the kingdom of Sussex, as several persons, Osmund, Aelfwald and Oslac, who had previously used the royal title, now sign with that of *dux*. In 825 the South Saxons submitted to Ecgberht, and from this time they remained subject to the West Saxon dynasty.

BIBLIOGRAPHY.—*Anglo-Saxon Chronicle*, ed. by Earle and Plummer (Oxford, 1899); Bede, *Historia Ecclesiastica*, ed. by C. Plummer (Oxford, 1896); W. de G. Birch, *Cartularium Saxonicum* (London, 1885–93); F. M. Stenton, *Anglo-Saxon England*, 2nd ed. (Oxford, 1947); R. H. Hodgkin, *History of the Anglo-Saxons*, 2 vol., 3rd ed. (Oxford, 1953).

SUSU, a tall and slender Negro people of western Africa, about 238,000 of whom dwell in Guinea and 50,000 in the northwestern districts of Sierra Leone. They live in confederate villages under paramount chiefs. Each individual belongs to a patrilineal clan, many of which have names common to the Mande-speaking peoples of this region. Within the village, the houses of particular lineages are grouped together. Marriage between cross-cousins is favoured. Polygynous unions are common and a man inherits his deceased brother's wives. Traditionally, the Susu were animists and had politico-religious confraternities known as *Simo*, but they have now largely adopted the Muslim faith, and, influenced by new economic forces, their way of life is changing.

Susu is also the name of the fresh-water dolphin of the Ganges; it is blind, fish-eating and pale in colour (*see* WHALE).

See A. Arcin, *La Guinée Française* (Paris, 1907); M. McCulloch, *Peoples of Sierra Leone Protectorate* (London, 1950). (M. P. BN.)

SUTHERLAND, EARLS AND DUKES OF. The first earl of Sutherland is generally taken to be a certain William (d. *c.* 1248), whose father, Hugh Freskin (d. *c.* 1214–22), acquired the district of Sutherland about 1197. Probably about 1235 William was created earl of Sutherland. His descendant William, the 5th earl (d. 1370), was a person of some importance in the history of Scotland; he married Margaret (d. *c.* 1346), daughter of King Robert Bruce. His descendant John, the 9th earl, died unmarried in 1514.

John's sister and heir, Elizabeth (d. 1535), countess of Sutherland in her own right, married Adam Gordon (d. 1538), a younger son of George Gordon, and earl of Huntly, and a grandson of King James I, and before 1516 Gordon became earl of Sutherland by right of his wife. He was succeeded by his grandson John (1525–67), who was poisoned at the instigation of George Sinclair, 4th earl of Caithness. His great-grandson John, the 5th earl of the Gordon line (1609–1679), was a strong Covenanter, being called by his associates "the good Earl John"; he fought against Montrose at Auldearn in 1645, but afterward he rendered good service to Charles II. John Gordon (1661–1733), who became the seventh earl in 1703, supported the revolution of 1688 and was a commissioner for the union of England and Scotland. He was a Scottish representative peer in four parliaments, president of the board of trade and manufactures, and lord lieutenant of the eight northern counties of Scotland. He was active in putting down the rising of 1715. This earl, who took the name of Sutherland instead of that of Gordon, was succeeded by his grandson William (1708–50), a representative peer, who helped to suppress the rebellion of 1745. William, the next earl, died without male issue in 1766, his daughter Elizabeth (1765–1839) claiming the peerage. Her title thereto was confirmed by the house of lords in 1771.

Established in the possession of the title and estates of the earldom, the countess of Sutherland was married in 1785 to George Granville Leveson-Gower (1758–1833), who succeeded his father as 2nd marquess of Stafford in 1803. In addition to the estates of the marquessate of Stafford, Leveson-Gower inherited the Bridgewater canal and estates from his maternal uncle, Francis Egerton, 2nd duke of Bridgewater, and these properties, together with his wife's estates, which included almost the whole of the county of Sutherland, made him a "leviathan of wealth," as he is called by Charles Greville. In 1833 he was created duke of Sutherland. Leveson-Gower was a member of parliament from 1778 to 1784, and again from 1787 to 1799 when he was called to the house of lords as Baron Gower. He was British ambassador in Paris from 1790 to 1792. From 1799 to 1801 he was joint postmaster general. He was a collector of paintings, and purchased Stafford house, London. He was responsible for the construction of about 450 mi. of road and of many bridges, but his policy of removing a large number of his tenants from the interior to the coast aroused bitterness and criticism. However, he reduced rents and brought thousands of acres into cultivation. He died at Dunrobin castle on July 5, 1833.

See G. E. C(okayne), *Complete Peerage*, vol. x (London, 1952); Sir Robert Gordon and George Gordon, *Genealogical History of the Earldom of Sutherland . . . to 1651* (Edinburgh, 1813); and also STAFFORD; BUCKINGHAM, EARLS, MARQUESSES AND DUKES OF.

SUTHERLAND, GEORGE (1862–1942), U.S. supreme court justice from 1922 to 1938, and often spokesman for the court when it nullified social legislation, was born in Buckinghamshire, Eng., on March 25, 1862. Brought to Utah soon after he was born, Sutherland attended Brigham Young university and the University of Michigan law school. He was admitted to the bar in 1883. Subsequently he settled in Salt Lake City and became active as a Republican in Utah politics, becoming in 1896 a member of the state senate. He was widely supported by members of the Church of Jesus Christ of the Latter-day Saints (Mormons) although he himself was not of their number. From 1901 to 1903 he was a member of the U.S. house of representatives and from 1905 to 1917 he served in the U.S. senate, his principal legislative accomplishment being the impetus he gave to the 19th (woman suffrage) amendment to the U.S. constitution. He was an ardent exponent of workmen's compensation and legislation for the benefit of seamen. In the 1920 presidential campaign and thereafter until his appointment to the court in 1922, he was an intimate adviser of Pres. Warren G. Harding. On the court, Sutherland quickly became identified with the conservative majority, writing the opinion outlawing the minimum wage (*Adkins* v. *Children's Hospital* 261 U.S. 525) (1923). Paradoxically, despite his hostility to governmental action, his most enduring contribution is probably his conception of the foreign relations power. It was, he successfully argued, not subject to the usual constitutional limitations (*U.S.* v. *Curtiss-Wright Export Corp.*, 299 U.S. 304 [1936]). He retired from the supreme court on Jan. 5, 1938, and died on July 18, 1942, in Stockbridge, Mass.

See J. F. Paschal, *Mr. Justice Sutherland* (1951). (J. F. PL.)

SUTHERLANDSHIRE, a county of northern Scotland, bounded north and west by the Atlantic ocean, east by Caithness, southeast by the North sea and south by the shire of Ross and Cromarty; pop. (1961) 13,442. The land area is 2,028 sq.mi. The western and northern shores are indented with beautiful bays and sea lochs, and terminate at many points in precipices and rugged headlands. Almost the whole county is mountainous; the summits are often of finer outline than those of the Grampians to the south, and the surface generally consists of desolate moorland. The highest point is Ben More in Assynt (3,273 ft.).

An irregular line from Loch Eriboll on the north coast to the neighbourhood of Cromalt near the southern boundary separates the two rock groups that form the foundation of the major portion of the county. On the western side of this line are ancient gneisses and schists (the Lewisian gneiss); these are penetrated by innumerable basic and acid dikes which generally have a northwest to southeast trend. On the eastern side of the line, occupying the whole of the remaining area except the eastern fringe of the county, is a younger series of metamorphic rocks, the Moine schists. In the northwest Torridonian breccias and sandstones rest unconformably on this gneiss, and Cambrian rocks upon the Torridonian; the white Cambrian quartzite capping the dark Torridonian rocks on some of the hills forms a striking scenic feature. Granite masses appear in the eastern schists, and patches of Old Red Sandstone form high land near the southeast coast. Evidence of glacial action is widespread.

The chief river is the Oykell, which, rising in Coniveall or Conival (3,234 ft.), a peak of Ben More, flows south and then south-

east for 25 mi. to Dornoch firth. Other rivers flowing to Dornoch firth are the Helmsdale (20 mi.), the Brora (26 mi.), preserving in its name (bridge river) the fact that its bridge was once the only important one in the county; and the Fleet (17 mi.), the estuary of which was embanked for 1,000 yd., in 1813 by Thomas Telford, whereby rich alluvial land was reclaimed. The Halladale (22 mi.), rising in Knockfin on the borders of Caithness enters the sea to the east of Portskerra. The district of Assynt is honeycombed with lakes and tarns, but the only large lake is Loch Assynt, 6¼ mi. long, 215 ft. above the sea. There are many waterfalls; those of Escallin, near the head of Glencoul, are among the finest in Great Britain.

Of the prehistoric inhabitants, there are considerable remains in the form of many *brochs* (or round towers) "Picts' houses," mounds, cairns and hut circles. Dun Dornadilla, in the parish of Durness, 4 mi. S. of Loch Hope is a tower, 150 ft. in circumference, still in good preservation, and 10 mi. N.W. of the angling resort of Altnaharra.

The Norse jarl Thorfinn overran this part of Scotland in 1034 and the Scandinavian colonists called it, in relation to their settlements in the Orkneys and Shetlands, *Sudrland* ("southern land"). After the conquest of the district by the Scottish kings, an area called Sutherland and including part of the modern county was conferred on Hugh Freskin (a descendant of Freskin of Moray), whose son William was created earl of Sutherland probably about 1235. On the southern shore of Helmsdale creek stand the ruins of the castle in which the 11th earl and his wife were poisoned by his uncle's widow in 1567, with a view to securing the title for her only child, who was next of kin to the earl and his son. Ardvreck castle, now in ruins, near Inchnadamph on the northern side of Loch Assynt, was the prison of the marquess of Montrose after his defeat at Invercarron (1650), from which he was taken by Neil Macleod of Assynt for execution at Edinburgh. In the graveyard of the old church of Durness is a monument to Robert Mackay, called Rob Donn (the brown), the Gaelic poet (1714–78).

Only about 1½% of the total area was arable land in 1938, the shire ranking lowest in Scotland in this respect. The great mass of the surface is grazing ground and deer forest. The best land adjoins Dornoch firth, where farming is in an advanced condition and there are fertile patches along the river valleys. At the beginning of the 19th century the first duke of Sutherland (then marquess of Stafford) adopted a policy of wholesale clearance of the crowded crofters of the interior to the coast. The duke incurred great obloquy, but persisted in his policy, which included reduction of rent, reclamation of land, and abolition of the tacksman or middleman. He also did much to open up the shire generally by the construction of roads and bridges. Attempts were made to repeople some of the glens (Strathnaver, for example) depopulated by the clearances. Crofters still largely predominate; over half the 2,196 holdings in 1955 were 5 ac. or under. The average size, 13 ac., is the smallest in Scotland, excepting the Shetland Islands. Oats, turnips and swedes, and potatoes are grown. The raising of sheep, mostly Cheviot, is the staple business of the county, and cattle are also kept. Deer forests belonging to the duke of Sutherland cover a large area. Next to agriculture, the deep-sea fisheries and the salmon fisheries in the rivers are the most important industries. Helmsdale and Golspie are fishing villages and fishing ports have been established at Loch Clash, Kinlochbervie, and at Culag, Lochinver, both on the west coast. Herrings are the principal catch, but cod, ling and other fishes are also taken. At Brora whisky is distilled, some woolens are manufactured, coal is mined and bricks are made. The exceptional facilities offered by the deer forests, moors and the many lochs and rivers attract large numbers of sportsmen; and Dornoch, Lochinver, Golspie and Brora are well known as holiday towns.

The railway enters the county at Invershin, goes to Lairg, Brora, Helmsdale, Kildonan and Forsinard, after which it leaves the shire. Glasgow steamers call at Lochinver about every ten days.

The population was 13,442 at the time of the 1961 census; the population in 1921 was 17,802. Gaelic and English were spoken by 3,286 persons in 1951. The county is the least densely populated in Scotland. Several islands lie off the west and north coast, but they are uninhabited. Dornoch (pop., 1951, 793), the county town, is the only royal burgh and small burgh. There are six county districts. The shire returns one member to parliament with Caithness, and forms a joint sheriffdom with Ross and Cromarty, with a sheriff-substitute resident at Dornoch.

SUTLEJ (*Zarados* of Ptolemy and Arrian; Vedic *Sutudri* or *Satadru*) one of the "five rivers" of the Punjab. It rises near the westernmost Manasarowar lake in Tibet, threads its way through the Himalayan gorges, crosses Himachal Pradesh, India, and enters Hoshiarpur district. Thence, it flows through the plains of the Punjab, receiving the Beas in the southwest of Kapurthala district and forming for 65 mi. the Indo-Pakistan frontier; it joins the Chenab, 40 mi. W. of Bahawalpur, Pakistan, after a 900-mi. course. The two rivers there form the Panjnad.

As the river flows through an area of precarious rainfall, dams have been built all along it: works include the Bhakra-Nangal multipurpose project; the Sirhind canal; and the Sutlej valley project. Of the five rivers it has the largest catchment area (18,500 sq.mi.) but in relation to its size it has the poorest discharge.

SUTO (Basuto), the occupants of Basutoland and neighbouring areas in South Africa. They are not all of one ethnic stock but consist of those who accept the authority of the paramount chief or king of Basutoland. The Basuto nation was created by the chief Moshoeshoe (known to Europeans as Moshesh) in the 19th century (*see* Basutoland: *History*). The majority are of southern Sotho linguistic stock, with a small minority from the Nguni-speaking tribes (about 15%). By the early 1960s they numbered about 500,000, though about 45% of the adult males were absent from the country as labour migrants.

The people live in small hamlets, which are grouped into parishes under hereditary headmen possessing administrative and judicial powers. Parishes are grouped into chiefdoms with hereditary chiefs, and these form provinces under chiefs, most of whom are close kin of the paramount. Authority at all levels is held by the genealogically senior man of the patrilineal line of descent that is regarded as the founding line of the settlement or chiefdom, the members of which regard their chief as a senior kinsman rather than as a mere administrative official.

The paramount is the senior patrilineal descendant from Mamohate, the senior wife of Moshoeshoe. He is advised both by a council of his close kin and by the national assembly, the *pitso*. He acts as head of the nation in all affairs, as final court of appeal and as ritual guardian of the country.

Corn (maize) is the staple food of the people; sorghum in the lowlands and wheat in the mountain areas are also important crops.

Initiation and formal instruction at puberty are practised, but the Basuto have no age-set system. Marriage is by transfer of bridewealth and preferred marriage is with a cross cousin, although marriage with a parallel cousin is also permitted.

Most Basuto have become Christian. The central feature of the pagan religion was the cult of the ancestors, who were worshiped by their descendants. Today, pagans believe in a Supreme Being, but this is probably a belief taken from Christian missionaries. *See* also Africa: *Ethnography (Anthropology)*: *Southern Africa;* Basutoland: *The People.*

Bibliography.—V. G. J. Sheddick, "The Southern Sotho," *Ethnographic Survey of Africa: Southern Africa*, part ii (1953); E. H. Ashton, *The Basuto* (1952); I. Schapera, *The Bantu-Speaking Tribes of South Africa* (1937), *Government and Politics in Tribal Societies* (1956). (J. F. M. M.)

SUTTEE, the Indian practice of concremation of a widow on the pyre of her deceased husband (Sanskrit *satī*, "true" wife). Whether the usage existed in Vedic India has been hotly debated. World-wide primitive usage compelled a widow to be the wife of her husband's brother or near kinsman, or, if he had been of princely standing, to be immolated with his concubines, slaves, steeds, etc., at his tomb, to maintain his dignity in the next world. This usage eventually found priestly support in the Hindu conception of matrimony as a bond eternally renewed when the married pair were reborn in succeeding lives—a concept hard to reconcile with polygyny. In late Hindu myth Sati was Siva's spouse who, resenting a slight put upon her lord by her father, destroyed herself but was reincarnated as Uma, the beauteous

wife of Siva. Here Sati's act is prompted by a very different motive, and the tale supports the view that the text of the *Rig-Veda* was tampered with, to support the Brahmanical ideal.

From early times reformers denounced the usage, and even the Brahmanical jurists sometimes deprecated suttee, allowing the widow the option of living an ascetic life. The Jains do not practice suttee. Buddhism probably discouraged it. Sikhism expressly forbade it, yet on Ranjit Singh's death, in 1839, several of his widows were burned. Outside British jurisdiction, suttee continued in families of high rank until late in the 19th century.

SUTTER, JOHN AUGUSTUS (1803–1880), U.S. pioneer on whose land the first great California gold discovery was made in 1848. Born of Swiss parents at Kandern, in the grand duchy of Baden, on Feb. 15, 1803, he moved to the U.S. in 1834 and bought land in Missouri. In 1839 he went to California, where he was given 49,000 ac. (19,830 ha.) of land on the Sacramento River by the Mexican governor and was appointed "Commissioner of Justice and Representative of the Mexican Government on the Frontier of the Rio del Sacramento." He built a fort, mill, tannery, distillery, blanket factory, smithy, and other shops. He imported Hawaiians and hired Indians to cultivate the land. He traded for furs, purchased boats that plied the Sacramento, acquired more land, and soon became wealthy. His settlement, "New Helvetia," became a rendezvous for adventurers, sailors, trappers, and hunters, including an increasing number of Americans. In 1846 when the Bear Flag revolt took place, Sutter at first maintained an appearance of fidelity to the Mexican government but made little objection when John C. Frémont's forces took control of his fort. After California was acquired by the U.S., Sutter felt the need of a new mill. It was while the millrace was being dug that an employee made the famous gold discovery on Jan. 24, 1848. Sutter tried to keep the strike a secret, but his property was soon overrun by gold seekers. Worse, the U.S. Supreme Court found the title to 98,000 ac. of his land invalid. Much of what was left him was used to reimburse those to whom he had granted subtitles, which he was forced to make good. Soon he was bankrupt. From 1871 he petitioned Congress for reimbursement. He died at Washington, D.C., on June 18, 1880, the evening after Congress had again adjourned without recognition of his claims.

SUTTNER, BERTHA FÉLICIE SOPHIE, Baroness von (1843–1914), Austrian novelist and pacifist, who won the Nobel Prize for Peace in 1905, was born at Prague on July 9, 1843, the daughter of Count Franz von Kinsky, Austrian field marshal. In 1876 she was hired as a secretary-housekeeper by Alfred Nobel, the munitions maker, but soon married Baron Arthur Gundaccar von Suttner, whose family disapproved of the match because she was seven years his senior and of a poor family. The couple lived at Tiflis in the Caucasus until 1885, when the Von Suttner family was reconciled to the marriage. Though she saw Nobel only twice more, she corresponded with him and is credited with influencing him in the establishment of the Nobel Prize for Peace.

Of her novels, *Die Waffen nieder!* (1889; best Eng. trans. by T. Holmes, *Lay Down Your Arms!*, 1892) was second in popularity in the 19th century only to Harriet Beecher Stowe's *Uncle Tom's Cabin*. *Das Maschinenzeitalter* (1899) reflects the scientific and free-thinking basis of her pacifism as derived from H. T. Buckle, Spencer, and Darwin. She founded the Austrian Society of Peace-Lovers in 1891 and edited the international pacifist journal *Die Waffen nieder!* from 1892 to 1899. She died in Vienna on June 21, 1914.

See Ellen Key, *Florence Nightingale and Baroness von Suttner* (1919).

SUTTON, one of the 32 London boroughs that came into existence on April 1, 1965, under the provisions of the London Government Act 1963, which reorganized local government in the metropolis and the area around it. These 32 boroughs constitute Greater London (*see* London). Sutton, in the south, is composed of areas that were previously in Surrey: the former municipal boroughs of Beddington and Wallington and Sutton and Cheam, and the urban district of Carshalton. *See* Beddington and Wallington; Carshalton; Sutton and Cheam.

SUTTON AND CHEAM, a municipal borough (1934) of Surrey, Eng., which under the London Government Act 1963 was amalgamated on April 1, 1965, with Beddington and Wallington, and Carshalton to form the new London borough of Sutton (*q.v.; see* also London). Sutton and Cheam Municipal Borough was a residential area on the northern edge of Banstead Downs, 11 mi. SSW of Charing Cross. Pop. (1961) 79,008. Area 6.8 sq.mi. It was a parliamentary constituency. In Cheam are many old wooden houses, including Whitehall (early Tudor) and the Old Cottage, a local museum. Cheam was divided by Archbishop Lanfranc into two manors. Henry VIII confiscated both and also Sutton Manor, which belonged to Chertsey Abbey.

SUTTON COLDFIELD, a municipal borough (1886) in the Sutton Coldfield Parliamentary Division of Warwickshire, Eng., 7 mi. NNE of Birmingham. Pop. (1961) 72,165. Area 21.8 sq.mi. It is a residential district with a few light industries, and the British Broadcasting Corporation's television transmitter mast (750 ft.) is a landmark. It was at Sutton Coldfield that John Wyatt, a native, invented a mechanical device in 1733 for spinning cotton. The parish church of Holy Trinity was built in the 14th century and contains the tomb of Bishop Vesey (d. 1555), a great benefactor of the town. The Vesey Memorial Gardens were laid out in 1938; his birthplace, Moor Hall, is a private house. Other old houses are New Hall at Walmley, the Moat House (Queen Anne) and the grove at Wishaw (14th century). Bishop Vesey's grammar school was founded in 1541 and the Roman Catholic seminary of Oscott College in 1785. Its first charter of incorporation (1528) gave the town the right to be described forever as the Royal Town of Sutton Coldfield and also gave 2,400 ac. (including 75 ac. of water) of Sutton Forest, now Sutton Park, to the town.

SUTTON HOO SHIP BURIAL, an early Anglo-Saxon king's grave containing a ship fully equipped for the afterlife (but no body), which came to light in 1939 with excavations at Sutton Hoo, an estate near Woodbridge, Suffolk, Eng. Richest Germanic burial found in Europe since the 17th century, it throws important light on the wealth and contacts of early Anglo-Saxon kings and is particularly valuable because burial of the dead in a ship is rare in England, although described in *Beowulf* (*q.v.*).

The ship's timbers had rotted away, but their impression in the sand and the rivets that remained showed it to be a mastless clinker-built rowboat more than 80 ft. long. The grave goods, laid out amidships, included a sword with a gold and garnet pommel, a gold-mounted shield, a visored helmet enriched with silver inlay, a mail coat, an iron-handled ax, three angons (throwing javelins), four spears, and a knife. Ceremonial objects, showing that the deceased was a king, were a wrought iron stand surmounted with a stag, which is explained as a unique Anglo-Saxon standard, and a great whetstone, 2 ft. long, carved with bearded faces.

The wealth of the obscure kingdom of East Anglia (*q.v.*) is shown by 41 items of solid gold, the fittings for a baldric, and ceremonial regalia. These include a magnificent purse lid decorated with seven ornamental plaques, and a pair of curved and hinged epaulettes whose ornamentation of interlaced beasts and checker patterns is reminiscent of Hiberno-Saxon manuscript art. Small buckles and strap mounts, buttons, and pyramid studs for the sword knot are all lavishly decorated with garnets, and there is an impressive gold buckle weighing more than 14 oz.

Besides the gold cloisonné jewelry, probably made in a single English workshop, there was a quantity of imported silverware. One great silver dish bears the control stamp of the emperor Anastasius I (491–518) and is the only demonstrably Byzantine object so far found in England. In addition, there are a fluted bowl, a set of nine shallow bowls, a plain silver bowl, a ladle cup, and two spoons, probably christening presents, inscribed in Greek with the names Paul and Saul. A bronze bowl of Near Eastern manufacture and six small vessels made from gourds show wide East Anglian trading connections. The skill of the Celtic west is represented by three hanging bowls ornamented with enameled escutcheons, and inside one of these was found the remains of a small rectangular harp, the earliest postclassical instrument known. Domestic objects included bronze cauldrons, ironbound buckets, drinking horns, a pottery bottle, and textiles.

There was no body or cremated remains in the grave, and it is

not certain whose cenotaph this was; its richness indicates a king, and Merovingian coins found in the grave date the deposit 650–670. In that period three kings of East Anglia died: Anna (d. *c.* 653), Aethelhere (d. 654), and Aethelwald (d. 662–663). A pagan burial, albeit with such Christian elements as the spoons, is, however, surprising for any of the kings, especially the latest. Anna was a notable Christian; his body was buried at Blythburgh, Suffolk, but the memorial could have been put up to placate some of his heathen followers. Aethelhere, however, seems the most likely candidate; he died fighting at the River Winwaed for Penda, the pagan king of Mercia, and his body was probably not recovered from the battle where Bede says many perished by drowning.

The rite of ship burial and certain items in the grave are most closely paralleled in Sweden and postulate a hitherto unsuspected Swedish origin for the East Anglian royal dynasty.

The grave goods are now in the British Museum.

BIBLIOGRAPHY.—R. L. S. Bruce-Mitford, *The Sutton Hoo Ship-Burial* (1947), "The Sutton Hoo Ship-Burial," *Proceedings of the Suffolk Institute of Archaeology*, vol. xxv, part 1 (1950), and "The Sutton Hoo Ship Burial," in R. H. Hodgkin, *A History of the Anglo-Saxons*, 3rd ed., vol. ii (1952); C. W. Phillips, "The Excavation of the Sutton Hoo Ship-Burial," *The Antiquaries Journal*, vol. xx (1940); Francis P. Magoun, Jr., "The Sutton Hoo Ship-Burial: a Chronological Bibliography," *Speculum*, vol. xxix (1954); Jess B. Bessinger, Jr., "The Sutton Hoo Ship-Burial: a Chronological Bibliography, Part 2," *Speculum*, vol. xxxiii (1958). (R. J. CR.)

SUTTON-IN-ASHFIELD, a market town and urban district in the Arkfield parliamentary division of Nottinghamshire, Eng., 12 mi. N.N.W. of Nottingham by road. Pop. (1961) 40,441. Area 16.4 sq.mi. The neighbourhood contains remains of Sherwood forest. Formerly the town relied solely on textiles but after the discovery of coal in the vicinity a variety of manufactures was introduced although its chief trade is still hosiery.

SU TUNG-P'O (regular name SU SHIH) (1036–1101) was one of China's greatest poets, painters and prose masters. He pioneered in the Sung school of Chinese painting. He excelled in dignified historical essays as well as in light, whimsical pieces. He was an authority in medicine who picked his own herbs, a great winebibber and maker of wines, a judge who sympathized with his prisoners, an inventor of such famous dishes as "Tung-p'o pork," an inkmaker, a renowned calligraphist whose every scribbled note was carefully preserved and whose family letters were inscribed in stone in his own lifetime. Wherever he was stationed as a magistrate, he dredged lakes, built canals, fought floods and constructed dams. He instituted hospitals and dispensaries, and was the first man to start the system of prison physicians because he could not stand the condition of the prisons. Alone, he fought the drowning of girl babies. He worked for famine relief and wrote successive letters to the empress dowager begging for forgiveness of tax to the poor. He forgave his friend who sent him to exile in Hainan Island, then almost outside the pale of civilization. All these actions made him indubitably the most loved and admired scholar of his and later days.

Of the eight prose masters of the T'ang and Sung dynasties, the Su family accounted for three, Tung-p'o himself, his brother Tse-yu and his father. When he went to the capital to take his examinations, Ou-yang Hsiu, the acknowledged leader of the scholars, said when he read his writings, "I must give place to this young man." His reputation was established. Later the court attendants reported that whenever they saw the emperor's chopsticks stop moving in the midst of eating, they knew it was a memorandum by Su Tung-p'o he was reading. It was his misfortune that the emperor listened to Wang An-shih, the social reformer, who banished all opposition to exile. Su Tung-p'o was first banished to Huangchow for four years (1080–84). There he took the poetic name of Tung-p'o, or Eastern Slope, the name of his farmstead; his regular personal name was Shih. The name Tung-p'o, with which he signed his famous paintings, became the one by which he was popularly known. With the emperor's death, during the Yuanyu regency of the empress dowager (1085–93), he was successively secretary to the emperor, governor of Chekiang, minister of civil service, of war, of education, and governor of Hopei. With the death of the empress dowager, the whole "Yuan-yu party" went again into exile (for seven years). Tung-p'o was banished to Huichow, in Kwangtung, and again to Hainan, whence he returned to China in 1101, one month before his death.

Tung-p'o's position and popularity came from a combination of literary brilliance and political integrity and his fearless championing of the people. His honesty with himself was the "bone" of his stature, while the charm of his personality, his wit, his gaiety and his literary accomplishments were the "skin and flesh" which gave to him the great charm which has endeared him to admirers throughout the generations.

See Lin Yutang, *The Gay Genius* (New York, London, 1947); *Selections from the Works of Su Tung-p'o,* trans. by C. D. LeGros Clark (London, 1931). (L. Y.)

SUVA, the capital, chief port and commercial centre, and largest town, of the British colony of Fiji (*q.v.*), near the mouth of the Rewa river on the southeast coast of Viti Levu, or "great Fiji," the principal island. First occupied by European settlers in 1870, it was selected for the new capital in 1877 although the actual move from Levuka on the island of Ovalau was delayed until 1882. From the original site on the western side of the Suva peninsula the town has spread to the north, west and south, while reclamation has provided space for the expansion of warehouses, factories, wharves and slipways. The rapidly increasing population reached 37,371 in 1956 and included 19,321 Indians, 9,758 Fijians, 3,394 Europeans, 2,094 part-Europeans and 1,647 Chinese.

Administration is by an elected municipal council. Secondary schools and a teachers' training college serve all ethnic groups. The central medical school provides medical and dental courses for native students from Pacific islands as far distant as New Guinea. Besides a general hospital there are special hospitals for mental and tuberculosis cases. Newspapers and periodicals are published in English, Hindustani and Fijian. Industry includes cigarette, soap and biscuit manufacture, copra-crushing and brewing. Electric power is generated locally. Shipping using the port ranges from tourist ships to small interisland vessels. Slipway facilities are widely used by vessels from surrounding island groups. A Royal New Zealand Air Force flying boat squadron is based at Lauthala bay. (R. M. FR.)

SUVÓROV, ALEXANDER VASILIEVICH, COUNT SUVÓROV RIMNIKSKY, PRINCE ITALYSKY (1729–1800), Russian field marshal, was born at Moscow on Nov. 24, 1729, the descendant of a Swede named Suvor who emigrated to Russia in 1622. He entered the army as a boy, served against the Swedes in Finland and against the Prussians during the Seven Years' War. After repeatedly distinguishing himself in battle he was made a colonel in 1762. He next served in Poland, dispersed the Polish forces under Pulawski, stormed Cracow (1768) and was made a major general. In his first campaigns against the Turks in 1773–74, and particularly in the battle of Kosludscki in the latter year, he laid the foundations of his reputation.

From 1787 to 1791 he was again fighting the Turks and won many victories; he was wounded at Kinburn (1787), took part in the siege of Ochakov, and in 1788 won two great victories at Focsani and on the Rimnik. For the latter victory, in which an Austrian corps under Prince Josias of Saxe-Coburg participated, Catherine II made him a count with the name Rimniksky in addition to his own name, and the emperor Joseph II created him a count of the Holy Roman empire.

On Dec. 22, 1790, Suvórov stormed Ismail in Bessarabia, and the sack and the massacre that followed the capture equals in horror such events as the "Spanish Fury" and the fall of Magdeburg. He was next placed at the head of the army which subdued the Poles, and repeated the triumph, and some of the cruelties, of Ismail at Warsaw. He was now made a field marshal, and was retained in Poland until 1795, when he returned to St. Petersburg.

His sovereign and friend Catherine II died in 1796, and her successor Paul dismissed the veteran in disgrace. But in February 1799 he was summoned by the tsar to take the field again, this time against the French Revolutionary armies in Italy.

The campaign (*see* FRENCH REVOLUTIONARY WARS) opened with a series of victories (Cassano, Trebbia, Novi) which reduced the French government to desperate straits and drove every French soldier from Italy, save for the handful under Moreau, which

maintained a foothold in the Maritime Alps and around Genoa. Suvórov himself was made a prince. But the later events of the eventful year went uniformly against the allies. Suvórov's lieutenant Korsákov was defeated by Masséna at Zürich, and the old field marshal, seeking to make his way over the Swiss passes to the Upper Rhine, had to retreat to the Vorarlberg, where the army, much shattered and almost destitute of horses and artillery, went into winter quarters.

Early in 1800 Suvórov returned to St. Petersburg in disgrace. Paul refused to give him an audience, and, worn out and ill, he died a few days afterward, on May 18, 1800. But within a year of his death the tsar Alexander I erected a statue to his memory in the Field of Mars, St. Petersburg.

Suvórov spared his own soldiers as little as he showed mercy to the population of a fallen city. He was a man of great simplicity of manners, and while on a campaign lived as a private soldier. But he had himself passed through all the gradations of military service. His gibes procured him many enemies. He had all the contempt of a man of ability and action for ignorant favourites and ornamental carpet-knights. Dragomirov (*q.v.*) avowed that his teaching was based on Suvórov's practice. Byron has given Suvórov a place in English literature in certain cantos of *Don Juan.*

See Anthing, *Versuch einer Kriegsgeschichte des Grafen Suworow* (1796–1799); F. von Smitt, *Suworows Leben und Heerzüge* (1833–1834), *Suworow und Polens Untergang* (1858); Von Reding-Biberegg, *Der Zug Suworows durch die Schweiz* (1896); Lieut.-Colonel Spalding, *Suvórof* (1890); G. von Fuchs, *Suworows Korrespondenz, 1799* (1835); *Souvorov en Italie,* by Gachot, Masséna's biographer (1903); and the standard Russian biographies of Polevoi (1853; Ger. trans., 1853); Rybkin (1874) and Vasiliev (1899).

SUWANNEE RIVER, the river of Stephen Foster's song "Old Folks at Home," rises in the Okefenokee swamp in southeast Georgia. As the river moves southward through Florida it is augmented by the discharge of many springs, and also receives the Withlacoochee, Alapaha and Santa Fe rivers. The channel is dredged 135 mi. inland from the Gulf to junction with the Withlacoochee; 250 mi. from its headwaters the Suwannee empties into the Gulf of Mexico, 22 mi. S. of Cross City, Fla. All but 35 mi. of the river's course is in Florida. In north Florida the Suwannee flows over shoals and boulders between high limestone banks lined with second-growth cedar. In its middle course it moves quietly through wild hammock and forested areas; in dry weather it forms only a narrow channel in a river bed of limestone strata which have been eroded into numerous curious forms. The banks of the river are heavily wooded with cypress and live oaks. Exotic flowers including gardenia, orchids and ixia are abundant. Bass, perch, catfish, bream and small game are plentiful.

In the 1780s the secluded bays and inlets of Suwannee sound were rendezvous points for pirates. Tales of buried treasure still persist. The Suwannee was the boundary between the Timucua and Apalachee Indian tribes when northern Florida was explored. They called it the Guasaca Esqui, meaning "River of Reeds." The present name is thought to be a Negro corruption of San Juanee, meaning "little St. John." (M. C. P.)

SUWARRO: *see* SAGUARO.

SUWAYDA', AS (Es SUWEIDA; SOUEIDA), town (administrative centre) and governorate (SUWEIDA) of the Jebel Druze (Jabal al Duruz) in the southern part of the Arab Republic of Syria. The town is 55 mi. S.S.E. of Damascus by road. Area (governorate): 2,143 sq.mi.; pop. (governorate, 1960) 99,512, (town, 1960) 18,121. The town, which stands on a hill on the site of a Roman settlement, is connected by road with Izra (Azra') and Damascus. The ancient name for the Jebel Druze was the Hauran. This area originally marked the boundary between Palestine and Syria. The northern part is known as El Lejja (the refuge), as robbers and rebels from the surrounding countries made it their hiding place. The volcanic nature of the soil makes it extremely fertile and wheat, barley, vetches and lentils are grown. Apples and grapes are marketed, but only locally because of poor communications.

The town, whose ancient name was Soada or Dionysias, is said to have been founded by the Nabataeans, but may date from an earlier period. The Nabataeans built a temple there to their god Dusares (later probably erroneously identified with Dionysus). The town came under Roman rule in the 1st century B.C. and Augustus handed it over to Herod the Great (reigned 40–4 B.C.). It remained in the possession of his family until A.D. 100. Herod Agrippa II encouraged the inhabitants to cultivate the soil and build houses. About 15 mi. S.S.W. of As Suwayda' is the ancient site of Bozrah (*q.v.*), a Roman colony under Alexander Severus. (M. V. S.-W.)

SUWON, a city in Kyŏnggi province, Republic of (South) Korea, about 26 mi. S. of Seoul on the Seoul-Pusan main railway. Population (1960) 90,806. Temperatures in the Suwon area range from an average of −5° C. (23° F.) in January to 27° C. (80° F.) in August. Average annual precipitation is about 46 in. Suwon's history as a town site goes back at least to the 5th century, though its present name dates from much later. In the 15th century Suwon was fortified as one of four bastions guarding the approaches to Seoul, and in 1794–96 new fortifications were constructed. Suwon was the provincial capital from 1897 to 1909 and again briefly during the Korean War. It was designated a city in 1949.

Strategically located in the fertile Kyŏnggi plain, Suwon naturally developed as a centre of Korean agriculture. Its economic life is largely concerned with handling the farm products of the surrounding countryside, such as rice, beans, wheat, vegetables, silk cocoons, pigs and, in particular, cattle. Suwon's factories produce mainly cotton and rayon goods, alcoholic beverages and bricks. Several agricultural research facilities are located in the city, such as the Institute of Agricultural Technology, the Sericulture Research station, and the College of Agriculture of Seoul National university. (E. W. WA.)

SUYUTI, AL- (ABU-AL-FADL ABD-AL-RAHMAN IBN ABI BAKR JALAL-AL-DIN AL-SUYUTI) (1445–1505), Arabic encyclopaedic author, whose works deal with almost every subject, the Islamic religious sciences predominating. He was the son of a judge of Asiyut in Egypt. Tutored by a Sufi friend of his father's, he was precocious, being already a teacher in 1462. In 1486 he was appointed to a chair in the mosque of Baybars in Cairo. From 1501 until his death he worked in seclusion on the island of Rauda.

His works number several hundred, and, though many are short booklets, others are large-scale works. He was co-author of *Tafsir al-Jalalain,* a word-by-word commentary on the *Koran,* the first part of which was written by Jalal-al-Din al-Mahalli (1389–1459). His *Itqan fi 'Ulum al-Koran* is a well-known work of koranic exegesis. Among his historical works is a history of the caliphs.

Suyuti was a compiler of genius rather than an original writer, but it is precisely his ability to select and abridge which makes his books so useful. This faculty, together with his keen, analytical mind, characterizes his best work, *al-Muzhir fi 'Ulum al-Lugha,* a guide to the Arabic science of lexicography or philology, in which the subject is treated in 50 aspects or categories. The work has been much used by European Arabists, especially by E. W. Lane (*q.v.*) in the introduction to his lexicon.

See C. Brockelmann, *Geschichte der arabischen Literatur,* vol. 2, pp. 143–158 (1902), and supplement to vol. 2, pp. 178–198 (1938). (J. A. HO.)

SUZDAL, PRINCES OF. In Russian history the area between the Oka river and the upper Volga was called the Suzdal land or *zemlya Suzdalskaya.* In the 10th and 11th centuries A.D. it was a remote corner of the Russian world, but during the 12th and 13th centuries it grew rapidly in political and economic importance, becoming a cradle of the Great Russian nationality. There the immigrants from the Russian south, which was devastated by nomadic invasions, found a quieter life and gradually merged with the scattered and peaceful indigenous Finnish tribes to form the bulk of the Russian peasantry.

The first important prince of the region was Andrew (*q.v.;* Russ. Andrei Yurievich) Bogolyubski, who ruled from 1157 to 1174. He disliked the old established cities of this area—Rostov and Suzdal with their influential boyar class and their old custom of popular assemblies (*veche*)—and moved his capital to Vladimir on the Klyazma river. He successfully colonized and organized the area,

becoming the most powerful among the contemporary Russian princes. His brother Vsevolod III (1176–1212) successfully continued Andrew's policies. Vsevolod's power and influence were felt by other Russian princes, as well as by the rich and proud city-state of Novgorod the Great, which depended on Suzdalia for grain supply. But the political structure erected by Andrew and Vsevolod did not endure. The Suzdal princes regarded their territories as their private property and, accordingly, divided them by testaments among their sons; these separated parts became the appanages or *udely*. During the 13th and 14th centuries Suzdalia thus disintegrated gradually into several lesser principalities, which recognized the seniority of the grand prince of Vladimir only nominally. On the other hand, the Tatar conquest (in 1237–40) made the grand prince of Vladimir, as all other Russian princes, an obedient vassal of the Tatar khan.

Prince Konstantin Vasilievich (1332–55) succeeded in building the grand principality of Suzdal and Nizhni Novgorod, with the capital in Nizhni Novgorod (now Gorki) from 1350; and his son Dimitri even became grand prince of Vladimir in 1359–62. Soon, however, the throne of Vladimir returned to the grand prince of Moscow (*q.v.*); and in 1392 Prince Vasili I Dimitrievich of Moscow, empowered by the khan Tokhtamysh, seized the region of Nizhni Novgorod and annexed it to Moscow's possessions.

See A. V. Ekzemplyarski, *Velikie i udelnye knyazya severnoi Rusi,* 2 vol. (1889–91). (S. G. Pu.)

SUZERAINTY was a term used in the middle ages to describe the relation between the feudal lord and his vassal. According to feudal law the vassal owed certain duties to the lord; he promised fidelity and service, and the lord was bound to perform reciprocal duties, not very clearly defined, to the vassal—*Dominus vassallo conjux et amicus dicitur* ("a lord is called the partner and friend of his vassal"). (*See* Feudalism and articles there cited.)

In modern times the term has been used by analogy to describe a relationship of dependency between states of unequal power, the superior state being the "suzerain" and the dependent state the "vassal." W. E. Hall described the relationship thus: "States under the suzerainty of others are portions of the latter which during a process of gradual disruption or by the grace of the sovereign have acquired certain of the powers of an independent community, such as that of making commercial conventions, or of conferring their exequatur on foreign consuls." (*International Law,* 8th ed., p. 32.) Suzerainty, however, is not to be confused with sovereignty. "It is a kind of international guardianship, since the vassal state is either absolutely or mainly represented internationally by the suzerain State." (L. F. A. Oppenheim, *International Law,* 8th ed., ed. by H. Lauterpacht, vol. i, p. 189 [1955].)

On the whole, usage favours a distinction between suzerainty and protectorate in that a protectorate flows from, or is a reduction of, the sovereignty of the protected state, while suzerainty is conceived as derived from, and a reduction of, the sovereignty of the dominant state. Moreover, a protectorate usually implies the existence of a treaty defining the relationship. It is not possible to ascribe any precise juristic significance to the term suzerainty, however. In any particular case the terms of the relationship must be deduced from the events or instruments by which it came into being, and not from the term used to describe it.

The subject is now of no more than historical interest because there are no longer any vassal states remaining. Examples of states under suzerainty were Bulgaria from 1878 to 1908 and Egypt from 1840 to 1914, during which periods both these countries were under the suzerainty of Turkey. The term was also used to describe the relationship which formerly existed between Great Britain and the Indian native states (*see* Interpretation act, 1889, s. 18). These states had no international relations whatever, either between themselves or with foreign states. Yet the rulers of these states successfully claimed in English courts the privilege and immunities which are by international law accorded to the heads of states when abroad; see *Statham* v. *Statham and the Gaekwar of Baroda* [1912] P. 92 and *Sayce* v. *Amir of Bahawalpur state* [1952] AU E.R. 32b. *See* Mandate; Protectorate; Sovereignty; Spheres of Influence.

Bibliography.—C. Stubbs, *Suzerainty: or the Rights and Duties of Suzerain and Vassal States* (1882); G. Jellinek, *Die Lehre von den Staatenverbindungen* (1882); C. Bornhak, "Einseitige Abhängigkeitsverhältnisse unter den modernen Staaten" in G. Jellinek and G. Meyer's *Staats- und völkerrechtliche Abhandlungen,* vol. i, pt. 5 (1896); J. Westlake, *International Law,* pt. I, pp. 25–27 (1904–07); A. G. Merignhac, *Traité de droit public international,* 2 pt. (1905–07); E. von Ullmann, "Völkerrecht" in G. Jellinek's *Das öffentliche Recht der Gegenwart,* Bd. 3 (1908); M. G. Georgantas, *De la notion de Souveraineté et de son évolution* (1921); L. Duguit, *Souveraineté et Liberté* (1922); L. F. A. Oppenheim, *International Law,* 8th ed., vol. i, pp. 188–191 (1955). *See* also *Correspondence Relating to the South African Republic,* c. 9507 (1899); L. Westlake, "Angleterre et la république sud-africaine" in *Revue de droit international,* p. 268 (1896); *Law Quarterly Review,* vol. xii, p. 113 (1896); *Law Magazine and Review,* 5th series, vol. xxv, p. 413 (1899–1900); *Journal of the Society of Comparative Legislation,* new series, vol. i, p. 432 (1899). (R. Y. J.)

SUZUKI, DAISETZ TEITARŌ (1870–), renowned interpreter of Zen Buddhism, especially to the west, was born Oct. 18, 1870, in Kanazawa, Japan. During his college days in Tokyo he was trained under two great Zen masters, Imakita Kōsen and Shaku Sōyen. Shaku Sōyen represented Mahayana Buddhism at the World Parliament of Religions, held in Chicago, Ill., in 1893, and there he met Paul Carus (1852–1919), a philosopher interested in oriental religions. Subsequently, Shaku Sōyen recommended Suzuki to become Carus' assistant, and Suzuki went to La Salle, Ill., where he helped Carus in the translation of the *Tao Te-Ching* and other works. During his 11 years at La Salle, Suzuki translated Asvaghosha's *Discourse on the Awakening of Faith in the Mahayana* (1900) and also completed his *Outlines of Mahayana Buddhism* (1907). He visited Europe in 1908. Upon his return to Tokyo in 1909, he taught English at the Peers' school as well as at Tokyo Imperial university.

In addition to his interest in Zen Buddhism, Suzuki was also influenced by Ralph Waldo Emerson. In 1921 he accepted the chair of religious studies at Ōtani university, Kyōto (emeritus 1960), where he also served for 20 years as the editor of an English-language journal, *The Eastern Buddhist.* In 1936–37 he was sent by the Japanese government to lecture on Zen and Japanese culture in Europe and America. Again after World War II Suzuki traveled abroad, participating in the Conferences on Philosophy East and West at Hawaii (1949 and 1959) and lecturing at U.S. and European universities. In 1960 he toured India as a guest of the Indian government.

Suzuki is the author of more than 130 major books and numerous articles, written in Japanese and English. Important among his English writings are: *Essays in Zen Buddhism,* three volumes (1927–34), *Studies in the Lankāvatāra Sūtra* (1930; his doctoral dissertation), *An Index to the Lankāvatāra Sūtra,* Nanjiō edition (1933 and 1934), *The Training of the Zen Buddhist Monk* (1934), *An Introduction to Zen Buddhism* (1934), *Manual of Zen Buddhism* (1936), *The Zen Doctrine of No-Mind* (1949), *Living by Zen* (1949), *Studies in Zen* (1955), *Mysticism: Christian and Buddhist* (1957) and *Zen and Japanese Culture* (1959). The Japanese government in 1949 conferred upon him a cultural medal and made him a member of the Japan academy. (S. Fu.)

SVALBARD is part of the kingdom of Norway and comprises all islands in the Arctic ocean between long. 10° and 35° E. and lat. 74° and 81° N. (Vestspitsbergen, Nordaustlandet [North East Land], Edgeøya [Edge Island], Barentsøya [Barents Island], Prins Karls Forland [Prince Charles Foreland] and a number of smaller islands), Bjørnøya (Bear Island), Hopen (Hope Island), Kong Karls Land (Wiche Islands) and Kvitøya (Gillis Land). The total area of Svalbard is 23,958 sq.mi. Spitsbergen (*q.v.*) covers 23,641 sq.mi., Bjørnøya 69, Kong Karls Land 128 and Kvitøya 102 sq.mi., respectively. Sovereignty over Svalbard was given to Norway by a treaty signed in Paris on Feb. 9, 1920, and the Norwegian colours were hoisted in Longyearbyen on Aug. 14, 1925. Svalbard is administered from Norway partly by the ministry of industry and partly by the ministry of justice. A governor (*sysselmann*) has his residence in Longyearbyen. He is also chief of police, notary public and assistant judge. An inspector of mines (*bergmester*) supervises mining activities. Norwegian criminal and civil law is in force. A mining ordinance gives the subjects of all signatory powers to the treaty equal rights to

exploit mineral deposits.

The name Svalbard means "the cold coast," an Old Norse name. According to the *Icelandic Annals,* Svalbard was discovered in 1194, four days' (*daegra*) sailing from Cape Langanes on Iceland. It is probable that this land was identical with Spitsbergen. The name Svalbard was introduced as an administrative name for all the islands when Norway assumed sovereignty. (A. K. O.)

SVEDBERG, THEODOR (THE) (1884–), Swedish chemist and Nobel prize winner in 1926, is distinguished for the development of physical chemical methods of studying colloids and macromolecules, particularly that of ultracentrifugation. He was born at Fleräng, near Valbo, Swed., on Aug. 30, 1884. From 1904 he studied chemistry at Uppsala university, where he became professor of physical chemistry (1912–49), professor emeritus (1949), director of the Institute of Physical Chemistry (1931–49) and director of the Gustaf-Werner Institute of Nuclear Chemistry (1949).

The development of the ultracentrifuge made a profound and timely contribution to the study of large molecules. In this instrument the migration of large molecules in solution under the influence of a strong centrifugal force (up to 100,000 times gravity) can be observed and related to the molecular size and shape. By this means Svedberg demonstrated that the molecules of certain pure proteins are all of one size. Thus, the ultracentrifuge can sometimes, in the characterization of a new protein, show the presence of contaminating substances. In the field of high polymers the behaviour of long chain molecules, particularly cellulose, in the ultracentrifuge has contributed to the elucidation of their properties. For these and other studies in colloidal systems Svedberg was awarded the Nobel prize. (K. A. SY.)

SVENDBORG, a seaport of Denmark, capital of the *amt* (county) of its name, on the island of Fyn. Pop. (1960) 23,824. The narrow Svendborg Sund separates Fyn from the lesser islands of Taasinge (245 ft.) and Turø. Inland there is also elevated ground, the Ovinehøi. The harbour is accessible to vessels drawing 20 ft. Near the town are the ruined castle of Ørkild, the watering place Christiansminde, and the orchards of Gammel Hestehave, where wine is produced.

SVENDSEN, JOHANN SEVERIN (1840–1911), Norwegian composer and conductor who helped establish the Norwegian national school of music. Born at Christiania (Oslo), Sept. 30, 1840, he started his career as a military bandmaster; a scholarship enabled him to study at the Leipzig conservatory. After conducting symphonic concerts in Oslo for several years, he became court *Kapellmeister* in Copenhagen in 1883, where he lived until his death on June 14, 1911. Svendsen, regarded as the most outstanding Norwegian conductor of his time, was certainly one of the most distinguished composers of Scandinavia. His chamber works include a string octet and his orchestral works four Norwegian rhapsodies, based on Norwegian folk tunes, the overture *Carnival in Paris* and the *Norwegian Artists' Carnival.*

SVERDLOVSK, an *oblast* of the Russian Soviet Federated Socialist Republic, U.S.S.R., was formed in 1934. Pop. (1959), 4,044,416. Area 74,440 sq.mi. The *oblast* is roughly triangular, with its western side lying along the crest line of the Urals, which within the *oblast* reach a maximum height of 5,148 ft. in Gora Konzhakovski Kamen. The *oblast* includes the eastern slope of the Urals and stretches out to the flat West Siberian plain (Zapadno-Sibirskaya Nizmennost). Its southwestern corner extends across to the western flank of the Urals in the upper Ufa basin. The main rivers draining eastward are the Tavda, with its headstreams Lozva and Sosva, and the Tura, with its tributaries Nitsa and Pyshma. The climate is sharply continental, January average temperatures ranging from −16° C. (3° F.) in the southwest to −20° C. (−4° F.) in the north, and July averages about 16°–17° C. (61°–63° F.). Rainfall is about 16 in. a year on the lowland, rising to 24 in. on the Urals. Almost the entire *oblast* is covered in taiga forest of pine, stone pine, birch and larch, with spruce, fir and oak appearing in the southwest. Only the highest summits of the Urals are bare; in the extreme south the *oblast* enters the forest-steppe zone, with alternating birch groves and grass steppe.

Of the 1959 population, more than three-quarters were urban dwellers, a reflection of the *oblast's* role as part of the Urals industrial region. The largest of the 39 towns and 88 urban districts are the administrative centre Sverdlovsk (778,602), Nizhni Tagil (338,501), Kamensk-Uralski (141,019), Serov (97,882), Asbest (*qq.v.;* 60,053), Pervouralsk (91,034), Krasnoturinsk (61,990), and Revda (54,870).

Like all the Urals region, the *oblast* is exceptionally rich in the range and quantities of minerals. Iron ore is mined in the Serov-Ivdel area, at Nizhni Tagil and at Kushva. Manganese occurs at Ivdel; chromium at Kluchevsk; cobalt, nickel, tungsten and vanadium near Sverdlovsk; beryllium near Asbest. Copper is mined at several places including Alapayevsk, Degtyarsk, Krasnouralsk and Kirovgrad. Asbest has the largest deposits of asbestos in the U.S.S.R. Bauxite is found at Kamensk-Uralski and Serov, and Polevskoi is a major producer of artificial cryolite. Platinum and gold are widespread, and zinc and sulfur deposits are worked. Coal and lignite are of inferior quality. Based on these minerals a large metallurgical industry developed in the 1930s and received further impetus during World War II. Nizhni Tagil has one of the largest iron and steel plants of the Soviet Union and smaller works are located at Serov, Kushva, Pervouralsk, Revda and Severski. Sverdlovsk and Nizhni Tagil are centres of large-scale heavy engineering. Nonferrous metals are processed in Krasnoturinsk, Krasnouralsk, Kirovgrad, Kamensk-Uralski and Revda (copper). There is a thriving chemical industry producing industrial chemicals, plastics, fertilizers and medical preparations. Timber working is widespread and well developed. The chief lack of the *oblast* is fuel. Coking coal is brought from Karaganda (Kazakh S.S.R.) and the Kuzbass (Kuznetsk basin) and natural gas is piped from Berezovo in the northern Urals. Agriculture is relatively unimportant and largely confined to the southeast (around Irbit), where grains and potatoes are grown. Market gardening is important around the towns.

All the main centres are connected by a rail network which is linked to the Trans-Siberian railroad. (R. A. F.)

SVERDLOVSK (formerly YEKATERINBURG), a town and administrative centre of the *oblast* of the same name in the Russian Soviet Federated Socialist Republic, U.S.S.R., lies on the eastern slope of the Ural mountains on both banks of the Iset river, a tributary of the Tobol, 850 mi. E. of Moscow. Pop. (1959) 778,602. The town has two 18th-century cathedrals and is the leading cultural and educational centre of the Urals region. It has a musical *conservatoire* and opera and dramatic theatres; a polytechnic; and institutes of mining, forestry, agriculture, medicine, law and teaching. It is also the seat of the Urals A. M. Gorki State university (founded 1920) and of the Urals branch of the U.S.S.R. Academy of Sciences.

Sverdlovsk is an important railway junction. Seven lines radiate from it to all parts of the Urals and the rest of the Soviet Union, including Moscow, Leningrad and Omsk. Its chief importance lies in heavy engineering, for which it is one of the largest centres of the U.S.S.R. Among its wide range of engineering products are machines for the metallurgical and chemical industries; mining and oil drilling equipment; turbines, diesels, ball bearings, electrical equipment, cables, precision instruments, refrigerators and bicycles. The chemical and rubber industries (plastics, tires) are well developed. Steel manufacture, utilizing pig iron from other centres and scrap, is still maintained but is less important than in other Ural centres. Light industrial products include footwear and other leather goods; clothing, tobacco and foodstuffs. A traditional craft is the cutting of diamonds and other precious stones.

Sverdlovsk has its origin in the establishment there in 1721 of the Verkhne-Isetski (Upper Iset) ironworks and fortress by V. N. Tatischev, historian, traveler and government functionary; and the settlement became the administrative centre for all the ironworks of the Urals. It was named Yekaterinburg in 1723 for Catherine I, wife of Peter I the Great. Other ironworks developed in the vicinity, and its industrial and administrative importance was increased from 1783 when the Great Siberian highway was built through it. The first railway arrived from Perm in 1878 and

several other lines were laid from it during the next fifty years. After the Revolution Yekaterinburg was the last place of imprisonment of the tsar Nicholas II and his family, who were shot there on the night of July 29–30 (N.S.; July 16–17 O.S.), 1918, to prevent their escape to Kolchak's White Russian army which was advancing toward the town. Yekaterinburg fell to the White troops a week later and was recaptured by the Red army in July 1919. It was renamed Sverdlovsk for the Bolshevist Yakov Sverdlov in 1924. Evacuation of industry to the Urals during World War II gave further impetus to expansion. (R. A. F.)

SVERDRUP, JOHAN (1816–1892), Norwegian statesman, was born at Jarlsberg on July 30, 1816. His father, Jakob Sverdrup, was a land steward, and the founder of the first school of agriculture in Norway. Johan entered the *storting* in 1850. He built up a strong political party, supported by the peasantry, and in 1872, secured a bill for the admission of ministers to the *storting*, which was a step to parliamentary control. King Charles XV refused his sanction to this bill, and on its third passing in 1880 Oscar II opposed his absolute veto. Sverdrup then proposed the proclamation of the law in defiance of the king's action. The retirement of Frederik Stang removed Sverdrup's chief political opponent from the field. He was aided in his campaign by Bjørnstjerne Bjørnson, and after a series of political crises he became prime minister in June 1884. He soon found himself at issue with Bjørnson on church matters; although during his term of office no fewer than eighty-nine measures, many of them involving useful reforms, became law, he failed to satisfy the extremists and was driven to rely on the moderate Liberals. He was compelled to retire in 1889, and died on Feb. 17, 1892, at Christiania.

SVYATOSLAV I (Svyatoslav Igorevich, d. 972), grand prince of Kiev from 945 and a hero of early Russian history, was the son of Igor and Olga (*qq.v.*). After coming of age, he dedicated all his efforts to bold military expeditions, leaving his mother, Olga, in Kiev to take care of the internal affairs of the state until her death in 969. The Russian chronicle describes the last heathen Russian prince, who "stepping light as a leopard . . . carried with him neither wagons nor kettles . . . nor did he have a tent. . . . He sent messengers to the other lands announcing his intention to attack them." Between 963 and 965 he defeated the Khazars on the lower Don and the Ossetians and Circassians in the northern Caucasus and stormed the capital of the Volga Bulgars. When the Byzantine government, engaged in a war against the Balkan Bulgars, asked Svyatoslav for help, he invaded Danubian Bulgaria in 967 and defeated the Bulgars. He then refused to cede his conquests to the Byzantines, declaring his intention of establishing his capital in Pereyaslavets (Pereyaslav-Khmelnitski) on the Danube and founding a Russo-Bulgarian empire. The Byzantine emperor John I Tzimisces, having gathered overwhelming forces, defeated Svyatoslav's relatively small army and forced him to abandon any claims to Balkan territory (971). In the spring of 972, while Svyatoslav was returning to Russia with a small retinue, he was ambushed by the Pechenegs near the cataracts of the Dnieper and killed.

See *The Russian Primary Chronicle*, Eng. trans. (1953); G. V. Vernadsky, *Kievan Russia* (1948). (S. G. Pu.)

SWABIA, Suabia or Suevia (Ger. *Schwaben*), one of the stem-duchies of mediaeval Germany, taking its name from the Suevi, a tribe who inhabited the district in the first century of the Christian era. They were joined by other tribes, and the district was called Alamannia, until about the 11th century, when the form Swabia began to prevail. In 496 the Alamanni were defeated by Clovis and governed by dukes dependent on the Frankish kings. In the 7th century the people were converted to Christianity, bishoprics were founded at Augsburg and Constance, and in the 8th century abbeys at Reichenau and St. Gall. The Alamanni had gradually thrown off the Frankish yoke, but in 730 Charles Martel again reduced them to dependence.

The duchy, bounded by the Rhine, the lake of Constance, the Lech and Franconia, was ruled by the counts of Raetia, one of whom, Burkhard, took the title of duke in 917 and was recognized as such by King Henry I., the Fowler, in 919. On his death in 926 he was succeeded by Hermann, a Franconian noble, who married his widow. When Hermann died in 948 Otto the Great gave the duchy to his own son Ludolf, who had married Hermann's daughter Ida; but he reduced the ducal privileges and appointed counts palatine to watch the royal interests. Ludolf revolted and was deposed, and other dukes followed in quick succession. During these years the Swabians were loyal to the kings of the Saxon house. In 1077, Rudolph, count of Rheinfelden, as duke, was chosen German king in opposition to the emperor Henry IV., but found little support in Swabia, which was given by Henry to his faithful adherent, Frederick I., count of Hohenstaufen. Frederick II. succeeded his father in 1105, and was followed by Frederick III., afterwards the emperor Frederick I. The earlier Hohenstaufen increased the imperial domain in Swabia, where they received steady support, although ecclesiastical influences were very strong. After numerous changes, the chief authority in Swabia in 1268 fell to the counts of Württemberg, the margraves of Baden, the counts palatine of Tübingen and the counts of Hohenzollern.

SWABIAN LEAGUE, an association of German cities, principally in the old duchy of Swabia. The cities had attained prosperity under the Hohenstaufen emperors, but on the extinction of that house in 1268, they were ill-defended against aggression by succeeding dynasties.

In 1331, 22 Swabian cities, including Ulm, Augsburg, Reutlingen and Heilbronn, formed a league at the instance of the emperor Louis the Bavarian, who in return for their support promised not to mortgage any of them to a vassal. The count of Württemberg was induced to join in 1340. Under Charles IV. the lesser Swabian nobles began to combine against the cities, and formed the *Schlegelerbund* (from *Schlegel*, a maul). Civil war ensuing in 1367, the emperor, jealous of the growing power of the cities, endeavoured to set up a league under his own control, for the maintenance of public peace. The defeat of the city league by Eberhard II. of Württemberg in 1372, the murder of the captain of the league, and the breach of his obligations by Charles IV., led to the formation of a new league of 14 Swabian cities led by Ulm in 1376. This league triumphed over the count of Württemberg at Reutlingen in 1377, and the emperor having removed his ban, it assumed a permanent character, set up an arbitration court, and was rapidly extended over the Rhineland, Bavaria and Franconia. Its professed aims were the maintenance of the imperial status of the constituent cities, the security against sale or mortgage and against excessive taxation, the protection of property, trade and traffic, and the power to suppress disturbances of the peace. There is no trace of co-operation with the Hanseatic towns.

For nearly a century there was no great effort at federation among the Swabian cities, though there were partial and short-lived associations. The growing anarchy in Swabia, where the cities were violently agitated by the constant infringement of their liberties (*e.g.*, the annexation of Regensburg by Bavaria in 1486), induced Frederick III., who required men and money for the Hungarian War, to conciliate the cities by propounding a scheme of pacification and reform. His commissioner, Count Hugo of Werdenberg, met the Swabian estates at Esslingen, and on Feb. 14, 1488, the Great Swabian League was constituted. There were four constituent parties, the archduke Sigismund of Austria, Count Eberhard V. of Württemberg, who became the first captain of the league, the knightly league of St. George, and lastly 22 Swabian imperial cities. The league received a formal constitution with a federal council consisting of three colleges of nine councillors each, a captain and a federal court with judicial and executive powers. The armed force which was to police Swabia consisted of 12,000 foot and 1,200 horse, each party contributing one-fourth. The league gained strength by the speedy accession of Augsburg and other Swabian cities, the margraves of Brandenburg-Ansbach, Bayreuth and Baden, the four Rhenish electors, etc., and in 1490 of Maximilian, king of the Romans, whom the league had helped to rescue from the hands of the Netherlanders in 1488. It did not render him the support he expected in his foreign policy, but it performed its primary work

of restoring and maintaining order with energy and efficiency. In 1492 it compelled Duke Albert of Bavaria to renounce Regensburg; in 1519 it expelled the turbulent duke, Ulrich of Württemberg who had seized Reutlingen, and it sold his duchy to Charles V; and in 1523 it defeated the Franconian knights who had taken up arms with Franz von Sickingen. In 1525 Truchsess, the league captain, aided by the forces of Trier and the palatinate, overthrew the rebel peasants of Königshofen on the Tauber and at Ingolstadt. The league, after several renewals, expired on Feb. 2, 1534, its dissolution being due to the Reformation. Futile attempts were made to renew it, in 1535 by the Bavarian chancellor, Eck, and in 1547 by Charles V.

See E. Osann, *Zur Geschichte des schwäbischen Bundes* (1861); K. Klüpfel, *Urkunden zur Geschichte des schwäbischen Bundes* (1846–53), "Der schwäbische Bund," *Hist. Taschenbuch* (1883–84).
(A. B. G.)

SWAHILI, *Wa-Swahili* from Arabic *sahil,* "coast," a term applied to the inhabitants of the East African coast from Mogadishu to Kilwa and of the islands of Zanzibar, Pemba and Comoro. They are a mixture of African and Arab, the African element being based on a local Bantu people with a large admixture of slaves from all over East and Central Africa. Their language, Ki-Swahili, has a purely Bantu grammar and a large Arabic vocabulary.

The Swahili are not a tribe, but a people united by a common language and religion (Islam). Though their culture is largely Islamic, there remain many traces of their African origin and some Indonesian and other non-African elements. The chief Swahili dialects are Ki-Unguja of Zanzibar (now the standard form); Ki-Mvita of Mombasa; and Ki-Amu of Lamu. Older forms are Ki-Ngozi and Ki-Kae. A large body of Swahili poetry and prose in Arabic comprises secular and religious works. (G. W. B. H.)

SWAHILI LANGUAGE. Swahili (Kiswahili) is a Bantu language spoken primarily on the east coast and islands of Africa between Lamu Island in the north and the mouth of the Ruvuma river in the south. It is spoken sporadically as far north as Mogadishu and there appear to be Swahili-speaking enclaves in Somaliland; it has also been heard as far south as Mozambique. A form of Swahili is also spoken on the west coast of Madagascar. People who speak Swahili as their sole mother tongue are usually referred to as Waswahili, but those who live in the larger towns prefer to be known as Waarabu ("Arabs"). There are about 15 main Swahili dialects as well as several pidgin forms in use.

Swahili as a lingua franca is current in: (1) Tanganyika, where it is the language of administration and primary education; (2) Kenya, where it is also the main language of administration and primary education, though alternative languages have been recognized in certain parts of the colony; (3) Uganda, where it is in competition with local languages; (4) Republic of The (former Belgian) Congo, where a form of Swahili (sometimes called Kingwana) is one of the four permitted languages of administration and where its use in education depends largely on mission society policy, although there are a few communities where Swahili appears to be the mother tongue. Swahili is regarded, with Nyika (including Giryama), as a language group closely related to the Taita language group.

Although modern Swahili contains an enormous vocabulary of Arabic borrowings (as well as borrowings from Persian, Hindi, Portuguese and English), in its grammatical behaviour it is characteristically Bantu; *i.e.,* (1) it has a large vocabulary of word roots traceable to a common Bantu stock; (2) its nouns are divided into classes, distinguished by typical prefixes, for the most part groupable in pairs denoting singular and plural; *e.g., m-tu,* pl. *wa-tu* ("person"), *ki-tu,* pl. *vi-tu* ("thing"), etc.; (3) verbs, adjectives, demonstrative and possessive forms are brought into agreement with nouns by means of prefixes; *e.g., wa-tu w-etu wa-le wa-kubwa wa-mekuja* ("those big people of ours have come"); (4) verb stems may be extended by means of suffixes, each with a particular nuance of meaning; *e.g., funga* ("shut"), *fungwa* ("be shut"), *fungika* ("become shut"), *fungia* ("shut for"), *fungisha* ("cause to shut"), *fungua* ("open"), etc.

Origin and History.—The name Swahili is derived from the Arabic *swahil* ("coasts"). Contact between Arabia and the east coast of Africa was made as early as the 1st century A.D. when, according to the *Periplus of the Erythraean Sea,* the people of Muza (in southern Arabia) "sent thither many large ships, using Arab captains and agents, who are familiar with the natives and intermarry with them." But the earliest known settlement at Pate is said to have been founded in A.D. 689. During the next 600 years other cities such as Lamu, Malindi, Mombasa and Kilwa were founded and reached a high level of civilization, until eclipsed by the Portuguese conquests in the early 16th century. After the Arabs had returned and ousted the Portuguese in the early 18th century and transferred their court from Oman to Zanzibar in 1832, Swahili came into its own, enhanced by the intermarriage of the Arab overlords with the coastal people.

During the early 19th century the spread of Swahili inland received a great impetus through its being the language of the Arab ivory and slave caravans, which penetrated as far north as Uganda and as far west as the upper reaches of the Congo (Lualaba) river. It was later adopted by Europeans. It has been challenged by local vernaculars in certain regions and there has also been a tendency among some Africans to prefer the metropolitan language as lingua franca. Others, however, look to Swahili as the future national language of East Africa.

Literature.—The earliest literature in Swahili was written in the Arabic script and a certain amount still exists in manuscript. Some has been published. In the latter half of the 19th century, the roman alphabet was successfully applied to Swahili by missionaries and the first complete Bible was produced by 1891 in the Mvita (Mombasa) dialect. For several years there was strong rivalry between this dialect and that of Zanzibar (Unguja), until the latter obtained official recognition as "standard" in the 1920s. Afterward the written language progressed apace, under the encouragement of the Interterritorial Language (Swahili) committee —later the East African Swahili committee—and the years after World War II saw an enormous development. Besides further Bible translations (including the "Union Version" of 1950) and a spate of government-inspired textbooks for schools and adult literacy campaigns, there are about 60 Swahili newspapers and broadsheets in East Africa and a dozen in the Congo area. Indigenous authors have gradually come to the fore, among whom Shaaban Robert is outstanding. *See* BANTU LANGUAGES.

BIBLIOGRAPHY.—E. Steere, *A Handbook of the Swahili Language,* rev. by A. B. Hellier (1943); J. L. Krapf, *Dictionary of the Suahili Language* (1882); W. E. Taylor, *African Aphorisms* (1891); C. Sacleux, *Grammaire des dialectes swahilis* (1909) and *Dictionnaire swahili-français et français-swahili,* 3 vol. (1939–41); C. Velten, *Suaheli Wörterbuch,* 2 vol. (1910–33); G. W. Broomfield, *Sarufi ya Ki-Swahili* (1931); F. Johnson, *Kamusi ya Kiswahili* (1935) and *Standard Swahili-English Dictionary, Standard English-Swahili Dictionary,* 2 vol. based on the dictionaries of A. C. Madan (1939); A. Seidel, *Suaheli Konversations-Grammatik* (1941); A. N. Tucker and E. O. Ashton, *Swahili Phonetics* (1942); E. O. Ashton, *Swahili Grammar* (1944); F. van den Eynde, *Swahili Spraakkunst* (1949) and J. Wegsteen, *Ki-Swahili* (1953) deal with Belgian Congo Swahili. (A. N. T.)

SWALLOW, the name for birds of the passerine family Hirundinidae. With graceful flight and much twisting and turning, the swallows seek their insect food in the air. The features that characterize the family are adaptatious for this: long pointed wings, small feet, short broad bill and wide gape. About 74 species of small to medium-sized birds comprise the family. Their colours are not gaudy, but many species have rufous or glistening white and iridescent deep-blue or green in the plumage. The range of the family is almost world-wide. The species breeding far north and south are strongly migratory, and some make long journeys across the equator to reach their winter quarters.

Swallows like company. They gather to feed in favourite spots; great numbers may assemble in the late afternoon to roost together during the night, and many species nest in colonies. Certain species come commonly about human habitations and nest on buildings or in nest boxes prepared for them. Though hardly songsters, their twitterings and chirping notes are pleasant, and this plus their familiar habits and their returning to temperate climates with the spring has given them a friendly place in people's minds and in literature. Their nesting is various: in a hole in a tree or bank, in a burrow dug by themselves in an earthen bank, a

saucer-shaped structure on a ledge of rock or a building, or a retort-shaped structure of mud pellets stuck to the side of a cliff or a building. There may be up to six eggs, pure white or spotted.

In a restricted sense "the swallow" is used in England for a familiar species, *Hirundo rustica,* called barn swallow in the United States, where it also is the best known species. The name martin (*q.v.*) is used for some swallows. (A. L. Rd.)

SWAMMERDAM, JAN (1637–1680), Dutch naturalist and biologist, who first described the red blood corpuscles and made the finest collection of microscopic observations that have ever been published, was born on Feb. 12, 1637, at Amsterdam. His father was an apothecary and naturalist who had a museum of curiosities, and Swammerdam started his career by helping with the museum. He went to Leiden university in 1661, graduating in medicine in 1667. Neglecting his practice, he became a pioneer of microscopy. He made a comparative study of insect life histories, which he classified according to the type of metamorphosis. He held that his observations supported the preformation theory of embryology. He went into minute details of the internal anatomy of the mayfly and the bee; observed the cleavage of the frog's egg and the subsequent segmentation and studied the anatomy of tadpole and adult; described the ovarian follicles of mammals in the same year as Regnier de Graaf (*q.v.;* 1672); and evolved improved techniques of injection which aided his study of human anatomy. Swammerdam discovered the valves of the lymphatics and was the first to describe the red blood corpuscles (1658). By a well-devised experiment he showed that the muscles alter in shape but not in size during contraction; this contradicted the current theory that a material fluid passed down the nerves to cause movement. He was the first to draw the sporangia of a fern. Deeply religious and somewhat unstable, he became for a time, after a quarrel with his father, a disciple of Antoinette Bourignon (*q.v.*).

Swammerdam died in Amsterdam on Feb. 15, 1680, after a period of deep spiritual and bodily misery.

His books, some of them magnificently illustrated were: *Tractatus physico-anatomico-medicus de respiratione usuque pulmonum* (1667); *Allgemeene verhandeling van bloedeloose diertjens* (1669); *Miraculum naturae seu uteri muliebris fabrica* (1672); and *Ephemerae vita* (1675). H. Boerhaave published a collection of Swammerdam's manuscripts, *Biblia naturae* (1737–38). (A. C. Ce.; X.)

SWAMP, a tract of saturated, miry, spongy ground occurring under conditions of impeded surface or internal drainage of substrata abundantly supplied with water from precipitation, ground water, tides or other sources and which may be seasonally or perennially subject to flooding. Swamps are abundant and extensive on the low flat-lying terrain of seacoasts, lake basins and stream valleys, may be widespread on uplands with low drainage gradients and often occur on steep slopes where ground-water seepage is abundant. Development of swamps is favoured by wet climates but they are common in arctic areas with permanent frost despite low precipitation.

In popular usage the term "swamp" indiscriminately includes bogs, marshes, fens, moors, heaths, muskegs and other related terrain features and may be distinguished from these commonly intergrading features only with difficulty even in local areas. Some characteristic properties of swamps include soft, mineral-rich substrata of well-mixed inorganic and organic fractions, little or no peat accumulation, absence of the floating mats of vegetation typical of bogs, relatively good aeration and filling by silting, or if by organic accumulation *in situ,* the materials added from the bottom upward.

Swamp vegetation is usually dense, of great variety and, depending upon the stage of development in relation to the changing environment, may consist principally of herbaceous species (cattails, grasses, sedges), woody shrubs (willow, alder, buttonbush), coniferous trees (white spruce, bald cypress) or broadleaved deciduous trees (willow, cottonwood, sycamore, cotton gum, red maple).

Examples in the United States are the mangroves of coastal Florida, the partially marshy Dismal swamp of Virginia and North Carolina and the Okefinokee swamp of Georgia. *See* also Marsh. (M. E. B.)

SWAN, SIR JOSEPH WILSON (1828–1914), English physicist and chemist who invented the dry photographic plate and a prototype of the electric light, was born at Sunderland, in County Durham, on Oct. 31, 1828. After serving his apprenticeship with a druggist in his native town, he became first assistant and later partner in a firm of manufacturing chemists in Newcastle.

Among its operations this firm included the manufacture of photographic plates, and thus Swan was led to one of the advances in photography with which his name is associated—the production of dry plates, the outcome of an original observation made by him on the effect of heat in increasing the sensitiveness of a gelatine bromide of silver emulsion. In 1862 he patented the first commercially practicable process for carbon printing in photography (*see* Photography). In 1879 Swan patented bromide paper.

In 1860 he produced an electric lamp with a carbon filament which was formed by packing pieces of paper or card with charcoal powder in a crucible and subjecting the whole to a high temperature. The carbonized paper thus obtained he mounted in the form of a fine strip in an evacuated glass vessel and connected it with a battery to Grove's cells, which though not strong enough to raise it to complete incandescence, were sufficient to make it red-hot. This was substantially the method adopted by Thomas A. Edison nearly 20 years later.

Subsequently Swan devised a cotton thread "parchmentized" by the action of sulfuric acid, and on Oct. 20, 1880, he gave at Newcastle the first public exhibition on a large scale of electric lighting by means of glow lamps. In another method devised by him, first exhibited in 1897, collodion was squirted into a coagulating solution and the tough threads thus obtained carbonized by heat. He also devoted attention to apparatus for measuring electric currents, to the improvement of accumulators and to the conditions governing the electrodisposition of metals.

Swan was knighted in 1904. Holding many honours, he died at Warlingham, Surrey, on May 27, 1914.

SWAN, the name for birds of a group that contains the largest of the waterfowl (family Anatidae). With long necks and short legs, swans are most at home on water, where they feed on aquatic plants. They also eat some grass and grain, though they walk somewhat awkwardly on the land.

A. W. AMBLER FROM NATIONAL AUDUBON SOCIETY

MUTE SWAN (CYGNUS OLOR) AND CYGNET

There are seven kinds of swans. Five, all white in colour, live in the northern hemisphere; a black-necked swan, *Cygnus melancoriphus,* lives in southern South America; and a black one, *Chenopis atrata,* lives in Australia and New Zealand, having been introduced into the latter country. The common swan of ornamental ponds everywhere is the mute swan, *Cygnus olor,* native of northern Europe and Asia and easily recognized by the knob on its forehead. It became a wild bird in Long Island, N.Y., after it escaped from estate ponds. Because of its pugnacious disposition and voraciousness, it poses a threat to other waterfowl.

American species are the arctic nesting whistling swan (*Olor columbianus*), a race of the Bewick's swan of Eurasia, and the rare trumpeter swan (*O. buccinator*) of Montana, Wyoming and western Canada. The latter was threatened with extinction, but has increased and extended its range into southern Alaska. The whistling swan winters in large numbers in the Currituck sound area of North Carolina and Virginia. On northward migration, flocks occasionally alight in the Niagara river and, unable to take flight as they are swept through the rapids, pass over the falls to their destruction. Swans have figured widely in song, story and mythology. Among the myths is the song of the dying swan. (Dn. A.)

SWANAGE, a seaside holiday town and urban district in the South Dorset parliamentary division of Dorset, Eng., 27 mi. E.S.E. of Dorchester by road. Pop. (1961) 6,062. Area 4.3 sq.mi. It lies in a sandy bay backed by precipitous chalk cliffs with numerous caves. Swanage bay was the scene of Alfred the Great's naval victory over the Danes in 877. Purbeck marble is still extracted from quarries first mentioned in 1250.

SWANSEA, Welsh *Abertawe,* a municipal, county and parliamentary borough and seaport of Glamorganshire, Wales, at the mouth of the river Tawe on the north side of Swansea bay. Pop. (1961) 166,740. Area 33.8 sq.mi.

The *Via Julia* from Nidum (Neath) to Loughor probably passed through the north of the borough, where Roman coins were found in 1835. The name Swansea stands for Sweyn's "ey" or inlet, and may have been derived from Sweyn Forkbeard, who certainly visited the Bristol channel. The earliest known form of the name is Sweynesse, which occurs in a charter granted by William, earl of Warwick, some time previous to 1184; in King John's charter (1215) it appears as Sweyneshe, and in the town seal, of medieval origin, it is given as "Sweyse." The Welsh name, Aber Tawy, first appears in Welsh poems of the beginning of the 13th century. The town grew up round the castle which Henry de Newburgh (or Beaumont) built in the early 12th century on the west bank of the Tawe. (*See* GOWER.) In the Great Rebellion the town was royalist until the autumn of 1645 when Colonel Philip Jones was made its governor. The older part of the town, being the whole of the municipal borough previous to 1836, occupies the west bank of the Tawe near its mouth and is now wholly given up to business. Stretching inland to the north along the river for some 6 mi. through Landore to Morriston, and also eastward along the sea margin toward Neath, is the industrial quarter, while the residential part occupies the sea front and the slopes of the Town hill (573 ft. high) to the west, stretching out to Sketty. The attractive seaside resort of the Mumbles or Oystermouth and a large part of the Swansea rural district were included in the borough in 1918. The east side of the river is known as St. Thomas and Port Tennant.

The 12th-century castle has entirely disappeared; but portions of the new castle, which was probably intended only as a fortified house, are incorporated in the buildings of the town. The remains include a tower, a so-called keep with the curtain wall connecting them, and a fine embattled parapet with an arcade of pointed arches in a style similar to that of the episcopal palaces of St. David's and Lamphey built by Henry Gower (d. 1347), bishop of St. David's, to whom the building of the new "castle" is also ascribed. Traces of St. David's hospital, built by the same prelate in 1331, are still to be seen at Cross Keys inn. The parish church of St. Mary was entirely rebuilt in 1896–98, only to be destroyed in 1941 by an air raid. Little now remains of the earlier work. Within the parish of St. Mary was St. John's, a church once owned by the Knights Hospitallers. This church, which was entirely rebuilt in 1820, was renamed St. Matthew in 1880, when a new St. John's was built within its own parish. The Royal Institution of South Wales, founded in 1835, possesses a museum in which the geology, mineralogy, botany, prehistoric and later antiquities of the district are well represented. Its library is rich in historical and scientific works relating to Wales and Welsh industries. The free library and Glynn Vivian art gallery (opened 1911) contain exhibits of local and general Welsh interest. Near the shore at the west end of the town is the new guildhall, opened in 1934. It is notable as containing Frank Brangwyn's 16 panels depicting the peoples, flowers, fruits and animals of the empire, originally painted for the royal gallery of the house of lords as part of a peers' war memorial. The grammar school, founded in 1682 by Hugh Gore (1613–91), Bishop of Waterford, was destroyed during an air raid (1941); a new Bishop Gore school was opened in 1952. The University College of Swansea was founded in 1920. Beau Nash was born in Upper Goat street in 1674. The town has about 3,000 ac. of open spaces. De Braose's charter of 1306 gave the burgesses the right to take from the lord's woods sufficient timber to make four great ships at a time and as many small vessels as they wished. Coal was even then worked in the district. Cromwell in his charter of 1655 recognized Swansea as "convenient for shipping and resisting foreign invasions."

From about 1768 to 1850 Swansea had a famous pottery. Beginning with earthenware which twenty years later was improved into "opaque china," it produced from 1814 to 1823 superior, beautifully decorated porcelain.

During the 18th century coal began to be worked at Llansamlet and copper smelting (begun at Swansea in 1717, though at Neath it dated from 1584) assumed large proportions. The coal was conveyed on the backs of mules and later by means of a private canal. Under an act of 1791 harbour trustees cleared the river bed and built a long pier on either side of it. A canal connecting the tidal part of the river Neath with the mouth of the Tawe was made in 1789, and also a small dock, Port Tennant or Salthouse dock, near the east pier, which continued to be used till 1875. Meanwhile in 1798 the entire coal-producing Swansea valley had been connected with the port by a canal 16½ mi. long (acquired by the Great Western railway in 1872). In 1845 the river was diverted eastward into a new channel (called the New Cut) and its old channel was locked and floated in 1852, thereby forming the North dock. (Area 11½ ac., and a half tide basin 2½ ac. Length of quays 5,500 ft.) The South dock, begun in 1852 and opened in 1859, is mainly used for shipping coal and general cargo and for discharging timber and fish. It has an area of 18½ ac.; length of quays 6,550 ft.; depth of water in lock 34 ft.

The next development was on the east side of the river where the shore of Fabian's bay, inside the harbour mouth, was extended for the construction of the Prince of Wales dock (authorized 1874, opened Oct. 1881, and subsequently extended), area 28 ac.; length of quay 6,872 ft. The very rapid increase in the demand for anthracite coal (for the shipment of which Swansea has practically a monopoly) soon necessitated still further accommodation and in July 1904 was begun the King's dock, which lies farther east and has an entrance direct from the bay. By means of the embankment made in connection with it, 400 ac. were reclaimed from the sea. King's dock and basin is 221 ac.; depth of water 40 ft. (high tide); length of quays 14,050 ft. The Queen's dock, exclusively an oil dock, occupies 150 ac. There are three dry docks and the entire harbour has some 270 ac. of deep water, over 6 mi. of quays and 130 mi. of railway sidings. Imports include crude oil, copper, nickel, iron, timber, grain and rubber. Exports include coal, patent fuel, refined oil, coke, tin plate and large quantities of manufactured goods.

Formerly acknowledged as "the metallurgical capital of Wales," Swansea is still an important seat of the copper, tin plate and spelter or zinc industries. Three-fourths of the tin plate manufactured in Great Britain and a very high proportion of the spelter are made in the district. There are also important steel works, and light industries have been developed, notably on the Fforestfach trading estate. In 1918 Swansea became an important centre for the distribution of oil, extensive storage and refining facilities (since enormously extended) being opened at Llandarcy.

Administration.—The town claims to be a borough by prescription because (although its charter of creation cannot be found or established) its status may be inferred from King John's charter of 1215 in which the inhabitants are addressed as "our burgesses of Sweyneshe." Its first grant of municipal privileges may have been in the charter of William, 3rd earl of Warwick, granted some time before 1184. In 1208 King John granted Gower by charter to William de Braose whose descendant gave Swansea and Gower various liberties by a charter of 1306 (*see* above). These were confirmed by the earl of Worcester's charter in 1532. Cromwell's charter of 1655 changed the title of portreeve into mayor, but this lasted only until 1659 when the title of portreeve was resumed.

From 1535 to 1832 (with the exception of 1658–59) Swansea was one of several contributory boroughs sending only one member to parliament. In 1832 Swansea, with four other contributory boroughs known together as Swansea District, was again represented by one member. In 1885 the district was split into two divisions, Swansea Town and Swansea District, each returning one member. In 1919 a different division was made which still

remains: Swansea West and Swansea East.

The assizes and quarter sessions for the administrative county of Glamorgan are held at Swansea and Cardiff alternately, but the borough has its own separate commission of the peace and quarter sessions. It became a county borough in 1889 and in 1923 the Church in Wales created a bishopric of Swansea and Brecon. In Feb. 1941 the whole business centre of the town (about 25 ac.) was destroyed by air raids.

SWARAJIST. The term *swaraj* (more properly *swarajya* from *swa*=own or self; *rajya*=rule or government) was originally employed in the ethical sense of control over self—self mastery. Prominence was first given to it in the political arena by Dadabhai Naoroji, who, in his presidential address to the Indian national congress in Calcutta in Dec. 1906 claimed as the right of India *swaraj* "or self government like that of the United Kingdom or the Colonies," and the word soon gained currency as the equivalent of dominion home rule. By a judicial decision of the Calcutta high court (1907, I.L.R., 34 Calc.), it was held not necessarily to mean "independent government" in the sense of government of the country to the exclusion of the present government. "It may mean, as is now well understood, government by the people themselves under the King and under British sovereignty." Its adoption as the label of a particular political party was the outcome of the lengthy and ever changing agitation set on foot by M. K. Gandhi in 1919. That movement commenced as one of passive resistance to the legislation known as the Rowlatt act, which was designed to strengthen the hands of the government in dealing with revolutionary crime, but, utilizing the resentment aroused among Mohammedans by the Turkish peace terms, which crystallized in the Khilafat movement, and the excitement which ran through the country as the result of disturbances in the Punjab and their sequelae, Gandhi, in the succeeding year, launched a campaign of protest under the guise of nonviolent nonco-operation, which was to include the boycott of the courts and government schools, the resignation of titles and government office, and abstention from participation in the forthcoming elections to the new legislative councils under the Reforms scheme.

In July 1920 the attainment of *swaraj* was included among the objects of the nonco-operation movement, and at the meeting of that year it was adopted by the Indian national congress as their avowed aim. But as to the precise meaning of *swaraj* there existed a vagueness which Gandhi did little to dispel. At one time he defined it as parliamentary government, whether within or without the empire; at another as dominion home rule; at a third as the universal employment of the spinning wheel; yet again as the triumph of the Khilafat party. On the other hand, at the first meeting of the legislative assembly in Feb. 1921, in the message from His Majesty the King Emperor, the ideal of *swaraj* was mentioned as the dream for years of patriotic and loyal Indians, while the reforms were cited as the beginnings of *swaraj* within the empire.

In the first elections under the 1919 act Gandhi's followers took no part, but an influential section of the congress, acting under the leadership of C. R. Das and Pandit Motilal Nehru, announced on Jan. 1, 1923, the formation of a Congress Khilafat Swarajya party, in whose program the capture of the councils and the obstruction of all business in them was the leading immediate item. After protracted dissensions and maneuvers this section was finally victorious within the congress, and at the elections of Nov. 1923 Swarajist candidates took the field as such and, noticeably as against the Moderates, achieved considerable success, though failing to secure (except in one province) the clear majority for which they had hoped. Throughout the term of the second reformed councils the Swaraj party continued their wrecking tactics, succeeding in two provinces (Bengal and the Central Provinces) in bringing about a temporary suspension of the reformed constitution.

Toward the end of that year the party was split by the defections of some of its members, notably in Bombay and the Central Provinces, in the direction of "responsive co-operation," including the acceptance of office under government, and in April 1926 an attempt was made to form an Indian national party of responsivists, independents and moderates "to prepare for and accelerate the establishment of *swaraj* or full responsible government in India, such as obtains in the self-governing dominions of the British empire." These differences bore fruit at the third election of the reformed councils at the end of 1926, when the Swaraj party definitely lost ground, noticeably in the assembly and in the United Provinces, Punjab and Central Provinces. The Indian national congress (which by that time was practically the Swaraj party) in Dec. 1927 declared for the complete independence of India.

SWARTHMORE, a borough of Pennsylvania, U.S., is a residential suburb of Philadelphia, located in Delaware county, about 11 mi. S.W. of the city. The population of less than 5,000 is composed principally of the families of commuters to offices in Philadelphia and Chester. The community grew up around Swarthmore college, a coeducational institution founded in 1864 by the Hicksite body of the Society of Friends and since 1911 nonsectarian. It is noted principally for its system of honours courses initiated in 1921 by Frank Aydelotte (1880–1956), president of the college from 1921 to 1940. Those who qualify for this program participate in honours reading and research seminars in three chosen fields of concentration during their junior and senior years. On the campus is the two-and-one-half story gray stone house in which the artist Benjamin West was born. It was built in 1724 in the characteristic Philadelphia style of that era.

For comparative population figures *see* table in PENNSYLVANIA: *Population.* (R. F. WE.)

SWASTIKA, a decorative and symbolic ornament consisting of a cross with equal arms, to the end of each of which is attached a line running at right angles to it, all these extensions occurring on the same relative side, so that all four extensions have the same rotary direction in relation to the centre. The swastika is one of the most ancient and widespread of all ornamental forms, appearing in both hemispheres. It is generally interpreted as a sun symbol. The swastika occurs in Aegean and archaic Greek pottery and in certain types of fret, found in Egypt and Greece. In some examples it appears as two identical S-curves, intersecting in the middle and set at right angles. Occasionally three arms are found, and the heraldic symbol of three running legs set radially, may be a development. Adolf Hitler made the swastika Germany's national symbol. (T. F. H.)

SWAT, a semi-independent state in the Peshawar division of West Pakistan adjoining the Mardan district, consisting of the valley of the Swat river above its confluence with the Panjkora, which is reached by the Malakand, the Shahkot and other passes from the south. It is divided into two distinct parts: Swat-Kohistan or mountain country, on the upper reaches of the Swat and the Swat proper lower down. The state has an area of 2,934 sq.mi. and a population (1951) of 518,596.

The Swat river (Sanskrit *Suvatu;* Greek *Suatos*) is formed by the junction of the Gabral and Ushu in the Kohistan at Kalam. The united stream sweeps in a great curve southwestward to Abazai, where it emerges to the north of Mohmand hill into the Peshawar plain. The river is utilized by canals to irrigate about 160,000 ac. The valley, which is very narrow in the higher mountains, contains a series of rich fertile tracts in the Swat proper and these are extensively cultivated. There are vast forests of deodar in the Kohistan region.

The Yousufzais (or Afghans) who form the largest group of the population are mainly concentrated in the lower valley while the Kohistan is peopled by Torwais and Garhwis. The state is known for its good administration in a traditionally turbulent area. Saidu Sharif, the capital, is 28 mi. from Malakand and is connected with it by a good motor road. Swat acceded to Pakistan soon after the partition of India and some of the powers of the ruler have been transferred to a partly elected, partly nominated council. In 1954 the administration of the adjacent territory of Kalam (area 822 sq.mi., pop. [1951] 9,702) was also entrusted to Swat.

Swat has great historic interest as a region bordering on Europe and Asia; it was one of the conquests of Alexander the Great and was also a remarkable stronghold of the Buddhist faith. It is recorded by Chinese pilgrims that in the 5th, 6th and 7th centuries there were as many as 1,400 monasteries in the region.

Excavations were begun in 1956 by the Italian Archaeological mission. At Mingora an important monastery has been found and at Udegram the mission has unearthed rock carvings which are probably prehistoric, part of the lower town and an immense staircase 25 ft. wide leading up from the valley to the mass of a castle, plans for the excavation of which were announced.

(K. S. Ad.)

SWATOW (also Shan'tow), a port in the province of Kwangtung, China, at the mouth of the main branch of the Han river. The population in 1953 was estimated at 200,000. Swatow was formerly a small fishing village, and the present town was built to a large extent on reclaimed land. English merchants settled on Double Island, in the river, as early as 1856 and began trading. Swatow is a great emigration port and was the scene of many kidnapping adventures on the part of foreigners in the early days. The region retained a full tribal organization until the latter half of the 19th century. Serious difficulties arose with the T'aip'ing rebellion and foreign interferences.

The city is the outlet for Chao-chow, about 30 mi. up the river, and the centre of a sugar cane district. Bean cake is manufactured and was also imported from Niuchwang, Shanghai and Hong Kong. Among usual leading exports are tea (after about 1872), sugar and oranges, paper, tobacco, grass cloth (manufactured at Swatow from the so-called Taiwan hemp), lace, drawn thread work embroidery. The imports ordinarily include cereals, peas, beans, kerosene oil and piece goods. The port was opened to foreign trade in 1869. The town, situated at the southern end of Formosa strait, is exposed to fierce typhoons. The Japanese occupied Swatow in 1939.

SWAYTHLING, SAMUEL MONTAGU, 1st Baron (1832–1911), British financier, was born at Liverpool on Dec. 21, 1832, of a Jewish family named Samuel. He took by royal licence the name of Montagu. He rose to be the head of the most important arbitrage house in London. He was Liberal M.P. for the Tower Hamlets from 1885 to 1900; he was a member of the Gold and Silver commission of 1887–90, being himself a bimetallist. He was created a baronet in 1894 and raised to the peerage in 1907. He was a zealous promoter of Jewish interests, founding the loan fund of the Jewish board of guardians, the Jewish workingmen's club and several synagogues.

Lord Swaythling succeeded Sir Julian Goldsmid as chairman of the Russo-Jewish committee, and also helped to establish a modern secular school for Jews at Jerusalem (1875). He died on Jan. 12, 1911, in London. His son, E. S. Montagu, was secretary for India.

SWAZI. The Swazi are an African tribe of the Embo-Nguni group of Bantu people. They speak a language closely allied to Zulu and Xosa. The political system consists of a hereditary paramount chief (*Ngwenyama*) who is the central figure of the tribe in political, social, military, religious and economic affairs. His mother (*Ndlovukati*) occupies a unique position and must be consulted in all tribal affairs. The *Ngwenyama* is advised by a small privy council (*liqoqo*) and by a general council (*libandla*) consisting of all chiefs and important men of the tribe. Any adult male, however, is entitled to attend and take part in its proceedings. Local government is in the hands of chiefs whose office is also hereditary. Each has a *liqoqo* and a *libandla* composed of his followers. The Swazi are horticulturists and have a culture which attaches considerable economic and ritual value to cattle.

(B. A. M.)

SWAZILAND, British protectorate in southern Africa, administered by the high commissioner for Basutoland, the Bechuanaland protectorate and Swaziland. It is a compact country of 6,704 sq.mi. surrounded on three sides by the Union of South Africa and on the east hemmed in from the sea by Portuguese East Africa, or Mozambique.

Physical Geography.—The territory consists of three fairly well-defined regions of about equal breadth running from north to south. The mountainous high country to the west is a spur of the Drakensberg which reaches altitudes of over 6,000 ft. The average rainfall in this region is 55.6 in. although it varies considerably from year to year and from place to place. This region provides good winter pasturage for sheep and is suitable for afforestation. The middle veld to the east is characterized by wide valleys with less rainfall (average about 36 in.) but suitable for mixed farming. There the tributaries which flow down steep gorges join to form the main river system which drains the territory into the Indian ocean. These rivers, the Komati, the Mbuluzi, the Usutu and the Ngwavuma, traverse the Lubombo flats which comprise the low country, the only considerable area of fairly level ground lying about 500 ft. above sea level and consisting of fertile soil with savannah type vegetation and good pastures. The rainfall is low (average about 20 in.) and badly distributed. The main rivers on their course eastward to the sea in Portuguese territory cut through the Lubombo range.

History.—According to their traditions the Swazi lived originally in southern Tongaland on the southeastern littoral of South Africa whence they migrated at the beginning of the 19th century to Shiselweni; *i.e.*, the country to the west of the Lubombo range between the Pongolo and Great Usutu rivers. There they remained until 1815 when, being attacked by the Ndwandwe chief, Zwide, the main portion of the tribe under their chief Somhlolo, otherwise Sobhuza I, fled northward to the present site under the Mdzimba mountain near Bremersdorp. Under Somhlolo's successor Mswazi, who gave his name to the tribe, they occupied considerable areas in what is now a portion of the Union of South Africa, and conquered and assimilated petty tribes or clans of Sotho origin. After their flight from Zwide the Swazi were occasionally raided by the Zulu and themselves made forays into neighbouring territory. Mswazi made an abortive attempt to check the depredations of the Zulu by seeking the protection of Queen Victoria through Theophilus (later Sir Theophilus) Shepstone.

More formal relations with Europeans were established with the signing of a treaty by Somcuba on behalf of Mswazi in July 1846 which purported to grant to *de Hollandsche Zuid Afrikaansche Natie* a large portion of territory included in the modern districts of Lydenburg, Middelburg, Barberton and Carolina in the Transvaal. In 1864 the Swazi assisted the Boers in a campaign against their traditional enemy Mapoko (Mapoch) and in 1880 assisted the British against Sekukuni, chief of the Bapedi.

By 1878 white settlers and prospectors had established themselves in considerable numbers in Swaziland, attracted by the fertile and well-watered terrain abounding in game and rich in the promise of minerals. They secured concessions over land, minerals and grazing as well as to exemptions from taxes, to railways, to telegraphs, to mining patents, collection of the revenue and trading.

In 1888 a charter of self-government was granted to the Europeans, and two years later, under a convention between the British government and the South African Republic of the Transvaal, a provisional government, consisting of representatives of the two powers and a representative of the Swazi, was set up with the consent of the latter. Three years later the provisional administration came to an end when the British government signed a new convention permitting the South African republic to acquire from the Swazi regent and her council an organic proclamation conferring on the republic rights of jurisdiction, legislation and administration without incorporation into the republic. The Swazi refused to sign the proclamation, but its provisions with some of those in the 1893 convention were embodied in the final convention of 1894 under which the South African republic exercised powers of protection, legislation, jurisdiction and administration in Swaziland subject to the limitations of the convention. After the conquest of the Transvaal all the rights and powers of the republic passed to Great Britain, and an order in council under the Foreign Jurisdiction act was issued in June 1903 providing that the governor of the Transvaal should administer Swaziland and conferring on him the right to legislate by proclamation. In 1906 these powers were given to the high commissioner. The main task of this administration was to settle the concessions question, which was achieved in 1907 by the issue of a proclamation under which land and grazing concessions suffered a deduction of one-third of their area for the exclusive use and occupation of the Swazi.

Population.—The population in 1956 was 237,041, an increase of 28.0% since 1946. The European population, which

in 1921 was 2,205, had risen to 5,932 in 1956, and in the same period the coloured or mixed population rose from 451 to 1,391. There are European villages at Mbabane, the administrative capital, and Bremersdorp, with European populations of between 800 and 1,000, and there are smaller villages at Emlembe, Piggs Peak, Goedgegun, Hlatikulu, Mankaiana and Stegi.

Religion.—Apart from indigenous religions, Christianity alone is represented by 23 denominations.

Education.—There were, in 1958, six government primary schools for Europeans, two government secondary schools and one other secondary school run by a voluntary agency. The total enrollment of European children in 1958 was over 1,000. Schools for Africans fall into the following groups: 95 unaided mission schools, 115 aided mission schools, 37 tribal schools, 3 national schools and 17 government-maintained schools. Total enrollment in 1957 was 26,589 pupils. In addition there are a teacher training school, a trades school and a nursing school. There were 4 mission schools for coloured children, of which 3 were government aided; the total enrollment in 1957 was 418.

There are three government and three mission hospitals in the territory (all of the latter are government aided) and two company hospitals treating employees only. There are also seven government health centres and one government cottage hospital.

Administration.—In the United Kingdom, responsibility for the administration of the territory rests with the secretary of state for commonwealth relations. A high commissioner has powers under the order in council of 1906 to administer and legislate for the territory, and he has similar powers in respect of Basutoland and the Bechuanaland protectorate. He resides, however, in the Union of South Africa, where he is also high commissioner for the government of the United Kingdom.

In Swaziland and directly responsible for its administration is a resident commissioner, who governs subject to the directions of the high commissioner under powers vested in him by proclamation and by his commission. He has a staff of administrative, technical and subordinate officials, and, in respect of European affairs, is assisted by an advisory council, first established in 1921, of ten elected Europeans. A system of administration was set up in 1950 by proclamation which provided for the administration of the Swazi by their paramount and other chiefs together with their councils in close conformity with traditional law and custom. A system of native courts was also established exercising criminal as well as civil jurisdiction over Africans.

Agriculture and Exports.—The principal occupation of the inhabitants, African and European, is agriculture, but many are also employed in mining, trading, building and forestry projects as well as by the government on public works, and in agricultural and veterinary services. Whereas the mining industry in Swaziland absorbs about 150 Europeans and 2,600 Africans, there are about 7,000 Swazi employed in the Union of South Africa, principally in gold and coal mines.

Of the 4,279,655 arable acres in the territory, 2,250,520 ac. are available for occupation by the Swazi; viz., 35 separate blocks of "native area" set aside for their sole and exclusive use and occupation, 240,000 ac. purchased by the Swazi nation from European landowners, 363,865 ac. of crown land and land purchased by the government for a Native Land Settlement scheme, and 12,883 ac. purchased by individual Africans. The remainder, apart from a few acres of crown land, is owned by Europeans who derive their title from concessions granted at the end of the 19th century by the then paramount chief Mbandeni.

Livestock raising is practised by Europeans and nearly all Swazi. The largest number of cattle (355,659 in 1957) are owned by Africans, but the number owned by others rose steadily during the 1950s, reaching 107,360 in 1957.

The staple diet of the Swazi is maize. They also grow sorghums which are usually converted into beer. Other crops are peanuts, sugar, tobacco, beans and rice. Tung trees do well, as do citrus, avocados, pineapples and other subtropical fruits.

The principal exports are asbestos, live cattle, cotton, hides and skins, butter, metallic tin and patulite (hard board). Gold is mined on a small scale and some barytes is produced. There are large unexploited coal fields in the low veld and rich iron deposits in the high veld.

Communication.—There are no railways in the territory, but 510 mi. of bus services are operated by the South African Railways administration.

Finance.—The main heads of taxation are income tax (levied on non-Africans only); customs and excise (under an agreement with the Union of South Africa, Swaziland receives a 0.149% share of the total collection of the Union); native tax; and base metal royalties (chiefly from asbestos production). The total ordinary revenue rose from £479,899 (1950–51) to £1,207,540 (1956–57). Similarly, expenditure increased from £559,410 (1950–51) to £1,052,374 (1956–57). £366,348 was received from colonial development and welfare funds in 1956–57.

A Swazi national treasury was established to handle the finances of the newly created native administration in 1950.

BIBLIOGRAPHY.—Hilda Kuper, *An African Aristocracy: Rank Among the Swazi of Bechuanaland* (Oxford, 1947) and *The Uniform of Colour . . .* (Johannesburg, 1947); B. A. Marwick, *The Swazi: An Ethnographic Account of the Natives of the Swaziland Protectorate* (London, 1940); annual reports on Swaziland in *Colonial Reports* series (H.M.S.O., London). (B. A. M.)

SWEARING. The common use of the word is for the uttering of profane oaths or curses. In English law, while blasphemy (*q.v.*) was at common law an indictable offense, cursing or swearing was left to the ecclesiastical courts. The Profane Oaths act, 1745, inflicted a sliding scale of fines for the use of profane oaths according to the rank of the offender: 1*s.* for a common labourer, soldier or seaman, 2*s.* for everyone below the rank of gentleman and 5*s.* for those of or above that rank; procedure under this act is regulated by the Summary Jurisdiction acts. By the Town Police Clauses act, 1847, the use of profane or obscene language is an offense punishable on summary conviction by a fine not exceeding 40*s.* or imprisonment not exceeding 14 days. The offense must be committed in a street. In the United States, blasphemy is a crime under either the common law or the statutes of substantially all of the states.

SWEATER, a jacket or short coat of a woollen material, usually knitted, and sometimes called a jersey, because it was worn by the sailors of the Island of Jersey. The first sweaters were heavy, dark-blue jackets, which were pulled on over the head, and were worn by the participants in athletic contests before and after games, to prevent cold.

SWEATING SICKNESS. A remarkable form of disease not known in England before, it attracted attention at the beginning of the reign of Henry VII. It was known indeed a few days after the landing of Henry at Milford Haven on Aug. 7, 1485, as there is clear evidence of its being spoken of before the battle of Bosworth on Aug. 22. Soon after the arrival of Henry in London on Aug. 28 it broke out in the capital and caused great mortality. This alarming malady soon became known as the sweating sickness. It was regarded as being quite distinct from the plague, the pestilential fever or other epidemics previously known, not only by the special symptom which gave it its name, but also by its extremely rapid and fatal course.

After 1485 nothing more was heard of it until 1507, when the second outbreak occurred, which was much less fatal than the first. In 1517 there was a third and much more severe epidemic. In Oxford and Cambridge it was very fatal, as well as in other towns, where in some cases half the population were said to have perished. There was evidence that the disease had spread to Calais and Antwerp, but with these exceptions it was confined to England.

In 1528 the disease recurred for the fourth time and with great severity. It first showed itself in London at the end of May, and speedily spread over the whole of England, though not into Scotland or Ireland. In London the mortality was very great; the court was broken up, and Henry VIII left London, frequently changing his residence. The most remarkable fact about this epidemic is that it spread over the continent, suddenly appearing at Hamburg, and spreading so rapidly that in a few weeks more than a thousand persons died. Thus was the terrible sweating sickness started on a destructive course, during which it caused fearful mortality throughout eastern Europe. France, Italy and the southern countries were spared. It spread much in the same way as cholera,

passing, in one direction, from north to south, arriving at Switzerland in December, in another northward to Denmark, Sweden and Norway, also eastward to Lithuania, Poland and Russia, and westward to Flanders and Holland, unless indeed the epidemic, which appeared simultaneously at Antwerp and Amsterdam on the morning of Sept. 28, came from England direct. In each place which it affected it prevailed for a short time only—generally not more than a fortnight. By the end of the year it had entirely disappeared, except in eastern Switzerland, where it lingered into the next year; and the terrible "English sweat" has never appeared again, at least in the same form, on the continent.

England was, however, destined to suffer from one more outbreak of the disease, which occurred in 1551, and with regard to this we have the advantage of an account by an eyewitness, the physician John Kaye or Caius.

The disease began suddenly with a sense of apprehension, followed by cold shivers (sometimes very violent), giddiness, headache and severe pains in the neck, shoulders and limbs, with great prostration. After the cold stage, which might last from half-an-hour to three hours, followed the stage of heat and sweating. The characteristic sweat broke out suddenly, and, as it seemed to those accustomed to the disease, without any obvious cause. With the sweat, or after that was poured out, came a sense of heat, and with this headache and delirium, rapid pulse and intense thirst. Palpitation and pain in the heart were frequent symptoms. No eruption of any kind on the skin was generally observed; Caius makes no allusion to such a symptom. In the later stages there was either general prostration and collapse, or an irresistible tendency to sleep, which was thought to be fatal if the patient were permitted to give way to it. The malady was remarkably rapid in its course, being sometimes fatal even in two or three hours, and some patients died in less than that time. Those who survived for 24 hours were considered safe.

The disease, unlike the plague, was not especially fatal to the poor, but rather, as Caius affirmed, attacked the richer sort and those who were free livers according to the custom of England in those days. "They which had this sweat sore with peril of death were either men of wealth, ease or welfare, or of the poorer sort, such as were idle persons, good ale drinkers and taverne haunters."

Causes.—Some attributed the disease to the English climate, its moisture and its fogs, or to the intemperate habits of the English people, and to the frightful want of cleanliness in their houses and surroundings which was noticed by Erasmus in a well-known passage, and about which Caius was equally explicit. But it must be concluded that climate, season and manner of life were not adequate, either separately or collectively, to produce the disease, though each may have acted sometimes as a predisposing cause. The sweating sickness was in fact, to use modern language, a specific infective disease in the same sense as plague, typhus, scarlet fever or malaria.

The only modern disease resembling sweating sickness is that known as *miliary fever* ("Schweissfriesel," "suette miliaire" or the "Picardy sweat"), a malady which has been observed in France, Italy and southern Germany, but not in the United Kingdom. It occurs in limited epidemics, not lasting more than a week or two (at least in an intense form). The attack lasts longer than the sweating sickness, is accompanied by eruption of vesicles and is not usually fatal. The first clearly described epidemic was in 1718 (though probably it existed before) and the last in 1861. Between these dates about 175 epidemics have been counted in France alone.

BIBLIOGRAPHY.—For history *see* Bacon's *Life of Henry VII*, and the chronicles of Grafton, Holinshed, Baker, Fabyan, etc. The only English medical account is that of John Caius, who wrote in English *A Boke or Counseill Against the Disease Commonly Called the Sweate, or Sweating Sickness* (London, 1552); and in Latin *De ephemera britannica* (Louvain, 1556; reprinted London, 1721). The English tract is reprinted in Babington's translation of Hecker's *Epidemics of the Middle Ages* (Syd. Soc., 1844). This also contains Hecker's valuable treatise on the English sweat, published in German (1834), and also printed in his *Volkskrankheiten des Mittelalters*, edited by Hirsch (Berlin, 1865). Grüner's *Scriptores de sudore anglico* (Jena, 1847) contains nearly all the original documents, including the two treatises of Caius. *See* also Hirsch, *Handbook of Geographical and Historical Pathology*, trans. by Creighton (New Syd. Soc., 1885).

SWEATING SYSTEM, a term used to describe oppressive industrial conditions in certain trades. The word sweater was used in England as early as 1850 to describe an employer or middleman who exacted monotonous work at very low wages. Charles Kingsley in his novel *Alton Locke* (1850) and his political pamphlets directed attention to London "sweaters"; and Henry Mayhew in 1851 talked of "the sweating system."

The system can be traced in certain trades in most countries both inside and outside Europe. It is closely associated with contracting. Individual workers or groups of workers contract to do a certain job for a certain price. Sometimes they carry out this contract themselves; in other cases they let it out again to subcontractors at lower prices. Contracting makes possible labour exploitation, often of women and children; and it produces erratic employment. When trade is brisk extremely long hours are worked in seriously overcrowded workrooms or dwelling houses. When trade is slack, subcontractors, whose overhead costs are far lower than those of factory employers, dismiss workers without scruple. One of the earliest objectives of factory and minimum-wage legislation is to eliminate such exploitation.

Contracting and subcontracting are common in the garment manufacturing industry. In the 19th century outwork existed in other industries, such as boot-and-shoe manufacture and soapmaking. Although legislation had by the middle of the 20th century either eliminated or carefully controlled sweating in most highly developed countries, the system was still operating in many Asiatic countries where larger numbers of people were engaged in home work than in industrial establishments.

Conditions tend to be worse in large cities, where sweatshops are hidden away in slum areas. They are aggravated by influxes of cheap labour from outside. To the native poor engaged in tailoring in London in the 1880s was added a large immigrant labour force of unskilled foreigners. In the United States the tireless search for cheap labour encouraged such immigration until 1917.

The first large-scale official inquiry into the sweating system was carried out by a select committee of the house of lords in England. Its report, published in 1890, described conditions in tailoring, bootmaking, furriery, shirt making, mantle making, cabinetmaking, cutlery and hardware manufacture, chainmaking and nail making, saddlery and harness making, military accoutrement establishments and dock labour. It claimed that sweating involved: (1) a rate of wages inadequate in terms of work done or basic standard of living; (2) excessive hours of labour; and (3) insanitary conditions in the houses or workshops in the houses in which the work was carried on.

The inquiry of the select committee did not bear fruit immediately. It was not until 1909 that the Trade Boards act set up boards to fix minimum wages in specified industries such as machine-made lace manufacture and chainmaking. In 1918 the scope of the act was extended to cover any trade where the ministry of labour considered that machinery for the effective regulation of wages was inadequate. Government interference was important, chiefly because it permitted the development of collective bargaining. By 1939, though many workers still received low wages, there were no longer any sweated trades in Great Britain.

In the United States most of the pressure to eliminate sweating was applied by organized labour. A number of acts, culminating in 1882, prohibited Chinese immigration. The first minimum-wage law was passed in Massachusetts in 1912. Numerous state laws improving factory and outwork conditions followed; and by 1939 practically all states had limited the hours of labour for women, and several had introduced legislation prohibiting night work.

Four main factors contributed in other countries as well as in Great Britain and in the United States to the control of sweated industries: (1) social idealism; (2) the pressure of trade unions; (3) the extension of the franchise and the growth of labour parties; and (4) the growing economies of factory production and the increased interest in human relations in industry. In 1928 the International Labour organization carried a recommendation concerning minimum-wage fixing that gave home workers and others subject to contracting and subcontracting some measure of protec-

tion; and further action was mooted after World War II.
(A. BRI.)

UNITED STATES

The sweating system grew in the United States during the Civil War, when soldiers' wives were employed to make uniforms with the relatively newly developed sewing machines. Its advance was slow, however, until the 1880s, when immigrants from eastern and southern Europe provided a large source of cheap labour. At that time the conditions necessary for a sweating system existed: a mass of unskilled, unorganized poverty-stricken men, women and children incapable of resisting the economic pressure to compete for work—women against their husbands, children against their parents; power-using factories or shops, where materials could be prepared for simple processes done in smaller shops or in the homes of workers or both; and mechanized transportation to and from the sweatshops and tenement dwellings, eliminating the long journeys on foot by heavily laden men, women and children. Motor trucks and the parcel post contributed to widen enormously the area within which the sweating system could be carried on.

As the first simple foot-powered sewing machines brought the sweating system to the soldiers' families, so the ceaseless evolution of the power-driven machines, first steam, later electric, served to concentrate the production of garments in vast skyscrapers near railway stations to which the employees travelled from their widely dispersed homes. That concentration facilitated labour organization within the garment industry.

Sweating-system evils also prevailed in other U.S. industries based on small workshops, homework and cheap labour, such as the making of cigars and artificial flowers. Inside and outside the garment industry the unions strove ceaselessly against sweating. In 1889 the tobacco workers obtained the passage of a New York statute forbidding all manufacture of tobacco in any tenement house, but it was held unconstitutional by the New York court of appeals and homework was prolonged for a number of years. The U.S. department of labour and various organizations, such as the National Consumers' league, also participated in the movement against the sweating system, and by the 1930s the conditions which had made it possible in the clothing industry had been virtually eliminated.

Beginning in 1912 in Massachusetts, a number of states passed laws creating minimum wage boards or commissions. The U.S. supreme court, however, in *Adkins* v. *Children's Hospital* (261 U.S. 525, 1923) held that a U.S. statute authorizing a wage board in the District of Columbia to fix minimum wages of women and minors violated the guarantees of the due process clause of the fifth amendment of the U.S. constitution, since freedom to contract between the employer and employee was impaired. The Adkins decision was reversed by the court in 1937 in the *West Coast Hotel Co.* v. *Parrish* (300 U.S. 379) decision. Many states passed minimum wage laws, and in 1938 a Federal Fair Labor Standards act covering industries engaged in interstate commerce was enacted.
(F. KE.; X.)

SWEDEN (SVERIGE), a kingdom of northern Europe, occupying a part of the Scandinavian peninsula. The length is about 990 mi., the extreme breadth (mainland) about 250 mi. and the total area, inclusive of inland waters, of which the four great southern lakes comprise 3,522 sq.mi., is 173,665 sq.mi. The boundary is estimated at 6,100 mi.; 4,737 mi. are coastal, the Norwegian frontier is 1,030 mi. and the Finnish 333 mi.

Physical Features.—The backbone of Sweden is an ancient mountain range of which the crest line forms the Norwegian boundary. The three main territorial divisions of Sweden are Norrland, Svealand and Gotaland. The country may be described in four main physical divisions: the northern mountains and lake region district, covering all Norrland and the western part of Svealand; the lowlands of central Sweden; the Smaland highlands, in the south and southeast; and the plains of Skane, occupying the extreme south of the peninsula. The first district is much the largest and highest and contains the finest scenery. The greatest heights lie along the boundary, and several exceed 6,000 ft. Kebnekaise (6,965 ft.) in northern Lapland, the highest point in Sweden, is a bold, somewhat isolated summit. Slightly south, beyond the Lule *alv*, is a lofty mass in which Sarektjakko (6,857 ft.), the second highest summit, is situated. Still farther south is the better-known Sulitelma (6,158 ft.); beyond is a progressive decrease in height southward until Areskutan (4,659 ft.) is reached. Near the southern flanks of Areskutan the railway from Trondheim, Nor., crosses into Sweden. South of the railway the highland line passes into Norway, but before doing so the mass rises into such heights as Helgasfjallet (5,892 ft.). There the snow line, which is about 3,000 ft. in the north, rises to 5,500 ft. The numerous rivers are linked with long, narrow, high-level (900–1,300 ft.) lakes, due mainly to dams of moraine. Among European countries only Finland exceeds Sweden in the number of its lakes. In Sweden the 96,000 lakes cover about 9% of the country, but the largest lie in the central lowlands. The northern rivers run through forests and enter the sea usually through long estuaries. The Torne, which, with its tributary the Muonio, forms the boundary with Finland, has a length of 354 mi. and drains Tornetrask (124 sq.mi.).

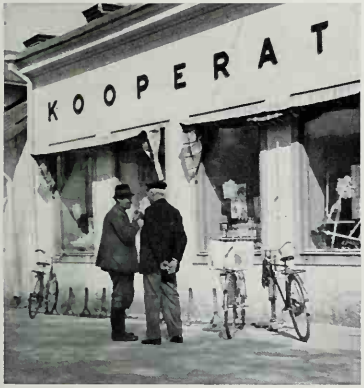
Along and near to the shores of the lake the railway from Gallivare runs to Narvik, Nor. The Kalix (267 mi.), with numerous rapids, is paralleled by the Lule, whose main stream is 279 mi. One of its branches, the Stora Lule, drains the Langasjaur and Luletrask, which together have a length exceeding 50 mi. but a total area of only 87 sq.mi. At the head of the Langasjaur is the Stora Sjofall (Great Lake falls; Lapp, Atna Muorki Kartje), 131 ft. In the southern mountain valleys of the region there are several beautiful falls, such as the Tannfors, not far from Areskutan. Still farther south of the Lule are the Pite river (230 mi.) and the Skellefte (255 mi.); the latter drains Horna *van* and Stora *van* (total area 163 sq.mi.). Horna *van* is a straight and sombre trough, but Stora *van* and the intervening Uddjaur are broad and picturesquely studded with numerous islets. The Ume (286 mi.) with its tributary the Vindel, of almost equal length, drains several lakes, including Stor Uman (63 sq.mi.). The further principal rivers of this region are the Angerman (279 mi.) with the magnificent Hallingsa fall within its drainage area, Indal (261 mi.) and Ljusnan (267 mi.). The Dal (323 mi.) embraces Lake Siljan (110 sq.mi.). The Klar has its upper waters—the Faemund—in Norway and reaches Lake Vanern after a course of 304 mi. The Torne has a drainage area of 15,521 sq.mi., but the Angerman has the largest basin (12,317 sq.mi.) of any river entirely in Sweden; the average for the other large northern streams is less than 8,000 sq.mi.

The lowlands resemble the coastal belt. There are fertile plains of clay, with innumerable lakes, including Vanern (Vener), Vattern (Vetter), Malar and Hjalmar, with areas respectively 2,141, 733, 440 and 190 sq.mi.; the Vanern is the largest lake in western Europe. The scenery of the lake district is quietly beautiful.

Granulite, also called eurite and halleflinta, is the most important of the Archaean formations and contains all the metalliferous deposits of Sweden. The rock is a very compact and fine-grained mixture of feldspar, quartz and mica, often grading into mica schist, quartzite and gneiss. The Cambrian and Ordovician strata occur in isolated patches. The deposits in most places are very little disturbed and form horizontal or slightly inclined layers. South of Vanern they are capped by thick beds of eruptive diabase. Conspicuous local features are the eskers or gravel ridges (*asar*), 100 to 200 ft. in height, in direction north-northwest. Their practical value lies in their excellent water supply and gravel.

South of the central lowlands the Smaland highlands lie in the heart of Gotaland and are a detached part of highland Sweden. They lie roughly south of Vattern and reach the southwest coast. The general elevation of this region exceeds 300 ft. and in the eastern part 600 ft.; the principal heights are Tomtebacken (1,237 ft.) and Galtasen (1,188 ft). There is much coniferous forest, but agriculture is profitable. The Smaland highlands abut southward upon the plains of Skane. Newer rocks of Cretaceous and Jurassic Ages, together with Recent marine deposits, have yielded level plains, with rich open meadows and cultivated lands, the monotony of which is in some parts relieved by beech woods. The Triassic formation (Rhaetic) in the northern part of Malmohus consists of

Strömmen harbour in Stockholm



A co-operative store in the town of Sigtuna on Mälaren (Lake Malar) in eastern Sweden

The main harbour at Smögen, a fishing port and resort on an island of the same name in the Skagerrak

A marginal quay at Göteborg's harbour, Sweden's chief seaport

A view of the southern Swedish port of Malmö showing its maritime industrial complex in the background

SWEDISH HARBOURS AND WATERFRONTS

PHOTOGRAPHS, (TOP, CENTRE RIGHT, BOTTOM LEFT, BOTTOM RIGHT) J. ALLAN CASH—RAPHO GUILLUMETTE, (CENTRE LEFT) KIT ROBBINS—RAPHO GUILLUMETTE

SWEDEN

A boom of logs for a sawmill in Västerbotten on the Gulf of Bothnia

A large paper mill in Skutskär, a sawmill centre in eastern Sweden

INDUSTRY OF SWEDEN

Housing settlement in Kiruna iron-mining centre on Luossajärvi (Lake Luossavaara), about 90 mi. from the Arctic circle

Automotive assembly line in Göteborg

An improvised lunchroom 600 ft. below ground in a mine at Persberg, western Sweden

BY COURTESY OF (CENTRE RIGHT) AB VOLVO, PHOTOGRAPHS, (TOP LEFT) REFOT—RAPHO GUILLUMETTE, (TOP RIGHT, BOTTOM RIGHT) CULLERS—RAPHO GUILLUMETTE, (BOTTOM LEFT) BELL, HOWARTH LTD. FROM BLACK STAR

sandstones and clay beds with coal.

On the whole the rivers of south Sweden are few and short. Lake Vanern is drained to the Kattegat (Cattegat) by the short Gota river, on which, near the lake, are the celebrated falls of Trollhattan.

Coast.—The coast of Sweden, though not indented with so many or so deep fiords as that of Norway, yet resembles it in having a fringe of islands which, throughout large parts of its length, shelters the coast of the mainland from the open sea. This *skargard* is interrupted only round the southern shore off Skane.

The island belt is widest (45 mi.) off the city of Stockholm (noted for its beautiful sea approach). Farther north, only the narrow Alands Hav lies between the Swedish coast and the vast Aland archipelago, which extends to Finland. The *skargard* islands as a whole are rugged and picturesque; in the Baltic several are well wooded, but the majority, particularly in the Gulf of Bothnia, are bare or heath-clad. For the large islands Oland and Gotland *see* separate articles.

Climate.—Several factors contribute toward variations of climate in Sweden's several regions. The land runs through more than 13½° of latitude; about 15% of its area is within the Arctic circle; the boundary mountains are sufficiently high to be much colder than the adjacent coastal lands, but are not high enough to shut out entirely the warming effects of those drifts of wind and water from the southwest which give Norway its remarkable climate; the eastern and southern borders are washed by the sea; the piercing winter winds from the great Russian pool of cooled atmosphere can sweep westward over the Swedish lowlands or can reach the mountain tops which, on the whole, offer but moderate gradients for ascents. February is the coldest month and has a mean temperature of below 32° F. over Sweden. Stockholm averages 25.7°, Goteborg 30.4°, Haparanda 10.6°, but the coldest region is in the heart of Lapland; there the average for Karesuando is 5.2°. In some northern parts frost occurs in every month of the year, and at Karesuando only five months have a mean temperature above freezing point. July is generally the warmest month. Goteborg has a July mean of 62.2°, Stockholm 62°, Haparanda 59° and Karesuando 54.2°. The summer is short but high maximums are sometimes recorded; *e.g.*, 88° at Karasjok in Lapland.

The relative length of the seasons shows contrasts resembling those of temperature. At Karesuando the last day-frost of spring occurs on an average on June 15 and the first of autumn on Aug. 27, while at Stockholm four and a half months, and at Hven in the sound more than six months are frostless. Ice forms in October in the north and breaks up in May or June, while in the midlands and south the corresponding months are late November and April. The ice covering of the lakes ranges from 100 days annually in the south to around 200 days in the north, though local increases of the ice period occur in the upper Smaland highlands. Drifting sea ice generally appears in the Kattegat in January, but the sound becomes obstructed later largely by drift ice from the Baltic. In exceptional winters (*e.g.*, 1941, 1942) the sound may be completely frozen over. Ice conditions in the Baltic vary considerably from year to year but usually navigation in the southern part of the Gulf of Bothnia is impeded from the end of November to the beginning of May; in the north the gulf is covered with ice from November to mid-May, and among the islands it may linger until June.

The length of the Swedish summer day varies between wide limits. At Karesuando in 68° 26′ N. lat. the sun is continuously above the horizon from May 26 to July 18; at Haparanda it may be seen for 23 hours and at Lund for 17½ hours at the summer solstice. Refraction increases the average length of the day by 30 min. in the north and by 15 min. in the south. More important is the effect of twilight, which results in daylight's lasting without interruption from June 16 to 27 as far south as Harnosand (62° 38′ N.)

The average annual rainfall for Sweden increases, on the whole, from north to south, reaching a maximum toward the southwest. Thus the average in the north of Norrland is 16.5 in. and in the south it is 22.5 in. At Boras, midway between the south end of Lake Vattern and the Kattegat, the average is 35 in. The minimum at Karesuando is 12.3 in. The period of maximum is generally the latter half of the summer and the minimum February and March. The proportion of total precipitation which falls as snow ranges from 36% in Lapland to 9% in Skane. Snow lies 47 days on an average on the plains of Skane, while in the north it lies from 140 to 190 days. The Swedish glaciers cover about 135 sq.mi. They occur in the northern highlands about Kebnekaise, Sarektjakko and Sulitjelma. The most southerly is on the slopes of Helagsfjall in lat. 62° 54′ N.

Vegetation.—The extensive range of latitude and altitude produces many local vegetation differences. Four main regions can be distinguished, but they tend to grade into one another near their adjacent fringes. In descending order of altitude and, largely, of latitude also are: (1) alpine regions, (2) a zone of birch woods, (3) a coniferous forest region and (4) a beechwood region.

The alpine region occupies only the upper flanks of the spinal mountain range, at an altitude above 1,600 ft. in the north and above 3,000 ft. in the south. The lower belt of this region consists of thickets formed by several species of shrublike willows and dwarf birch, as well as of dwarf shrub communities and meadows. With increasing altitude the dwarf shrub communities, with species of *Cassiope* and crowberry, become more dominant and are accompanied by grass heath types. The vegetation of calcareous ground is much richer in species than that of siliceous ground and comprises Lapps' rhododendron, tree nymph (*Dryas*), arctic poppy (rare), several characteristic species of sedges, saxifrages, whitlow grasses, etc. The higher tops of the mountains are characterized by a few flowering plants scattered in carpets of mosses and lichens, alternating with almost or, at the highest altitudes, even entirely naked soil.

The birch forests are much more extensive in the north, where they form a subalpine belt, separating the coniferous forest region from the alpine region. The dominant tree is the birch *Betula tortuosa*, which grows together with rowan, bird cherry, some willows and aspen (which rarely attains tree size). The undergrowth is rich in herbs; *e.g.*, wood crane's-bill, globeflower, goldenrod, etc. In dry situations, on gravel, it is poorer; there lichens, especially reindeer "moss" and *Nephroma arctica*, are dominant.

The coniferous forests are the most extensive and cover nearly all the country northeast of Lake Vanern. The important tree on the drier and higher ground is the Scotch pine (*Pinus sylvestris*), with cowberry in the undergrowth. More fertile soils are occupied by spruce (*Picea abies*), having bilberry in the undergrowth. The conifers extend from sea level to 2,500 or 3,000 ft.; and though the pine and spruce predominate yet there is a sprinkling of lowland birches, rowans and aspen in the north, while large numbers of such deciduous trees as oaks, ash, limes, maples and elms are found in the south, especially on fertile soil. These latter trees seldom extend beyond 1,600 ft. above sea level, and the oak ceases about lat. 61° N. (inland) or lat. 63° N. (coastal).

The beech forests (with occasional oak) are restricted to Skane. They are tending to be reduced by extended land cultivation. The flowering plants of southern and middle Sweden are largely immigrants since the last glacial epoch, and the number of species shows a decrease northward. Many plants in northern Sweden are considered to be glacial survivors.

Fauna.—The effects of the great latitudinal range of Sweden are also shown in the case of fauna. Some animals, such as the hare, various shrews, the ermine, the weasel, the squirrel, the fox and the elk, are common to the entire country. Formerly bounties were paid for the extermination of the larger animals, which were thus greatly reduced in number. The bear is now protected and, with the wolf, glutton or wolverine and lynx, is becoming restricted to the more solitary depths of the northern forests. The wild reindeer is extinct, but large domesticated herds form practically the sole source of livelihood of the Lapps. The elk is the most valuable game animal, but badgers, which exist in large numbers in the midland plains, otters and the rare pine martens of the coniferous forests are hunted for their skins. The roe deer is spread over the southern and central parts of the country with a tendency to spread northward. The arctic fox and the lemming are mountain animals, the former being confined to the highlands of the north. When the lemmings make their nonperiodical migra-

tions to the lowlands their vast hordes are the prey of bird, wild and domesticated animals. The common porpoise is the only cetacean occurring in the Baltic, and with the gray seal it is held to be responsible for much damage to fisheries.

Bird life is very abundant, especially in spring and summer; several types such as the teal, snipe, golden plover and wagtail are found throughout the country. In the northern mountains the ptarmigan is common; ducks and other waterfowl frequent the lakes both there and in the south; the golden eagle, certain buzzards, owls and the small Lapland bunting are found. In the coniferous forests, grouse, capercailzie and woodcock have long been the principal game birds but are now less common; the crane lives in marshy clearings; birds of prey, crows and many other passerines range over a wide area. In the midlands partridges and doves are fairly common. On the coast large numbers of gulls and terns are found, also the eider duck and the rare sea eagle. There are very few reptiles or amphibians and the viper is the only poisonous species of snake. The rivers and lakes are generally well stocked with fish, such as salmon, trout, char, pike and perch. Both fresh-water and sea fisheries are important, especially on the west coast, where herring, cod, flatfish, mackerel and sprat are caught. Great numbers of a small herring called *stromming* are fished in the Baltic together with cod, eel, salmon and flatfish. In the brackish waters of the east coast salt-water and fresh-water forms are found. The crayfish occurs in many places in central and southern Sweden. Pearls are at times found in a fresh-water mussel, and oysters are cultivated to a small extent on the west coast. Among the lower marine animals a few types of arctic origin are found, not only in the Baltic but even in lakes Vanern and Vattern, having survived the changes consequent on the separation of the Baltic sea and the Arctic ocean.

Insect fauna is rich, even in the north. In summer in the northern lowlands the gadfly is common enough to cause annoyance to the reindeer, which it drives to the upper pastures and snow fields.

See M. Zimmerman, *Les Etats scandinaves* (Paris, 1933), vol. iii of *Géographie universelle*, ed. by P. Vidal de la Blache and L. Gallois; C. A. M. Lindman, *Svensk Fanerogamflora*, 2nd ed. (Stockholm, 1926).

(W. E. Wh.; X.)

HISTORY

The first historical notice relating to Sweden is found in Tacitus, *Germania*, ch. 44, in which we find the name of the chief people of the peninsula, the Swedes proper, Suiones (Old Norse, Sviar; Swedish, Svear; Anglo-Saxon, Sweon), who eventually gave their name to the whole country. According to Tacitus they were governed by a king, whose power was absolute, and possessed a strong fleet which secured them from hostile incursions. Their original territories lay on both sides of Lake Malar, in the provinces later known as Uppland, Sodermanland and Vastmanland. Other early Roman writers, Mela and Pliny, mention the country. (*See* Scandinavian Civilization.)

The people next in importance to the Suiones in the peninsula (Swed., Gotar; O.N., Gautar; A.S., Geatas) are first mentioned by Ptolemy (under the form "Goutai" for Gautoi). He puts them in the southern part of the country. Gotaland consisted of Vastergotland and Ostergotland divided from one another by Lake Vattern, together with Smaland. In early times Vastergotland seems to have been the most important; Varmland, the district to the north of Lake Vanern, and the whole of the country to the north of Svealand seem to have been of small importance. Jamtland was always considered a part of Norway. After the time of Ptolemy we hear no more of Sweden until the 6th century, when a surprisingly full account of its peoples is given by the Gothic historian Jordanes. He speaks of trade in furs of arctic animals which were sold by merchant to merchant until they reached Rome.

For the same period information is to be found in the Anglo-Saxon poem *Beowulf*. The hero himself belonged to the Geatas (*i.e.*, in all probability, Gotar), his mother being the daughter of their king Hrethel. Haethcyn, Hrethel's son, is said to have perished in a disastrous battle against the Svear, but his fall was avenged by his brother Hygelac in a subsequent engagement in which the Swedish king Ongentheow was killed. This Hygelac is clearly identical with that Chochilaicus wrongly described as a Danish king by Gregory of Tours (iii, 3) who died in battle with the Franks under Theodberht about A.D. 520. About the time of Hygelac's death strife broke out in the royal family of the Svear, between Onela, the son and successor of Ongentheow, and Eanmund and Eadgils, the sons of his brother Ohthere. The latter fled for protection to the Gotar and the war which ensued cost the lives of Eanmund and of Heardred, the son and successor of Hygelac. According to the poem, Beowulf himself became king of the Gotar, but Beowulf's later history has a mythical character, and for three centuries after this time there is no reference to Swedish affairs in English or other foreign authorities. The kingdom of the Gotar probably ended soon after the 6th century.

At a later time the kings of Norway claimed descent from the ancient royal house of the Svear, and their traditions, though in great part mythical, prove at least its high antiquity. They trace its origin to the god Frey, son of Niorðr, who is said to have founded Uppsala, the ancient capital of Sweden. Among his descendants Athils, the Eadgils of *Beowulf*, is a prominent figure; in general, the account given of him agrees with *Beowulf*.

Early Kings.—Four generations after Athils' time a king named Ingialdr was overthrown by a prince from Skane, called Ivarr Viðfaðmi. His son Olafr Tretelgia withdrew to Varmland, which he brought into a state of cultivation, though he was subsequently sacrificed by his subjects in a time of famine. The sons of Olafr Tretelgia moved westward into Norway, and if we may trust later traditions Sweden passed out of their family. The subsequent kings of Sweden are said to have been descended from Ivarr Viðfaðmi. About 830 the missionary bishop Ansgar made his way to Birca on Lake Malar. For more than a century after Ansgar's death no serious mission seems to have been attempted.

During the 9th century extensive Scandinavian settlements were made on the east side of the Baltic. The famous expeditions of Rurik and Askold which resulted in the origin of the Russian monarchy appear to have taken place toward the middle of the century, but it is not possible to connect these names with any families known from Swedish tradition. Many names in Russian chronicles seem to be Swedish.

The continuous history of Sweden begins in the early 10th century, when a king named Eric son of Emund was reigning at Uppsala. Before the end of the century his descendant Olaf Skottkonung had created what for the moment was the strongest kingdom in the north. Under him Christianity was established in Sweden. As a leading ally in a coalition which included Sweyn, king of Denmark, and Eric, jarl of Lade, he overthrew Olaf I Tryggvesson, king of Norway, and annexed for a time part of Trondheim and the district now called Bohuslan, which he handed over to Sweyn Jarl, brother of Eric Jarl, as a marriage portion for his daughter Holmfrið. Several years later another Norwegian prince, Olaf II Haraldsson the Fat, returning to Norway as king, put an end to the Swedish and Danish supremacy and in 1015 forced Sweyn Jarl to leave the country. An attempt to arrange an alliance between Olaf Skottkonung and Olaf Haraldsson (who married a Swedish princess) was unsuccessful, and their relations were still strained when the former died, probably in the winter of 1021–22.

Anund, his son, early in his reign allied himself with Olaf Haraldsson against Canute of Denmark, who had demanded the restitution of the rights possessed in Norway by his father, Sweyn. The allies took advantage of the Danish king's absence to harry his land. On his return an indecisive battle was fought at Helgi Å, and Anund returned to Sweden. Olaf was driven from Norway by the Danes, but in 1030 he raised a small army in Sweden only to meet his death at the battle of Stiklestad. Anund, who died about 1050, according to Adam of Bremen, was succeeded by his brother Emund the Old, who had been previously passed over because his mother was unfree, the daughter of a Slav prince captured in war. This king had become a Christian, but soon quarrelled with Adalbert, archbishop of Bremen, and endeavoured to secure the independence of the Swedish church, which was not obtained for another century. Under Emund there was a rectification of boundaries which assigned Blekinge to Denmark.

With Emund's death in 1060, the old family of Swedish kings died out. He was succeeded by his son-in-law Steinkel, a noble of

Vastergotland, who was warmly attached to the Christian religion, though he refused to destroy the old sanctuary of idols at Uppsala. During his reign grants of land in Varmland made by the king to the Norse jarl Haakon Ivarsson led to a successful invasion of Gotaland by Harald III Hardraade of Norway. Steinkel also had disputes with Denmark. His death in 1066 was followed by a civil war, and for the next 15 years the history of Sweden is very obscure. In 1081 we find the sons of Steinkel, Inge and Halstan, reigning. Inge's attachment to Christianity caused him to be expelled after a short time by his brother-in-law Sweyn or Blotsweyn, so called from his revival of the old sacrifices, described by Adam of Bremen. Sweyn retained the kingship for only three years before Inge returned and slew him.

Organization of the Kingdom.—Under Blotsweyn's grandson, King Sverker (1134–55), who permanently amalgamated the Swedes and Goths (each of the two nations supplying the common king alternately for the next 100 years), Sweden began to feel the advantage of a centralized monarchical government. Eric IX (1150–60) organized the Swedish church on the model prevalent elsewhere and undertook a crusade against the heathen Finlanders, which marks the beginning of Sweden's overseas dominion. Under Charles VII, the archbishopric of Uppsala was founded (1164). But the greatest mediaeval statesman of Sweden was Birger Jarl, of the powerful Folkungar family. He practically ruled the land from 1248 to 1266. He founded Stockholm, but he is best known as a legislator who prepared the way for the abolition of serfdom. The increased dignity which the royal power owed to Birger was still further extended by King Magnus Ladulas (1275–90). Both these rulers adopted the dangerous expedient of creating a number of almost independent duchies in Sweden, but the danger of weakening the realm by partition was averted, though not without violent complications. In 1319 the severed portions of Sweden were once more reunited. Meanwhile the political development of the state was steadily proceeding. The formation of separate orders, or estates, was pronounced by Magnus Ladulas, who extended the privileges of the clergy and founded a hereditary nobility (Ordinance of Alsno, 1280). In connection with this institution we now hear of a heavily armed cavalry as the kernel of the national army. The knights too became distinguishable from the higher nobility. To this period belongs the rise of a prominent burgess class, as the towns began to acquire charters. At the end of the 13th century and the beginning of the 14th provincial codes of laws appeared and the king and his council executed legislative functions.

Union of Kalmar, 1397.—The first union between Sweden and Norway occurred in 1319, when the three-year-old Magnus, son of the Swedish royal duke Eric and of the Norwegian princess Ingeborg, who had inherited the throne of Norway from his grandfather Haakon V, was in the same year elected king of Sweden (Convention of Oslo). A long minority weakened the royal influence in both countries, and Magnus lost both his kingdoms before his death. The Swedes, irritated by his misrule, superseded him by his nephew Albert of Mecklenburg (1365). In Sweden, however, Magnus' partialities and necessities had led directly to the rise of a powerful landed aristocracy and, indirectly, to the growth of popular liberties. Forced by the unruliness of the magnates to lean upon the middle classes, Magnus summoned in 1359 a primitive form of the *riksdag*, on which occasion representatives from the towns were invited to appear along with the nobles and clergy. His successor, Albert, was forced to go a step farther and, in 1371, to take the first coronation oath. (X.)

The Swedish nobles, however, remained dissatisfied with the rule of Albert of Mecklenburg. In 1388, therefore, they called in Margaret (*q.v.*), Valdemar IV Atterdag's daughter and successor on the Danish throne, who became queen of Sweden also in 1389. Norway was already united with Denmark, and Margaret's plan was to bring all the three kingdoms together into a single union under Denmark. An act of union was drawn up at Kalmar in 1397 but was not proceeded with as the queen was dissatisfied with some of its clauses. At the same time her relative Eric of Pomerania was elected future king of the three kingdoms. After Margaret's death in 1412 he assumed the government and proceeded to try to change Sweden into a vassal country with the help of a system of Danish bailiffs. The consequent oppression became so hard that the Swedish peasants rose in 1434 under the leadership of Engelbrekt Engelbrektsson, a noble from Bergslagen. The national movement spread swiftly and the Danish strongholds were conquered. In the course of this fight for freedom Engelbrekt called together at Arboga, in Jan. 1435, the first *riksdag* in which nobles, priests, burghers and peasants were all represented. In 1436 Engelbrekt was murdered. For several decades a party in favour of the union, made up of the more important nobles and the clericals, with Danish support, proceeded to contend against a national democratic party under native leaders taken from the nobles. Danish kings were on several occasions recognized as rightful sovereigns also of Sweden, among them, for instance, Christopher of Bavaria (1440–48), in whose reign there was enacted in 1442 a new general law which, with additions, held good until 1734, and Christian I (1457–64). In between, a Swedish nobleman, Karl Knutsson, was recognized king (as Charles VIII) during the years 1448–57, 1464–65 and 1467–70, being twice expelled and twice reinstated. During the years that followed no native-born sovereigns were set up against the Danish kings, but the national party was under the leadership of *riksforestandare*, or regents, in the persons of members of the Sture family. After *riksforestandare* Sten Sture the Elder (1470–97) came the Danish king Hans (1497–1501), then Sten Sture again, (1501–03), and then Svante Sture (1504–12) and Sten Sture the Younger (1512–20). The times were troubled and there were continual struggles with the Danes. Many Germans lived in the towns. After a Swedish victory over the Danes at Brunkeberg in 1471 an old ordinance by which the Germans should have half the seats in the town councils was abolished. The country now became impoverished, but the national feeling grew stronger than ever, concurrently with a hatred of Denmark caused by Danish intrigues against Sweden. After 1512 the aristocracy, led by Erik Trolle, sought reconciliation with Denmark. Things came to a climax when the Danish king Christian II conquered the country; and in 1520 Archbishop Gustaf Trolle, Erik's son, had the foremost men of the national party in Stockholm executed (the "Stockholm blood bath"). Among noteworthy events in this period may be mentioned the founding of Sweden's first university in Uppsala in 1477.

Gustavus I Vasa.—The Scandinavian union was represented by the Danish king living in Copenhagen. Denmark and Norway held the south, and foreign trade lay in the hands of the Hanseatic league. Sweden was a land of peasants, divided by provinces and inspired by all the local patriotism that troublous times evoke. Uppsala university had deteriorated, and the affairs of the Catholic Church had long been neglected by the Holy See.

A young Swedish nobleman, Gustavus (Gustaf) Eriksson Vasa, who had been treacherously carried off to Denmark as hostage and prisoner, in 1519 made his escape to Lübeck, where he won good friends. At the end of May 1520 he landed again in Sweden near Kalmar and quickly learned of the massacre in Stockholm, in which his father and brother-in-law had perished. He made his way up to Dalecarlia, where he raised his first peasant troops. He had neither arms nor money, but one province after another came to his support. Archbishop Trolle, his foremost opponent, had to take flight from Sweden. Gustavus formed an engagement with Lübeck, which sent ships with arms and soldiers. In June 1523 Gustavus was elected king of Sweden by an assembly in Strangnas, and Stockholm surrendered to him.

But it was not easy to guard what had been won. Lübeck had secured for itself a monopoly of Sweden's foreign trade as well as exemption from tolls and bonds for very considerable sums in payment for services rendered. Drastic taxation had to be imposed, and the churches and monasteries also were requisitioned for help. Archbishop Gustaf Trolle was a fugitive, and the Swedish church found itself in a state of dissolution. Gustavus Vasa's first appeals to the pope were for bishops of Swedish birth, a new archbishop and church reformation. No favourable response came from Rome, and in these circumstances the king was influenced by Lutherans. The New Testament was translated into Swedish. At the *riksdag* held in Vasteras in 1527 the church settlement decreed

that the bishops had to hand over their castles to the king, and the chapters and canons, moreover, their superfluous incomes; that Peter's pence should no longer be paid to the pope; and that the pope's authority should not be invoked for the election of bishops. The year afterward the king was crowned in the cathedral of Uppsala.

Gustavus Vasa had to defend the interests of the kingdom against the individual claims of the provinces. The peasants had become disturbed by the novelties of the Reformation, and an adventurer came upon the scene, declaring he was Sten Sture's son. Disturbances broke out but the king put them down with severity. Christian II had been driven out of Denmark also, and this produced a *rapprochement* between Sweden and Denmark. Finally Christian was incarcerated in a Danish prison in 1532. Meanwhile, Lübeck's privileges had come to be intolerable. In 1531 every church was called upon to hand over its largest bell to supply means for meeting the Lübeck liabilities. Lübeck declared war on Denmark and Sweden; with the help of German allies its forces took Copenhagen and Malmo. They could not hold out, however, and in 1536 peace was declared. In the following year a treaty was framed between Sweden and Lübeck regarding certain exemptions from tolls but without trade monopoly. Sweden's debts were declared to be settled.

King Gustavus had now to organize administration. He heard talk about Roman law as the basis of all law. He consequently called in a great number of Germans, so that the years 1538–44 came to be known as the "German period"; but the Germans were for the most part adventurers and the results were not satisfactory. The newcomers caused the king to distrust his earlier helpers. There was a dangerous outbreak in Smaland, led by a peasant, Nils Dacke, who was killed in 1543. The influence of the Germans then ceased, and Gustavus ruled alone. He devoted great attention to mining, trade and agriculture. An influential landed nobility was restrained. At the *riksdag* meeting of 1560 the king presented his testament. He died on Sept. 29 of that year, aged 64. He was the founder of the new Sweden.

Eric XIV.—Gustavus Vasa's eldest son succeeded. Scholarly, display-loving, wayward, he ruled well at first but became insane. Gustavus had bestowed on his other sons dukedoms which entailed discords between them. On the occasion of his coronation Eric created the ranks of counts and barons, but he was mistrustful of the nobles, upon whom he imposed great exactions. He entered into negotiations for marriage with Queen Elizabeth I and afterward with Mary Stuart, as well as with two princesses on the continent, but met with refusals. Toward the close of his reign he married his low-born mistress, a good-hearted woman, Karin Mansdotter.

When the Teutonic Knights of Estonia and Livonia ceased to be, the town of Tallinn and the neighbouring region went over of their own accord to Sweden. This was the first step toward a Swedish sovereignty over the Baltic south of the Bay of Finland. The situation pointed ominously to coming conflicts, for Russia took Narva, the Danish duke Magnus took Osel and Poland took a large portion of the Teutonic Knights' country.

A break with Poland resulted in internal dissensions, as Eric's brother John, who was duke of Finland, married Catherine, the sister of the Polish king Sigismund II, and as security for a loan to his brother-in-law got some Polish castles in Livonia. A brief civil war took place. Eric's troops took Abo, and John and Catherine were put in prison, after being tried and sentenced by the *riksdag*. Negotiations were entered into with Ivan IV the Terrible in Russia and ended with a proposal for a treaty containing the condition—a blot upon Eric's name—that Duke John's wife should be handed over to the tsar, evidently to be used as a means of enforcing concessions from the Polish king. The shameful clause may be ascribed to Eric's madness, and it was never enforced. A severe conflict ensued with Denmark, which wished to have the union preserved and kept Sweden's "three crowns" in its coat of arms. Eric wanted to break the cordon in the south and would not recognize Denmark's claims to supremacy over the Baltic. He ostentatiously added the Danish and Norwegian kingdoms' arms to his own. Other disputes followed, and war resulted in 1563. The Swedish fleet won victories but the Danes had the mastery in the war on land, which was waged on both sides with grim ferocity. Lübeck and Poland allied themselves with Denmark. In 1567 Eric's insanity became manifest. He had some of the foremost nobles imprisoned, murdered one of them himself and had others of them assassinated. He became penitent and set free Duke John, who, in alliance with his younger brother Charles and the nobles, dethroned and imprisoned Eric in 1568. He died in prison in 1577.

John III.—John, an amiable man, learned in theology but irresolute and weak, was acclaimed king. The Danish war died a natural death through the economic attrition of both powers, and in Dec. 1570 peace was declared at Stettin through Polish and French mediation; the frontiers remained the same as before the war. Gotland was allotted to Denmark; Alvsborg on the west coast was to be surrendered by Sweden with a large sum of money; the question of the three crowns was left for later decision, but in the meantime Denmark was free to incorporate them in its arms; the Danish and Norwegian arms were to be taken out of the Swedish coat of arms. Sweden's gain consisted in the fact that from this time forward it was to be considered Denmark's equal as a Baltic state. Between Sweden and Poland good relations sprang up at once, but war broke out with Russia, waged with changing fortunes and occasionally interrupted by armistices. As a result the whole of Estonia, with Narva, was incorporated with Sweden.

The great religious struggle in Europe was followed with interest by John, who sought to take up a mediative role and entered into negotiations with the pope. Both papal and Spanish envoys came to Sweden, and it looked as though the king would go over to the Catholic Church, but when the pope refused to accept his offer of mediation he changed his tone and entered into definite opposition to Rome. He had previously given out a liturgy which evoked general displeasure in Sweden by its approach to the Catholic form of religious services; taken as a whole, his religious measures caused great unrest. His son Sigismund adopted the Catholic faith and was elected king of Poland in 1587. The internal administration of the country was marked by continual lack of funds (despite high taxation) and at the same time by extravagance over new buildings. Falls in the currency made the financial situation worse. The nobles were rewarded by special privileges for their part in the overthrow of Eric—a departure from the more democratic policy of Gustavus and Eric. The nobles now showed a tendency to aggrandize their powers, but the peasants never became serfs and the *riksdag* of the four estates continued to be called together. There were serious complications when John died in 1592.

Charles IX.—The crown went to Sigismund, who was scantily gifted, obstinate and a devoted servant of the Catholic Church; his uncle Charles on the other hand was a practical man of affairs, passionate and hard. The evangelical learning had taken still stronger root, and at a church meeting in Uppsala in 1593 it was decided to abide by the Holy Scriptures, the three great symbols and the unmodified Augsburg Confession. Sigismund was forced to agree. On the day previously, however, he had protested to the papal nuncio against the undertaking, and he considered himself entitled to disregard it. He returned to Poland in 1594. Duke Charles and the council soon fell out, with the result that the leading members of the council went off to see Sigismund. In 1598 Sigismund landed with a small Polish army, but he suffered defeat and returned hastily to Poland. This caused the Swedish *riksdag* in 1599 to declare him dethroned. Charles took a sanguinary revenge. The captive members of the council were executed.

Duke Charles was not crowned until 1607. He aimed at a monarchical-democratic conception of government. During the years 1561–93 the *riksdag* or other estates were convened 15 times; during Charles's much briefer reign, 1594–1611, they were called together 16 times. New church dissensions were caused by the fact that Charles cherished Calvinistic ideas while the clergy and the people expressed themselves more and more strongly in favour of evangelical tenets. By the peace with Russia concluded at Teusina in 1595 Sweden's ownership of Estonia and Narva had been recognized, but the troubles in Russia evoked new conflicts.

Swedish troops marched into Moscow, and the Russian crown was offered to one of Charles's sons. Novgorod was conquered. War with Poland was waged in Livonia, where the Poles won certain advantages without being able to turn them to account. Because of Charles's endeavour to extend the region of Finnmark up to the arctic under the rule of Sweden, and from other causes, war broke out with Denmark in 1611 and Kalmar was taken by the Danes. When Charles died in 1611 he left to his son, Gustavus Adolphus, an inheritance of three unfinished wars.

Gustavus II Adolphus.—Gustavus Adolphus, born in 1594, was declared of age on his father's death. He had been given a thorough classical education and spoke Latin and several modern languages. He was an organizer of great capacity both in military and in internal administration. His personal influence was attributable to a strong and good character and a clear and far-sighted vision. In all his work he had assistance of extraordinary value from his chancellor of state Count Axel Oxenstierna (*q.v.*).

The Danes were successful in the war and took the harbours on the west coast. Through the mediation of King James I of England peace was concluded in 1613. The peace conditions were hard: Alvsborg was to be handed back to Sweden in return for a payment of 1,000,000 dalers within six years; the Finnmark region was given to Denmark-Norway; and the right of the Danish king to use the Swedish three crowns in his coat of arms was recognized.

Sweden next desired peace and a fixed frontier with Russia. The war was continued until the conclusion of peace at Stolbova —also with King James's mediation—in 1617. Sweden won eastern Karelia and Ingermanland, whereby land connection was obtained between Finland and Estonia, but gave up Novgorod. During and after these wars a great task of organization was carried out in Sweden. The council acquired a stronger position as the king's advisers at the head of the administration. Ministerial offices were set up after the French-Burgundian model; a method of judicial procedure was instituted and courts of appeal were established. The nobility was incorporated in the social system as officials of the government, and much was done to improve local administration. The king promised not to begin war or conclude peace or make a treaty with other powers without the sanction of the *riksdag*. The ransom of Alvsborg made heavy taxes necessary, and even so the town of Amsterdam had to help with a loan. The financing of the state and the raising of the loan were based in a high degree on the copper ore in the Falu mines. In Jan. 1619 the ransom sum was paid and Alvsborg was returned to Sweden. Gustavus Adolphus endeavoured to promote commerce. New towns were founded, of which Goteborg on the west coast was the most prominent (1621). A source of future difficulties lay in the privileges of the nobles. Their graduated freedom from taxes—the reward of the servants of the state—was to diminish thereby the state's revenue from taxation; this danger, however, did not present itself until later. The nobility secured the stability which it needed through the founding of the *riddarhus*, or house of the nobles, in 1626, in which records were inscribed of the noble families. Higher education was furthered. The University of Uppsala got from the king in 1624 a great donation of lands. Great attention was given to the equipment of the army and navy. Relations with Poland were settled by armistices until the year 1617, when negotiations led to no result. Sigismund had never abandoned his claim to the Swedish crown and therein lay a menace for the future. Open war began again and Gustavus Adolphus took personal part in it. Livonia and Riga were conquered, and for several years the Swedes carried the war into Polish Prussia. Through the mediation of Brandenburg, France and England, a six years' truce was agreed to at Altmark in Sept. 1629; some Prussian seaports continued to be Swedish possessions, the Swedes retaining the right to levy tolls.

Meanwhile the great conflict that was to be known as the Thirty Years' War (*q.v.*) had started within the German kingdom, and Sigismund's efforts were menacing to the religious peace of Sweden. The emperor's troops were sent to Prussia to help Poland. The royal houses of Sweden and Denmark were related to German princes who asked for help. Catholic France sought to check the emperor's increasing influence and negotiated with Sweden and Denmark; the English government was working to the same end. Christian IV of Denmark entered into the war but was completely defeated, and Wallenstein's imperial army took up its position on the Baltic. Wallenstein was made duke of Mecklenburg and imperial admiral over the Atlantic ocean and the Baltic; an imperial fleet was to be built. When Stralsund was threatened, Gustavus Adolphus sent help (1628), and after long and many deliberations with the council and the *riksdag* he made his way with 13,000 men to Pomerania in June 1630; he preferred to have Germany as his theatre of war rather than to await an attack in Sweden. When success attended his invasion of the country, France concluded a subsidiary treaty with him at Bärwalde in Jan. 1631. The evangelical princes of Germany took up a dubious attitude at first, and the Swedes were blamed when they did not succeed in relieving the town of Magdeburg, which was taken by the imperial general Tilly. After various negotiations, a decisive victory was won at Breitenfeld in Sept. 1631.

A series of brilliant Swedish successes followed. Gustavus Adolphus held his court in Mainz and Frankfurt, crossed the river Lech and made his entry into Munich. The emperor had dismissed Wallenstein in 1630, but now took him back into favour, and soon the latter was once again in the field with a strong army. The Swedes vainly stormed his fortified camp at Nuremberg, and the brisk war which ensued led to the battle of Lützen in which Gustavus Adolphus met his death on Nov. 16 (O.S. 6), 1632.

It may be thought a matter for wonder that Sweden, a country with a population relatively so small, should have been capable of so great a political and military expansion as was shown in the Thirty Years' War. The explanation lies partly in an extraordinary national effort to avert an imminent danger. Sweden's armies were made up also in an increasing degree of foreign mercenaries, German, English and Dutch. In the Polish war a number of excellent Swedish generals had learned their business, and Gustavus Adolphus was one of the foremost military leaders of the time.

Christina.—The death of Gustavus Adolphus produced a crisis in Germany. His daughter Christina, still a minor, was recognized as queen (1632–54) and the regency was undertaken by Axel Oxenstjerna. It was well that a new treaty was concluded with France in 1633 and that Wallenstein was murdered in Feb. 1634, but in the autumn the Swedes suffered a defeat at Nördlingen which lowered their prestige. In 1635 Saxony concluded peace with the emperor, and its example was followed by Brandenburg in 1636; both declared war against Sweden. Johan Banér (*q.v.*) restored the good name of the Swedish troops by victories at Wittstock in 1636 and Chemnitz in 1639. His successor, Lennart Torstensson (*q.v.*), thrice invaded the emperor's dominions and won the second battle of Breitenfeld in 1642. Almost everywhere Swedish interests were opposed by Denmark. Torstensson was ordered, therefore, to leave Bohemia for Denmark in 1643, and Jutland was overrun by Swedish troops. Christian IV was forced to conclude a peace at Bromsebro in 1645 by which Sweden won Jamtland, Harjedalen, Gotland, Osel and, for 30 years, Halland; exemption from the Oresund tolls was accorded for goods to Sweden and the Baltic provinces. By the treaty of Westphalia (*see* WESTPHALIA, PEACE OF) Sweden acquired important German possessions—portions of Pomerania, Wismar, the bishoprics of Bremen and Verden as fiefs of the German empire, and prize money as rewards for the soldiers. A number of Swedes had in the course of the war amassed fortunes; the influence of the nobles in this way had been increased and a chasm had grown between them and the unprivileged classes. This was intensified by the circumstance that many foreigners had entered into Swedish service and had been rewarded with landed estates. Through them and through the impressions which Swedes had gathered in foreign countries, new ideas regarding the social status of the nobles had become prevalent.

The exercise of governmental power and the administration of the state in general were regulated by the constitution of the year 1634. Christina came of age in 1644, but laid down the crown ten years later. Richly endowed by nature, she had acquired much learning and many accomplishments but she was flighty and arbitrary, wont to overload with gifts those who were momentarily

in her favour. She went over to the Catholic church and died in Rome in 1689 after a life not lacking in excitement.

Charles X Gustavus.—Her successor was Charles Gustavus, son of Charles IX's daughter Catherine, who had married John Casimir of Palatinate-Zweibrücken. He had been brought up in Sweden, becoming *generalissimus*. The costs of the court were cut down, and in the *riksdag* of 1655 the nobles agreed that the endowed estates necessary for administration of the kingdom, for defense of the realm and for working the mines should return to the crown, that certain estates bestowed as gifts should acquire the character of fiefs with right of return to a new king, while a fourth of such estates as had been bestowed after the year 1633 should be returned to the crown. This measure was put into effect at once but ceased to operate during the strenuous war years that followed.

Charles X's short reign was occupied by war. With Poland there was as yet no peace; as the Russians had directed an attack against Poland, the king and the *riksdag* decided to settle the matter by means of a war. Charles X took Warsaw and Cracow (1655). Poland looked as though lost, but a national rising followed in favour of King John Casimir and two years of fruitless fighting resulted. Charles X concluded peace with the rising power, Brandenburg, and the Swedish and Brandenburgian armies together won a three days' battle at Warsaw in 1656. In that year the Russians began war and the emperor joined Sweden's enemies, and the following year Denmark, Holland and Brandenburg followed suit. If all these powers could have co-operated closely Sweden's position would have been extremely perilous, but they could not. Charles turned resolutely upon Denmark, took Jutland and led his troops first over the ice to Fyen and then across the islands to Sjalland. The credit of this military exploit belongs chiefly to the quartermaster general of the army, Erik Dahlberg. Denmark was forced to make peace in 1658 at Roskilde, Sweden receiving Skane, Halland, Blekinge and Bornholm, Bohuslan and Trondheims Lan. Both kingdoms undertook to make common cause to keep enemy fleets out of the sound. The king now thought of confronting his other enemies, but as certain negotiations with Denmark did not quite take the turn he expected and the Dutch were busily agitating in Denmark, he guessed that hostilities from this country would follow the moment he had become engaged in the east. He resolved to avoid this danger by completely crushing Denmark and began war anew in Aug. 1658. His policy, however, failed, as in Poland, through a contingency with which the statesmen of those days seldom reckoned—the kindling of a national patriotic movement. Copenhagen withstood his attack and Holland sent a fleet which after a fierce struggle with the Swedes relieved the Danish capital. The situation was anxious, and a *riksdag* was called together at Goteborg. Charles X Gustavus died on the night of Feb. 12–13, 1660, during its session.

Charles XI.—Charles XI, a boy of four, succeeded his father as king at a moment when Sweden was surrounded by enemies. The widowed queen and five officials became regent. The first thing done was to secure peace: with Poland, the emperor and Brandenburg at Oliva in 1660, whereby Sweden's right to hold Livonia was recognized and John Casimir renounced all claim to the Swedish throne; with Denmark the same year at Copenhagen, whereby Trondheims Lan and Bornholm went back to Denmark, compensation being given in the case of Bornholm; and with Russia in 1661 at Kardis, confirming the peace of Stolbova. Magnus Gabriel de la Gardie, the chancellor of state, was weak, and none took the lead in the way Axel Oxenstjerna had done after the death of Gustavus Adolphus. The nation's defenses were allowed to decline although its foreign policy was adventurous. It was desired to obtain subsidies from other powers, and a rash treaty was concluded with France in 1672. In this year the king came of age, having reached his 17th year. His education had been neglected and he grew up awkward in his bearing, but he was an ardent patriot and extremely industrious.

The general decadence of Sweden became manifest when its foreign policy led to war with Brandenburg (1674) and Denmark (1675). The war with Brandenburg was ill managed, and Sweden's military prestige was lowered by a defeat at Fehrbellin in 1675; most of its German possessions were lost. The Danes invaded Skane where they continued to have many friends. The king took the leadership. On the sea the Danes were successful, but Swedish victories at Lund in 1676 and other places saved the southern provinces. Peace treaties were signed in 1678 and 1679 with the different enemies. France accepted terms which in some respects were humiliating for her ally Sweden. Some small regions in Germany were lost, but there was no question of any cession of land to Denmark.

Great portions of the country had been laid waste, the fleet had been destroyed, trade and commerce had suffered and finances were in disorder. A great task lay before the young king and he accomplished it, but it was by the methods of an absolute monarchy. The *riksdag* was summoned several times—1680, 1682–83, 1686, etc.—but its attitude toward the king grew more and more complaisant and it gradually came to leave practically all matters of legislation and taxation to him. An inquiry was held into the conduct of the members of the regency, and they were sentenced to restore or pay large sums of money; a very drastic confiscation of private estates was taken in hand whereby land and incomes were made to fall to the treasury. Charles then put the army and navy in good trim. The officers were remunerated by small farms; cavalry was raised by a certain number of assessed farms and the infantry by the landowners. In the year 1681 the Swedish national debt amounted to 44,000,000 dalers in silver, but on Charles's death it was only 11,500,000 dalers. Charles married the Danish princess Ulrica Leonora, but the two northern states soon fell apart again. In 1681 Sweden entered into treaties with Holland at The Hague with a view to safeguarding the frontiers fixed by the peace treaties of Westphalia and Nijmegen. The emperor gave his signature to the compact the following year. Sweden wanted to draw Holstein into its sphere of interest so as to secure less interrupted communications between its different German possessions; Charles therefore gave what support he could to his maternal uncle Christian Albert, duke of Holstein-Gottorp, in his disputes with Denmark. For Denmark, conversely, it was important to hinder the Holstein plans of expansion lest Denmark should be hemmed in on the south. Charles XI died in April 1697, leaving one son, Charles, and two daughters.

Charles XII.—For the third time in the 17th century a regency was called on to act, but in Nov. 1697 the *riksdag* declared the 15-year-old Charles XII to be of age. At the coronation he placed the crown on his own head and gave no *kungaforsakring* (literally "king's guarantee"). Charles was a gifted and precocious youth who had been grounded in book learning and military science. He had high morality, but was inclined to obstinacy. In his early youth he was addicted to wild pranks and perilous sports. Within three years of his accession, however, the Great Northern War (1700–21) had broken out.

Poland, Denmark and Russia were united against Sweden. In March 1700 the Poles marched on Riga and the Danes against Holstein. The young king showed unusual power and decision. Protected by the English and Dutch fleets he landed on Sjalland (Zealand) and enforced on Denmark the peace of Traventhal in Aug. 1700. In November the Swedes, under the personal leadership of Charles, won a momentous victory over the Russians at Narva. In the summer of 1701 Charles marched south, drove the Russians and the Saxons over the Duna and invaded Poland. Charles called upon the Poles to dethrone Augustus II the Strong who had begun the war. This stirred the national feeling of Poland and won Augustus help from other states. Charles's route went through Kaunas in Lithuania to Warsaw and Cracow which he took (1702). After continual but fruitless victories for the Swedes a Polish parliament, not constitutionally formed, declared Augustus dethroned and elected as his successor Stanislaus Leszczynski, a weak man who lacked the strength to defend his own status as king. The Polish parliament concluded peace with Sweden in 1705. In this way Poland became a kind of subject state to Sweden, but Charles miscalculated Sweden's capacity for maintaining such a relationship. During these Polish conflicts the tsar Peter took town after town in Livonia and Ingermanland, among them Dorpat and Narva. Charles, in the meantime, held

that the most important thing was to force Augustus to conclude peace before he took up his stand against Peter, and he carried war into Saxony. The Swedish army was now one of the most famous in Europe, and Charles's power was very great. Leipzig was taken, and Augustus, by the peace signed for him by his ministers at Altranstädt in Sept. 1706 and ratified by him in October, renounced the Polish throne and allowed Charles to remain some time in Saxony with his army. Augustus sought by intrigue to overrule the peace treaties but had to reaffirm them in Jan. 1707. From different quarters came efforts to benefit by Swedish aid; nearly a dozen German princes sought Charles out, and he was visited by 30 envoys from various courts. The most notable of them was the duke of Marlborough, who took note that the Swedish king did not intend to place himself on the side of France in the War of the Spanish Succession as rumour had declared. Quarrels with the emperor began, but with the help of English and Dutch mediation a peace was concluded in 1707 which secured for Protestants in Silesia the church rights promised them in the peace of Westphalia.

Charles was anxious to come to conclusions with Russia, but he had given the tsar too long to prepare. The Russians had forced their way into Poland but were driven out again, and after a victory at Holowczyn (July 1708) Charles crossed the Dnieper. Adam Ludwig Levenhaupt was to have come with an army from Livonia to his support but was delayed and experienced a severe defeat, arriving eventually with a diminished force and without commissariat. Devastations by the Russians made it difficult for the Swedes to provide themselves with necessaries, and when the Cossack Ivan Mazepa-Koledinsky offered his co-operation Charles concluded a treaty with him in Oct. 1708. Mazepa did not manage to get his Cossacks to join with him, but Charles advanced into the Ukraine. In an unusually hard winter thousands of Swedish soldiers perished. Charles began to lay siege to Poltava; the battle which followed (June 28, 1709) ended in a Swedish defeat, largely because of the fact that Charles was wounded in the foot and that his generals were not united. Charles made his way as a fugitive to Turkey and his proud army capitulated at Perevolochna on July 1, 1709.

The unfortunate Russian campaign made an immense impression both in Sweden and in other countries. Augustus declared the peace with Sweden invalid and marched into Poland. The Danes crossed the sound and sought to overrun Skane. The government in Stockholm put forth all its powers to defend southern Sweden, and Magnus Stenbock won a decisive victory at Helsingborg in Feb. 1710. The Danish troops were taken back to Denmark. The Russians occupied the Baltic provinces completely after they had taken Riga. Meanwhile Charles was busying himself in Bender trying eagerly to bring about a war between the Turks and Russia. His stay in Turkey is difficult to explain; probably he did not want to return until the Russians had been either conquered or weakened. Thrice the Turks began war but without any gain to Sweden. Meanwhile Stenbock made his way to Germany, but though he defeated the Danes, who had overrun Bremen and Verden, at Gadebusch (1712) he was forced to capitulate with his army at Tönningen in 1713. The Russians and Saxons failed to take Stralsund, but in 1713–14 the Russians swarmed over Finland.

From Feb. 1713 Charles was held captive by the Turks, but after a swift journey he arrived at Stralsund in Nov. 1714. The country was in a sad plight. The population was only 1,250,000 and the burden of the great war was too heavy. The people were oppressed by heavy taxes and services imposed by the state and worn out by conscription.

During the years that followed Charles XII strengthened Sweden's defenses against the Danes, who received help from a Russian army and threatened an invasion of Sweden. The plan was not carried out, and Charles turned toward Norway to force the Danish king to make peace. While there he was killed by a bullet on Dec. 11 (O.S. Nov. 30), 1718.

In Stralsund Charles had made acquaintance with the Holstein baron von Görtz, who afterward became his confidant and associate. Unscrupulous and resourceful, Görtz discovered new means of taxation, simplified the state administration, lowered the value of the currency and conducted an adventurous policy abroad. The poverty thus brought about evoked a hatred against him which after Charles's death led to his being tried and executed. With Charles the Swedish supremacy over the Baltic came to an end. Swedish forces were withdrawn at once from Norway, one brigade which was sent against Trondheim experiencing terrible losses in the mountains. The *riksdag* assembled for the special purpose of putting an end to the absolute power of the sovereign.

Ulrica Leonora.—Charles's sister, the younger Ulrica Leonora, married to Frederick of Hesse-Cassel, was declared not to be the successor to the throne but was elected queen (1718–20) on condition that the *riksdag* alone should have the right to draw up the constitution. The changes effected in the constitution drawn up in 1719 and 1720 and in the *riksdag* decrees of 1723 were drastic. When the *riksdag* was not in session the council ruled together with the king who on most questions had only two votes. While the *riksdag* was sitting, decisions on foreign affairs and other matters were made by a secret committee which was formed of members of the three higher estates, the nobles, the clergy and the burghers. The peasants were left in the background but they had to be heard on questions of taxation. Each estate chose its own speaker; in the house of nobles the speaker was called *landmarskalk*. The membership of the council could be revoked only by law, and this made parliamentary government very primitive because it led to political lawsuits.

In 1719 peace was concluded with Hanover on the basis of Sweden's giving up Bremen and Verden and receiving a sum of money in return. In 1720 an arrangement was come to with Prussia, which acquired a large part of Pomerania and also paid a sum of money. In 1720 also peace was made with Denmark, which renounced the regions it had conquered in return for being allowed to keep Schleswig. Russian fleets, however, were harrying the coasts of the Baltic, and an attack upon Stockholm was with great difficulty warded off. Peace was not concluded with Russia until 1721, at Nystad, when Sweden gave up Livonia, Estonia, Ingria and parts of Finland; northern and western Finland was restored to Sweden.

Frederick I.—In 1720 the crown was transferred to the queen's consort Frederick I (1720–51). The leading statesman until the end of the 1730s was Count Arvid Horn, who pursued a prudent peace policy. The state's debts amounted to 60,000,000 or 70,000,000 silver dalers, and the emergency currency (*nodmynten*) was working much harm. A state bankruptcy was inevitable, and after this the emergency currency disappeared. With the English Navigation act as a model, a proclamation was issued in 1724 prohibiting foreign vessels from conveying to Sweden anything except the produce of the countries to which they belonged. The Swedish commercial fleet increased, an East Indian company was formed in 1731 (which caused bad blood in England), new manufactories were founded with support from the state, in particular the textiles industry, thanks to Jonas Alstromer, who created the manufactory at Alingsas. A new general law was framed in 1734.

During the first years following the peace of Nystad Russia sought to intervene in Sweden's internal affairs. Horn therefore in 1726–27 made Sweden a member of the alliance between England, France and Prussia. Indeed, foreign subsidies began to have their ill effects in Sweden. Against Horn's prudent leadership there grew up an opposition which gained ground in 1734 and became victorious in the *riksdag* of 1738–39. Horn gave up office in Dec. 1738. The victorious party was called Hattar, "the Hats" in contradistinction from Mossorna, "the Caps."[1] The program of the Hats favoured war with Russia with a view to winning back the Swedish provinces and power and a more intensive industrial policy; Sweden was to become a great power politically and economically.

The party of the Hats controlled the fortunes of the country until the 1760s, although sometimes strongly opposed by the Caps.

[1]These party names originated in the Progressives' nicknaming their opponents—who sacrificed everything to perpetuate an inglorious peace—the "Nightcaps," they themselves receiving their name from the three-cornered hats that they wore as officers.

They dominated the council from 1739 onward with Count Karl Gyllenborg as their leader. Conditions seemed favourable for a war with Russia in 1741 and war began. The campaign was badly conducted and the Swedish troops took to flight. Discontented thousands of peasants marched to Stockholm and could be dispersed only with difficulty. At the beginning of the *riksdag* session in 1742 the Caps were in power, and the peasants now won for themselves a place in the secret committee. An inquiry was conducted into the actions of the instigators of the war, and two of the generals were condemned to death and executed. The Hats now raised the question of the succession to the throne, the royal couple being childless, and the question soon became bound up with that of the Russian peace, as the Russian empress Elizabeth wished to see Duke Adolphus Frederick of Holstein elected. The peasants particularly wanted to have the Danish crown prince, and a great conflict raged in *riksdag* circles, much bribing being done by foreigners. In June 1743 peace was concluded at Abo. Russia retained only a small portion of its conquests in Finland, and Adolphus Frederick was chosen as heir to the Swedish throne. Denmark threatened war and Sweden's impotence became manifest when a Russian army came to the country to protect it against the Danes. While the *riksdag* was in session there was a split among the Caps, and the council came to include both Hats and Caps. Adolphus Frederick came to Sweden and in 1744 married Ulrica (Ulrike), sister of the Prussian king Frederick II. Russia's disposition to control the destinies of the country now took on a singularly offensive aspect. Foreign gold filled the coffers of the parties before the *riksdag* of 1746–47 assembled. The Hats opposed the Russian proceedings resolutely and in this were supported by many of the Caps. Through new elections to the council, the Hats now came definitely into power again, with Count Anders Johan von Hopken as head of the government. Some members of the Caps party were charged with treason and the matter was investigated by a commission. This had the effect of diminishing their opposition to the Hats. An alliance was contracted in May 1747 with Prussia and France which improved the situation. The successful party rewarded its friends lavishly and came down heavily upon its foes.

The years which followed were comparatively quiet. The brilliant young noble Count Carl Gustaf Tessin was president of the council, which was the strongest known within the period of the *riksdag's* predominance; national defenses were strengthened, customs tolls were raised to help industry, which was also supported by loans and premiums; agriculture was supported also by loans on easy terms. The improvement secured by these measures was not entirely real, and great credits were extorted from the Riksbank, which had to increase its issue of notes. Notes depreciated in value and exchange offices were set up to keep the currency at the desired height.

Adolphus Frederick.—King Frederick died in 1751 and was succeeded by Adolphus Frederick (1751–71), who was under the influence of his stronger-minded wife. Controversial pamphlets and journals came out in numbers. Tension between the king and the council increased. A rash effort at a coup d'état was made during the *riksdag* of 1755–56; the royal couple was humiliated and an autograph stamp was introduced in order to dispense with the king's personal intervention in the government. Count Axel von Fersen came to the front as the foremost man in the Hats party; he was chosen as *landmarskalk* (*see* above), leading the bureaucracy of the nobles.

Opposition grew therefore among the unprivileged classes against the nobility and the bureaucracy. The state's finances were in a bad way, yet in spite of this the country was let in for a war, on the side of France and Austria, against Frederick II of Prussia (1757), largely because of the French subsidies. The war was badly managed, but peace was made with Prussia without loss of territory in May 1762. The Caps took office in the *riksdag* of 1765–66. Russia and Prussia together decided to support the Swedish constitution, to prolong its general state of weakness and at the same time to work against the French influence. A Russo-Danish alliance was also formed with the purpose of maintaining the Swedish constitution. England was on the same side and it is calculated that these powers spent 3,000,000 silver dalers in order to ensure the victory of the Caps, while France's contribution to the Hats amounted to 2,000,000 dalers. The corruption was shameful. Against the wishes of the royal couple, the crown prince Gustavus married the Danish princess Sophia Magdalena.

The court had helped the Hats during the *riksdag*, and by the king's direct intervention the council was obliged to call together the *riksdag* which sat from 1769 till 1770. The Caps were now in a minority and their men were left out of the council which was controlled by the Hats. The court had been promised that the king's power would be increased, but the Hats did not keep their word. Nor did the *riksdag* succeed in solving the great financial problems. In Feb. 1771 Adolphus Frederick died. Somewhat previously Russia and Denmark had come to an agreement that, should any change be made in the constitution of Sweden in the direction of increasing the king's power, they would uphold the existing constitution (Dec. 1769) by force of arms. Sweden's constitution therefore seemed to be held in greater value by its enemies than by its own people.

Gustavus III.—The crown prince was in Paris at the time of his father's death. He returned home to meet the *riksdag*. The Hats and the court party were receiving money from France, the Caps from Russia and England. The parties were of about equal strength. Gustavus now decided to precipitate a change of the constitution. Col. Jakob Magnus Sprengtporten crossed to Finland to stir up a rising, in which he succeeded but not so quickly as had been expected. Johan Kristoffer Toll (*q.v.*), a superintendent of forests, went to Skane and won over the garrison at Kristianstad. Prince Charles, the brother of the king, also busied himself in Skane. The council got wind of the danger and the king had to act speedily. On Aug. 19, 1772, he collected around him the officers of the guard and arrested the council in the royal castle. The Stockholm troops declared for him and the people acclaimed him in the streets. The king became again head of the government, with the right to appoint officials. It was an attempt to divide the balance of power between the sovereign and the *riksdag*.

Russia and Denmark were furious but did not begin war, being much influenced by the resolute attitude of France. Gustavus surrounded himself with capable men. He was unusually gifted, but too much of a visionary. He had a real passion for the theatre but took up his business as king with honest intent; an inquiry into the conditions of governmental offices was initiated and a number of reforms in the administration were instituted as a result. In order to help the finances distilling was decreed a state monopoly and a number of distilleries were founded on behalf of the crown. This displeased the peasants, who wanted to be free to distil, and it increased drunkenness. Another cause for dissatisfaction lay in the king's efforts to transform the nobility into a court nobility and to manage promotion so as to break the power of the old bureaucracy.

A dispute with his mother discouraged the king, who next visited Italy and Paris; a subsidy treaty was concluded with France but Gustavus had no great success there. During the *riksdag* of 1786 the situation became worse; most of the king's measures were rejected.

Then Turkey's war with Russia in 1787 seemed to offer favourable conditions for a successful foreign policy. He began war with Russia in June 1788, taking command himself. But the sea fight at Hogland was undecisive, and a conspiracy, known as the Anjala league, was formed by certain officers. The conspirators opened negotiations with the Russian empress, to bring about peace. News came that Denmark had entered the war as Russia's ally, and Gustavus was obliged to hurry back to Sweden. The Swedish people became animated by patriotism and the king talked to the peasants as Gustavus Vasa had done of old. England and Prussia brought pressure to bear on Denmark to leave Sweden in peace and a truce was concluded. By the beginning of the following year, 1789, Gustavus was strong enough to have the leaders of the Anjala league imprisoned. The *riksdag* was called together at the beginning of February. The meeting was

a stormy one, the king, supported by the unprivileged estates, driving through the so-called Act of Union and Security, which embodied a number of far-reaching constitutional innovations: the council ceased to exist; the king acquired full powers over the administration of the state and the appointing of officials; and the *riksdag* lost its initiative in legislation. Once again the authority of the king became dominant. The *riksdag* became responsible for the national debts, the management of which was placed in the hands of the *riksgaldskontor*, which was dependent on the *riksdag* and which was given the right to take up new loans and to issue credit notes. A number of the leaders of the opposition had been imprisoned in the course of the session. Such was the king's second revolution. (*See* GUSTAVUS III; WALLQVIST, OLAF.)

The Russian war became a war of exhaustion on land. During the years 1789–90 there were several sea fights, the second of which, at Svensksund in 1790, constituted a fine victory under the king's command. In Aug. 1790 peace was concluded without modification of frontiers. But Sweden's finances were destroyed and the internal strife brought to a head. One result was a conspiracy against the king's life. He died on March 29, 1792, from the effects of a murderous attack by Captain J. J. Anckarstrom at a ball in the opera house on March 16, 1792.

Gustavus IV.—At the age of 13 Gustavus IV (1792–1809) became king of Sweden with his uncle Duke Charles as regent. Duke Charles's closest associate was Baron Gustaf Adolf Reuterholm, a man of great powers of work but pedantic and of little intelligence. The regency brought order into the higher grades of government service and sought above all to further the prosperity of agriculture. Reuterholm was not lacking in appreciation of the latest reforms: the freedom of the press was at first increased, but when the horrors of the French Revolution broke out it was restricted. The repute of the regency did not stand very high. It wavered between a French and a Russian alliance. The king, the duke and Reuterholm visited the Russian court in order to arrange a marriage between the king and a Russian princess, but the whole thing fell through when the king refused to give a guarantee of the princess's religious freedom. On the basis of a treaty concluded in 1794 between the two countries, the fleets of both Denmark and Sweden sought to protect neutral trade against privateers of England and other nations.

In 1796 Gustavus IV took the government into his own hands. He was scantily gifted but he knew how to make himself felt and combined obstinacy with passionate temper. So bad were the finances that the town of Wismar had to be mortgaged to Mecklenburg in 1801. In the course of a long journey in Germany in the years 1803–04 Gustavus came into touch with French *émigrés*, who inspired him with hatred of Napoleon. He took part in the coalition against France in 1805–07, with the result that Pomerania and Stralsund were lost. In the treaty of Tilsit Napoleon and the emperor Alexander agreed to attack Sweden unless it declared war against England and prevailed on Denmark to do the same. The army in Finland gave way before the Russians. The strong fortress of Sveaborg was treacherously surrendered without any real struggle. In the spring the Swedes marched again to the south of Finland and won several victories, afterward immortalized by the poet J. L. Runeberg. In Dec. 1808 the Swedish forces evacuated Finland. Gustavus IV became more and more difficult and he annoyed even England, his only ally. The feeling grew that the king must be dethroned. Gen. Georg Adlersparre started a revolutionary movement in Varmland, but before he reached Stockholm Adj. Gen. Karl Johan Adlercreutz had secured possession of the king's person (March 13, 1809). The king died in exile in 1837.

Charles XIII.—The *riksdag* was called together on May 1 by the provisional government and it decided to draw up a new constitution and elect a king. The form of government which resulted was based on a division of power between the ministry, the representatives of the people and the judicature. The king was to be advised by a ministry the members of which were to be appointed by the king but to be answerable to the *riksdag*. General legislature was to be the work of the king and the *riksdag*, and the *riksdag's* control over taxation was confirmed. As soon as this constitution had been adopted Charles XIII was recognized as king (1809–18). The credit of drafting the new constitution belongs chiefly to Hans Jarta.

King Charles was childless and it was necessary to choose a successor to the throne. Public opinion in Sweden at this period occupied itself with Norway's union with Sweden as a compensation for Finland. The *riksdag* chose as successor to the throne Prince Christian Augustus of Augustenburg, who was commander in chief in Norway, manifestly in the hope that the choice would help to bring this union about. The prince declared that he could accept this offer only after a declaration of peace. During the late summer of 1809 the Russian war was extended to the north of Sweden and the Swedish forces met with two defeats. The Russians in the meantime had become weary of the war and in September peace was declared at Fredrikshamn (Hamina); Sweden lost Finland and Aland. Soon treaties of peace followed with its other enemies. Pomerania was returned to Sweden.

The heir to the throne, who took the name of Charles, died suddenly in 1810, and a *riksdag* at Orebro had again to deal with the question of a successor. Opinions were divided when a suggestion was made that they should elect one of Napoleon's marshals, Jean Bernadotte, prince of Ponte Corvo. It seemed certain that Napoleon would approve, and accordingly Bernadotte was chosen and assumed the name of Charles John, to become the founder of a new Swedish dynasty, the succession to the throne having been provided for in an ordinance passed in 1810. The old King Charles XIII had shown courage and patriotic feeling during Gustavus III's war with Russia, but he was not very intelligent and he lacked strength of character. The crown prince therefore began to exert considerable influence, with stimulating and beneficent results.

In 1811 Napoleon prepared to break down Russia's resistance. A great war began, and the crown prince perceived that Sweden could not possibly go into it with France against England and Russia. He seems early to have thought out his plans for the conquest of Norway, and in 1812 he brought about a complete change in Swedish policy with this goal in view. Napoleon's occupation of Swedish Pomerania without declaration of war facilitated this change, and Russia and England favoured the idea that Norway should pass over to Sweden from Napoleon's ally Denmark. After the defeat of the French armies in Russia Charles John took part in the final conflict as commander of the northern army, with Swedish as well as Prussian and Russian troops under him. With these the crown prince, after the battle of Leipzig, made a movement against Denmark, which by the peace of Kiel (Jan. 1814) was forced to hand over Norway in return for Pomerania.

The Norwegians themselves, desiring independence, held an assembly at Eidsvold in May 1814 where they framed a constitution and elected the Danish prince Christian Frederick as king; but they were forced as the result of a short war to surrender. Christian Frederick laid down his crown and a *storting* endorsed the union with Sweden. Special conditions affecting the union were framed in the act of union or *riksakt* passed by the parliaments of both countries in 1815.

Charles XIV.—Bernadotte succeeded to the throne, as Charles XIV John, in 1818. The industrial economic revolution, evolving new classes which claimed their share in the government and burst asunder the old methods of representation, began in Sweden during his reign and embittered his later years. He was a strong personality, capable and lovable but distrustful of new ideas and apprehensive of too great changes. A liberal opposition was formed in the *riksdag*, above all among the burghers, and took a strong line against the government's repression of newspapers, against bureaucratic formalities and against the king's tendency to disregard the advice of his ministers. From the 1830s onward there was a question of reform in the matter of the representation of the people, and a great many bills were put forward without accomplishing much. In 1840 the state council was changed by the creation of special departments. With the king's active cooperation the state's foreign debts were paid while internal debts were diminished, and in 1834 a regulation of the exchange was effected. Much was done for the canal system; one section of the

Gotha canal was opened in 1822 and the rest of it in 1832. In 1842 the board school system was changed.

Oscar I.—Charles XIV was succeeded in 1844 by his son, a man of gifts and of more open mind but of less strength of will than his father. During the earlier years of his reign he stood for a policy of reform but later changed noticeably toward conservatism. During the last two years of his life he was ill, and the administration was taken in hand by his son Charles. The restrictions on agriculture and commerce were relaxed; the guild system was done away with in 1846; the old rules confining trade to the towns were removed. By decisions reached in the *riksdag* for 1844–45 the right to suppress newspapers was definitely abrogated. In 1847 an important poor-law system was introduced. One of the greatest questions of that period was that of communication: railways had to be built, but there were differences of opinion as to the laying down of the main lines through the sparsely inhabited land. In the *riksdag* of 1853–54 it was decided that the main lines should be built by the state, and the direction of the work was placed in the hands of Col. Nils Ericson. Friends of reform concentrated more and more on the demand for a thoroughgoing change in the formation of the *riksdag*, and many proposals were submitted without any of them being accepted. The revolutionary movement which marked the year 1848 on the continent aroused feeling also in Sweden and some street disturbances occurred in Stockholm. The government produced a new scheme of representation, which, however, was rejected by the *riksdag* of 1850. A strong Scandinavian movement manifested itself (especially in Denmark and Sweden, although the Norwegians also took part in it); great meetings of students were held, but no program for the unity of the Scandinavian countries with any real life in it was ever framed. When the Schleswig question led in 1848 to war between Denmark and the German states, a Swedish army was collected together in Skane and troops were carried over to Fyen, in case Jutland should be invaded; but they never had occasion to come into action. During the Crimean War Sweden and Norway maintained neutrality. In Nov. 1855 they reaped the benefit: a treaty was concluded with England and France by which the union states undertook not to cede any portion of their territories to Russia while the two other powers undertook to prevent any Russian efforts in that direction (the November treaty). Side by side with the peace treaty in Paris there was concluded a separate agreement between Russia, France and England in which Russia undertook not to fortify the Aland Islands.

The increasing consumption of spirits, as a result of the right to distil for household needs, called forth anxiety and led to a temperance movement conducted by persons such as Peter Wieselgren, a pastor. Private distilleries were done away with by the *riksdag* of 1853–54 and the distilling of spirits was regulated by the state. Ever since the middle of the 19th century the question of temperance has affected parties.

Charles XV.—The crown prince had conducted the government for two years when he became king under the name Charles XV (1859–72). He was a fascinating man who by his liveliness of mind, personal friendliness and artistic temperament won great popularity, but he lacked the thoroughness and perseverance in work and the seriousness which are called for in a statesman. His undeniable gifts were bound up with an extravagant imagination which in certain situations was a danger. This became most noticeable in the German-Danish war of 1863–64, when he gave the Danes distinctly to understand that they could count on an alliance between Denmark and Sweden-Norway, an idea which had never had the sanction of the state councils of either of the union countries and which was never to be realized. A number of Swedes, however, did enter the Danish army as volunteers.

The social changes which had come about in the life of Sweden necessitated a new form of electoral representation, and increasingly the general feeling was that it must be solved by giving up the division into four estates and that a *riksdag* with two chambers should be founded instead. The minister of justice, Louis de Geer (*q.v.*), took the lead in the matter and the government submitted a bill in the *riksdag* of 1862–63. The new constitution was finally accepted by all four estates in the *riksdag* of 1865–66, and in Jan. 1867 the first *riksdag* of the new kind was assembled. It soon became evident that it was the peasants who had benefited most by the reform: within two years they dominated the second chamber and their party—the so-called Lantmanna party—was a power to reckon with.

Through the influence of the finance minister, Baron J. A. Gripenstedt, Sweden was led more in the direction of free trade, and its treaty of commerce with France in 1865 was an epoch-making event in the development. All duties on grain were removed. The right of trading was extended in 1864. Among more important reforms may be mentioned the criminal law and the maritime law of 1864. The wars which were in progress on the continent caused the *riksdag* to grant large sums for armaments but the government's proposal for a reorganization of the national defenses was rejected.

Oscar II.—After a prolonged illness Charles XV died in Sept. 1872 and was succeeded by his brother Oscar II. The most noteworthy event in the latter's reign was the dissolution of the union with Norway.

The union plans of 1815 were never quite clear, and great doubts had always existed regarding the documents in question and their precise bearing. The Norwegians, apprehensive of Swedish domination, had sought always to give the narrowest possible application to the provisions of the act of union, while the Swedes, generally speaking, sought to give them a wider application, although with increased powers for Norway. The first great crisis arose over the question of the post of *stathallare*, or governor, of Norway. At first this post had been held by Swedes, but as this had invited opposition they were succeeded by Norwegians. In 1859 the *storting* resolved to abolish the post altogether; but in Sweden it was contended that as the post was provided for in the act of union it could not be rejected by Norway alone. The question was shelved for a time, and in 1873 Oscar II sanctioned the resolution of the *storting*. The first great union committee had met in the 1840s and had been followed by several smaller committees and in the years 1865–67 by a great committee which worked out a plan for a new agreement between the two kingdoms. This, however, had been rejected by the *storting* in 1871. The position of the sovereign of the two kingdoms was now becoming extremely difficult, because modern parliamentarianism had progressed much further in Norway than in Sweden and the powers of the sovereign had consequently become much less extensive in Norway. Thus the king had been forced to make continual concessions to the Norwegians. In the year 1835 it had been decided that the premier or another member of the Norwegian government should be called into the Swedish-Norwegian ministerial council whenever questions concerning the foreign affairs of Norway alone or both kingdoms should be under discussion. In the 1880s a dispute arose about this joint ministerial council. From 1891 onward the *storting* demanded Norwegian consulates for Norway. Also, it reduced the annual grant for the court. Because of the protectionist demand of Sweden a Swedish-Norwegian joint law (dealing with reciprocal freedom from customs duties) ceased to exist in 1897, and in 1898 the king ratified a decision of the *storting* to remove the joint union emblem in the upper right-hand corner of the Norwegian flag; it disappeared afterward from the Swedish flag. A small union committee put forward in 1902 a proposal for separate consular services but joint foreign office. Negotiations took place and in 1903 understanding seemed to be in sight. The Swedish plan provided that the foreign minister might be either a Norwegian or a Swede. But from both sides there came charges of breach of faith, and in Norway all parties demanded dissolution of the union. The Swedes were astonished when on June 7, 1905, the *storting* declared the union with Sweden to be dissolved. The *riksdag* acquiesced in the dissolution of the union upon certain conditions, such as that Norway should destroy certain frontier fortresses erected near the Swedish frontier, that the regions adjoining the frontier should constitute a neutral zone without fortifications and military manoeuvres, etc. King Oscar laid down the Norwegian crown on Oct. 26, 1905.

During the first years of King Oscar's reign, however, the question of national defense had been the dominant one, the first

chamber contending for an improved system and the peasant members in the second chamber making demands which could not for a long time be conceded. A number of army organization proposals were rejected, and during the 1880s the question of customs duties caused bitter disputes. The increasing exportation of grain from the United States and Russia together with the strong industrial competition among all countries resulted in a protectionist movement in Germany and France which extended also to Sweden. The second chamber was protectionist and the first in favour of free trade; in 1887 the majority in the first chamber was so small that it would have been possible to bring about a protectionist system by the collective vote of the two chambers. The king then dissolved the second chamber, which returned with a majority in favour of free trade. The *riksdag* of 1888 on the other hand accepted the customs system, as the general election in the autumn of 1887—through a technical mistake in the election of the free-trade candidates in Stockholm—had given the protectionists a small majority. This issue had profound consequences both in the formation of parties in the second chamber and in other respects. New men came forward, the most conspicuous of whom was the landed proprietor E. G. Bostrom, premier from 1891 to 1900 and 1902 to 1905. The question of the national defenses was provisionally settled by the *riksdag's* decisions in the autumn of 1891 (at an extraordinary session) and in 1901. In the latter year the Swedish military tenure established was completely abolished; the training time of the men liable for military service was increased from 90 to 240 days in the infantry and a longer period in the case of the navy and special branches of the army.

At this period new legislation was necessitated by industrial development and the great increase in the number of industrial workers. In 1881 the conditions of labour for children in all factories were regulated and in 1889 a law was passed on providing against injury while at work; the former law was amplified in 1900 by one dealing with the condition of industrial work for women and children. Trade unions were formed and collective agreements governing labour conditions were arranged between employers and workers. In 1889 a Social Democratic party was formed and in the autumn of 1896 Karl Hjalmar Branting was elected to the second chamber as its first Social Democratic member. The Liberals and Social Democrats demanded an extension of the suffrage for the second chamber, but it was not until the severance of the union with Norway had given a strong impetus to a closer national union on new lines that the question came to be decided. The severance of the union led to the formation of the coalition ministry of Christian Lundeberg, a manufacturer, whose task it was to carry out the dissolution of the union. He was followed at the close of the year 1905 by a Liberal ministry with Karl Staaff, a solicitor, as premier. The *riksdag* of 1906 rejected his suffrage bill, and the ministry resigned. Rear Adm. Arvid Lindman succeeded, with a Conservative ministry; he found a way, in 1907, of solving the suffrage question on the basis of universal and equal suffrage (with certain exceptions) together with the democratization of the communal scale of voting and proportional voting for both chambers. Among other important decisions may be mentioned the monopoly in the issuing of bank notes accorded to the Riksbank in 1897, the restriction of the right of companies to acquire land in northern Sweden in 1906, and the part ownership of the state in the Grangesberg company which owned the Lapland iron ore mines.

Gustav V to World War I (1914).—When Oscar II died on Dec. 8, 1907, the crown fell to his eldest son Gustav V. Modest and simple in his character, avoiding needless display, but capable of bearing himself with the dignity required by his position, he proved a good king for a democratic epoch. Reforms were carried through, as, for instance, the laws on civil marriages in 1908 and for the protection of the peasants in Norrland in 1909. In the summer of 1909 the country witnessed a great trial of strength between the industrial workers' trade unions and employers' associations, which culminated in an unsuccessful general strike.

In the autumn of 1911 the election for the second chamber was held with a doubled electorate and the Lindman ministry gave up office. Karl Staaff formed his second ministry. The Lindman ministry had approved a cruiser of a new type but Staaff's government set aside its decision. A movement to cover the costs of its construction by private subscription produced a sum of 17,100,000 kronor, and the ship was built. The people, indeed, had been disquieted by the way in which the Russians were arming in Finland and with widespread espionage by the Russians in Sweden, and new evidence of alarm was forthcoming in Feb. 1914 when 30,000 peasants met in Stockholm to request the king that the whole question of national defense should be handled simultaneously. The king's answer was favourable, and the Liberal ministry (which had passed a new law for workmen's protection in 1912 and introduced old-age pensions in 1913) resigned.

The king commanded the *landshovding* Hjalmar Hammarskjold (*q.v.*) to form a new ministry. K. A. Wallenberg became minister of foreign affairs. The second chamber was dissolved and the *riksdag* reassembled after the election of May 1, 1914. The new government had a majority in the first chamber but not in the second. (K. Hr.)

World War I.—By World War I Sweden found itself more or less cut off from the western world, from which it had bought a large proportion of its food supplies, raw materials and finished goods. The war put an end to some of the party strife that had rent the country, particularly in connection with the strengthening of defense. In due course Sweden declared its neutrality, measures for the reorganization of defense were passed, and Swedish neutrality was preserved all through the war despite external and even some internal pressure by activists favouring one or the other side. At the same time a common policy of neutrality bound the three Scandinavian states together and erased some of the bitter feelings between Sweden and Norway.

At first industry and commerce took an upward swing as a result of large purchases by the belligerents; soon, however, this intensive buying exhausted stocks which could no longer be replenished, and the export of some goods was made subject to licence. At the same time the state confiscated important goods at fixed prices, distributing them by rationing, and took over certain imports. Nevertheless, there was at times extreme shortage of some foodstuffs.

The Hammarskjöld government continued in office until 1917, when it resigned over the supply position. A Conservative ministry carried on until the general elections, which brought considerable gains for the Social Democrats. A Liberal-Socialist coalition under the Liberal Nils Eden took over, and Sweden's most famous Socialist, Hjalmar Branting, entered the government for the first time. The new government solved the supply problem by two agreements with Great Britain, the second one being the so-called *modus vivendi* agreement concluded in March 1918. The Russian Revolution, followed by the German, seriously affected Swedish thought and led to a revival of demands for a democratization of the constitution. As a consequence, a special session of the *riksdag* (*urtima*) passed a radical proposal that included *inter alia* votes for women. One effect of the bill was to make the Social Democrats the strongest party in the *riksdag;* the Conservatives were the second strongest, while the Liberals, who lost some seats, split into two groups in 1923—the National Liberals (Folkfrisinnade) and the Liberals—thereby further weakening their position. A small Communist party and a Farmers' party were also formed.

Another consequence of the Russian revolution, Finland's secession from the Russian empire, affected the Swedish scene. While the government recognized Finland's independence in Jan. 1918, it did not dare grant military aid to the new nation lest this should involve Sweden in the war or lead to difficulties at home—the sympathies of left-wing Social Democrats being with the soviets while a strong bourgeois opinion demanded aid for the tsarists. The government's decision adversely affected Finno-Swedish relations, and Finnish resentment was further increased by the dispatch of a Swedish naval expedition to the Åland Islands in February, when the islands were threatened by a Russian landing. Finnish circles interpreted this action as an attempt to express practical sympathy with an earlier demand by the Ålanders to be absorbed into the Swedish realm. Not until Sweden's calm acceptance of the League of Nations decision in favour of Finland's

retention of the islands did this resentment gradually die down.

The Interwar Period.—The great advance of the Social Democratic party was the most striking development between World Wars I and II. In 1920 the Liberal-Socialist coalition was followed by the first purely Social Democratic government under Branting, but the autumn elections resulted in a small setback for his party, whereupon a "business" ministry of a moderately Conservative character carried on until after the elections of 1921, when Branting formed a new government. This remained in office until 1923, to be followed by a series of short-term governments from the right or from the left, including a Liberal government under Carl Ekman that commanded only 33 votes and nevertheless carried on from June 1926 to the general elections of 1928. During this period the Social Democrats succeeded in passing a bill which greatly reduced armaments. The Conservative ministry which followed Ekman's government was defeated over its proposal to increase the corn duties and another Liberal government took office in 1930. Ekman resigned following the financial failure and suicide of the prominent Swedish businessman, Ivar Kreuger. In the autumn elections of 1932 the Social Democrats and the Farmers made considerable gains and the Liberal government resigned.

In Oct. 1932 Per Albin Hansson formed his first ministry. Its program included strong measures to overcome the depression, aiming however at financial expansion rather than economies. The Social Democratic party was not strong enough to force these measures through the *riksdag*, and an arrangement with the Farmers—the so-called *Kohandel* ("underhand deal")—was concluded. Against the combined votes of the two parties most affected by the depression, the opposition had to yield. Gradually general developments in industry began to justify the government measures, and unemployment began to fall steeply. Apart from a few months in 1934, when the Farmers governed alone, the Social Democratic party—either alone or in coalition with others—remained in office throughout the ensuing period. In the elections of 1934 it nearly attained an over-all majority. The government's policy was always prevention rather than cure of social and economic ills, and the country prospered accordingly. At the same time the feeling of security of tenure that their strong position gave to the Social Democratic governments was reflected in the increasing atmosphere of calm that prevailed in labour relations and led to a standardization of the so-called collective agreements which by 1940 were covering about 1,000,000 workmen.

Sweden shared in the economic effects of the post-World War I crisis. Early in 1922 unemployment rose to 160,000 and was fought by large-scale relief work. By the close of 1922, however, the nadir had been reached; Sweden's recovery kept step with that of the world outside. At the same time the democratization of the country increasingly affected the economic field. Labour had earlier been organized into trade unions and the co-operative movement flourished, but revolutionary ideas no longer prevailed and a policy of "progress by peaceful negotiation" was embarked upon. The eight-hour day was introduced without wage reduction, and despite shorter working hours production increased. State railways were gradually electrified, road traffic motorized and the merchant fleet substantially increased. The production of pulp was aided by improved forestry; agriculture benefited by research on seed selections; and highly developed engineering works produced increasing quantities of goods for export. In addition, Sweden joined the countries which exported capital.

During the depression in the early 1930s an agricultural crisis was followed by one in industry and unemployment assumed unusual proportions, leading to a number of grave labour conflicts. Sweden followed Great Britain in abandoning the gold standard, which relieved the position temporarily; but when the Social Democrats resumed office in 1932 it was still serious and improved only gradually.

After World War I, Sweden supported the League of Nations, prepared even to abandon something of its policy of neutrality. However, the gradual deterioration of the international situation, the advent of Nazi Germany and the failure of the League to save Abyssinia had a sobering effect. Defense once again became a primary consideration, although a law passed in 1936 for the reorganization of the army aimed rather at modernization than at expansion. When the Germans invaded Austria, Hansson proposed immediate increases in defense in parliament, while the League was advised that Sweden proposed to resume its strict neutrality policy. The fate of Czechoslovakia pointed the danger still further. The government declined Germany's offer of a nonaggression pact, unsuccessfully attempted to bring about closer relations with other Scandinavian governments and also tried to co-operate with Finland in a possible defense of the unfortified Åland Islands, which attempt foundered against Soviet opposition.

World War II.—On the outbreak of World War II the Social Democratic government reiterated the country's official neutrality and strengthened defense. Negotiations were opened with Great Britain resulting in a comprehensive war trade agreement in Dec. 1939. After the outbreak of the Russo-Finnish war a coalition government of all parties was formed. "Finland's cause is ours" became the general watchword. The material aid given to Finland was of unparalleled generosity and endangered Sweden's own security, but Finnish requests for regular troops were refused, although volunteers for Finland were helped and encouraged. Allied demands for the transit of British and French troops to Finland were also met with a refusal from Sweden as they were from Norway.

The German invasion of Denmark and Norway isolated Sweden from the west. The German government put heavy pressure on Sweden to allow the transit of troops and armaments; not until after the end of the war in Norway did Sweden enter into the agreement permitting German soldiers on leave and goods, including war material, to pass over certain Swedish railways. The outbreak of the second Russo-Finnish war forced Sweden to a still greater concession. On June 25, 1941, permission for the transit of a whole armed division from Norway to Finland had to be granted. These agreements aroused bitterness and widespread criticism in Sweden and were the subject of sharp Allied protests. The traffic came to an end in the late summer of 1943, after Germany's position had deteriorated and Swedish military strength increased. Until then Sweden had been forced to supply Germany with iron ore and other important materials in exchange for the import of essential goods; now stringent reductions in Swedish exports to Germany were made, and a new war trade agreement concluded with Great Britain and the United States.

Despite its precarious position Sweden succeeded in steering clear of all German attempts to persuade it to participate in the German "new order" policy and all through the war it was a shelter for refugees from German oppression, not only Norwegians, Danes and Finns, but also Balts and Jews of all nationalities. After the war, relief, largely in the form of credits, amounting to about $600,000,000 was extended mainly to Norway and Denmark, and medical supplies, food and clothing were sent wherever needed. The Swedish Red Cross under Count Folke Bernadotte was actively engaged in aiding prisoners of war and had the government's generous support in these activities.

The Postwar Period.—After World War II the Social Democrats dissolved the coalition and governed alone. The first postwar elections in 1948 confirmed them as the strongest party. The unexpectedly slow recovery of devastated Europe, the cumulative effect of unfavourable world trends, Europe's growing economic dependence on the United States and the temporary elimination of Germany from world trade seriously affected Swedish economy. To aid in the revival of Europe, Sweden adopted a generous foreign credit policy. When the expected revival did not materialize a serious economic crisis ensued. Large cuts had to be made in imports and expenditure was reduced, although these reductions never affected social welfare or defense. Gustav V died in 1950 and was succeeded by his son Gustav Adolf (*see* GUSTAVUS VI).

In Oct. 1951 Tage Erlander, prime minister after the death of Hansson, formed another coalition with the Farmers. The economic position gradually improved. The country as a whole blamed the coalition for the high food prices, and the elections of 1952 resulted in slight losses for both government parties.

The government continued to improve the social services and on

Jan. 1, 1955, a compulsory health insurance scheme for the whole nation came into force. On Oct. 1, 1955, the rationing of alcohol came to an end. The general elections in 1956 resulted in further losses for the coalition government. The position in the *riksdag* was then as follows: Social Democrats, 106 (110 in 1952); Liberals, 58 (58); Conservatives, 42 (31); Farmers, 19 (26); Communists 6 (5). The Social Democrats had to make important concessions to the Farmers to preserve the coalition. In Oct. 1957 a national referendum took place on the subject of an additional national pension scheme, in which the plan advocated by the Social Democrats proved victorious. Thereupon the Farmers resigned from the government, and on Oct. 31 Erlander formed a Social Democrat minority government.

Sweden joined the United Nations in 1946. The increasing tension between east and west confronted the government with serious problems, and the proposed formation of the North Atlantic treaty led to much heart searching. Sweden proposed in 1948 a common defense pact to Denmark and Norway. When this failed (largely because the United States was unwilling to supply arms to such an organization), Sweden decided to revert to the traditional policy and not to join the North Atlantic Treaty organization. It was fully understood, however, that only a strong defense could protect such neutrality, and increasing amounts were spent on armaments. In all other respects Sweden supported international cooperation. It joined the Council of Europe and took an active part in the work of the Organization for European Economic Cooperation.

In April 1954 the tercentenary of the first treaty of friendship and commerce between Sweden and Great Britain was celebrated and a British-Swedish chamber of commerce with its headquarters in Stockholm was inaugurated. At the end of 1954 a cyclotron with a capacity of 200,000,000 v. was completed at the Swedish Nobel institute. Relations with the U.S.S.R. and eastern European countries were affected in 1955 by the discovery in Sweden of cases of espionage in which members of the embassies of Communist countries in Stockholm were involved. In March 1956 an agreement between Sweden and Norway was signed under which Sweden was allowed greater facilities for its transit traffic through Norwegian harbours in Trondheim fjord, which would prove useful in time of war. The agreement also provided for the construction of a Swedish oil port in Trondheim fjord. In Oct. 1957 the great Arctic iron ore company, Luossavaara-Kirunavaara, was nationalized at a cost of 825,000,000 kronor.

BIBLIOGRAPHY.—Ingvar Andersson, *Sveriges Historia* (Stockholm, 1943); R. Svanstrom and C. F. Palmstierna, *Short History of Sweden* (London, 1934); C. Hallendorff and A. Schück, *History of Sweden* (Stockholm and London, 1938); the Tidens Forlag, *Svenska Arbetarklassens Historia* (Stockholm, 1941 ff.); John Lindgren, *Per Albin Hansson i Svensk Demokrati* (Stockholm, 1950); Royal Institute of International Affairs, *The Scandinavian States and Finland* (London, 1951). For a further selective list *see* S. E. Bring, *Bibliografisk Handbok till Sveriges Historia* (Stockholm, 1934); for a more extensive one, K. Setterwall, *Svensk Historisk Bibliografi* (Uppsala, 1907, 1923, 1939 ff.). (A. H. Hs.)

POPULATION

The population of Sweden is low in comparison with the area of the country (44 per square mile) and unevenly distributed.

Table I shows the distribution of population in 1960 in the administrative districts or counties (*län*), which differ from the former division into provinces (*landskap*).

The population according to the 1960 census (7,495,129) was about four times as large as it had been in 1750 (1,780,000) and twice what it had been in 1850 (3,480,000).

The average annual excess of births was 7.86 per 1,000 in the 19th century, reaching a maximum of 10.39 in the years 1841–60; in the 1930s it had fallen to 2.78, with a minimum of 2.09 in 1935. There was a rise during World War II culminating with 9.57 in 1945. After that it diminished again, being 6.43 in 1950 and 5.17 in 1956.

Emigration, especially to the U.S., increased rapidly from 1860 onward, with a maximum during the period 1881–90 (average 38,000 a year). Then the U.S. legislation on immigration and also the attainment of a comparatively high standard of living in Sweden in the years after World War I caused a decline in the number of emigrants, and in the 1940s there were practically none. In 1956 there were 14,737 emigrants, including 4,258 Swedish citizens.

TABLE I.—*Area and Population*

Län (county)	Area (sq.mi.)	Population 1960 census			
		Urban	Per cent of total	Total	Density (per sq.mi.)
Älvsborg	4,928	152,410	40.6	375,006	76.1
Blekinge	1,173	59,313	41.0	144,468	123.1
Gävleborg	7,614	118,176	40.3	293,070	38.5
Göteborg och Bohus	1,986	484,757	77.6	624,762	314.6
Gotland	1,225	15,626	28.8	54,322	44.4
Halland	1,904	73,068	43.0	170,060	89.3
Jämtland	19,903	24,777	17.7	139,800	7.0
Jönköping	4,436	139,250	48.8	285,271	64.3
Kalmar	4,487	79,408	33.7	235,770	52.5
Kopparberg	11,724	78,849	27.6	285,862	24.4
Kristianstad	2,478	59,125	23.1	256,475	103.5
Kronoberg	3,827	33,268	20.9	158,977	41.5
Malmöhus	1,878	425,595	68.0	625,667	333.1
Norrbotten	40,879	81,655	31.2	261,672	6.4
Örebro	3,492	135,018	51.5	262,239	75.1
Östergötland	4,278	207,438	58.0	357,693	83.6
Skaraborg	3,262	87,466	35.0	250,180	76.7
Södermanland	2,645	135,781	59.7	227,615	86.0
Stockholm (city)	72	808,484	100.0	808,484	11,228.9
Stockholm	2,997	203,072	44.0	461,082	153.8
Uppsala	2,084	90,556	53.9	167,735	80.5
Värmland	7,497	107,233	36.8	291,085	38.8
Västerbotten	22,834	51,029	21.3	239,625	10.5
Västernorrland	9,924	77,202	27.0	285,620	28.8
Västmanland	2,615	132,964	57.2	232,589	89.0
Totals	170,143*	3,861,520	51.5	7,495,129	44.1

*Total excludes 3,522 sq.mi. of large inland lakes not apportioned to any *län*.

The immigrants were at earlier periods considerably fewer, but they exceeded the emigrants during World War I and after 1930. The highest figures after World War II were approximately 33,000 in 1948, 32,000 in 1951 and 30,000 in 1955. Besides the large number of refugees, many immigrants were attracted by the favourable conditions for obtaining employment, and skilled labourers (particularly from Italy and Germany) were offered employment through the intermediacy of the government.

In 1960 there were 3,738,696 male and 3,756,433 female inhabitants. The population is racially more homogeneous than in most other countries. There were in 1950 about 34,000 Finns living in the frontier region and 5,300 Lapps in the far north.

Population is denser in the south than in the north: in 1960 Malmöhus *län* had 333 persons per square mile and Norrbotten *län* 6. The annual excess of births over deaths in 1951–55 was about 40,000 or 5.52 per 1,000. The low birth rate of the 1930s was one of several reasons for expecting the working population to be stationary up to 1960, when the increase in the birth rate that took place during and after World War II began to have its effect upon the labour market.

The urban population was 51.5% of the total in 1960. Up to 1860 this percentage had remained nearly stationary at about 11, growing subsequently with the progress of the industrialization. Stockholm, the capital of Sweden, had in 1960 a population of 808,484, Göteborg 404,349 and Malmö 228,878. Norrköping, Helsingborg, Örebro, Uppsala, Västerås, Borås, Linköping, Eskilstuna and Gävle had each between 50,000 and 100,000 inhabitants. In all, there were 133 towns, 100 of which had more than 10,000 inhabitants.

Religion.—More than 99% of the total population belongs to the Established Church of Sweden, which is Lutheran and has the king as its supreme head. Among the minorities the Jews (about 6,700 in 1950) and the Roman Catholics (20,000) are the most important. There are also certain rather important "free-church movements" (numbering about 400,000 adherents), many of which are recognized within the Established Church. Freedom to leave the Established Church without joining a "foreign religious community" (*i.e.*, without becoming a Jew, a Roman Catholic, a Methodist, Episcopalian or a Swedenborgian) was introduced at the beginning of 1952 but did not seem to bring about any immediate consequences.

With the archbishop of Uppsala as primate, the Established Church of Sweden is divided into 13 dioceses each headed by a bishop. The church parishes number 2,500.

Education.—Public elementary education became free and compulsory in 1842. In 1954 compulsory schooling was increased from eight to nine years. Besides the two old universities in Uppsala (founded in 1477) and Lund (founded in 1668), there are universities in Stockholm and in Göteborg. Stockholm has also special graduate schools of technology, forestry, commerce, dentistry, veterinary science, pharmacy, arts, music, etc.; Göteborg has graduate schools of technology and commerce; Malmö has a graduate school of dentistry; Umeå, in the north of Sweden, has a graduate school of dentistry and one of medicine was established in 1958. For advanced studies in social work and public administration, there are special institutes in Stockholm, Lund and Göteborg.

Self-education is an outstanding feature in Swedish cultural life: nearly 65,000 study circles were in action in 1957, with students numbering about 690,000. Considerable government subsidies are given for this form of adult education as well as for popular lectures. People's colleges, the first of which date from the 1860s, provide a special type of voluntary education for persons about 20 years of age. Organized as boarding schools, many of them are sponsored by political parties, by bodies concerned with the labour market or by popular movements.

In 1956 there were 8,464 primary schools with 816,533 pupils and 41,885 teachers, 142 higher elementary and intermediate schools with 45,346 pupils, 252 secondary schools with 141,684 pupils and 537 vocational schools with 118,724 pupils. There were 26 teachers' training colleges with 4,250 students. In 1956 the four universities had together 16,935 students with teaching staffs totaling 1,433 professors and lecturers. The remaining 11 institutions of higher education had 6,882 students in 1956.

Social Welfare.—About 10% of the national income is used for welfare and social security, including compulsory health insurance, maternity benefits, child allowances, old-age pensions, rent rebates to large families, free school lunches and schoolbooks and equipment. About 29% of the state budget expenditure goes on social security and public health.

Conditions on the labour market are often regulated by voluntary decisions taken jointly by leading organizations concerned (the central body of trade unions, the central body for the professions, certain agricultural organizations and the employers' union). In this way it was agreed in 1946 that "enterprise councils" should be introduced into industry. Working hours, vacations (three weeks' paid holiday), safety arrangements and other provisions for protecting the workers, collective bargaining and rights of association are settled by law.

Traditions and Recreations.—In some parts of Sweden—Dalecarlia, Värmland, Gotland—old customs are well preserved and picturesque peasant costumes still sometimes worn at festivals. Christmas is the main festival of the year. People dance around Maypoles on Midsummer's Eve and light bonfires on the eve of May day. Of later origin are the celebrations for the "queen of lights" (St. Lucia) on Dec. 13.

Special dishes include hard rye bread in flat cakes (*knäckebröd*) and *smörgåsbord*, an hors d'oeuvre consisting of different kinds of snacks (herring, salmon, eel, salads, slices of meat, etc.) and often served at the beginning of the meal with bread and butter and, sometimes, a glass of liquor (*snaps*). In view of the social evils associated with drunkenness prohibition was suggested by the temperance movement in 1910, but was rejected by a plebiscite in 1922.

Ivan Bratt's system of rationing, introduced in 1917, subjected the purchase of liquor to severe control by law. This system was abandoned in 1955, the one remaining regulation being that in restaurants a light meal must be ordered in connection with alcoholic beverages (except wine).

Swedish gymnastics from the time of Per Henrik Ling have had a world-wide following. The National Athletic association, formed by the various clubs, has a total membership of about 600,000. Other favourite forms of sport are skiing, skating, ice hockey, bandy (an ice game with a ball), sailing, rifle shooting and track finding. There are 41 golf courses. Lawn tennis had a special patron in King Gustavus (Gustaf) V and became increasingly popular. Association football is the most popular sport.

GOVERNMENT, JUSTICE AND DEFENSE

Sweden is a limited monarchy, the constitution resting primarily on the Constitution act (*Regeringsformen*) of 1809. June 6, the day of enactment and also the day when Gustavus Vasa entered Stockholm in 1523, is now celebrated as Sweden's national day. The throne is hereditary but may be occupied only by male members of the royal family. The executive and judicial authority is vested in the king alone, but his resolutions must be countersigned by a member of the cabinet (*statsråd*). The cabinet is appointed by the king and is responsible to the parliament (*riksdag*), without the consent of which it cannot act. The government consists of a prime minister (*statsminister*); the heads of the 11 departments of administration, which are justice, foreign affairs, defense, social affairs, communications, finance, church and education, agriculture, commerce, interior affairs, civil services; and three or four ministers without portfolio. Administrative posts are equally open to men and women. Holders of government offices are appointed by the king. Apart from provincial governors and a few others, who are as a rule appointed for a period of six years, Swedish civil servants are undismissable except in case of default and after trial and judgment.

The central administration outside the ministries is on the whole independent of politics, which usually play little part in appointment to government offices. The cabinet shares legislative powers with the *riksdag* and introduces new bills in it; and their members have also the right to introduce new bills. In order to take effect legislative bills must be approved by the government as well as by the *riksdag*. By the 1950s the general tendency in the Swedish political life seemed to be toward increasing the influence of the government, which for the two preceding decades had been based on a majority in the *riksdag*.

The *riksdag* is bicameral. Members of the upper chamber (150 seats), who sit for eight years, are elected by the town and county councils, and one-eighth of their number stands for election every year. The 230 members of the lower chamber are elected every fourth year by direct vote of all men and women over 21 years of age (universal suffrage dates from 1921). Candidates must be 23 years old to be eligible. The two chambers have equal authority and their agreement is required to enact a bill.

Amendments to the constitutional law, in which among other things the freedom of the press is guaranteed, must be decided twice by the *riksdag* and the cabinet in two different sessions with a general election intervening. There are nine standing joint committees for different fields of legislation.

Local Government.—Sweden is divided into 25 administrative districts or *län* (*see* Table I). The chief official of the *län* is the governor (*landshövding*), appointed by the king in council. Under him is the administrative board of the district with executive, fiscal and civil defense functions and control over the "local" police (under the ultimate authority of the ministry of the interior). Whereas these are the local branches of the state administration, there is on the other hand a far-reaching local self-government in the parishes, dealing among other things with health and sanitation, poor relief, education, fire and certain police work. Matters relevant to larger areas are decided upon by the representative bodies of the *län* (*landsting*). An average parish has a few thousand inhabitants; the areas represented by the *landsting* have several hundred thousand. The local units have the right to decide upon their own taxes to raise what money they need.

Justice.—Justice is administered by tribunals of three instances. In the first instance, there are 116 rural judicial circuits (*domsagor*), the court of which (*häradsrätt*) consists of a judge and from seven to nine lay assessors (*nämndeman*), who assist the judge with information and advice regarding the case. If seven out of eight or nine assessors present are unanimously of a different opinion from the judge, they can outvote him; but this seldom occurs. The town courts, called *rådhusrätt*, consist of the mayor (who in Sweden is a judge, not a chief of the local government) and at least two aldermen; they have assessors only in serious criminal cases. Second, there are six courts of appeal (*hövrätter*),

which are established in Stockholm, Jönköping, Malmö, Umeå, Göteborg and Sundsvall. Third, the supreme court (*högsta domstolen*) passes sentences in the name of the king, who is nominally the highest judicial authority. Cases dealing with administration and administrative law are eventually referred to the supreme administrative court (*regeringsrätten*), but there is no court trial of the constitutionality of administrative decisions. After 1948, when legal proceedings were reformed, argument was [in principle] oral rather than written.

Defense.—Sweden's armed forces consist of three services, the army, the navy and the air force. The command of all armed forces is held by the king and, under him, by the supreme commander, who has the defense staff at his disposal, co-ordinates the activity of the three services and directs military planning. Military service is compulsory for all males between 20 and 47 years of age; about 40,000 conscripts become liable every year.

In the army the preparatory training lasts 10 to 11 months (1952), to be supplemented later by three operational training courses each of one month's duration and with an average interval of six years. In wartime the army comprises about 600,000 men mobilizable in a short time. The home guard is an integral part of the army and recruitment is voluntary, its members being mostly drawn from the age groups under 20 or over 47.

The navy, which includes the coastal artillery, consisted in 1957 of 3 cruisers, 23 destroyers and fast frigates, 33 motor torpedo boats, 24 submarines and about 100 small surface vessels. The coastal artillery is responsible for the defense of harbours, anchorages and potential invasion areas. Its permanent installations are as a rule blasted into the rock. The air force in 1957 comprised 50 squadrons, 30 of which consist of day fighters, 3 of night fighters, 12 of attackers and 5 of reconnaissance aircraft. By the middle of the 1950s, transition to jet propulsion was completed, the core being formed by the Swedish J29 ("Flying Barrel") jet fighters. Some wings have Hawker Hunter and De Havilland Venom fighters. Among the attackers is the Swedish Saab design A32 "Lansen," considered to be one of the fastest fighter-bombers in the world. The "Flying Barrel" would in due time be succeeded by the new Saab design, the delta-winged supersonic fighter J35 "Draken."

The armed forces are supported by various voluntary defense associations comprising several hundred thousand men and women; they make increasing use, for example, of the Swedish women's volunteer service. Equipment is to the greatest possible extent designed and constructed by Swedes. Every Swedish citizen between 16 and 65 years of age who does not belong to the armed forces may have to serve in the civil defense, which is trained in peacetime to be responsible in war for first aid, clearing and fire services as well as for protection against chemical and biological warfare. Air-raid shelters were being blasted into the rock in an endeavour to give protection even against atomic bombs.

ECONOMY

An important change in the structure of Swedish economic life took place after the 1870s. In those years about 70% of the population earned its living from agriculture, but in 1950 only 23% did so. During the same period industry increased its proportion from 15% to 41% of the population, becoming in the process the most important source of national wealth. The corresponding figures for commerce and communications are 5% (1870) and 22% (1950).

Agriculture.—Less than 10% of the whole area of Sweden is under cultivation, with the main agricultural regions in the south and around the big lakes in central Sweden; natural meadowland covers about 3%. More than half the land area is covered by forests; the rest is mountainous and barren. Sweden is practically self-sufficient as far as agricultural products are concerned. The yield per acre is high and exceeded by few countries in Europe. Wheat cultivation, through scientific plant breeding, has been extended almost to the Arctic circle and is superseding the cultivation of rye. Sugar beets are grown in southern Sweden and on the islands of Öland and Gotland, the crop being nearly equal to the domestic demand. Nearly 50% of all agricultural income is derived from milk and milk products.

TABLE II.—*Agricultural Production*
(In 000 metric tons)

Crop	1934–38	1948–52	1953	1954	1955	1956
Wheat	696	713*	988	1,020	716	952
Rye	409	281*	297	301	170	269
Barley	213	232	468	360	407	612
Oats	1,257	809	945	861	596	1,149
Sugar beets	1,888	1,777	1,997	1,848	1,663	1,786
Potatoes	1,847	1,865	1,727	1,429	1,285	2 079

*1948–50.

TABLE III.—*Livestock*
(In 000 head)

Livestock	1937	1947	1951	1956
Cattle	2,986	2,625	2,532	2,433
Pigs	1,425	1,195	1,374	1,573
Sheep	353	349	256	157
Horses	633	497	384	283
Chickens	10,980	14,023	11,654	11,601

After 1930 and particularly after World War II a large-scale mechanization and modernization took place within Swedish agriculture: the number of tractors in use grew from 2,100 in 1930 to 19,100 in 1939 and from 21,300 in 1945 to 107,458 in 1954, when there were, moreover, about 10,000 light tractors; and between 1944 and 1951 there were notable increases in the numbers of combines (from 407 to 6,200), mechanical binders (from 71,000 to 86,900) and milking machines (from 40,000 to 108,000) in use. The government encourages small farms to pool their machinery, and nearly every farmer is a member of the producers' co-operative groups for marketing agricultural products. About 90% of the arable area is privately owned, and not more than 20% of it is held on lease. Of the 414,000 farms, about 75% had in 1944 less than 25 ac. of arable land, while less than 2% had 125 ac. or more. In central and northern Sweden the farms mostly include some forest. The principal aim of Swedish agricultural policy is to further the creation of larger operative units.

Industry.—The bases of Swedish industry are the forests, the iron ore deposits, the water power and the technical skill of the workers. Between 1935 and 1950 the volume of industrial production was doubled (*see* Table IV), and between 1900 and 1950 it grew sevenfold. Small enterprises predominate. By far the greater part of Swedish industry is in private hands. Of the industrial workers only 5% are employed in government-owned factories and 4% in co-operative enterprises; thus privately owned firms employ 91% of industrial workers. The number of factories in 1950 amounted to 17,034 with a total of 663,000 workers,

TABLE IV.—*Industrial Production, 1934–56*
(In 000 metric tons, if not otherwise stated)

Industry	1934	1938	1946	1950	1953	1954	1955	1956
Coal	415	431	488	399	285	267	282	294
Iron (Fe content)	3,232	8,411	4,308	8,283	10,176	9,285	10,452	11,302
Pig iron	325	668	686	890	1,001	939	1,176	1,332
Crude steel	879	995	1,208	1,439	1,762	1,842	2,129	2,401
Copper, primary	7.9	10.7	14.5	16.7	17.0	16.6	17.3	16.9
Lead ore (Pb content)	8.2	8.6	21.3	22.7	25.5	30.0	32.2	32.7
Zinc ore (Zn content)	29.1	34.6	37.8	37.1	45.1	58.4	58.8	66.0
Aluminum, primary	0.3	1.9	3.6	4.0	9.6	10.6	10.0	12.5
Pyrites (sulfur content)	40	84	137	202	192	197	194	243
Electricity (000,000 kw.hr.)	6,030	8,162	14,197	18,177	22,437	23,958	24,721	26,631
Cement	583	993	1,462	1,936	2,352	2,465	2,550	2,495
Lumber, sawn (000 cu.m.)	...	5,952	5,398	6,220*	7,250	7,700	8,400	7,950
Wood pulp	2,870	3,061	2,736	3,188	3,200	3,647	3,903	4,119
Newsprint	259	258	270	330	341	342	366	412
Other paper	...	410	643†	693	759	886	935	...
Sulfuric acid	127	167	262	333	342	395	397	399
Superphosphates (20% P_2O_5)	...	260	301†	433	499	536	494	501
Nitrogenous fertilizers (in terms of N)	...	8.4	17.4	25.6	29.1	26.5	25.4	27.2
Merchant vessels (000 g.r.t.)	50	166	147	348	494	550	523	495
Motor vehicles (passenger cars / commercial)	11,485	26,690	2,582 / 10,516	19,490 / 9,392	32,366 / 11,577	44,716 / 17,553	47,282 / 19,114	51,608 / 20,675

*1951. †1947.

but only 55 factories employed more than 1,000 workers, and only 539 from 201 to 1,000 workers.

Timber and Wood Products.—Forests, mainly pine, cover about 87,000 sq.mi. or more than 50% of the surface area of Sweden. Current annual growth available to the timber industry is estimated at 60,000,000–70,000,000 cu.yd. Cutting and conservation are regulated by law, and large-scale scientific research is conducted by experimental stations and institutes with a view to increasing the yield per acre. One-half the forest area is owned by private persons, one-fourth by companies, one-fifth by the state and the rest by the church and local communities. After the outbreak of World War II the industry was concerned to supply not only the traditional timber, pulp, paper and board but also more valuable products such as rayon, alcohol, plastics, dyes, glue, resins, turpentine, etc.

Iron.—Sweden's iron ore deposits are among the richest in the world. They consist of the long-worked mines in central Sweden (Bergslagen; estimated deposits 220,000,000 tons) and of the later workings in northern Lapland (Kirunavaara, Gällivare, etc.; estimated deposits 2,000,000,000 tons). The domestic iron industry uses mainly ore from Bergslagen, which contains a singularly pure iron. The Lapland mines, the ore of which is phosphorus bearing but rich in iron content (60% to 70%), are worked for export. More than 90% of the output of iron ore is exported; about 17,000,000 tons were exported in 1956. Of Europe's total resources in ore of high iron content, more than 90% is to be found in the Lapland mines. There are also some ore deposits containing gold, silver, copper and arsenic; also zinc ore is mined.

Fuel and Power.—Sweden's coal deposits, which are found only in the south, normally produce not more than 300,000 tons a year, together with firebrick and fire clay, though output was doubled during World War II. Imports of coal and coke were estimated to amount to 7,200,000 tons per year in the 1950s, chiefly from Great Britain, Poland and the German Federal Republic. In addition, approximately 1,000,000 tons of crude oil and 2,000,000 tons of motor and fuel oil were imported. Wood and peat are also used as fuel. But electric power has more and more replaced fuel in homes as well as in industry.

After Switzerland and Norway, Sweden has the world's highest consumption of electricity per person. The installed capacity amounted in 1956 to 6,590,000 kw., including 5,240,000 kw. in hydroelectric plants. The output of energy rose between 1939 and 1957 from 6,030,000,000 kw.hr. to 29,124,000,000 kw.hr., after the harnessing of about half the existing water-power resources. More than half of the electric power is supplied by private companies. The trunk transmission line from Harsprånget in the far north to central Sweden, opened in 1952, first in the world to have a voltage as high as 380,000, has a transmission capacity between two and three times as high as that of lines with a voltage of 200,000.

Foreign Trade.—Sweden's foreign trade is relatively high. Including earnings from shipping, its value in 1952 was about $498 per head of the population, a figure exceeded by only a few countries. This was explained by the growing industrialization of the country's economy, requiring the import of many raw materials and semifinished products, fuel, machinery and chemicals. On the other hand Sweden's own products, such as lumber, pulp, iron ore, high quality steels, ball bearings, etc., were in great demand in the markets of the world. (*See* Table V.)

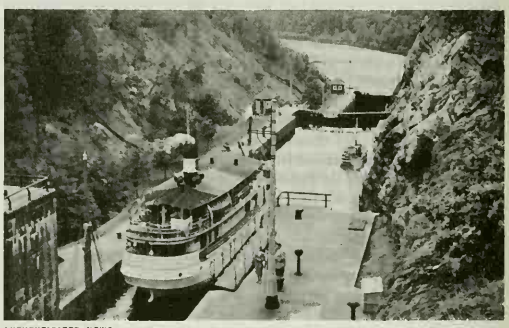
In 1951–54 about 71% of the value of Sweden's total foreign trade was with the European Payments union countries (including 22% with the sterling area), 11% with the dollar area and 18% with the rest of the world. Traditionally the largest buyer of Swedish goods was the United Kingdom (16.5% in 1952). From 1948 to 1951 it was also Sweden's foremost supplier, but in 1952 imports from the German Federal Republic exceeded those from the United Kingdom.

TABLE V.—*Foreign Trade*
(in 000,000 kronor)

Item	1934	1938	1946	1950	1954	1955	1956
Imports (c.i.f.)	1,307	2,087	3,309	6,117	9,192	10,337	11,428
Exports (f.o.b.)	1,303	1,844	2,548	5,708	8,195	8,933	10,047
Balance	−4	−243	−851	−409	−997	−1,404	−1,381

Communications.—More than 50% of Sweden's imports and exports are transported on Swedish vessels. The merchant navy grew considerably in the course of the 20th century. With a gross tonnage of 2,959,000 in 1956, it put Sweden in tenth place among the world's seafaring nations or, according to the relation of the tonnage to the population, in fourth place (after Norway, Great Britain and Denmark). In 1954 more than half the tonnage was engaged in ocean trading, on regular runs to all parts of the world. Sweden is one of the foremost shipbuilding nations. The amount of tonnage launched in 1957 amounted to 673,000 gross register tons, 70% of which was for foreign owners. The canal network is important.

In 1946 the Swedish, Norwegian and Danish commercial airlines were merged to form the Scandinavian Airlines system (S.A.S.) to operate services to North and South America. In 1951 an agreement was signed extending the merger to cover all the operations of the three companies. This merger agreement was made retroactive to 1950. In 1954 the S.A.S. was the first to fly over the north pole.



AUTHENTICATED NEWS

THE LOCKS AT TROLLHÄTTAN ON THE GÖTA CANAL

Sweden possesses a highly developed network of railways, the oldest of which dates from the 1850s. From the beginning the state took an active part in the construction of trunk lines: The total length of the lines amounts to about 10,000 mi. (more in relation to the population than that of any other European country), of which about 40% has been electrified (this portion however accounted for 86% of the total traffic in 1951). An act of parliament of 1939 provided for the gradual nationalization of railways that were still privately owned. Ferries connect the railways with continental lines through Denmark and Germany.

The total length of public roads is about 58,000 mi., to national highways covering 2,700 mi. and provincial highways 12,400. Motor vehicles registered in Dec. 1956 comprised 734,533 passenger cars and 120,615 trucks and buses. The density of population being low, bus lines are naturally of great importance to Sweden.

In 1957 Sweden had 2,312,000 telephones, or 315 per 1,000 inhabitants, which was the second highest proportion in the world. The number of radio licences (2,535,000 in 1956) in relation to the population (345 per 1,000) was surpassed only by the United States. There was no regular television broadcasting until Oct. 1956; the number of television licences was 70,000 in Nov. 1957. There is only one sound and television broadcasting system (Radiotjänst); it is partly state-owned and nonpolitical and carries no advertising in its programs. There are about 250 newspapers, of which approximately 150 are dailies (around 20 appear also on Sundays). Their total circulation is 5,300,000 copies. The Liberal press covers more than half the circulation.

Finance.—The state budget is divided into two parts, the budget of current expenditure and the capital budget, the latter being in principle devoted to government investments, which are covered by national loans. (*See* Table VI.) The gross national product increased between 1938 and 1956 from 11,060,000,000

TABLE VI.—*Current Budgets, 1938–1958**
(in 000,000 kronor)

Item	1938–39	1952–53	1953–54	1954–55	1955–56	1956–57	1957–58†
Revenue	1,442	8,058	8,085	8,607	10,072	10,091	12,183
Expenditure	1,349	7,647	8,130	8,619	9,691	10,978	11,294
Balance	+93	+411	−45	−12	+381	−287	+889

*Years ending June 30. †Estimates

kronor to 44,445,000,000 kronor (*i.e.*, four times) while the index numbers of cost of living (1935 = 100) rose from 106 in 1938 to 244 in Dec. 1957. The public debt rose between Dec. 1939 and Dec. 1957 from 2,664,289,000 kronor to 19,095,109,000 kronor.

The unit of currency is the krona (plural, kronor). Its exchange rate against the pound sterling was 14.5 kronor = £1 in 1958 (*cf.* 16.92 kronor = £1 before 1946); against the U.S. dollar the rate was 5.18 kronor = $1 in 1958 (*cf.* 3.60 kronor = $1 before devaluation in 1949).

The Swedish credit market is characterized by the multiplicity of types of credit institutions, official, semiofficial and private, among which the distribution of the functions and of the operating fields is not always clearly defined. The Bank of Sweden (1657), which is the oldest national bank in the world, is subordinated to the *riksdag* and not to the government. There were 19 commercial banks in 1958. Two of these, Svenska Handelsbanken and Skandinaviska Banken, held each about one-quarter of the assets of the whole commercial banking system; the third largest Sveriges Kreditbank, holding about one-tenth, is state-owned; two others, Stockholms Enskilda bank and Göteborgs bank, held about one-tenth. Whereas these five have a more or less nationwide network of branch offices, the other commercial banks are regional or local. Savings banks numbered about 450 in 1955, the total of deposits in them amounting to not much less than that in the commercial banks. Special state organizations supervise the activities both of the commercial and of the savings banks. Mortgage loans are advanced by a number of special institutions (for housing, agriculture, shipping, etc.).

BIBLIOGRAPHY.—H. W. Ahlman *et al., Sverige Nu: atlas över folk, land och näringar* (Stockholm, 1949); Royal Institute of International Affairs, *The Scandinavian States and Finland* (London, 1951); Swedish Social Welfare Board, *Social Sweden* (Stockholm, 1952); W. G. R. Howell, *Sweden* (London, H.M.S.O., 1954); E. F. Heckscher, *An Economic History of Sweden* (Stockholm, 1954). Royal Central Bureau of Statistics, *Statistisk Årsbok for Sverige* (Stockholm, annual, 1914 ff.). *See* also N. Afzelius, *Books in English on Sweden*, 3rd ed. (Stockholm, 1951). (G. HN.; C.-M. B.; F. BN.)

SWEDENBORG, EMANUEL (1688–1772), Swedish scientist, philosopher and theologian, one of the greatest and most learned men of his country, was born in Stockholm on Jan. 29, 1688, the second son of Jesper Swedberg, who was later appointed bishop of Skara. The family assumed the name of Swedenborg when it was ennobled in 1719. Emanuel was educated at Uppsala university; and after completing his course there he traveled abroad for five years (1710–14), visiting England, Holland, France and Germany in pursuit of scientific and mechanical knowledge. At that period his chief interests were mathematics and astronomy, and he was cordially received by John Flamsteed and by Edmund Halley. He was anxious to see and to hear Sir Isaac Newton, whose work he studied avidly, but there is no evidence that the two met. Swedenborg's inventive and mechanical genius flowered early, and his plans, practical and otherwise, ranged from a method of finding terrestrial longitude by the moon to new methods of constructing docks and even to tentative suggestions for the submarine and the airplane.

On returning home, Swedenborg began to publish a scientific periodical, *Daedalus hyperboreus* (Uppsala, 1716–18). Charles XII of Sweden was interested in the venture but would provide no funds, and after six issues the periodical was obliged to cease publication. Charles, however, in 1716 appointed Swedenborg assessor extraordinary to the Royal College of Mines, in which office, and later as assessor, he devoted himself for 30 years to the development and improvement of Sweden's metal-mining industry. But he continued to seek fresh knowledge in this and other spheres and from time to time traveled in Europe and in England in pursuit of learning in physics, chemistry, mechanics and philosophy. He had already published in 1718 the first work on algebra in the Swedish language; and in 1721 he published at Amsterdam his *Prodromus principiorum rerum naturalium* in which he offered a geometrical explanation of the phenomena of chemistry and physics. In 1734 he published his *Opera philosophica et mineralia*, 3 vol. (Leipzig), the first volume of which constitutes his *Principia rerum naturalium*. In this work, though without the facts that later scientists have demonstrated, he reached by inductive argument conclusions that are close to the theories of modern nuclear physicists. The first product of the infinite is conceived of as "pure motion" or conatus-to-create, and the subsequent creation is a complex of motion. The atom is composed of vortex particles, but these are compositions of motions more interior. The pattern strongly resembles the pattern of the atom as described in terms of nucleus and electrons. In the same work and from the same general principle Swedenborg suggested the nebular hypothesis to account for the formation of planets.

During the next ten years Swedenborg was engaged in anatomical and physiological studies. As he had sought to find the "soul" of creation in pure motion, he now sought to understand the soul of man and to find it in its own kingdom, which is the body. He published *Oeconomia regni animalis*, 2 parts (London and Amsterdam, 1740–41), and *Regnum animale*, 3 parts (The Hague, 1744–45). Although in this field of study he made many experimental investigations, he used them only sparingly, because he recognized the danger of trusting too much to his own observations and of "originating a whole series of inductive arguments from my particular discovery alone. . . . I therefore laid aside my instruments . . . and determined rather to rely on the researches of others than to trust to my own" (introduction to *Oeconomia*). In both works he therefore begins each section with numerous quotations from the accepted experimental anatomists and then proceeds to reason upon them. Gustaf Retzius acknowledged that Swedenborg had been the first to indicate the real nature of the cerebrospinal fluid. In the *Oeconomia* Swedenborg also showed, long before any other physiologist, that the motion of the brain synchronizes with that of the lungs and not with that of the heart and circulatory system; and he discerned the relationship between various regions of the cerebral cortex and corresponding motor regions of the body.

It is, however, in the field of theology that the name of Swedenborg is chiefly known, and the circumstances in which he turned to theological studies have tended to obscure his reputation as a great scientist and philosopher. He asseverates that his entry upon this new study was in response to a divine vision and call; that his spiritual senses were opened so that he might be in the spiritual world as consciously as in this world; and that the long series of exegetical and theological works that proceeded from his pen constituted a revelation from God for a new age of truth and reason in religion. Furthermore, he held that this new revelation of the Lord was what was meant by the second advent. Because of his otherworld experiences Swedenborg has often been regarded as a spiritualist "medium," yet it may be observed that the normal idea of mediumship is reversed, the man being admitted into the spiritual world, not the spirits into the material world. Nor can he properly be classed among the mystics, for in his dry, matter-of-fact accounts of the spiritual world and in his acutely reasoned theology there is little in common with mysticism. His manner is still that of the scientific and philosophic investigator.

Philosophically he had consistently recognized the infinite as the indivisible power and life within all creation, and he had never wavered in his belief that this infinite was God. In his theology he maintains the absolute unity of God in essence and person. A trinity of persons in one God is, he says, an impossibility of thought, but since God is the Divine Man there is necessarily in Him, as in every person, a trinity of love, wisdom and activity; and in God these are infinite and divine. It was this God who came into the world as Jesus Christ for the redemption and salvation of men. The inmost soul of Jesus was the Divine itself, the one God; yet, in order that there might be an intimate contact of God with fallen man, the Divine assumed from Mary a human nature comprising all the planes of human life. But during the course of Jesus' life on earth, by resistance to every possible temptation and

by living to their divine fullness the truths of the Word of God, all that was actually from Mary was expelled, and the planes of human nature were "glorified" or made divine from the indwelling Divine Soul and became the divine embodiment of that soul. The Trinity is therefore a Trinity in one Person, the Father being the originating Divine itself, the Son the manifestation and embodiment of that Divine and the Holy Spirit the outflowing life and activity of the "Divine Human."

In rejecting tripersonalism Swedenborg necessarily rejects the orthodox conceptions of redemption. The redemption of mankind was a deliverance from the domination of evil. The hells, which are the communities of the spirits of evil men in the spiritual world, were aspiring to enforce themselves upon men's minds, destroying their freedom to discern between truth and falsity and therefore between good and evil. By admitting into Himself their temptations and by complete resistance to them, Jesus broke their power; and the inflowing of the Inmost Divine into the human plane thus perfected interposed an eternal and infinitely powerful barrier between the hells and men. The "Divine Human," to which belongs "all power in heaven and on earth," stands forever between man and the forcible imposition of themselves by the hells. Man is thus free to know and obey the truth. His salvation depends on acceptance of and response to divine truth.

Swedenborg, accepting that all creation has its origin in the divine love and wisdom, accepts that all created things are necessarily forms and effects of specific aspects of that love and wisdom and so "correspond," on the material plane, to spiritual realities. The material realm is one of effects whose causes are spiritual and whose purpose is divine. The Scriptures, he says, are written according to this principle of "correspondence," which becomes the key to unlock their spiritual teaching. In *Arcana coelestia*, 8 vol. (London, 1749–56), he presents a detailed interpretation of Genesis and Exodus according to the principle of correspondence; and in the posthumously published *Apocalypsis explicata*, 4 vol. (London, 1785–89), he does the same with the Book of Revelation and incidentally with many other parts of the Scriptures. The theological teaching in his works was collated as a systematic theology in *Vera Christiana Religio* (Amsterdam, 1771). The philosophy of the creation and maintenance of the universe is dealt with in *Sapientia angelica de Divino Amore et de Divina Sapientia* (Amsterdam, 1763) and *Sapientia angelica de Divina Providentia* (Amsterdam, 1764). Swedenborg's best-known work is probably *Heaven and Hell* (*De Coelo et ejus mirabilibus, et de inferno*, London, 1758).

Swedenborg died in London on March 29, 1772, and was buried in the Swedish church. In 1908 the Swedish government had his body removed to Uppsala cathedral.

Swedenborg neither preached nor founded a church. The New Jerusalem Church, founded by a few disciples about 1784, remains small but has branches throughout the world. The Swedenborg society was founded in London in 1810 to publish his works.

Bibliography.—R. L. Tafel, *Documents Concerning the Life and Character of Emanuel Swedenborg*, 3 vol. (London, 1875–77); G. Trobridge, *Swedenborg, Life and Teaching*, 4th ed. (New York, 1944); E. Swift, *Swedenborg, the Man and His Works* (London, 1883; new ed., 1932); C. S. Sigstedt, *The Swedenborg Epic* (New York, London, 1953); also the *Transactions* of the International Swedenborg congress, 1910 (London, 1911). (A. Cm.)

SWEDISH LANGUAGE AND LITERATURE.

The Swedish language is about 1,000 years old. It began in the viking age when the Scandinavian national states first emerged, and the characteristic features which separate the modern Scandinavian languages developed. Old Swedish was originally spoken in Sweden itself—excepting the most northerly and southerly parts and certain western areas—in the coastal belts of Finland, Estonia and Livonia, with their islands, and in certain parts of Russia conquered by vikings, who here and there maintained themselves until the 13th century. In the modern world it is used by about 7,750,000 people; *i.e.*, by 6,750,000 living in Sweden, by 370,000 Swedish-speaking Finns (less than 10% of the Finnish population) and by about 630,000 Swedish immigrants in North America.

The ultimate ancestor of Swedish is Common Norse, one branch of the family of Germanic languages commonly divided into three groups: (1) East Germanic (Gothic); (2) West Germanic (Dutch, English, Frisian and High German); and (3) North Germanic—*i.e.*, Scandinavian (Swedish, Danish, Icelandic, Norwegian and Faroese). Swedish with Danish form the East Scandinavian group; the three others are West Scandinavian. Modern Swedish, though still more inflected than contemporary Danish or Norwegian, has dropped a great many of the inflections current, in some instances, into the 18th century (*see* below). The verbs are divided into four conjugations, the nouns into five declensions and the gender of these nouns into common (*den*) and neuter (*det*). Only two case forms have survived for the nouns: nominative and genitive. The genitive is formed by adding *s* to the nominative both in the definite and indefinite form, both in singular and plural; thus *dag, dags, dagens, dagar, dagars, dagarnas*. As an example of the changed forms compare singular nominative definite *dagen* with the classical Old Swedish *daghr-in, dagens* with *daghs-ins, dagarna* with *dagha-nir*, dative form *daghum-in* and accusative *dagha-na; cf.* Anglo-Saxon *daeg* (modern English "day"), German *Tag* and Danish *dag*. The adjectives are declined according to gender, number and definite and indefinite forms. Examples of the sound mutations between Danish and Swedish are Danish *gade* and Swedish *gata* ("street"), Danish *bog* and Swedish *bok* (book).

Runes.—Runes are the oldest form of Swedish writing. About 2,500 runic inscriptions survive from the earliest period of *fornsvenska* (*c.* A.D. 800–1225); most of them are found in Uppland and date from the 11th and 12th centuries. Cut in stone, often put into metrical form, they commemorate the genealogy and achievements of the dead. The most remarkable stone, Rökstenen in Östergötland, dates from the 9th century and is unique in the length and interest of its inscriptions, written in seven different types of script. Other interesting inscriptions have been found in Hälsingland, including the Nalstadsten, in the so-called "staffless" script. Unfortunately, save for the runes, nothing of the earliest Swedish literature has survived, but it is likely that there was some sort of literary culture akin to that of Iceland.

CLASSICAL PERIOD OF OLD SWEDISH

We possess a greater and more reliable knowledge of Swedish after the religious orders (first Cistercians, later Dominicans) had introduced the Latin alphabet, originally for the transcription of religious texts, and had founded schools to assist their missionary work. The earliest examples of Latin letters are on coins of the time of Olaf Skötkonung (993–1022). Then follows the classical period of Old Swedish (*c.* 1225–1375), with the beginnings of a native literature, inferior in quality and volume to that of Iceland. The purest Swedish language of this time is highly inflected, uses simple constructions and is surprisingly free from foreign influences, the few Greek or Latin loan words generally relating to church matters. Nouns have four cases, the suffixed article is inflected, and there is also much inflection of verbs.

The oldest extant manuscript in classical *fornsvenska* is not a translation from Latin, but part of a native legal code, *Västgötalagen*, compiled in the 1220s. A number of such local codes survive; *e.g.*, *Östgötalagen*, *Dalalagen*, records of provincial laws set down by a magistrate and preceding the lawbooks like *Upplandslagen* (1296) and *Södermanslagen* (1327), compiled on official initiative and ratified by the king. These codes are forceful and full of concrete images; they often use alliteration and solemn prose rhythm to make their pronouncements more memorable. They offer the happiest examples of the vernacular literature in their age, and culminate in Magnus Eriksson's *Landslag* (*c.* 1350) and in a similar work, *Um styrilsi kununga höfdhinga* (*On the Conduct of Kings and Princes*), probably written about 1330 for the guidance of the same ruler. This reflects both Swedish respect for law and the influence of Christian culture, even of scholasticism.

Dialects.—Differences of dialect occur in the runic inscriptions and in these legal codes, as in other products of this old literature. Two main groups of Swedish dialects, each comprising a number of subdivisions, have been classified: the Svea group, including the northern areas, Uppland, Hälsingland, etc., and parts of Öster-

götland; the Göta group comprising Värmland, Västergötland, parts of Småland, etc., and more recently the south Swedish provinces, which till long after the Reformation were Danish-speaking regions. The archaic language of Gotland is excluded from this classification. Even in the period of Old Swedish it differed so fundamentally from the Swedish of the mainland that *forngutniskan* ranked as a separate language. It may be studied in about 200 runic inscriptions and a 13th-century *Gutalag*.

Religious and Secular Literature.—The importance of Christianity in the cultural development of Sweden is immense, yet the new religion prevailed only gradually, and pagan rites existed well into the middle ages and in remote districts for much longer. Slowly, too, the primitive society of viking times was replaced by a more complex organization, dominated by the nobility and the clergy, patrons and providers of literature. Latin eclipsed the vernacular as a literary medium, and this helps to explain the derivative nature of mediaeval Swedish letters. Thus, though the poetry of chivalry reached Sweden, the Swedish writings inspired by chivalrous ideals were mere free adaptations of French or German originals. *Eufemiavisorna*, the best-known of these works, three verse romances translated anonymously between 1303 and 1312, are in rhymed doggerel, a metre again employed in the chronicle romance *Erikskrönikan*, a record of the fortunes of the Swedish kings between 1220 and 1320, probably compiled soon after the latter date. An anonymous and hybrid work, it is the most powerful single poem of mediaeval Sweden.

Only in religious writing, from among unknown Latin hymn writers, compilers of legendaries, etc., do a few personalities emerge. The Dominican Petrus de Dacia (*c.* 1235–89), with his *Vita benedictae virginis Christi Christinae*, is often considered the first Swedish author; Bishop Brynolf Algotsson (*c.* 1250–1317) is the first distinguished Latin hymn writer in Sweden. But both were overshadowed by St. Bridget (*c.* 1302–73), the founder of Vadstena monastery. Only the Latin version of her text, the *Revelationes*, survives, first printed at Lübeck in 1492.

YOUNGER OLD SWEDISH

The downfall of the Folkunga dynasty and the death of Bridget mark the end of a period. Under the impact of the stormy *Unionstiden*, with its foreign rulers, the character of society and of language changes, and we get the *yngre fornsvenskan* (*c.* 1375–1526), a language diluted with German loan words—mostly technical and commercial terms—and with Danicisms, which sheds a number of its inflections. At the same time there occurs a standardization of the literary language, largely under the influence of Vadstena *kloster*, the main focus of culture in Sweden until the Reformation. Both Latin and German affected the syntax of the vernacular.

Literature.—The insecurity of these times had an adverse effect on literature, yet with the growth of a reading public, translations and adaptations multiplied. They included versions of books of the Bible, a long doggerel romance of Alexander (*c.* 1380) and a political allegory, the poem *Schacktavelslek* (*The Game of Chess, c.* 1465). In the rhymed chronicles, too—*e.g.*, *Karlskrönikan, Sturekrönikan* and *Lilla Rimkrönikan* (which treats events up to 1520)—political interest prevails. Of greater literary value are the Latin hymn *Rosa rorans bonitatem* by Bishop Nicolaus Hermanni (1326–91) and the two short lyrics *Frihetsvisan* and *Trohetsvisan* that have earned for Bishop Thomas Simonsson (*c.* 1380–1443) the title of the first Swedish national poet. Yet the most accessible genre of Swedish mediaeval letters—the *folkvisor* (ballads or folk songs)—is anonymous and undated. The oldest probably go back to 1300, but they were not transcribed until much later. Though outnumbered by the Danish, and often deriving from Danish or Norwegian sources, many of the Swedish *folkvisor* have a strong and lasting appeal.

EARLY NEW SWEDISH (c. 1526–1700)

As Gustavus Vasa's accession (1523) and the breach with Denmark mark the beginning of modern Swedish history, so the printed translation of the New Testament (1526) is considered to usher in *den äldre nysvenskan*. Yet in both this work and the complete translation of the Bible (*Gustav Vasas Bibel*, 1541), conservatism is visible in the word order and in the maintenance of inflections already shed in less solemn contexts. The translators excluded the Danicisms typical of the *kanslispråk* (clerical style), but were much influenced by Luther. Still, a standardized literary *rikspråk* does grow up during this period, and an increase of interest in grammar, vocabulary and orthography is visible during the 17th century. The centralization of government in Stockholm and the influence of the court affected both the written and spoken norms, while Swedish intervention in the Thirty Years' War led to the assimilation of many foreign words.

Literature of the Reformation.—The Reformation stimulated no immediate growth of letters; such books as appeared were still mainly religious and didactic. Closely involved in the Bible translations were Olaus Petri (*c.* 1493–1552), his brother Laurentius (d. 1573), the first Lutheran archbishop, and Laurentius Andreae, at one time Gustavus' chancellor. Of these, Olaus seems to have had the most robust personality. After studying at Wittenberg he returned to Sweden full of reforming zeal, which won him the king's favour; later, his independence and critical spirit nearly cost him his life. His vigorous approach is revealed in his published sermons, in his hymn translations, in catechism and in legal writings; it is shown, too, in his *Swensk krönika*, the first authentic historical work in Sweden, far surpassing the official chronicle of Peder Swart. Petri may have written the biblical *Tobie comedia* (1550), the first complete extant Swedish play.

Uppsala university (founded 1477), closed in 1517 because of political unrest, was not revived until late in the 16th century and became a focus of learning and letters only in the 17th century. Johannes Rudbeckius, one of the first Swedish educators, was associated with it, and so was Johannes Messenius (1579–1636), who wrote several mythical-historical plays, such as *Disa* (1611) and *Signill*. Although formless and crude these plays enjoyed great popularity, as did Magnus Olai Asteropherus' more successful *Tisbe* (1610), the first nondidactic drama in Swedish.

In the first half of the 17th century literature remained limited in scope and quantity. Some Latin poetry, the sermons of Rudbeckius and a handful of plays—including *Judas Redivivus* (1614) by J. Rondeletius, *Gustaf den förste* by A. Prytz and some comedies by J. Chronander—are worth recording. A unique contribution, however, was made by the vagrant adventurer Lars Wivallius (1605–69), whose lyrics, many of them written in prison, reveal a feeling for nature new to Swedish poetry.

Literature of Stormaktstiden, 1650–1730.—With its intervention in the Thirty Years' War and the ensuing military triumphs, Sweden established itself as a European power (*Stormakt*). This led to a great development of national pride and culture that is revealed in the literature of this epoch. The first and probably the best work is Georg Stiernhielm's allegorical epic *Hercules*, written about 1647 in vigorous hexameters. Stiernhielm (1598–1672) was the greatest literary figure of the 17th century. His followers include the two brothers Columbus, of whom Samuel (1642–71) was Stiernhielm's secretary and wrote *Odae sueticae* and the prose *Målroo eller Roomål*, a charming collection of anecdotes which illumine Stiernhielm's character. A rival to Stiernhielm was the unidentified "Skogekär Bergbo," whose *Wenerid* (1650; published 1680) was the first sonnet cycle in Swedish. Another unsolved problem of this period is the authorship of the epithalamium *Bröllopsbeswärs Ihugkommelse*, a realistic poem long attributed to Stiernhielm. Several of Stiernhielm's admirers, including Urban Hiärne (1641–1724), formed a coterie at Uppsala in the 1660s. Hiärne led a troupe of student actors who in 1665 performed his play *Rosimunda* before the young Charles XI. He also wrote lyrics and the pastoral *Stratonice* (1666–68), the first Swedish "psychological" narrative.

Stiernhielm was eager to revive old Swedish words and eliminate foreign loans. Columbus also demanded a more vigorous, flexible language for prose and poetry, as did "Skogekär Bergbo" in *Thet swenska språketz klagomål*. National pride and religious feeling are combined in the works of the bishops Haqvin Spegel (1645–1714) and Jesper Swedberg (1653–1735) father of Emanuel Swedenborg. Both were hymn writers and ardent philologists. Spegel

wrote a religious epic, *Guds werk och hwila* (1685) and sponsored the plan for a new hymnbook which appeared in 1695 and became the poetry book of the Swedish people, and, like *Gustav Vasas Bibel,* of lasting influence. Even Lars Johansson (1638–74; pseudonym "Lucidor") was represented therein. This poet, though one of the most attractive lyricists of the period, otherwise stands apart from the general development; some of his poems have survived in a posthumous collection, *Helicons blomster.*

At Uppsala meanwhile the scholar Petrus Lagerlöf (1648–99) worked on poetics and attempted to impose classical standards, and O. Verelius (1618–82) edited and translated Icelandic sagas. Olaus Rudbeckius (1630–1702), a man of vast energy and learning, became interested in Verelius' work on antiquities and enthusiastically developed the theory that Sweden was the lost Atlantis and had been the cradle of western civilization. These theories he launched in a massive work. *Atland eller Manhem* (1679–1702), deliberately written in Swedish to prove that the language was suited to scientific purposes. The work was translated into Latin as *Atlantica* and attained European fame.

Rudbeckius had encouraged Hiärne's actors. Lagerlöf now sponsored a second Uppsala company, which moved to Stockholm in 1686 and acted in Lejonkulan (the "Lion's Den"). This became Sweden's first regular (though short-lived) theatre. The "Lejonkula" plays show little dramatic talent but reveal how Lagerlöf's teaching had produced more correct verse.

Baroque tendencies are found in the work of some late 17th-century poets, notably Gunno Eurelius (ennobled as Dahlstierna; 1661–1709), who dedicated his elaborate *Kungaskald* (1697) to Charles XI. J. Runius (1697–1713) expresses a Christian stoicism found among the Swedish people during the disastrous early decades of the 18th century. J. Frese (1691–1729) is a gentler and more intimate poet; his lyrics and hymns contain some of the emotional pietism that became a feature of 18th-century thought.

MODERN NEW SWEDISH (c. 1700 AND AFTER)

Den yngre nysvenskan is the form of Swedish now current; it is intelligible (except for a few words and expressions) to a modern reader without commentary. Olof von Dalin's *Then swänska Argus* is the first important literary work in this form. During the 18th century several grammars and dictionaries contributed to standardization, and the Swedish academy (founded 1786) was given the task of "working for the purity, strength, and sublimity of the language." It issued a grammar in 1836, but its dictionary (*Svenska Akademiens ordbok*) did not appear until the 1890s. Meanwhile, C. G. af Leopold of the academy had published (1801) a treatise on spelling involving the radical Swedicizing of foreign words, especially the numerous French loans absorbed during the previous century. Few changes in orthography occurred during the 19th century; the most recent spelling reform was made in 1906. The importance of the written language increased after 1850; it was much affected by the press (dating, in a modern sense, from the foundation of L. Hierta's *Aftonbladet* in 1830), and the gap between the spoken language of educated men, modelled on the written norm, and the printed word was gradually reduced. An important aspect of this tendency was the speech usage whereby the singular form of the verb was employed with a plural subject; this became the norm of the press in the 1940s.

Literature, 1730–1809.—After the death of Charles XII (1718) and the collapse of his empire, a more utilitarian attitude to life and letters gradually set in. Olof von Dalin (*q.v.*) (1708–63) is the outstanding popularizer of the new ideas of French and English Enlightenment. Educated at Lund university (founded 1668), he later went to Stockholm and began to publish, anonymously, *Then swänska Argus,* modelled on Joseph Addison and Justus van Effen, almost the first serious journalistic venture in Sweden. In forceful, supple prose, Dalin ridiculed the foibles of the capital. *Sagan om hästen* (*The Story of the Horse,* 1740) showed his mastery of allegorical satire, occasional verses and songs his skill as a lyric poet. He became a court poet and favourite and a member of Queen Lovisa's Vitterhetsakademi (1753). In spite of the increasing secularization of literature during *Frihetstiden* (1718–72), the drama still lacked life; neither Count Karl Gyllenborg's comedy *Den svenske sprätthöken* (*The Swedish Fop,* 1737) nor O. Celsius' tragedy *Ingeborg* has more than historic interest.

With the second phase of the Enlightenment, marked by the deepening influence of France, are associated Hedvig Charlotta Nordenflycht (1718–63), Count Gustaf Creutz (*q.v.*) (1731–85) and Count Gustaf Gyllenborg (1731–1808). Fru Nordenflycht published in 1743 *Den sörjande turturdufvan* (*The Sorrowing Turtledove*), in which she lamented the death of her husband in short lyrics. Creutz was a more sophisticated personality. He wrote little, devoting his life after 1763 to diplomacy, but his few writings, of which *Atis och Camilla* is the most important, reveal a mastery of form and Epicurean views. Gyllenborg did not achieve his friend's formal elegance, but his best poems, *Världsföraktaren* and *Vinterkväde,* have nobility and feeling for nature.

Prose—particularly the novel—developed more slowly. The first genuine novel was *Adalrik och Giöthilda* (1742–44) by J. Mörk and A. Törngren, showing the influence of the Icelandic sagas and of Madeleine de Scudéry. But the raciest prose of this period is found in J. Wallenberg's *Min son på galejan* (*My Son on the Galleon,* 1781), a humorous, realistic diary of voyages on an East Indiaman. The most notable historian was S. Lagerbring, and J. Ihre enriched historical philology with his *Glossarium sviogothicum* (1769). The only two 18th-century Swedes with European reputations were scientists: Carl von Linné (Linnaeus; *q.v.*) (1707–78) and the versatile and mystic Emanuel Swedenborg (*q.v.*) (1688–1772). Linné's work at Uppsala much enhanced the prestige of the university, and his Swedish travel books were later praised for their understanding of nature and freshness of style.

The Gustavian period (1771–1809) takes it name from the brilliant Gustavus III (1746–92) and his intolerant son Gustavus IV. Gustavus III saw himself as the patron of art and letters; at his court he collected the best writers of the time, and by founding the Swedish academy he gave them official status. He was especially interested in drama and opera; thanks to his patronage a proper theatrical tradition was begun and the leading authors—often against their will—had to produce texts for performance. Gustavus himself sketched out some of these works, of which the opera *Gustaf Vasa,* thanks to Johan Kellgren's verses, has survived longest. The comedies of the Gustavian era were modelled on the French; the tragedies were pseudoclassical.

Johan Henrik Kellgren (*q.v.*) (1751–95) was the greatest academic poet of this epoch and the arbiter of taste. In his best poems, such as *Den nya skapelsen* (*The New Creation*), thought and emotion are fused. In *Stockholmsposten,* the main organ of literary opinion in the capital, Kellgren used his polemical wit against Thomas Thorild (1759–1808), the truculent champion of individual genius, who wrote a long poem, *Passionerna* (1781), but is chiefly remembered for his polemical prose. After Kellgren's death the controversy was continued by C. G. af Leopold (1756–1829), who still imposed pseudoclassical standards on the academy and applied them in his rhetorical odes and his plays. Johan Gabriel Oxenstierna (1750–1808) did his most original work in 1770–74 while a diplomat in Vienna; *Skördarne* (*The Harvests*) and *Ode till Camilla* reveal an elegiac, pre-Romantic feeling for the beauty of nature. Unfortunately he revised these poems before publishing them. Bengt Lidner (1759–93) was a follower of Thorild but showed much more poetic inspiration than Thorild in his play *Erik XIV,* his opera *Medea*—both lyrical rather than dramatic in tone—and in his most successful work, the ode *Grevinnan Spastaras död* (1783). The influence of both Milton and the Bible may be traced in his use of language.

Carl Mikael Bellman (*q.v.*) (1740–95) stands apart from the conflicting ideals of his time. A poet and musician, he fused his gifts in the lyrics of *Fredmans epistlar* (1790) and *Fredmans sånger* (1791). His combination of stylized realism with humour and the most delicate sense of language and rhythm is unique in Swedish. He is the greatest Swedish lyricist of the 18th century.

The dissertation *Om upplysning* (*On Enlightenment*) by Nils

von Rosenstein, the first secretary of the academy, nobly expressed the highest ideals of the Gustavian epoch; and various memoirs, by G. G. Adlerbeth, G. J. Ehrensvärd, Count Fredrik Axel Fersen, Duchess Hedwig Elizabeth and others, evoke the witty but artificial atmosphere of Gustavus III's court and illustrate the development of prose. Gustavus IV did not encourage literature; however, the energetic Anna Maria Lenngren (1754–1817), wife of *Stockholmsposten's* editor, wrote some of her best verse satires between 1795 and 1800, many of her anonymous shafts being aimed at aristocratic foibles. The sentimental idylls of F. M. Franzén (*q.v.*) (1772–1847) also belong to these years; they are full of a pre-Romantic idealism imbibed from German and English sources and later toned down under pressure from the academy.

LITERATURE, 1800–70

The Romantic Movement.—The loss of Finland in 1808, the deposition of the king and the new constitution of 1809 all greatly affected Swedish literature, for out of them an ardent national spirit emerged as one of the motifs of Swedish romanticism. The teaching of B. Höijer (1767–1812), admirer of Immanuel Kant and F. W. J. von Schelling, at Uppsala laid the basis for the idealism which soon became the core of Swedish romanticism.

Then the impact of Goethe and Schiller and of the German Romantics was felt, and Swedish romanticism became as much dominated by German culture as the previous era had been by French. Student societies became the focuses of the new movement, and the periodicals issued by them and their adherents—*e.g.*, *Polyfem* (1809–12) and *Fosforos* (1810–13)—attacked the academy. *Fosforos*, from which the Fosforister took their name, played an especially important initial role in the controversy between the old school and the new during 1810–20; in it Per Daniel Amadeus Atterbom (1790–1855), the most gifted poet of the Fosforister, first revealed his talents in his *Prolog*. V. F. Palmblad, publisher and prose writer, led the practical polemics at Uppsala; L. Hammarsköld, with J. C. Askelöf and C. J. Livijn, carried on the fray at Stockholm, while J. D. Valerius and P. A. Wallmark championed the old school, as for a time did J. O. Wallin, who later became the greatest hymn writer of the century (hymnbook of 1819).

Meanwhile another society, Götiska Förbundet, had been founded in Stockholm in 1810, with the object of raising the moral tone of society by study of the heroic, "Gothic" past. Erik Gustaf Geijer (*q.v.*) (1783–1847), one of the founders, suggested the publication of a periodical, and the first issue of *Iduna* consisted largely of his own poems, *Vikingen*, *Odalbonden*, etc. Most of Geijer's poetry, including the first Swedish translation of Shakespeare (*Macbeth*), was written before 1816. Esaias Tegnér (*q.v.*) (1782–1846), professor of Greek at Lund, also contributed to *Iduna*, though he refused to engage in literary controversy, and P. Ling (1776–1839), too, was associated with it.

A brief golden age of Swedish letters occurred in the 1820s. Atterbom published his most remarkable work, the verse drama *Lycksalighetens Ö* (*The Isle of Bliss*) in 1824–27; here, in often obscure but beautiful language, he expressed his philosophy in the allegory of Astolf and Felicia. His other verse drama, *Fågel blå* (*The Blue Bird*), was never finished, and his later years were devoted to a lyric cycle called *Blommorna*, to university work and to *Svenska Siare och skalder* (1841–55), which in a sense laid the foundations of literary history in Sweden. In 1820–25 Tegnér published *Frithiofs saga*, an epic on an Icelandic theme; this became the most popular poem of the first half of the century, and Tegnér, poet, orator and letter writer, who combined Gustavian wit and sense of form with idealism, was the greatest personality of the age. Erik Stagnelius (1793–1823), a complex and brilliant writer, published his *Liljor i Saron* in 1821 and his verse drama *Bacchanterna* in 1822, but most of his work appeared posthumously. Though influenced by the Romantics, Stagnelius held aloof from them and died in deliberately sought isolation.

A number of minor Romantic writers should be mentioned—*e.g.*, A. Törneros (1794–1839); P. Elgström (1781–1810); S. Hedborn (1783–1849), author of several attractive hymns and ballads and of an autobiography, *Minne och poesi;* A. Afzelius (1785–1871), collector of folk songs (*Svenska folkvisor från forntiden*, 1814–16); and E. Sjöberg (Vitalis).

The Romantics cultivated verse rather than prose, though Palmblad certainly contributed something to prose with his long short story *Amala* (1817) and his novels, including *Familien Falkenswärd*. Malla Montgomery Silfverstolpe's *Memoarer* (published 1908–11) gives a naïve picture of the Romantic circle at Uppsala. Geijer's cumbersome historical works—*Svea Rikes Häfder* (1825), etc.—do not reveal his stimulating personality, but *Minnen* (*Memories*) of 1834 are full of the fresh charm of his poetry.

The Coming of Realism.—The most complex personality among the later Romantics was Karl Jonas L. Almqvist (*q.v.*) (1793–1866), who combines an extravagant imagination, as in *Drottningens juvelsmycke* (*The Queen's Jewel*), with the realism which is increasingly prominent in his works after 1836. Almqvist, a master of prose style, was at his best in the long short story (the *novell*); of his works in this genre *Det går an* (1839), an outspoken attack on marriage, foreshadows Strindberg's method of raising problems for debate. Almqvist was enormously productive and versatile; most of his works were published in the volumes of *Törnrosens bok* (1832–51). Meanwhile the realistic novel had made some headway. F. Cederborgh (1784–1835) wrote *Uno von Trasenberg* (1809–10) and *Ottar Tralling*, light-hearted satires on life and manners; but the novel as an accepted genre was really established by Fredrika Bremer (*q.v.*) (1801–65), whose "sketches from ordinary life" appeared from 1828. Her novels, *Grannarne*, etc., have dated considerably, but were formerly very popular in both Europe and America. Sofie von Knorring (1797–1848) wrote novels principally about aristocratic families, and the very popular Emilie Flygare-Carlén (1807–92) produced a long series of stories, including *Rosen på Tistelön* (1847), usually dealing with life on the west coast.

Realism, then, made only slow headway, in spite of the social consciousness of the liberal paper *Aftonbladet* and the example of the great Finnish-Swedish poet J. L. Runeberg (*q.v.*). The literature of the 1840s and 1850s is mainly an aftermath of romanticism, though the critic B. E. Malmström imagined that he opposed this movement. His elegy *Angelica* (1840) had kinship with the graceful but insipid poetry of his contemporary C. V. Böttiger and the older K. A. Nicander. There was a fresher feeling about G. Wennerberg's *Gluntarne* (1847–50), songs about Uppsala student life, but there is nothing new in them. The movement known as Scandinavianism produced a good deal of verse: C. V. A. Strandberg (*Talis Qualis*, 1818–77) was the fieriest poet of this type. Later he did excellent translations of Byron. O. P. Sturzen-Becker (*Orvar Odd*, 1811–69) was a versatile writer with talents in verse and the short story (*Samlade Arbeten*, 1861–62). Popular reading was provided by August Blanche (1811–68), particularly in the successful *Bilder ur verkligheten* (1863–65), which depicts Stockholm life with humour and vivacity. He also wrote popular dramas. Frans Hedberg (1828–1908) held the field at the Royal theatre with pompous historical plays such as *Bröllopet på Ulfåsa*. The quasi-historical novels of M. J. Crusenstolpe (1795–1865), a brilliant and unscrupulous polemical writer, were much appreciated in their day. The most influential historian was A. Fryxell (1795–1881), with his *Bilder ur svenska historien;* and the dominating philosopher was the idealistic C. Boström (1797–1866).

Poetic realism became the official program of Signaturskalderna (the "Pseudonym Poets"), a small group of the 1860s influenced by the Norwegian critic L. Dietrichson. They included C. D. af Wirsén, E. Bäckström and P. Wikner, but only Count C. Snoilsky (1841–1903) had sufficient temperament and poetic gift to carry out the program. Some of Snoilsky's best work is found in *Dikter* (1869) and *Sonetter* (1871). Meanwhile Wirsén had established himself as the conservative literary critic, and from his post as secretary of the academy he launched formidable opposition against all innovators, including Snoilsky. A. V. Rydberg (*q.v.*) (1828–95) exemplifies this rather uneasy transition from idealism to naturalism even more patently than does Snoilsky. Novelist,

poet and critic, he began as a radical journalist and ended as a professor and as author of the philosophical *Dikter* (1882–89). His most important early work was the ideological novel *Den siste Athenaren* (1859), and his treatise *Bibelns lära om Kristus* (1862) had considerable influence through its attack on the narrow orthodoxy of the Swedish Church and prepared the way for scientific rationalism.

LITERATURE FROM 1870

Strindberg.—Four influences combined to free Swedish literature from its petrifying conventions: English thought in the writings of Charles Darwin, Herbert Spencer and J. S. Mill; French naturalism as taught and practised by Émile Zola; the drama of the Norwegians Henrik Ibsen and B. Bjørnson; and the criticism of Georg Brandes. Modern Swedish literature begins with the work of Johan August Strindberg (*q.v.*) (1849–1912), for the modern drama dates from his *Mäster Olof* (1872) and the modern novel from his *Röda rummet* (*The Red Room*, 1879). By his originality, his mastery of prose and dialogue, his great energy and the influence of his work and personality on Swedish, not to say European, literature he is by far the greatest writer Sweden has yet produced.

The 1880s.—Strindberg overshadowed all other writers of the 1880s. G. af Geijerstam (1858–1909), author of *Erik Grane* (1885) and of several other novels and short stories, longed to lead his age, but could not do so. A. M. Edgren-Leffler (1849–92), later duchess of Cajanello, wrote a number of stories called *Ur livet* (*From Life*) in which she satirized conventions adversely affecting women. The more gifted Viktoria Benedictsson ("Ernst Ahlgren," 1850–88) died—by her own hand—before she had fulfilled her promise; she too, but more boldly, was concerned with the position of women in society. Her stories, *Från Skåne* (1884), have lasted better than her novels, *Pengar* and *Fru Marianne*, and reveal how much regional characterization had entered into the new style of literature. With her was associated Axel Lundegård (1861–1931), who soon turned his talents to the historical novel. Alfhild Agrell (1849–1923) wrote her best stories about Norrland. Regional poetry was written by A. U. Bååth (1853–1912) of Skåne; it was surpassed in beauty and originality by the work of another Skåning, Ola Hansson (1860–1925), novelist, critic and poet. Hansson's early poems (*Notturno*, 1885) and his sketches, *Sensitiva Amorosa* (1887), reveal the boldness in imagery and the interest in psychological nuances which are more typical of the next decade.

1890–1910.—Though, except for the work of Strindberg, little of lasting value was achieved in the 1880s, it must be remembered that the generation just discussed accustomed the public to the debating of problems in literature. The reaction against utilitarianism was begun in 1888 by Verner von Heidenstam (1859–1940) with his volume of verse *Vallfart och vandringsår* (*Years of Pilgrimage and Wandering*). Heidenstam's later poetry (*Dikter*, 1895, and *Nya Dikter*, 1915), with his historical tales *Karolinerna* (*King Charles's Men*) and *Folkungaträdet* (*The Tree of the Folkungs*), won him the Nobel prize in 1916. Although much of his work now seems precious and dated his short contemplative lyrics are very fine. Oscar Levertin (1862–1906), stimulated by his friend Heidenstam's example, wrote poetry full of rich colours and the lore of the past (*e.g., Legender och visor*, [1891]); these and other poems show his affinity with the pre-Raphaelites, his knowledge of French symbolism and his awareness of his own Jewish heritage (*Kung Salomo och Morolf*). His criticism published in *Svenska Dagbladet* and elsewhere did much to mould contemporary taste. Gustav Fröding (1860–1911), too, was encouraged by Heidenstam; the title of his first volume of verse, *Guitarr och dragharmonika* (*Guitar and Concertina*) shows the two different strains in the poet, melancholy and gay, which constantly intermingle and out of which he wrested such original music. Despite the threat of madness he managed to publish four more collections of poems before his reason finally gave way, in 1898, by which time he had established himself, especially with *Stänk och flikar* (*Rags and Splashes*), as the greatest Swedish lyric poet since Bellman.

With Fröding, who was from Värmland, regionalism entered into neo-Romantic poetry, and with the work of Erik Axel Karlfeldt (1864–1931) Dalarna came into its own. Karlfeldt's mature poems, such as *Fridolins visor* (1898), *Fridolins lustgård* (*Fridolin's Garden*, 1901) and *Flora och pomona* (1906), reveal his individual strength and the richness of orchestration which he could give to Swedish. He especially loved describing the beauties of autumn, and his erotic poetry is full of the symbolism of country life. He was posthumously awarded the Nobel prize.

Meanwhile Selma Lagerlöf (1858–1940) had developed the prose tale. Her long series of novels and short stories, beginning with *Gösta Berlings saga* (1891), reached an international public; she was, indeed, the first Swede to win the Nobel prize for literature (1909). She often used old Värmland tales for material, though one of her best novels, *Jerusalem I* (1901), is set in Dalarna. Like Karlfeldt but unlike Fröding she is a traditionalist. Per Hallström (1866–1960) is a more skilful writer of short stories than of novels; like Lagerlöf, he did his significant work before 1910 (*Briljantsmycket, Thanatos, De fyra elementerna*). His pessimism led him to depict the futility of human passions, often against a mediaeval or Renaissance background. Romantic, too in his love for the skerries, was Albert Engström (1867–1940), story writer, cartoonist, founder of the popular humorous paper *Strix* and one of the most beloved of Swedish writers. Tor Hedberg, son of Frans Hedberg (*see* above), began with the style of the 1880s; his most lasting work was in the tragedy *Johan Ulfstjerna* (1907) and in the lyric.

Both Hjalmar Söderberg (1869–1941) and Bo Bergman (1869–) had links with the rationalism of the 1880s and with the aestheticism of the 1890s, yet both stand apart from the neo-Romantics. Both, like Strindberg, loved Stockholm, which perpetually recurs as the setting of their works. Söderberg's forte was the short story (*Historietter*), in which his psychological subtlety and irony are happily combined; but his novels, including *Martin Bircks ungdom* (1901) and *Doktor Glas* (1905), have an appreciation of mood and an awareness of the faults of contemporary society which make them unique in Swedish fiction. Söderberg, especially, is one of the masters of Swedish prose. Bergman also achieved memorable short stories (*Drömmen*, 1904; *Skeppet*, 1915; etc.), but his real medium was the lyric; he developed and refined his talent in a series of collections from *Marionetterna* (1903) to *Riket* (1944). Like Söderberg, he is a passionate enemy of hypocrisy and brutality, but his characteristic tone is subdued realism.

1910 Onward.—From the beginning of the 20th century industrialization, the growth of the press and the spread of education increasingly influenced letters. The reading public continued to grow, and the volume of writing similarly increased, a marked feature being the development of the novel, with which are associated Gustaf Hellström (1882–1953), Ludvig Nordström (1882–1942), Elin Wägner (1882–1949) and Sigfrid Siwertz (1882–). Hellström's work as journalist took him to the continent and to the United States, and long residence in England greatly influenced him. His irony and careful realistic detail emerge in his best-known novel, *Snörmakar Lekholm här en idé* (*Lacemaker Lekholm Gets an Idea*), set in a small provincial town. Siwertz is a more elegant stylist—like so many Swedes he began as a lyricist—and a decisive influence upon him was the philosophy of Henri Bergson, reflected in *En Flanör* (1914). His weightiest works are the family saga *Selambs* (*Downstream*, 1920), a novel of Stockholm during World War I, and *Jonas och Draken* (1928); he is also a skilful writer of short stories. Nordström came from Norrland, and his birthplace—Härnösand—is perpetuated in his work as "Obacka." Overflowing with vitality and gifted with keen but grotesque humour, he did some of his best work in *Landsorts-bohème* (1911) and *Planeten Markattan* (1937), but his most satisfying productions are the short stories in *Fiskare* (1907), *Obackabor* (1921), etc. Elin Wägner was an ardent pacifist and feminist; her horror of militarism is shown in her novel *Släkten Jerneploogs framgång* and her feminism in many of her witty short stories. Her most powerful work is the peasant novel *Asa-Hanna* (1918).

The outstanding novelist of the second decade of the 20th century is Hjalmar Bergman (1883–1931). Gifted with a fantastically vivid imagination and restlessly energetic, Bergman wrote a long series of stories many of which are set in "Wadköping" (his native Örebro) and its surroundings, others in Italy. In *Loewenhistorier* (1913), as in other works, Bergman depicts an irrational, impulsive, unsuccessful hero, in *Farmor och Vår Herre* (*Thy Rod and Thy Staff*) one of the dominating female personalities who fascinate him. The satire *Markurells i Wadköping* (*God's Orchid*, 1919) and *Swedenhielms*, one of the few living Swedish comedies, are his most widely known works. Bergman's witty dialogue and rapid characterization, and his inventiveness, both tragic and burlesque, make him one of the most original of Swedish writers, and as dramatist the only native successor to Strindberg.

Meanwhile the "proletarian" novel had been developed—not always by working-class writers—by G. Hedenvind-Eriksson (1880–), M. Koch (1882–1940) and I. Lo-Johansson (1901–), who is particularly concerned with the miseries of the *statarens* (farm labourer's) life. There is particularly harsh and telling criticism of working-class life in *Lars Hård* and other stories by Jan Fridegård (1897–). V. Moberg (1898–) wrote both realistic and historical novels of peasant life. The autobiographical element, strong in many of these writers, and their contemporaries, is evident, too, in the works of Agnes von Krusenstjerna (1894–1940). In her novel cycles (*e.g., Tonyböckerna* and *Fröknarna von Pahlen*) she describes her own aristocratic and upper-middle-class environment, and analyzes a degenerate and perverted psychology. Her works constituted a new departure in Swedish literature.

The greatest Swedish writer of this period is undeniably Pär Fabien Lagerkvist (1891–), who won the Nobel prize in 1951. He was a vital force in literature after World War I and influenced both form and

content. His anguished search for a personal philosophy recalls Strindberg; it is the basis of plays such as *Himlens hemlighet* (1919) and *Han som fick leva om sitt liv* (1928). His best lyrics, such as *Hjärtats sånger*, combine intensity of feeling with subtle rhythm, and his stories have visionary power (*Bödeln, The Hangman*, 1933) and an elemental simplicity, as in *Barabbas* (1950).

Both Eyvind Johnson (1900–) and Harry Martinson (1904–) were self-educated proletarians who turned more and more to experiment with form. Johnson's best work was probably his autobiographical *Romanen om Olof* (1934–37). Martinson wrote both poetry and prose but achieved his most striking effects in prose, in *Kap Farväl* (*Cape Farewell*) and in *Vägen till Klockrike* (1949). Writers of great popularity, too, were Olle Hedberg (1895–), a witty novelist of middle-class life, F. Nilsson-Piraten (1895–), a Skåne humorist, and his fellow-Lundensian Frans G. Bengtsson, author of the historical-humorous novel *Röde orm* (1941–45) and a masterly essayist. Of the youngest authors, Stig Dagerman (1923–) showed particular promise in his play *Skuggan av Mart* (1948).

Many other names occur in connection with the development of lyric poetry. V. Ekelund (1880–1949) developed the dithyrambic form to a high pitch of perfection, then turned to aphoristic prose; A. Österling (1884–) and others became poetic realists; Dan Andersson (1888–1920) developed the ballad; B. Sjöberg (1885–1929) and Karin Boye (1900–41) made the lyric a vehicle for their personal conflicts, Boye's contribution being outstanding. Bertil Malmberg, Hjalmar Gullberg and Johannes Edfelt were among the most successful of modern poets. G. Ekelöf (1907–) wrote effective surrealist verse; and of the "generation of the '40s" (*fyrtiotalisterna*) E. Lindegren and K. Vennberg may be mentioned.

Henrik Schück (1855–1947), with his emphasis on the historical approach to literature, was the father of modern Swedish criticism. Levertin, A. Blanck (1881–1950) and M. Lamm (1880–1950) were his most distinguished pupils.

BIBLIOGRAPHY.—Standard history: H. Schück and K. Warburg, *Illustrerad svensk litteraturhistoria*, vol. i–vi, 3rd ed., rev., H. Schück (ed.) (1926–30); G. Castrén, *ibid.*, vol. vii, literature, 1870–1914 (1932); E. H. Linder, *ibid.*, vol. viii, literature, 1910–40 (1948).

Shorter histories: E. N. Tigerstedt, *Svensk litteraturhistoria* (1948); H. Schück, *Sveriges litteratur intill 1900*, rev. by Ö. Lindberger (1952); E. Bredsdorff, B. Mortensen, R. Popperwell, *An Introduction to Scandinavian Literature* (1951).

Anthology: H. Schück, R. G. Berg and F. Böök (eds.), *Sveriges national-litteratur, 1500–1900*, 30 vol., 2nd ed. (1921–22); G. Bergman, *A Short History of the Swedish Language* (1947). (B. M. E. M.)

SWEELINCK or SWELINCK, **JAN PIETER** (also known as JAN PIETERSZOON) (1562–1621), Dutch organist and composer, was probably born at Amsterdam, although his family came from Deventer. It has been persistently stated that he was a fellow pupil with Giovanni Gabrieli and Leo Hassler of the great Andrea Gabrieli in Venice, but this was disproved by F. H. J. Tiedeman in the *Vereeniging voor Nederlandsche Muziekgegeschiedenis* (Amsterdam, 1876). Sweelinck was the son of the organist of the Old church, Amsterdam, and lived in Holland from about 1577 onward. At a date not later than 1581 he was appointed organist to the Old church, his father having died in 1573, and this post he held until his death on Oct. 16, 1621.

A complete edition of his works, edited by Max Seiffert with prefaces, was brought out by Breitkopf and Härtel in 1894–1901.

SWEET, HENRY (1845–1912), English linguist, who was among the first to emphasize the importance of physical, articulatory considerations in historical sound change, was born Sept. 15, 1845, at London, and was educated at King's college (London), Balliol (Oxford) and Heidelberg. He combined competence in historical linguistics with strong interest in the emerging science of phonetics, and a Readership in Phonetics at Oxford (1901) was created especially for him. Sweet died at Oxford on April 30, 1912.

His publications on history and sounds of English, on general phonetics and editions of Old and Middle English texts include: *Handbook of Phonetics* (1877); *New English Grammar* (1892–98); *History of Language* (1900); *Anglo-Saxon Reader* (1908); and *Sounds of English* (1910). His *Collected Papers* were edited by H. C. Wyld (1913). (E. PM.)

SWEET ALYSSUM (*Lobularia maritima*), the ever-popular little white-flowered (sometimes violet), fragrant, half-hardy perennial of the mustard family that is usually grown as an annual. It is wide-spreading on the ground, with branches rising 6 to 12 in., and numerous small flowers distributed in lengthening racemes. It occurs in both large and small forms, some with white-edged leaves and some with double flowers. It is an excellent edging plant that blooms over a long season. (J. M. BL.)

SWEETBREAD, a popular term for certain glands of animals, particularly when used as articles of food; these are usually the pancreas, called stomach sweetbread by butchers, and the thymus, or neck sweetbread. The term is also sometimes used to include the salivary and lymphatic glands. *See* LYMPH AND LYMPHATIC SYSTEM; PANCREAS.

SWEETBRIER (*Rosa eglanteria*), a tall, stiff-stemmed rose, called also eglantine, native to Europe, widespread in Great Britain and extensively naturalized in eastern North America. The erect, somewhat branching stem, armed with stout hooked prickles, grows usually from five to eight feet high. It bears fragrant leaves, of five to seven leaflets, and solitary white or pinkish flowers, one to two inches across. While the sweetbrier is infrequently grown for ornament, various hybrids with other roses are in cultivation. *See* ROSE.

SWEET CLOVER, or MELILOT, the common name for plants of the leguminous genus *Melilotus* comprising about 20 species, native of southeastern Europe and Asia Minor. They are annual or biennial, erect herbs with three-parted leaves of usually small narrow leaflets and small yellow or white flowers borne in slender axillary clusters (racemes). Most of the species contain coumarin, resulting in a sweet odour and bitter taste.

Biennial white sweet clover (*M. alba*), biennial yellow sweet clover (*M. officinalis*) and the bitter clover (*M. indica*), a winter annual adapted only to the southern and southwestern U.S., are of agricultural importance for pasture, soil improvement and hay in the U.S. and Canada. The first two species have become widely naturalized throughout the U.S. and Canada, while bitter clover is widely scattered in the western U.S. *M. suaveolens*, a yellow-flowered species, is grown in a limited way in the southern prairies of Canada, and occasional plants of *M. altissima* have been collected in the eastern states. Several varieties have been developed such as Madrid for the corn belt, Erector for the Great Plains and Willamette for the northwest.

See U.S. Department of Agriculture, *1948 Yearbook of Agriculture*, p. 360 (1948). (E. A. H.; N. TR.)

SWEET CORN: *see* CORN; VEGETABLE.

SWEET FERN, a common name applied to the genus *Comptonia* of the wax myrtle family, also known as the sweet gale family. The genus contains a single species, *C. peregrina*, and the variant asplenifolia, sweet-smelling deciduous shrubs up to five feet, with finely cut alternate leaves and inconspicuous flowers in catkins, fairly common in sandy or rocky soils in eastern North America. (J. M. BL.)

SWEET FLAG: *see* ACORUS CALAMUS.

SWEET GALE (*Myrica gale*), a shrub of the family Myricaceae, called also gale, sweet willow, bog myrtle and Dutch myrtle, widely distributed in the north temperate zone. It grows usually from two to four feet high, with numerous twiggy branches; narrow, short-stalked, fragrant leaves; inconspicuous, unisexual flowers, borne in short spikes, and small, waxy-resinous, drupelike fruits attached to the persistent flowering bracts. The sweet gale occurs throughout the northern U.S. and Canada and on bogs and moors in the British Isles.

ROCHE

SWEET GUM FRUIT CLUSTERS (LIQUIDAMBAR STYRACIFLUA)

See WAX MYRTLE.

SWEET GUM (*Liquidambar styraciflua*), a handsome North American tree of the witch hazel family (Hamamelidaceae), called also red gum, liquidambar (*q.v.*) and bilsted, found from Connecticut to Missouri and southward to Florida and Mexico. It has a straight trunk, sometimes 140 ft. high and 5 ft.

in diameter, with rough bark, corky-winged branches, large, deeply lobed leaves, and drooping, spiny, globose, capsular fruiting heads containing winged seeds.

The hard reddish-brown wood is known to the lumber trade as red gum or satin walnut. The tree yields liquidambar, called also copal balsam, a fragrant gum resin very similar to the storax produced by *L. orientalis* of Asia Minor.

SWEET PEA, *Lathyrus odoratus,* a hardy annual climbing herb of the pea family (Leguminosae) commonly grown outdoors in the cool temperate regions of the world, especially North America and Europe. The sweet pea is a native of Italy, and hundreds of varieties have been developed under domestication. It is one of the most widely used commercial cut flowers because of its excellent durability and the ease with which the plants can be grown in greenhouses.

The fragrant flowers have an unusual colour range: from pure white through cream and various intermediate shades to deep orange; from pink to deep scarlet; and from blue shades to dark purple. In the home garden, sweet pea is often used as a climber in dense rows up to six feet tall or, especially in the dwarfed nonclimbing forms (var. *nanellus*), as a colourful compact border. The plants are frost resistant. In mild climates, seeds may be planted in the late fall about ½ in. deep and 2 in. apart in rich garden soil for growth during the cool winter and for very early spring bloom. In less mild climates, seeds should be similarly planted in early spring after severe frosts are over and the soil can be properly worked.

These sun-loving plants are usually mulched to conserve soil moisture during their growth. (R. C. R.)

SWEET POTATO (*Ipomoea batatas*), a food plant of the family Convolvulaceae (*q.v.*), a native of tropical America, widely cultivated in tropical and the warmer temperate climates. It is a tender perennial grown as an annual in regions where frost occurs. The stems are long, trailing, sometimes short, with entire to palmately lobed leaves of quite variable shape. The flowers, borne in cymes in the axils of the leaves, are funnel shaped, tinged with rose-violet or pink, and are similar in size and shape to those of the common bindweed. The edible part is the much enlarged tuberous root, varying in shape from fusiform to oblong or pointed oval. Root colours range from white through cream to orange, and occasionally purple inside, and from light buff to brown or rose, copper and purplish red outside. The solids of the root consist largely of starch and the orange-fleshed varieties are high in carotene.

The sweet potato rarely produces flowers under common cultural conditions in latitudes higher than about 35°, and produces no seed above about 30° except under special cultural treatment. The seeds produce plants that are unlike the parent plant, therefore the sweet potato is propagated vegetatively by sprouts arising from the roots, or by cuttings of the vines. Seeds, however, are produced commonly in the field in the tropics. They have a very hard seed coat that may retard germination as much as a year, unless the seed coat is nicked or abraded, after which the seeds germinate promptly.

The sweet potato is now widely grown in the United States, in tropical America, the warmer islands of the Pacific, in Japan, and after about 1930 its culture increased rapidly in the U.S.S.R. It is best adapted to light, friable soils such as sandy loams and sandy soils, producing excessive vine growth and small yields or irregularly shaped roots on very rich, heavy soils. At least four to five months of warm weather are required for large yields. In Japan the crop long has been grown for drying and for manufacture of starch and alcohol. In the United States it is grown only for food.

In the United States the sweet potato is propagated usually by bedding the roots close together in sand or light soil in a hotbed heated by manure, by flues or electricity, about four to five weeks before the frost-free date. The sprouts which emerge are pulled from the roots and transplanted to the field about 1 to 1½ ft. apart on ridges 8 to 12 in. high, 3 ft. apart, after the weather and soil have become warm. The roots are harvested mostly about the time of the first frost in autumn since cold soils cause them to deteriorate. The crop responds well to heavy applications of chemical fertilizers low to medium in nitrogen and high in phosphorus and potash. Heavy applications of animal manures or green manures are conducive to the formation of elongated, irregular roots and to root surface injury by scurf, a fungus disease. If the roots are to be stored, they should be placed in the storage house immediately upon harvesting, at a temperature of 85° F. and a relative humidity of 85% to 90% which is maintained for a week to ten days. Then the temperature is lowered to 50°–55° where it is kept as long as the roots are in storage. The "curing" period at high temperature and humidity is essential to hasten the healing of wounds caused in harvesting, thus reducing danger of loss from decay.

The eating quality—and also the market quality and storage life—are greatly improved by curing. If diseases are controlled, roots that are properly cured and stored remain in good edible condition six to seven months.

The sweet potato is very susceptible to several diseases, control of which requires great care in field and storehouse sanitation and management, seed-stock treatment with disinfectants, and crop rotation.

In the United States the principal producing states are ordinarily Louisiana, North Carolina, New Jersey, Virginia, South Carolina, Mississippi, Texas and California. (V. R. B.)

SWEET SHRUB, a common name for *Calycanthus,* a genus which comprises three or four species of deciduous North American shrubs, noted for their sweet-smelling chocolate-coloured flowers, so fragrant in fact that they are variously called sweet shrub, sweet-scented shrub, strawberry shrub and sometimes merely shrub. One species, *C. floridus,* native from Virginia to Florida, is known as Carolina allspice; another, *C. fertilis,* is found from Pennsylvania to Georgia and Alabama; while a third, *C. occidentalis,* occurs in California. All of these are cultivated for their fragrant blossoms and their value as ornamental shrubs. (J. M. Bl.)

SWEETSOP (Sugar Apple; *Annona squamosa*), a small tree or shrub of the family Annonaceae. It has thin oblong-ovate leaves, solitary greenish flowers and a yellowish-green fruit, resembling a shortened pine cone, with a tubercle corresponding to each of the carpels from the aggregation of which it has been formed. The fruit is three to four inches in diameter and contains a sweet creamy-yellow custardlike pulp. Native of the West Indies and tropical America, it has been widely introduced into the eastern hemisphere and is much prized as a fruit. *A. muricata* is the soursop (*q.v.*). *A. reticulata* is the custard apple (*q.v.*), and *A. glabra* the alligator apple.

SWEET WILLIAM, *Dianthus barbatus,* an old-fashioned garden favourite of the pink family (Caryophyllaceae) and one of the most widely cultivated of all the 80 species of *Dianthus* grown in gardens. A relative of the carnation (*q.v.*), it was first introduced into cultivation about 1575 and has since been grown throughout the temperate world. The original European plant and some modern forms are still grown as perennials, but nearly all the finest modern varieties are best treated as biennials. Some strains bloom from seed in a single season and are grown as annuals.

ROCHE

SWEET WILLIAM (DIANTHUS BARBATUS)

Sweet william, 12–20 in. high, has flat opposite leaves lanceolate to ovate-lanceolate in shape, and a stem slightly swollen at the joints. The flowers, borne in a profuse flattish cluster (corymb), have toothed petals that are red, white-dotted or white-striped toward the slightly bearded base. The plant is variable as to colouring, especially in some modern varieties.

Among the many horticultural varieties are excellent scarlet and pink strains that come true from

seed and double-flowered strains which usually do not seed true. Another sort, known as auricula-eyed, has white- or pale-centred flowers with a rim of contrasting colour on each petal.

Seeds of the biennial sort are best sown in June, then potted for carrying over the winter in a cold frame. Many biennials and all the perennials often persist from self-sown seed. (N. Tr.)

SWELLENDAM, a municipality of Cape Province, South Africa, third oldest settlement in the Cape. It is 192 mi. by rail from Cape Town. Pop. (1960) 4,896 all races, 2,511 white. It was earmarked for settlement by Van Imhoff in 1745 and laid out along the Koringlande, a tributary of the Breede, in a fertile valley below the Langeberg mountains. The Drostdy (now a museum) and, across the river, the church—the third to be built in South Africa outside Cape Town—were begun in 1746. Swellendam was the centre of a magisterial district which stretched to the Gamtoos river and until 1786, when Graaff-Reinet was created, it was the frontier outpost of the Dutch East India company's rule. Early English travelers tended to discount it as a small village, but in the 18th century it became a flourishing commercial centre with wainwrights, blacksmiths and traders to serve the farmers when they trekked in from the frontier. The closing of the frontier outpost affected Swellendam, and economic as well as political reasons lay behind the revolt of 1795 (*see* Cape of Good Hope). After the first British occupation (1795–1803) the development of Graaff-Reinet (1786) and the creation of smaller districts both east and north buffered Swellendam from frontier impacts, and its links with Cape Town improved. It was hard hit by the emancipation of the slaves (1834), but continued to be a thriving community, mainly based on farming with a wide variety of crops, viz., wool, oranges, wheat, barley and tobacco. (W. A. Ml.)

SWETCHINE, MADAME (1782–1857), Russian mystic, whose maiden name was Soymanof, was born in Moscow, and under the influence of Joseph de Maistre became a member of the Roman Catholic Church in 1815. In the following year she settled in Paris, where, until her death, she maintained a famous salon remarkable no less for its high courtesy and intellectual brilliance than for its religious atmosphere. Though not physically beautiful, she had a personality of rare spiritual charm.

Her *Life and Works* (of which the best known are "Old Age" and "Resignation") were published by M. de Falloux (2 vol., 1860) and her *Letters* by the same editor (2 vol., 1861).

SWETE, HENRY BARCLAY (1835–1917), English Biblical and patristic scholar, was born at Bristol on March 14, 1835, and educated at King's college, London, and Caius college, Cambridge. He was ordained in 1858, and became dean of Caius college in 1865. In 1877 he accepted the college living of Ashdon, Essex, and in 1882 was made professor of pastoral theology at King's college, London. In 1890 he succeeded B. F. Westcott as regius professor of divinity at Cambridge, retiring with the title of emeritus professor in 1915. He died at Hitchin on May 10, 1917.

Swete's works on Biblical texts are of great importance. In 1887 he published the first volume of his edition of the Greek text of the Old Testament, completing the series in 1894; in 1898 the Greek text of the Gospel of St. Mark with notes and introduction; and in 1906 that of the Apocalypse of St. John. He was the editor of *Cambridge Theological Essays* (1905), *Cambridge Biblical Essays* (1909) and *Essays on the Early Church and Ministry* (posthumous, 1918). His historical and critical works include *The Apostles' Creed in Relation to Primitive Christianity* (1894); *Church Services and Service Books Before the Reformation* (1896); *Patristic Study* (1902); *The Appearance of our Lord After the Passion* (1907); and *The Last Discourse and Prayer of Our Lord* (1913).

SWEYN I, King of Denmark (d. 1014), son of Harold Bluetooth, the christianizer of Denmark, by his peasant mistress Aesa, according to the Jomsvikinga Saga, though more probably his mother was Queen Gunild, Harold's consort. The lad was a born champion and buccaneer. His first military expedition, in alliance with the celebrated Jomsborg Viking, Palnatoke, was against his own father, who perished during the struggle (*c.* 986). Six years later he conducted a large fleet of warships to England, which did infinite damage, but failed to capture London. During his absence, Denmark was temporarily occupied by the Swedish king, Eric Sersel, on whose death (*c.* 994) Sweyn recovered his patrimony. About the same time he repudiated his first wife Gunild, daughter of duke Mieszko of Poland, and married King Eric's widow, Sigrid. This lady was a fanatical pagan of a disquieting strength of character. Two viceroys, earlier wooers, were burned to death by her orders for their impertinence, and she refused the hand of Olaf Trygvessön, king of Norway, rather than submit to baptism, whereupon the indignant monarch struck her on the mouth with his gauntlet and told her she was a worse pagan than any dog. Shortly afterwards she married Sweyn, and easily persuaded her warlike husband to unite with Olaf, king of Sweden, against Olaf Trygvessön, who fell in the famous seafight off Svolde (1000) on the west coast of Rügen, after a heroic resistance immortalized by the sagas, whereupon the confederates divided his kingdom between them. After his first English expedition Sweyn was content to blackmail England instead of ravaging it, till the ruthless massacre of the Danes on St. Brice's day, Nov. 13, 1002, by Ethelred the Unready (Sweyn's sister was among the victims) brought the Danish king to Exeter (1003). During each of the following eleven years, the Danes, materially assisted by the universal and shameless disloyalty of the Saxon ealdormen, systematically ravaged England, and from 991 to 1014 the wretched land is said to have paid its invaders in ransoms alone £158,000. Sweyn died suddenly at Gainsborough on Feb. 13, 1014.

The data relating to his whole history are scanty and obscure, and his memory has suffered materially from the fact that the chief chroniclers of his deeds and misdeeds were ecclesiastics.

His most remarkable exploit, Svolde, was certainly won at the expense of Christianity, resulting, as it did, in the death of the saintly Olaf. It is not surprising, therefore, that Adam of Bremen, and the monkish annalists who follow him, describe Sweyn as a grim and bloody semi-pagan, perpetually warring against Christian states. But there is another side to the picture. Viking though he was, Sweyn was certainly a Christian viking.

He built churches, he invited English bishops to settle in Denmark (notably Godibald, who did good work in Scania), and on his deathbed he earnestly commended the Christian cause to his son Canute. He was cruel to his enemies but he never forgot a benefit. Thus he rewarded the patriotism of the Danish ladies who sacrificed all their jewels to pay the heavy ransom exacted from him by his captors, the Jomsborg pirates, by enacting a law whereby women were henceforth to inherit landed property in the same way as their male relatives.

Of his valour as a captain and his capacity as an administrator there can be no question. His comrades adored him for his liberality, and the frequent visits of Icelandic skalder to his court testify to a love of poetry on his part. Indeed one of his own strophes has come down to us.

He had a long cleft beard, whence his nickname of *Tiugeskaeg* or Fork-Beard.

See *Danmarks riges historie. Oldtiden og den aeldre middelalder,* pp. 364–381 (1897–1905). (R. N. B.; X.)

SWEYN II or **ESTRITHSON,** King of Denmark (d. 1076), was the son of Estrith, sister of Canute the Great, and of the Earl Ulf. When his father was murdered by Canute, Sweyn fled to Sweden, where he remained until the accession of Magnus the Good to the throne of Norway, at which time Magnus named him earl of Jutland and viceroy of Denmark. Sweyn, however, was not satisfied with this position, aspiring to become king of Denmark. He therefore involved Denmark and Norway in a war which was only terminated in 1047 with Magnus' death, caused by a fall from a horse as he rode in pursuit of the defeated and fleeing Sweyn. His dying wish, however, was that Sweyn be made king of Denmark, and Sweyn thus began the Estrith dynasty, known as the second or middle Danish dynasty, which ruled for nearly 300 years.

Sweyn was almost immediately embroiled in conflict with Harald Haardraade, Magnus' heir to the throne of Norway, who wished also to become king of Denmark. For 17 years, until

1064, the battle between the two countries was pursued sporadically and at the end of that time, although the Norwegians had decisively defeated the Danes in the naval battle of Nissa, the two monarchs declared a truce and the boundaries of Norway and Denmark were declared to be the same as when the war started. After Harald's death in 1066, Sweyn felt free to turn his attention to some other project than the defense of his country. In 1069, therefore, he sent a fleet to assist an internal rebellion against William I of England, a country which he, as nephew of Canute, felt he was entitled to rule. William, however, was able to bribe Sweyn's lieutenant and thus to forestall that threat. In 1075 Sweyn made another similar effort which proved equally abortive.

Sweyn was a devout Roman Catholic, and although he was at one time embroiled with the archbishop of Bremen over Sweyn's marriage to his own stepdaughter, he did much to encourage the development of the church in Denmark. He died in 1076. He left numerous sons, perhaps as many as 14, all illegitimate. Five of them subsequently became kings of Denmark.

SWIDNICA (German SCHWEIDNITZ), a town of Poland in the former Prussian province of Silesia, Ger., on the Weistritz river, 28 mi. S.W. of Breslau by rail. Pop. (1957 est.) 34,000. Swidnica, dating from about the 11th century, received civic rights in 1250. About 1278 it became the capital of a principality, with an area of about 1,000 sq.mi., which belonged to Bohemia from 1353 until 1741, when it passed into the possession of Prussia. The "Pölerei of Schweidnitz" is the name given to the riotous revolt of the town (1520–22) against a royal edict depriving it of the right of coining its own money. The town contains several old churches, one of which was built in the 14th century, and an old town hall. The surrounding country is fertile and highly cultivated, and large quantities of flax and hemp raised encourage the weaving industry. Beetroot for sugar, grain and fruit are also grown. The manufacture of linens, hosiery, furniture, gloves, machinery, margarine, organs, carriages and other hardware goods is carried on. The beer of Swidnica has long been famous.

SWIFT, GUSTAVUS FRANKLIN (1839–1903), U.S. pioneer in the meat-packing industry, was born June 24, 1839, at Sagamore (West Sandwich), Mass. When he was 14 he left school to take a position at a salary of $1 a week working for his brother, a butcher. By 1859, at the age of 20, he had gone into business for himself as a buyer and slaughterer of cattle on a small scale. In that year he opened his own butcher shop in Eastham and soon became the manager of a number of small shops, which he opened himself and placed in charge of other men. In 1872 he became the partner of James A. Hathaway in the firm of Hathaway and Swift. Three years later Swift, who was the cattle buyer for the firm, transferred his headquarters to Chicago, where the centre of the cattle market had shifted. He soon determined that the meat packing business would be substantially more profitable than the meat selling business if some method were devised for shipping fresh meat from Chicago to the east, instead of sending live cattle to be slaughtered on arrival, as was the custom. He therefore hired an engineer to design a refrigerator car which would make this possible, and in 1877 he successfully shipped the first refrigerator carload of fresh meat to the east. In 1878 he formed a partnership with his brother and in 1885, with a capitalization of $300,000, the firm of Swift & Company came into being, with Swift as first president. At his death less than 20 years later, the capitalization had increased to $25,000,000; more than 80 times the original size. In addition to his promotion of refrigerator cars, which was one of the most important factors in the development of the meat packing business, Swift was a leader in the field of turning previously unused parts of slaughtered animals into valuable by-products. He died March 29, 1903, in Chicago.

See Louis F. Swift and Arthur Van Vlissingen Jr., *Yankee of the Yards* (1927).

SWIFT, JONATHAN (1667–1745), dean of St. Patrick's, Dublin, British satirist, was born at No. 7 Hoey's court, Dublin, on Nov. 30, 1667, a few months after the death of his father, Jonathan Swift (1640–1667), who married about 1664 Abigaile Erick, of an old Leicestershire family. His grandfather, Thomas Swift, vicar of Goodrich near Ross, appears to have lost his possessions by taking the losing side in the Civil War and died in 1658 before the restoration could bring him redress. He married Elizabeth, niece of Sir Erasmus Dryden, the poet's grandfather. The young Swift was supported by his uncle Godwin, a Tipperary official; at the age of six he entered Kilkenny school, where Congreve was a fellow student, and he completed his education at Trinity college, Dublin. Here he exhibited few signs of precocious genius, and it was only by special act of indulgence that he obtained his degree; then, on the death of his uncle, he left Ireland and sought counsel of his mother in Leicester.

His first employment commenced towards the close of 1689, when he became secretary to Sir William Temple, who had retired to Moor Park, near Farnham, after the rapid decline of his influence during the critical period of Charles II's reign. It was there that Swift's acquaintance with Esther Johnson, the "Stella" of the famous *Journal,* was begun. Stella's mother was living at Moor Park, as servant of Temple's sister, Lady Giffard. Swift was 22 and Esther 8 years old at the time, and a curious friendship sprang up between them. He taught the little girl how to write and gave her advice in reading. On his arrival at Moor Park, Swift's duties were merely those of account-keeper and amanuensis, but his ability gradually won him the confidence of his employer and he was entrusted with some important missions.

In 1694, however, Swift (who had in the meantime obtained the degree of M.A. *ad eundem* at Oxford) quitted Temple, who had, he considered, delayed too long in obtaining him preferment; but it was only after five months' delay, when Swift had unwillingly begged the favour of a testimonial from his discarded patron, that he was able to obtain the small prebend of Kilroot near Belfast (Jan. 1695).

In the meantime he had grown tired of Irish life and was glad to accept Temple's proposal for his return to Moor Park, where he continued until Temple's death in Jan. 1699. During this period he wrote much and burned most of what he had written. He read and learned even more than he wrote. Moor Park took him away from brooding in Ireland and brought him into the corridor of contemporary history, an intimate acquaintance with which became the chief passion of Swift's life. His *Pindaric Odes,* written at this period or earlier, in the manner of Cowley, indicate the rudiments of a real satirist, but a satirist struggling with a most uncongenial form of expression. Of more importance was his first essay in satiric prose, which arose directly from the position which he occupied as domestic author in the Temple household. Temple had in 1692 published his *Essay upon Ancient and Modern Learning,* transplanting to England a controversy begun in France by Fontenelle. Incidentally Temple had cited the letters of Phalaris as evidence of the superiority of the ancients over the moderns. William Wotton's criticism of Temple's general conclusions caused Swift to write *Battle of the Books* in 1697 in refutation. Boyle's *Vindication* and Bentley's refutation of the authenticity of Phalaris came later. Swift's aim was limited to co-operation in what was then deemed the well-deserved putting down of Bentley by Boyle, with a view to which he represented Bentley and Wotton as the representatives of modern pedantry, transfixed by Boyle in a suit of armour given him by the gods as the representative of the "two noblest of things, sweetness and light." Though written in 1697, the satire remained unpublished until 1704, when it was issued with *The Tale of a Tub.*

After Temple's death, Swift suffered several disappointments in attempting to find employment, but he eventually secured the rectory of Agher in Meath with the united vicarages of Laracor and Rathbeggan, to which was added the prebend of Dunlavin in St. Patrick's—the total value being about £230 a year. He was now often in Dublin, at most 20 mi. distant, and through Lady Berkeley and her daughters—Lord Berkeley was now a lord justice of Ireland—he became the familiar and chartered satirist of the fashionable society there. But he very soon began to grow tired of Ireland again and to pay visits in Leicester and London. His resolution to exchange divinity for politics must appear fully justified by the result. The *Discourse on the Dissen-*

sions in Athens and Rome (Sept. 1701), written to repel the tactics of the Tory commons in their attack on the Partition treaties "without humour and without satire," and intended as a dissuasive from the pending impeachment of Somers, Orford, Halifax and Portland, received the honour of being generally attributed to Somers himself or to Burnet, the latter of whom found a public disavowal necessary In April or May 1704, appeared a more remarkable work. Clearness, cogency, masculine simplicity of diction, are conspicuous in the pamphlet, but true creative power told the *Tale of a Tub*. Although it lacks coherence and attains no conclusion, it is the most strikingly original of Swift's satirical works. In his kindred productions he relies mainly upon a single element of the humorous—logical sequence and unruffled gravity bridling in an otherwise frantic absurdity, and investing it with an air of sense. In the *Tale of a Tub* he lashes out in all directions. The humour, if less cogent and cumulative, is richer and more varied; the invention, too, is more daringly original and more completely out of the reach of ordinary faculties. The book is throughout a mine of suggestiveness. At the same time, it lacks unity and coherence, it attains no conclusion, and the author abuses his digressive method of composition and his convenient fiction of hiatuses in the original manuscript.

In Feb. 1701, Swift took his D.D. degree at Dublin, and before the close of the year he had taken a step destined to exercise a most important influence on his life, by inviting two ladies to Laracor. Esther, daughter of a merchant named Edward Johnson, a dependant, and legatee to a small amount, of Sir William Temple's (born in March 1680), whose acquaintance he had made at Moor Park in 1689, and whom he has immortalized as "Stella," came over with her companion Rebecca Dingley, a poor relative of the Temple family, and was soon permanently domiciled in his neighbourhood. Meanwhile the sphere of his intimacies was rapidly widening. He had been in England for three years together, 1701 to 1704, and counted Pope, Steele and Addison among his friends. In 1708 he wrote the finest example of his irony, the *Argument to prove that the abolishing of Christianity in England may, as things now stand, be attended with some inconveniencies*, and about this time, too (Nov. 1707), he produced his best narrative poem, *Baucis and Philemon*.

The next few months witnessed one of the most amusing hoaxes ever perpetrated against the quackery of astrologers, the victim being a Protestant alarmist and plot vaticinator styled John Partridge. In Jan. 1708, Swift, under the name of Isaac Bickerstaff, issued a solemn prediction that the notorious almanac maker, Partridge, would die at 11 o'clock P.M. on March 29, and on March 30 he published a letter confirming this prophecy. Partridge's fatuous denial and reply to Bickerstaff elicited Swift's amusing *Vindication of Isaac Bickerstaff, Esq.*, in April 1709. The episode has left a permanent trace in literature, for when, in 1709, Steele was to start the *Tatler*, it occurred to him that he could secure the public ear in no surer way than by adopting the name of Bickerstaff.

From Feb. 1708, to April 1709, Swift was in London, urging upon the Godolphin administration the claims of the Irish clergy to the first-fruits and twentieths ("Queen Anne's Bounty"), already granted to their brethren in England. His having been selected for such a commission shows that he was not yet regarded as a deserter from the Whigs, although the ill success of his representations probably helped to make him one. By Nov. 1710, he was again domiciled in London, and writing his *Journal to Stella*, that unique exemplar of a giant's playfulness. In the first pages of this minute record of a busy life we find him depicting the decline of Whig credit and complaining of the cold reception accorded him by Godolphin, whose penetration had doubtless detected the precariousness of his allegiance. Within a few weeks he had become the lampooner of the fallen treasurer, the bosom friend of Oxford and Bolingbroke, and the writer of the *Examiner*, a journal established as the exponent of Tory views (Nov. 1710). He was now a power in the state and the associate of ministers on a footing of perfect cordiality and familiarity. For a brief time he seemed to resume the whole power of the English press in his own pen and to guide public opinion as he would—his services to his party as writer of the *Examiner*, which he quitted in July 1711, were even surpassed by those which he rendered as the author of telling pamphlets. We need not suppose that he was consulted respecting the great Tory strokes of the creation of the twelve new peers and the dismissal of Marlborough (Dec. 1711), but they would hardly have been ventured upon if *The Conduct of the Allies* and the *Examiners* had not come first.

Generous men like Oxford and Bolingbroke cannot have been unwilling to reward so serviceable a friend, especially when their own interest lay in keeping him in England. Swift still had formidable antagonists in the archbishop of York, whom he had scandalized, and the duchess of Somerset, whom he had satirized. Anne was particularly amenable to the influence of priestly and female favourites, and it must be considered a proof of the strong interest made for Swift that she was eventually persuaded to appoint him to the deanery of St. Patrick's, Dublin. In June 1713 he set out to take possession of his dignity, and encountered a very cold reception from the Dublin public. The dissensions between the chiefs of his party speedily recalled him to England. He found affairs in a desperate condition. The queen's demise was evidently at hand and public opinion was turning towards the Whigs when the Tories manifestly could not be trusted to maintain the Protestant succession. Bolingbroke's brain teemed with the wildest plans. Swift's mediation was unavailing.

When the discord of Oxford and Bolingbroke had become patent to all the nation, Swift, foreseeing, as is probable, the impending fall of the former, retired to Upper Letcombe, in Berkshire, and there spent some weeks in the strictest seclusion. This leisure was occupied in the composition of his remarkable pamphlet, *Some Free Thoughts on the Present State of Affairs*, which indicates his complete conversion to the bold policy of Bolingbroke. The utter exclusion of Whigs as well as Dissenters from office, the remodelling of the army, the imposition of the most rigid restraints on the heir to the throne—such were the measures which, by recommending, Swift tacitly admitted to be necessary to the triumph of his party.

Bolingbroke's daring spirit, however, recoiled from no extreme, and, fortunately for Swift, he added so much of his own to the latter's ms. that the production was first delayed and then, upon the news of Anne's death, immediately suppressed. This incident only just anticipated the revolution which, after Bolingbroke had enjoyed a three days' triumph over Oxford, drove him into exile and prostrated his party. Almost the first acts of Bolingbroke's ephemeral premiership were to order him £1,000 from the exchequer and despatch him the most flattering invitations. The same post brought a letter from Oxford, soliciting Swift's company in his retirement; and, to the latter's immortal honour, he hesitated not an instant in preferring the solace of his friend to the offers of Bolingbroke. When, a few days later, Oxford was in prison and in danger of his life, Swift begged to share his captivity; and it was only on the offer being declined that he finally directed his steps towards Ireland, where he was very ill received. The draft on the exchequer was intercepted by the queen's death.

During these four busy years of London life, Swift had entered deeply into the literary life of the period. He was treasurer and a leading member of the Brothers, a society of wits and statesmen which recalls the days of Horace and Maecenas. He promoted the subscription for Pope's *Homer*, contributed some numbers to the *Tatler*, *Spectator*, and *Intelligencer*, and joined with Pope and Arbuthnot in establishing the Scriblerus Club, contributing to *Martinus Scriblerus*, his share in which can have been but small, as well as *John Bull*, where the chapter recommending the education of all blue-eyed children in depravity for the public good must surely be his. His miscellanies, such as *A Meditation upon a Broomstick*, and the poems *Sid Hamet's Rod*, *The City Shower*, *The Windsor Prophecy*, *The Prediction of Merlin*, and *The History of Vanbrugh's House*, belong to this period. A more laboured work, his *Proposal for Correcting, Improving and Ascertaining the English Tongue* (1712), in a letter to Harley, suggesting the regulation of the English language by an academy, is chiefly remarkable as a proof of the deference paid to French taste. His

History of the Four Last Years of the Reign of Queen Anne is not on a level with his other political writings.

To Swift the change from London to Dublin meant for the time the fall from unique authority to absolute insignificance. All share in the administration of even Irish affairs was denied him; every politician shunned him; and his society hardly included a single author or wit. Before and after his elevation to the deanery of St. Patrick's, Esther Johnson and Mrs. Dingley continued to reside near him, and superintended his household during his absence in London. He had offered no obstacle in 1704 to a match proposed for Stella to Dr. William Tisdall of Dublin. Whatever the cause, his conduct proved the fatal embitterment of his life and Stella's and yet another's. He had always been unlucky in his relations with women. Esther Vanhomrigh, "Vanessa" (b. Feb. 14, 1690), the daughter of a Dublin merchant of Dutch origin, had become known to Swift at the height of his political influence. He lodged close to her mother, and Vanessa insensibly became his pupil, while he became the object of her impassioned affection.

But Swift was devoid of passion. Of friendship, even of tender regard, he was fully capable, but not of love, and Vanessa's ardent and unreasoning display of passion was beyond his comprehension. Yet Vanessa assailed him on a very weak side. The strongest of all his instincts was the thirst for imperious domination. Vanessa hugged the fetters to which Stella merely submitted. Flattered to excess by her surrender, yet conscious of his binding obligations and his real preference for Stella, he could neither discard the one beauty nor desert the other. When Vanessa's mother died (1714), she followed him to Ireland, taking up her abode at Celbridge within ten miles of Dublin Unable to marry Stella without destroying Vanessa, or openly to welcome Vanessa without destroying Stella, he was thus involved in the most miserable embarrassment; he continued to temporize. Had the solution of marriage been open Stella would undoubtedly have been Swift's choice, but some mysterious obstacle intervened.

Meanwhile Swift's efforts were directed to soothe Miss Vanhomrigh, to whom he addressed *Cadenus* [*Decanus*] *and Vanessa,* the history of their attachment and the best example of his serious poetry, and for whom he sought to provide honourably in marriage, without succeeding either in his immediate aim or in thereby opening her eyes to the hopelessness of her passion. Worn out with his evasions, she at last (1723) took the desperate step of writing to Stella or, according to another account, to Swift himself, demanding to know the nature of the connection with him, and this terminated the melancholy history as with a clap of thunder. Stella sent her rival's letter to Swift, and retired to a friend's house. Swift rode down to Marley Abbey with a terrible countenance, petrified Vanessa by his frown, and departed without a word, flinging down a packet which only contained her own letter to Stella. Vanessa died within a few weeks. She left the correspondence for publication, but it was suppressed until it was published by Sir Walter Scott. Five years after Vanessa's death Stella died, on Jan. 28, 1728.

Between the death of Vanessa and the death of Stella came the greatest political and the greatest literary triumph of Swift's life. Although he was not an Irish patriot in the strict sense of the word, his pride and sense of equity alike revolted against the stay-at-home Englishmen's contemptuous treatment of their own garrison, and he delighted in finding a point in which the triumphant faction was still vulnerable. His *Proposal for the Universal Use of Irish Manufactures,* published anonymously in 1720, urging the Irish to disuse English goods, became the subject of a prosecution, which at length had to be dropped. A greater opportunity was at hand. A patent for supplying Ireland with a coinage of copper halfpence was accorded to William Wood on such terms that the profit accruing from the difference between the intrinsic and the nominal value of the coins, about 40%, was mainly divided between him and George I.'s favourite duchess of Kendal, by whose influence Wood had obtained the privilege. Swift now had his opportunity, and the famous six letters signed M. B. Drapier (April to Dec. 1724) soon set Ireland in a flame. Every effort was used to discover, or rather to obtain legal evidence against, the author, but none could be procured; the public passion swept everything before it; the patent was cancelled.

The noise of the *Drapier Letters* was followed by the anonymous publication of *Travels Into Several Remote Nations of the World,* in four parts, by Lemuel Gulliver, first a surgeon and then a captain of several ships (Benjamin Motto, Oct. 1726), the work being well advanced, it would seem, by 1720. The keenness of the satire on courts, parties and statesmen certainly suggests that it was planned while Swift's disappointments as a public man were still rankling and recent. Although he was afraid of the reception the book would meet with, especially in political circles, the world chose to be diverted by it. In the first two parts the misanthropy is quite overpowered by the fun. The third part, equally masterly in composition, is less felicitous in invention; and in the fourth Swift has indeed carried out his design of vexing the world at his own cost.

Swift's grave humour and power of enforcing momentous truth by ludicrous exaggeration were next displayed in his *Modest Proposal for Preventing the Children of Poor People from being a Burden to their Parents or the Country,* by fattening and eating them (1729). The *Directions to Servants,* a satire on domestics, was first published in 1745, while *Polite Conversation,* written in 1731 was published in 1738. Little beyond occasional verses—trivial and often indecent—followed, but the delightful *Hamilton's Bawn,* and the verses on his own death (1731) are exceptions, and in *The Legion Club* of 1736 he composed the fiercest of all his verse satires. His popularity remained as great as ever (he received the freedom of Dublin in 1729), and he governed his cathedral with great strictness and conscientiousness; but the attacks of giddiness to which he had always been subject increased upon him, and he grew more and more capricious and morbidly suspicious. In March 1742 it was necessary to appoint guardians of Swift's person and estate. In September of the same year his physical malady reached a crisis, from which he emerged a helpless wreck, with faculties paralysed rather than destroyed, and he eventually sank into the dementia which preceded his death on Oct. 19, 1745. He was interred in his cathedral at midnight in a coffin by the side of Stella's. His epitaph was written by himself.

An object of pity as well as of awe, Swift is one of the most tragic figures of English literature. His master passion was imperious pride—lust for despotic dominion; place, profit and literary fame were comparatively indifferent to him. Contemptuous of the opinion of his fellows, he hid his virtues, paraded his faults, affected some failings from which he was really exempt, and, since his munificent charity could not be concealed from the recipients, laboured to spoil it by gratuitous surliness. "To think of him," says Thackeray, "is like thinking of the ruin of a great empire." Among those influenced by Swift may be mentioned Chesterfield, Smollett, Cobbett, Hazlitt, Scott, Borrow, Newman, Belloc.

BIBLIOGRAPHY.—The contemporary lives of Swift, most of which contain a certain amount of apocrypha, are those of Lord Orrery (1751); Patrick Delany's *Observations on Orrery* (1754); Deane Swift's *Essay Upon the Life of Swift* (1755); and Thomas Sheridan's *Life* (1785). John Hawkesworth, in the life prefixed to his edition of the *Works* in 1755, adds little of importance. Samuel Johnson's *Life* in "Lives of the Poets," vol. 3 (1781), is marred by manifest prejudice. J. Barrett produced an *Essay Upon the Early Life* of some value (1808). Six years later came the useful biography of Sir Walter Scott and, in 1819, appeared the elaborate *Life* by W. Monck Mason in the form of an appendix to his ponderous *History of St. Patrick's.*

A new epoch of investigation was inaugurated by John Forster, who began a new scrutiny of the accumulated material and published his first volume in 1875. Invaluable in many respects, it exhibited the process as well as the result of biography, and never got beyond 1711. The *Life* by Sir Henry Craik (1882 and reissues) continued to hold the field.

Valuable monographs were produced by Sir Leslie Stephen (Men of Letters and the Memoirs, in the *Dictionary of National Biography*), by Thackeray in his *English Humourists,* by W. R. Wilde in his *Closing Years of Dean Swift's Life,* by Lecky in his *Leaders of Public Opinion,* by G. P. Moriarty, J. Churton Collins (1893), Max Simon (1893), Henriette Cordelet (1907) and Sophie Shilleto Smith (1910). The anecdotes of Swift related in Spence's *Laetitia Pilkington,* Wilson's *Swiftiana,* Delany's *Autobiography,* etc., though often amusing, can hardly be accepted as authentic.

Later biographical and critical works include Carl Van Doren, *Swift* (1930; 1931); S. L. Gwynn, *The Life and Friendships of Dean Swift* (1933); W. D. Taylor, *Jonathan Swift* (1933); Ricardo Quintana, *Mind and Art of Jonathan Swift* (1936; reprinted, 1953), *Swift: an Introduction* (1955); Bertram Newman, *Jonathan Swift* (1937); Bernard Acworth, *Swift* (1948).

The collective editions of Hawkesworth (various issues, 1755–79), T. Sheridan (1785), John Nichols (1801, 1804, 1808), Scott (1814, 1821) and Roscoe (2 vol., 1849) were in most respects superseded by the edition in Bohn's Standard Library in 14 vol. (including the two subsequently issued volumes of *Poems;* 1897–1910) arranged as follows: i, biographical introduction by W. E. H. Lecky, *Tale of a Tub, Battle of the Books, Critical Essay Upon the Faculties of the Mind, The Bickerstaff Pamphlets,* etc., ed. by Temple Scott; ii, *Journal to Stella,* ed. by F. Ryland (two portraits of Stella); iii and iv, *Writings on Religion and the Church,* ed. by Temple Scott; v, *Historical and Political Tracts—English,* ed. by Temple Scott; vi, *Historical and Political Tracts—Irish,* ed. by Temple Scott; vii, *The Drapier's Letters,* ed. by Temple Scott; viii and xi, literary essays, including *Gulliver's Travels,* ed. by G. R. Dennis, *A Proposal for Correcting, Improving and Ascertaining the English Tongue, Hints Towards an Essay on Conversation, Character, Directions to Servants* and *Autobiographical Fragment,* ed. by Temple Scott; ix, contributions to the *Examiner, Tatler, Spectator,* etc., ed. by Temple Scott; x, historical writings, including the *Four Last Years, Abstract of English History* and *Remarks on Burnet,* ed. by Temple Scott; xii, essays on the portraits, etc., bibliography by W. Spencer Jackson and index; twelve portraits of Swift are included in the work, in addition to two portraits of Stella and one of Vanessa; xiii and xiv, poems, ed. by W. Ernst Browning.

Translations and editions of *Gulliver's Travels* have been numerous; it was included in vol. 36 of "Great Books of the Western World" (1952), ed. by Robert M. Hutchins. "Valuable Notes for a Bibliography of Swift" were published by S. Lane Poole in *The Bibliographer* (Nov. 1884). *See* also H. Teerink (ed.), *Bibliography of the Writings in Prose and Verse of Jonathan Swift* (The Hague, 1937).

(R. G.; X.)

SWIFT, a bird so called from the speed of its flight. Despite its appearance, the swift (*Apus apus*) is in no way related to the swallow and is not even a passerine bird, its affinities being with the hummingbirds. The common swift is a summer visitor to Europe, arriving in May and returning to Africa at the end of August. It nests in holes in buildings, less frequently in cliffs and quarries. On the wing it gives the idea of a bow and arrow; it is a larger bird than the swallow and its flight is swifter and steadier. Except for a grayish-white patch under the chin, the plumage is entirely sooty black. Near sunset the birds often sweep round in screaming bands, and sometimes then ascend out of sight. It has been stated that they may spend the night on the wing at great heights; further observation on this point is needed, as also on the assertion that swifts actually pair in mid-air. A larger species, *A. melba,* with white under parts, inhabits the mountains and cliffs of southern Europe and is very numerous in the old water tower at Lucerne.

The allied genus *Collocalia* constructs nests of saliva and these are used by the Chinese for soup. The birds breed in caves, and range from north Madagascar to the Marquesas, one species occurring in the hill country of India. The chimney swallow of the United States (*Chaetura pelagica*) is also a swift, cementing its nest with saliva. It is migratory.

SWIMMING, the action of self-support and self-propulsion on or in the water; though used by analogy of inanimate objects, the term is generally connected with animal progression and especially with the art of self-propulsion in the water as practised by man. It is one of the most useful of man's physical accomplishments. There have been many cases in which beginners have demonstrated some ability in swimming upon their first immersion in the water, but generally speaking it is a skill that must be acquired.

Types of swimming strokes commonly used are the breast, side, crawl and back strokes. Each of these has been popular at different periods. The Australian crawl, comparatively recently popularized, is probably the best known stroke. With the increased acceptance of swimming as a recreation came its organization as a national and international competitive sport. Swimming was one of the sports included in the revived Olympic games at Athens in 1896, and it continues to hold an important place in the Olympics.

This article is divided into the following sections:

1. History
2. Teaching Swimming
3. Swimming Strokes
4. Competitive Swimming
5. Floating
6. Diving
7. Channel Swimming
8. Open-Water Swimming

1. History.—Swimming was rated highly in the days of Greece and Rome, practised widely and stressed in the training of warriors. But it fell into almost complete disuse throughout Europe in the middle ages, when the belief became general that outdoor bathing helped to spread the epidemics that so often swept the continent. It was not until the second half of the 19th century that the prejudice was overcome sufficiently for aquatic activities to be resumed. Competition was resumed, first in Great Britain, then elsewhere, and the study of swimming technique led to gradual improvement in methods.

As knowledge grew, there came steadily increasing appreciation of the value of swimming in protecting life as well as in physical education, for it had gained recognition as the most beneficial of all forms of exercise. Physical educators agreed that there is no physical activity that can contribute as much as swimming to a person's physical well-being.

Teaching of swimming was taken up by the schools; sectional and national learn-to-swim campaigns were fostered in most countries. During World War II nearly all belligerent nations included swimming as a part of the training of servicemen, and after the war the gains in swimming instruction and promotion that had been made were consolidated and furthered by national organizations such as the U.S. Conference for National Cooperation in Aquatics. This organization, operating informally and meeting annually, sought to further the cause of swimming specifically because of its safety value but also because of its value as an all-around exercise.

2. Teaching Swimming.—Many methods of swimming instruction were tested as changes in stroke took place, but in the long run the great majority of teachers enrolled in one of two predominant schools, the European, whose protagonists believe the breast stroke should be taught first, and the U.S., whose followers assert that the crawl is mastered far more readily than any other stroke. Both schools favour preliminary drills on land or in shallow water, so that beginners may gain some knowledge of the required movements before attempting to swim.

The greatest handicaps in learning are failure to relax and haste in performing the action. Fear of the water, or the mere nervousness experienced by most nonswimmers, generally causes muscle tension and engenders the feeling that it is necessary to thrash wildly with arms and legs in order to keep afloat. Every effort must be made to overcome these natural tendencies, for rapid progress is dependent upon complete relaxation and unhurried movements. Instructors usually help pupils to become accustomed to immersion by having them play in shallow water, ducking the head and trying to pick up objects from the bottom, before starting schooling.

During World Wars I and II, when great numbers of men in their late 20s and early 30s had to be taught and speed in training was essential, a method of starting the pupil on his back was developed. In this position floating was facilitated, developing confidence in the water; normal breathing was utilized in contrast with the breathing difficulty in face-down strokes, and a very simple propulsive action was taught—a frog kick with a synchronous short arm stroke, the arms being underwater to prevent splash, moving from side, horizontal, downward to the thighs.

3. Swimming Strokes.—In the early stages of swimming competition, the standard racing style was the side stroke, in which both arms remain submerged throughout, while the legs execute a wide scissor kick, opening and closing in movements resembling those of walking. The first advance over this was made when swimmers conceived the idea of minimizing resistance by thrusting the upper arm forward above the surface, instead of underwater. The new stroke was called the single overarm. The double overarm followed, with both arms recovering above the surface. This stroke was made prominent in 1893 by an Englishman, J. Trudgen, who told of picking it up in South America. It took his name, misspelled Trudgeon in some countries, and served later to introduce the basic arm action of the crawl, the most efficient mode of swimming developed so far.

The changes in leg drive were marked by a consistent trend to-

ward narrower and faster movements. Following the advent of the trudgen, the wide scissor kick of the side stroke, originally performed with legs bent excessively, was reduced in scope and executed with the legs a good deal straighter. Then, with the introduction of the crawl, the legs began to whip up and down alternately in a gradually quickening and diminishing thrash, speeding up by degrees from two to ten proportionately smaller

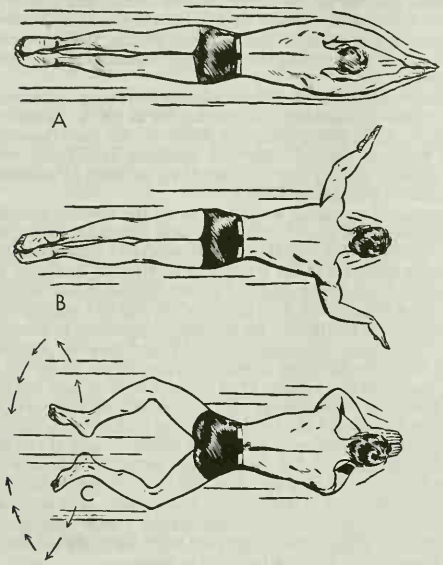


FIG. 1.—THE BREAST STROKE

A. Position at start and finish of stroke. B. Drive of arms, swimmer inhaling. C. Start of recovery of arms; legs ready for kick; swimmer exhaling under water. Arrows indicate continuation of leg drive

scissors to each cycle of the two arms.

Breast Stroke.—Two styles of breast stroke have been evolved, the ancient and orthodox variety, believed to be the oldest of all strokes, and the butterfly, a racing adaptation brought out in 1935 by Henry Myers of the United States. The orthodox breast stroke, used for teaching beginners, is a stroke of great practical value. It affords the means to assist a tired swimmer or support a drowning person with comparatively little effort; it enables the swimmer to advance safely, head up, through waters littered with flotsam or wreckage; it is the only stroke that makes it possible for soldiers to swim under the burden of heavy packs and arms. It also has the advantage, during war, of being noiseless, as the arms and legs are underwater in recovery and propulsion. The butterfly is a competitive style only.

The two styles are distinguished by differences in the action of the arms. In both the arms start their drive from full extension in front, hands close together on the surface, palms down. But in the orthodox breast stroke the arms propel by pressing down and outward at a slant that places them at an angle of about 80° on reaching the shoulder line, when the elbows are brought to the sides, the hands to the chest, again palms down, and the arms shot forward underwater to their original position. In the butterfly the arms pull under the body all the way to the thighs, then emerge and fling forward above surface in circular motion. The leg action of both strokes, known as the frog kick, is performed by drawing up the legs in a manner somewhat like that of squatting, knees apart, feet close and pointed; then the feet are turned outward, and, with a continuous sweep, the legs are kicked vigorously back and out to an angle of about 90° and snapped sharply together.

The closing movement occurs as the arms shoot forward in recovery. After the kick the body is allowed to glide motionless for a brief spell, forming a straight line from hands to pointed toes. The orthodox stroke may be used comfortably head up, but greater speed is attained in both styles by raising the head to inhale during the drive of the arms and dipping the face to exhale underwater during their recovery. The butterfly is the faster by a wide margin.

The butterfly was revised by the Fédération Internationale de Natation Amateur (Helsinki congress, 1952) to permit a synchronous up-and-down action of the legs characterized by the term "fishtail kick," or "dolphin," which further increases the speed of this type of breast stroke.

Side Stroke.—The side stroke is valued chiefly for lifesaving purposes, as it is used in the most efficient carries devised for the rescue of the drowning. But it is liked for open-water pleasure swimming also, because it is faster than the breast stroke and the face remains constantly above water. The body stays steadily on its side, as the name implies, and the arms propel alternately. The underarm starts from comfortable extension overhead and pulls below the body to the waist, then bends upward and pushes forward; the upper arm drives a little outside the body line, sweeping to the thigh, and recovers close to the surface. The legs open slowly for the scissor kick, the under leg moving backward, the upper leg forward, both knees bent a little, feet pointed; when at an angle of about 50° they whip smartly together, the closing movement timed with the finish of the upper arm drive. This stroke is also excellent for survival swimming or for emergency situations where it is imperative that the arms in all actions be kept underwater.

Crawl Stroke.—The crawl is the fastest of all known strokes, yet it requires little effort, as attested by the fact that many of the greatest feats of endurance achieved by men and women stand to its credit, including the crossing of the English channel and swimming the length of the Panama canal. Its remarkable speed has made it not only supreme in the competitive field but also extremely valuable in saving drowning persons, getting away from a sinking ship and battling strong tides or currents in the effort to reach safety.

There is evidence that the crawl was well known in some of the South sea islands long before Richard Cavill of Australia introduced it about 1900 and broadcast its merits by breaking the world record for 100 yd. with it. It was discovered that the stroke had been popular in the Solomons as far back as the oldest islanders could remember, and long before it was introduced in Australia sportsmen in Honolulu frequently arranged swimming races for young Hawaiian swimmers at Waikiki beach in which the boys used a stroke very similar to the one later demonstrated with

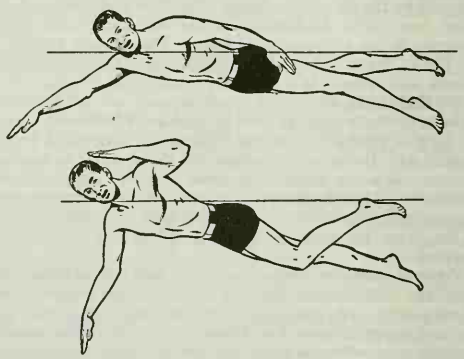


FIG. 2.—THE CRAWL STROKE, INTRODUCED INTERNATIONALLY IN 1908

such success by Duke Kahanamoku. The name of crawl was coined as the result of a press account of the initial showing of the stroke, in which the writer stated that Cavill appeared literally to crawl over the water. Cavill performed two kicks per full stroke,

and his action was adopted universally in Australia, so the leg drive remained unchanged until the stroke made its way to the United States.

The earliest knowledge of the crawl obtained in the U.S. was derived from hazy newspaper accounts and descriptions of Cavill's strokes. In attempting to imitate the action, U.S. swimmers unwittingly developed a faster leg drive, executing four scissoring kicks per stroke instead of the two performed by Cavill. The new leg drive was combined with the double overarm action of the trudgen, and the resulting stroke was called the four-beat crawl, to distinguish it from the two-beat variety.

Meanwhile, experiments had led to the belief that the four-beat leg thrash was most effective when composed of one comparatively wide and three very narrow scissoring motions. This style, termed the four-beat trudgen crawl or the four-beat single rhythm crawl, became the accepted racing stroke. The combination of movements was responsible for so notable an improvement in average performances that technical experts conceived the possibility of obtaining still greater speed from a six-beat drive set to a double rhythm—two major and four minor kicks—which would minimize resistance and afford more even distribution of propelling power. But coaches were unanimous in declaring that the swift thrash would prove entirely too laborious for distances longer than 100 yd., and the subject was dropped temporarily.

Late in 1917 two young swimmers of the New York Women's Swimming association, Charlotte Boyle and Claire Galligan, determined to give the six-beat a trial, and by the summer of 1918 they broke records with it over the longer regulation courses, 880 yd. and one mile. So convincing was this demonstration that it caused a sudden change of mind among the coaches and competitors. The six-beat crawl immediately won favour in the U.S., presently in other countries, and within a few years it had become the recognized stroke the world over, not only for racing but also for all-around purposes. Another Women's Swimming association member, Ethel McGary, successfully developed in turn the eight- and ten-beat varieties of the crawl, but as the eight- and ten-beat strokes were found to require exceptionally limber legs in time the six-beat crawl came to be accepted as the standard action. Nevertheless, a great many women and some men used the eight-beat successfully. It enabled Gertrude Ederle of the U.S. to cross the English channel in 14 hr. 34 min. and break the men's record for the course, 16 hr. 33 min. No men and very few women have made their mark with the ten-beat crawl.

The extent of the progress made in the U.S. through the adoption and development of the crawl is best illustrated by the lowering of the U.S. records in regulation competitive events—the 100 yd., 220 yd., 440 yd. in 25-yd. pools; and 880 yd. and 1 mi. over long courses (50 yd. and upward). By 1956 the men's record in the 100 yd. had dropped from 58.6 to 48.9 sec.; in the 220 yd. from 2 min. 25.4 sec. to 2 min. 3.4 sec.; in the 440 yd. from 5 min. 22.2 sec. to 4 min. 26.7 sec.; in the 880 yd. from 11 min. 24.2 sec. to 9 min. 30.7 sec.; and in the mile from 25 min. 24.4 sec. to 19 min. 40.4 sec.

The answer to what will be the ultimate in speed swimming, a question asked often because of the frequency of record breaking, seems to be that the development of speed swimming is still in its infancy, as witnessed by the advances made in skill since the introduction of the crawl stroke. With increased skill in swimming as a result of widespread and ever-expanding instruction and facilities, it is difficult even to guess what the future holds.

The correct position for swimming the crawl is like that assumed in standing erect on tiptoe. The arms move alternately, so timed that one will start pulling just before the other has completed its drive, thereby making the propulsion continuous. Each arm catches in front of the body in line with the shoulder on the same side, almost fully extended with elbow and shoulder slightly flexed. Such slight flexion affords greater strength and arm position than does full extension. The elbow on catch should be slightly higher than the hand, forearm turned slightly inward, wrist straight, hand palm down, fingers and thumb close but not cupped. The power of the arm stroke can be thought of as pressing downward and backward during the first third of the stroke, pulling during the middle third and pushing through and backward but not upward during the final third of the action, completing an arc of roughly 180°. The direction of pull should be on a line from the shoulder moving slightly under the body and out past the thigh on recovery.

On recovery the elbow is bent, the elbow quickly flexes and the forearm circles quickly out and around until it points toward the place of entry. Then the elbow is extended for the catch of the arm on a line with the shoulder.

The legs are held fairly straight but fully relaxed. The legs whip up and down alternately in an undulating action so that in both the up kick and down kick there will be a backward push against the water with the lower leg and more so with the foot. The scope of the thrash must vary somewhat according to the individual length of leg, the number of beats used per stroke and the distance to be covered.

Breathing is effected by turning the head to either side during the recovery of the upper arm, for the purpose of a gasping inhalation, and then dropping the head squarely face downward, exhaling through the nose or nose and mouth. One breath is taken for every full stroke (right and left arm). The skilled swimmer learns to breathe on either side, but the regular breathing pattern is established on one side or the other.

Crawl balance is attained by working the arms independently of the shoulders so that the body lies perfectly flat even during the action of inhalation, when the head turns without distorting the trunk position. The shoulders are level at all times; the water line of the head is at the eyebrows when the face is down, and any distortion of the trunk in the shoulder or pelvic girdle should be avoided.

Backstroke.—The style of the backstroke in vogue prior to 1912 was aptly described as an inverted breast stroke. But that year Harry Hebner, the U.S. champion, amazed spectators in the Olympic race at Stockholm with his supine crawl, never before seen in Europe, winning in record time. That exhibition soon drove the old stroke (double overarm on the back with a frog kick) into discard, for the back crawl is a good deal faster, requires less effort and is easier to learn. In the back crawl the swimmer's position is such that the body is in a horizontal plane, head lifted slightly upward looking at the feet, with hips slightly lower than the shoulders. The arms work alternately and are intertimed as in the prone crawl or free style. Each arm stretches above the head and enters the water six inches to a foot outside the shoulder line, pulling and driving to the thigh with a circular outward motion at a depth sufficient only to keep the arm underwater in a plane

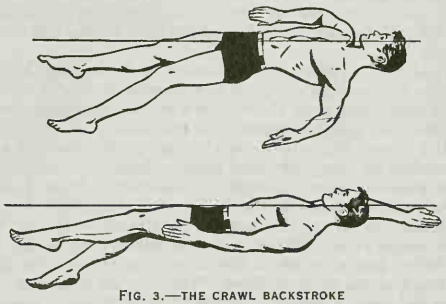


FIG. 3.—THE CRAWL BACKSTROKE

parallel to the surface of the water. The arms in recovery are carried straight or with a partially flexed elbow to the catch, either in line with the shoulder or in a semicircular motion sideward and upward. The shoulders and trunk are kept flat without any rolling. The legs execute an alternate thrash as in the prone crawl; the knees bend more but are kept underwater at all times. This makes for a much deeper leg kick (leg and foot) than in the prone crawl. In the back crawl, with the face out of water all the time, the breathing presents no difficulties as it does in the breast stroke and prone crawl.

The back crawl appeared on the scene last of all of the competitive styles and has undoubtedly made the greatest strides in record breaking.

4. Competitive Swimming.—Swimming joined the ranks of standardized competitive athletics in 1869, when the Amateur Swimming association was founded in England. Other countries followed the lead at intervals, the U.S. launching the Amateur Athletic union in 1878, and before many more years nearly all western nations had formed governing bodies and taken measures to promote the sport. Eventually, national championships for men and women, as well as intervarsity and interscholastic title tests, were instituted everywhere. International competition was inaugurated with the holding of the revived Olympic games at Athens in 1896 when four swimming events were placed on the program. Activities grew apace thereafter, and frequent intercountry encounters were added to the quadrennial Olympiads, making increasingly evident the desirability of unifying rules and regulations.

To this end, the Fédération Internationale de Natation Amateur, composed of controlling national organizations, was formed in London in 1908. It undertook to draft a code, supervise and register world's records and outline programs for Olympic games. As membership was made conditional to the observance of all F.I.N.A. rules, universal standards were soon established. The federation chose swimming, diving and water polo as the branches of aquatic sports over which it would assume authority. The three classes of swimming events already popular were retained: free style, permitting the use of any kind of stroke; backstroke, regulated by the one provision that the entire course must be covered in supine position; and breast stroke, confined to this mode of swimming and governed by strict rules concerning the execution of the action.

Free-style races have always furnished the bulk of the competition in the Olympic schedule. In national championships, free-style predominance is even greater. Although long-distance races are an integral part of summer activities, most competition takes place in pools. The building of enclosed and heated pools speeded the growth of aquatic contest during the winter months.

In the 1896 Olympics, Hungary took top honours with two wins to one each for Greece and Austria. The II Olympiad in 1900 saw Australia, Germany and Great Britain win two events each and France one. The United States took its first Olympic swimming victories at St. Louis in 1904, winning two of the free-style events and a relay. Germany, with four victories, however, captured top honours. Great Britain, in 1908, showed its superiority by winning four of the six swimming events, but the 1912 Olympics, marked by the initial appearance of women contenders, was practically a standoff with Canada, Germany and the U.S. winning two events each.

World War I caused the 1916 games to be canceled, but by 1920 the U.S. was ready to profit from earnest study of the crawl. In the 1920 Olympics the U.S. took five of seven swimming events, the other two going to Sweden. At Los Angeles in 1932, while the U.S. women successfully defended their laurels, the men were defeated by the vastly improved Japanese, winners of all but one of the contested swims. Competition was close at Berlin in 1936 in the men's swimming events, which resulted in three victories for Japan, two for the U.S. and one for Hungary. The Netherlands' women's team won four of the five women's races—the 100- and 400-m. free style, the backstroke and the relay—losing only the breast-stroke event to Japan. This was the first time that Japan had taken a first place in the Olympic games in women's swimming and the first time since World War I that the U.S. women's team had been unable to win a first place.

In the London Olympic games of 1948 the U.S. men's team won first place in all six events, the first time since the Olympic games revival that any nation had achieved such a sweep in the Olympics. In women's swimming, the team from Denmark, which had been sharing honours with the Dutch team in international swimming circles since 1936, won two first places—the 100-m. free style and the backstroke. The team from the Netherlands won the breast-stroke event; and the U.S. team with Ann Curtis won the 400-m. free-style race and the relay.

The U.S. maintained its leading position in men's swimming at the Helsinki Olympic games, 1952, winning four races—the 100- and 1,500-m. free style, the backstroke and the relay. France, after great advances during several Olympiads, won a magnificent victory in the 400-m. free style, and Australia won the breast stroke. Women's swimming supremacy saw another great shift, from northern Europe, the Netherlands and Denmark to central Europe. The women's team representing Hungary won four of the five races in a most outstanding and convincing manner—the 100- and 400-m. free style, the breast stroke and the relay—and the backstroke race was won by South Africa, the first time this team had ever won an Olympic swimming competition.

The 1956 Olympic games at Melbourne, Austr., served notice that the Australians had arrived as the world power in swimming. Australian competitors took all free-style titles in both men's and women's events in a magnificent display of form and power. New records were established for every event on the Olympic program except for the men's and the women's 200-m. breast stroke events.

Two years later, in Australian state and national competition in the spring of 1958, 31 new world records were reported. Fifteen-year-old Latvian-born John Konrads and his 13-year-old sister Ilsa accounted for 16 of the new marks.

At Rome, for the 1960 Olympics, Australia and the U.S. took four events each in the men's swimming events while in the women's competition five of the seven races went to the U.S. swimmers and one each to Australia and Great Britain.

With increased swimming facilities the world over, and means of communication and transportation making possible more and more international meeting in addition to the Olympic games, excellence in performance follows no one pattern or nation, which a checking of the record books clearly indicates.

A summary of first places in aquatic events (swimming, diving and water polo) in the Olympic games indicates the broad spread of the sport since the first of the modern games in 1896. In the men's events the order through 1960 was: U.S. 56, Australia 15, Germany 12, Japan 10, Great Britain 10, Hungary 9, Sweden 6, France 3, Canada 2, Italy 2, Argentina, Austria, Greece and Mexico 1 each. First place results for women were: U.S. 37, Australia 6, Netherlands 6, Germany 4, Great Britain 4, Hungary 4, Denmark 3, Sweden 2, Japan 1 and South Africa 1.

5. Floating.—One of the most useful accomplishments of a swimmer is floating. It is a matter of buoyancy, which varies with individuals. Buoyancy cannot be taught, but body balance in the water can be acquired through knowledge. In learning to float, the beginner experiences difficulty in keeping the legs in the horizontal plane, and this can be remedied by the counterbalance of extending the arms beyond the head, whether in face down or back floating position.

Floating motionless can be performed with proper breath control and body balance. Bending the knees to shorten the axis often improves balance in the water. Supine or back floating with a minimum of motion of the arms and legs is a valuable safety skill for the swimmer as well as for the novice. Wartime research by naval authorities showed that this type of floating conserves energy and, in areas where detonations may occur, minimizes percussion danger, facilitating survival with fewer internal injuries.

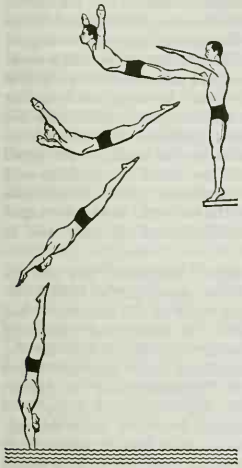
Movement of the legs in an alternating walking motion and downward pressure of the hands and forearms is the only motion required to keep the body comfortably afloat in the back floating position.

6. Diving.—The method of entering the water with either head foremost or feet foremost is called diving. Shortly after man learned to swim, the dive became closely related to swimming, perhaps only as a quick way of getting into the water in escaping from wild animals or fire. Diving for pleasure and exercise, as it is done today, can most easily be performed by standing on the edge of a pool or dock, holding the arms overhead and bending forward at the hips until the balance is lost. The entry into the water will be smooth and without ill effect.

Diving, plain and fancy, which is today a part of all aquatic contests, developed rapidly through the first half of the 20th

century. Many gymnastic and tumbling stunts can be performed before entry is made into the water, and the thrill of this combination is a challenge to young persons to achieve perfection in this stimulating sport.

Diving became a part of the swimming program of the Olympic games as early as 1904. The program at that time was plain high diving and variety high diving, with contestants performing from both the 5-m. and 10-m. firm platforms. The contest was composed of compulsory dives and voluntary dives. Springboard diving, plain and variety, was performed from an elastic board at heights of 1 m. and 3 m. above the water. It was necessary to have judges to mark the dives performed, with attention paid to the take-off, the bearing of body in the air and the entrance into the water.



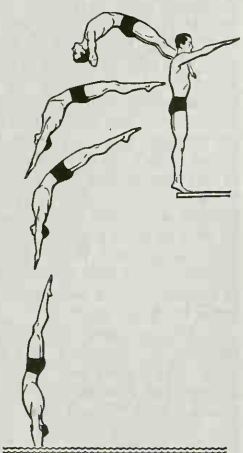
ILLUSTRATIONS FROM OFFICIAL A.A.U. SWIMMING HANDBOOK BY COURTESY OF AMATEUR ATHLETIC UNION OF UNITED STATES

FIG. 4.—FORWARD HEADER DIVE (RUNNING)

The Olympic committee composed a table listing the type of dives allowed, their description, and a value for each dive based on the difficulty of performance. These early tables comprised 14 dives from the high firm platform and 20 from the springboard, as compared with the 46 and 47 of mid-20th century.

To indicate the advances made in diving, it can best be described by comparing the dives of yesteryear with today's complex stunts. The report following the 4th Olympiad suggested elimination of the double somersault because it was believed that a diver could not control the execution of such numerous spins without chance of injury. Today, divers from most sections of the world are confidently somersaulting 2½ times backward and 3½ times forward from both firm and springboard elevations. The combinations of spins and twists are often done, and a forward 1½ somersault is often combined with a double twist, performed with great skill and ease.

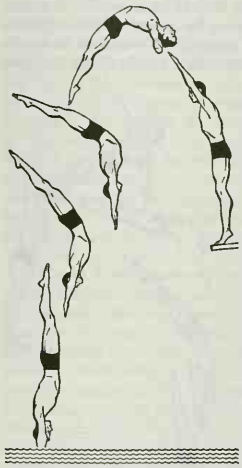
Beginning with the Olympic games of 1920, which excluded German athletes, U.S. divers dominated international competition. This supremacy can be credited to three U.S. coaches: Ernst Brandsten, an outstanding Swedish diver in the 1912 games whose chief contribution was his thoroughness in teaching fundamentals and his success in standardizing equipment; Fred Cady, a sports-minded artist whose sensitiveness to grace and poise had great influence; and Mike Peppe, a midwest coach who developed more national and international diving champions than any other coach. The combination of ideas and tuitions represented by such men as these produced U.S. diving champions, in all Olympic events after 1920 except for the men's 10-m. platform event in 1956 and both events in the women's diving competition in 1960.



ILLUSTRATIONS FROM OFFICIAL A.A.U. SWIMMING HANDBOOK BY COURTESY OF AMATEUR ATHLETIC UNION OF UNITED STATES

FIG. 5.—BACKWARD HEADER DIVE

The Groups of Competitive Dives.—Competitive dives are divided into five groups encompassing springboard dives and fixed platform dives, with the addition of arm stand dives done from fixed platforms. The first includes the forward dives (fig. 4), in which the person dives out away from the platform and rotates ½ turn to 3½ turns before entering the water. The second group comprises backward dives (fig. 5), in which the diver stands at the edge of a platform and rotates backward ½ turn to 2½ somersaults. The third group (fig. 6), the forward motion and reverse spin dives, is commonly called the reverse group. In this form of diving the diver can rotate in reverse from ½ turn to 2½ turns.



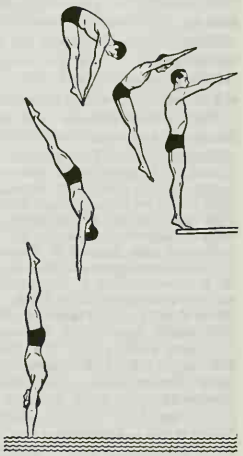
ILLUSTRATIONS FROM OFFICIAL A.A.U. SWIMMING HANDBOOK BY COURTESY OF AMATEUR ATHLETIC UNION OF UNITED STATES

FIG. 6.—HALF GAINER DIVE (RUNNING)

The fourth group is that of inward dives (fig. 7). In this category the diver stands on the outer edge of the platform and jumps backward with a forward rotation from ½ turn to 2½ somersaults. The fifth and last classification is that of twisting (fig. 8). In this movement the diver rotates the body on the long axis, from ½ twist to 2½ twists, while revolving on the short axis from ½ turn to 1½ somersaults. With the possibilities of combining all groups with twisting, it can easily be seen why the last grouping offers more dives than any other from which the diver can choose his optional dives.

The majority of dives included in the present roster may be executed in three distinct positions; namely, straight, pike or tuck. In the first position (straight), the body must be held extended, and no flexion at the hips or knees is allowed. In the second position (pike), there may be a bend at the hips, with no knee flexion. In the third position (tuck), both the knees and hips are flexed and the position of the body resembles a ball. In all positions the feet must be pointed.

Diving Board.—The surfaces of diving boards, whether firm or flexible, must be covered with cocoa matting or some other nonskid material the full length of the run. Springboards may be made of wood, aluminum, steel or other resilient material. Standard equipment for springboards is the adjustable fulcrum, which enables the diver to adjust the flexibility of the board to his liking.



ILLUSTRATIONS FROM OFFICIAL A.A.U. SWIMMING HANDBOOK BY COURTESY OF AMATEUR ATHLETIC UNION OF UNITED STATES

FIG. 7.—BACK-JACK-KNIFE DIVE

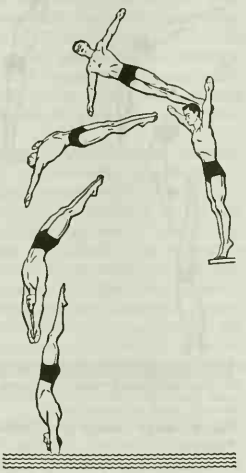
7. Channel Swimming.—Much interest has centred in the many attempts to swim across the English channel. The first recorded attempt was made on Aug. 24, 1872, by J. B. Johnson, who started from Dover but remained in the water only 65 min. On Aug. 12, 1875, Capt. Matthew Webb made his first attempt. He started from Dover and remained in the water 6 hr. 49 min., until the weather became too rough for him to continue. It is estimated that he was about 13½ mi. across when he had to give up. On Aug. 24 and 25, 1875, Webb swam the channel, diving from the Admiralty pier, Dover, and touching Calais sands after swimming for 21 hr. 45 min. The accomplish-

ment created a great sensation in England, and many swimmers have since repeated Webb's performance.

8. **Open-Water Swimming.**—The enjoyment and pleasure derived from bathing and swimming in salt or lake waters make it among the most popular of all sports and recreational activities the world over, but proper caution in water activities should be taken.

Swimming in unsupervised areas should be avoided. A person should never swim outward from shore for any great distance without an accompanying boat. If he is caught in sea grass or weeds, he should not struggle against it but rather gently shake the arms and legs until the grass floats free.

Tides, riptides and undertow must be noted in offshore swimming, and knowledge of safety skills in such circumstances can save lives. Riptides are riverlike masses of water caused by the contour of the bottom. They are bound on both sides by eddies that swing back toward shore. A swimmer should never attempt to fight a riptide but rather swim diagonally with it and to its edge, then swing into the eddy and be carried in toward shore. Undertow or backwash is simply the receding portion of the water that has been carried in by the breakers. Because of its movement outward from the shore, it is very disconcerting. Fortunately the water travels in two directions at two levels. The upper level travels inward and the swimmer should push off the bottom and emerge at the surface and go with the breakers toward shore.



ILLUSTRATIONS FROM OFFICIAL A.A.U. SWIMMING HANDBOOK BY COURTESY OF AMATEUR ATHLETIC UNION OF UNITED STATES

FIG. 8.—HALF-TWIST, FORWARD DIVE (RUNNING)

Tides may cause currents running diagonally toward the beach, and the best procedure is to swim across or with them until shallow water is reached.

Due to the increased popularity of such activities as sailing, boating, skin diving, underwater spear fishing, water skiing and surfboarding, more people are taking to the open water as a form of recreation. A large percentage of drownings could be prevented if each person exercised caution and gained a thorough understanding of his own abilities before engaging in these sports.

For Olympic records in all aquatic events *see* OLYMPIC GAMES. *See* also DROWNING AND LIFESAVING; WATER POLO.

BIBLIOGRAPHY.—The literature on the subject of swimming is considerable; the most useful work of general reference is Ralph Thomas' (1904), which contains a bibliography from the earliest time. Other chief works on the technique of swimming that may be mentioned are: Thevenot, *The Art of Swimming* (1789); Steedman, *Manual of Swimming* (Melbourne, 1867); W. Wilson, *The Swimming Instructor* (1883); A. Sinclair and W. Henry, *Swimming* (1893); C. M. Daniels, L. de B. Handley and O. Wahle, *How to Swim and Save Life* (1907); H. R. Austin, *How to Swim* (1914); L. de B. Handley, *Swimming and Watermanship* (New York, 1918); R. J. H. Kiphuth, *Swimming* (New York, 1942); Frances A. Greenwood, *Bibliography of Swimming* (New York, 1940); T. K. Cureton, *How to Teach Swimming and Diving* (New York, 1934); U.S. Naval Institute, *Swimming* (Annapolis, Md., 1944); R. J. H. Kiphuth and H. M. Burke, *Basic Swimming* (New Haven, Conn., 1950; London, 1951); John A. Torney, Jr., *Swimming* (New York, London, 1950); D. A. Armbruster and L. E. Morehouse, *Swimming and Diving*, 2nd ed. (St. Louis, 1950); Max Madders (ed.), *Swimming: the Official Coaching Book of the E.S.S.A.* (London, 1951); *Handbook of the Amateur Swimming Association* (annually); *Amateur Athletic Union of the United States Official Swimming, Water Polo and Diving Rules* (annually); National Collegiate Athletic Association, *Official NCAA Swimming Guide* (annually).

(L. DE. B. H.; R. J. H. K.; K. B. ML.)

SWINBURNE, ALGERNON CHARLES (1837–1909), English poet and critic, was born in London on April 5, 1837. He was the son of Admiral Charles Henry Swinburne (of an old Northumbrian family) and of Lady Jane Henrietta, a daughter of George, 3rd earl of Ashburnham. It may almost be said to have been by accident that Swinburne owned London for his birthplace, since he was removed from it immediately, and always felt a cordial dislike for the surroundings and influences of life in the heart of a great city. His own childhood was spent in a very different environment. His grandfather, Sir John Edward Swinburne, bart., owned an estate in Northumberland, and his father, the admiral, bought a beautiful spot between Ventnor and Niton in the Isle of Wight, called East Dene, together with a strip of undercliff known as the Landslip. The two homes were in a sense amalgamated. Sir Edward used to spend half the year in the Isle of Wight, and the admiral's family shared his northern home for the other half; so that the poet's earliest recollections took the form of strangely contrasted emotions, inspired on the one hand by the bleak north, and on the other by the luxuriant and tepid south. Of the two, the influences of the island are, perhaps naturally, the stronger in his poetry; and many of his most beautiful pieces were actually written at the Orchard, an exquisite spot by Niton bay, which belonged to relatives of the poet, and at which he was a constant visitor.

After some years of private tuition, Swinburne was sent to Eton, where he remained for five years, proceeding to Balliol college, Oxford, in 1857. He was three years at the university, but left without taking a degree. Clearly he must have cultivated while there his passionate and altogether unacademic love for the literature of Greece; but his undergraduate career was unattended by university successes, beyond the Taylorian prize for French and Italian, which he gained in 1858. He contributed to the "Underground Papers," published during his first year, under the editorship of John Nichol, and he wrote a good deal of poetry from time to time, but his name was probably regarded without much favour by the college authorities. He took a second class in classical moderations in 1858, but his name does not occur in any of the "final" honour schools. Jowett was his firm friend after he left Oxford, and Swinburne visited him repeatedly at the seaside and in the country.

He left Oxford in 1860, and in the same year published the two dramas, *The Queen Mother* and *Rosamond*, a wonderful performance for so young a poet, fuller of dramatic energy than most of his later plays, and rich in harmonies of blank verse.

After spending some time with the Bell Scotts at Newcastle, Swinburne came to London, and began the intimate friendship with D. G. Rossetti which was to last for the next ten years. In 1861 he was introduced to Monckton Milnes (Lord Houghton, *q.v.*), who took an interest in his work and invited him to Fryston. There Swinburne created an unforgettable impression in the brilliant circle which gathered there, by his wide and varied acquaintance with ancient and modern literature and his astonishing memory. In 1862 he met Meredith, and a review of one of Victor Hugo's books resulted in correspondence. In the autumn of that year he lived with Meredith and the Rossettis at 16 Cheyne Walk. Swinburne showed symptoms of something approaching epilepsy, and it was thought better that he should be with friends. In 1864 he made a tour in France and Italy. At Florence he met Mrs. Gaskell and Walter Landor. For Landor he had a great admiration, but their meetings were few and short. On his return to London he took lodgings in Dorset street, where he made his home for many years.

In 1865 appeared the lyrical tragedy of *Atalanta in Calydon*, followed by *Chastelard*, and the famous *Poems and Ballads* in 1866. The *Poems and Ballads* were vehemently attacked, but *Dolores* and *Faustine* were on everyone's lips: as a poet of the time has said, "We all went about chanting to one another these new, astonishing melodies." By the close of his 30th year, in spite of hostility and detraction, Swinburne had placed himself in the highest rank of contemporary poets, and had even established himself as leader of a choir of singers to whom he was at once master and prophet.

Meanwhile, his life was disturbed by domestic sorrow. A favourite sister died at East Dene, and was buried in the churchyard of Bonchurch. Her loss overwhelmed Admiral Swinburne with grief, and he could no longer tolerate the house that was so

full of tender memories. So the family moved to Holmwood, in the Thames Valley, near Reading. Swinburne entered on a course of gaiety in London which seriously affected his health and alarmed his friends. At the same time he saw much of the Pre-Raphaelite movement, and for the next few years he was involved in a rush of fresh emotions and rapidly changing loyalties.

Works.—It is indeed necessary to any appreciation of Swinburne's genius that one should understand that his inspiration was almost invariably derivative. His first book is deliberately Shakespearian in design and expression; the *Atalanta,* of course, is equally deliberate in its pursuit of the Hellenic spirit. Then, with a wider swing of the pendulum, he recedes, in *Poems and Ballads,* to the example of Baudelaire and of the Pre-Raphaelites themselves; with the *Song of Italy* (1867) he is drawing towards the revolt of Mazzini (whom he had met at the house of George Howard, afterwards earl of Carlisle); by the time *Songs before Sunrise* are completed (in 1871) he is altogether under the influence of Victor Hugo, while Rome has become to him "first name of the world's names." But, if Swinburne's inspiration was derivative, his manner was in no sense imitative; he brought to poetry a spirit entirely his own, and a method even more individual than his spirit. It is well to distinguish clearly between the influences which touched him and the original, personal fashion in which he assumed those influences, and made them his own. The spirit of Swinburne's muse was always a spirit of revolution. In *Poems and Ballads* the revolt is against moral conventions and restraints; in *Songs before Sunrise* the arena of the contest is no longer only the sensual sphere, but the political and the ecclesiastical. The detestation of kings and priests, which marked so much of the work of his maturity, is now in full swing, and Swinburne's language is sometimes tinged with extravagance and an almost virulent animosity.

With *Bothwell* (1874) he returned to drama and the story of Mary Stuart. The play has fine scenes and is burning with poetry, but its length transcends all possibilities of harmonious unity. *Erechtheus* (1876) was a return to the Greek inspiration of *Atalanta;* and then in the second series of *Poems and Ballads* (1878) the French influence is seen to be at work, and Victor Hugo begins to hold alone the place possessed, at different times, by Baudelaire and Mazzini. At this time Swinburne's energy was at fever height; in 1879 he published his eloquent *Study of Shakespeare* and *Poems and Ballads* (second series) and in 1880 no fewer than three volumes, *The Modern Heptalogia,* a brilliant anonymous essay in parody, *Songs of the Springtides,* and *Studies in Song.* It was shortly after this date that Swinburne's friendship for Theodore Watts-Dunton (then Theodore Watts) grew into one of brotherly intimacy. After 1880 Swinburne's life remained without disturbing event, devoted entirely to the pursuit of literature in peace and leisure.

The conclusion of the Elizabethan trilogy, *Mary Stuart,* was published in 1881, and in the following year *Tristram of Lyonesse,* a wonderfully individual contribution to the modern treatment of the Arthurian legend, in which the heroic couplet is made to assume opulent, romantic cadences of which it had hitherto seemed incapable. Among the publications of the next few years must be mentioned *A Century of Roundells,* 1883; *A Midsummer Holiday,* 1884; and *Miscellanies,* 1886. The current of his poetry, indeed, continued unchecked; and though it would be vain to pretend that he added greatly either to the range of his subjects or to the fecundity of his versification, it is at least true that his melody was unbroken, and his resplendent torrent of words inexhaustible. His *Marino Faliero* (1885) and *Locrine* (1887) have passages of power and intensity unsurpassed in any of his earlier work, and the rich metrical effects of *Astrophel* (1894) and *The Tale of Balin* (1896) are inferior in music and range to none but his own masterpieces. In 1899 appeared his *Rosamund, Queen of the Lombards;* in 1908 his *Duke of Gandia;* and in 1904 was begun the publication of a collected edition of his poems and dramas in 11 volumes.

Later Years.—Meanwhile, in the autumn of 1879 the state of his health seriously alarmed his friends, and, with the consent of Lady Jane Swinburne, Watts-Dunton took him to live at his house, The Pines, Putney, where the last 30 years of his life were spent in complete retirement. He very rarely made the short journey into London, although his health was gradually restored. The friends visited Paris in 1882. In 1888 a flash of his old excitability involved him in a quarrel with Whistler.

Besides this wealth of poetry, Swinburne was active as a critic, and several volumes of fine impassioned prose testify to the variety and fluctuation of his literary allegiances. His *Note on Charlotte Brontë* (1877) must be read by every student of its subject; the *Study of Shakespeare* (1880)—followed in 1909 by *The Age of Shakespeare*—is full of vigorous and arresting thought, and many of his scattered essays are rich in suggestion and appreciation. His studies of Elizabethan literature are, indeed, full of "the noble tribute of praise," and no contemporary critic did so much to revive an interest in that wonderful period of dramatic recrudescence, the side-issues of which have been generally somewhat obscured by the pervading and dominating genius of Shakespeare. His prose work also includes an early novel in the form of letters, *Love's Cross-currents,* which had appeared serially in the Tatler and was revised for publication in 1905.

Place in Literature.—The service which Swinburne rendered to the English language as a vehicle for lyrical effect is simply incalculable. He revolutionized the entire scheme of English prosody. Nor was his singular vogue due only to this extraordinary metrical ingenuity. The effect of his artistic personality was in itself intoxicating, even delirious. He was the poet of youth insurgent against all the restraints of conventionality and custom. The young lover of poetry, when first he encounters Swinburne's influence, is almost bound to be swept away by it; the wild, extravagant license, the apparent sincerity, the vigour and the verve, cry directly to the aspirations of youth like a clarion in the wilderness. But, while this is inevitable, it is also true that the critical lover of poetry outgrows an unquestioning allegiance to the Swinburnian mood more quickly than any other of the diverse emotions aroused by the study of the great poets. It is impossible to acquit his poetry entirely of the charge of an animalism which wars against the higher issues of the spirit—an animalism sometimes of love, sometimes of hatred, but, in both extremes, out of centre and harmony.

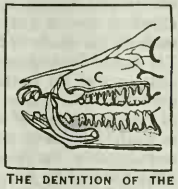
Yet, when everything has been said that can be said against the unaesthetic violences of the poet's excesses, his service to contemporary poetry outweighed all disadvantages. No one did more to free English literature from the shackles of formalism; no one, among his contemporaries, pursued the poetic calling with so sincere and resplendent an allegiance to the claims of absolute and unadulterated poetry. Some English poets have turned preachers; others have been seduced by the attractions of philosophy; but Swinburne always remained an artist absorbed in a lyrical ecstasy, a singer and not a seer. His personality was among the most potent of his time, and his artistic influence was both inspiring and beneficent. He died at Putney on April 10, 1909.

See R. H. Shepherd, *The Bibliography of Swinburne* (1889); Theodore Wratislaw, *Algernon Charles Swinburne, a Study* (1900); G. E. Woodberry, *Swinburne* (1905); Edmund Gosse, article in *Dict. Nat. Biog.,* 2nd supp., 1901–11; Edward Thomas, *Algernon Charles Swinburne, a Critical Study* (1912); John Drinkwater, *Swinburne, an Estimate* (1913); Clara J. Watts-Dunton, *The Home Life of Swinburne* (1922); J. V. Nash, *The Religious Life of Swinburne* (1923); Paul Dottin, *Swinburne et les dieux* (1925); J. W. Mackail, *Studies of English Poets* (1926); H. Nicolson, *Swinburne* (English Men of Letters Series, 1926); G. Lafourcade, *La jeunesse de Swinburne, 1837–67* (1928); S. C. Chew, *Swinburne* (1929). (E. G.; X.)

SWINDON, a municipal borough and parliamentary division of north Wiltshire, England, 78 mi. W. of London by road. Pop. (1951) 68,932. Area 9.9 sq.mi. Until 1841 it was a small market town (now Old Swindon) with a population of less than 2,000, but after the Great Western railway had built its main engineering works to the north the population rapidly increased to 45,000, and in 1900 it received its charter. More than 21,000 people are employed in the railway works and other industries include engineering, clothing and tobacco. Under the Town Development act, 1952, provisional approval was obtained from the ministry of housing and local government to accept overspill population and industry from the Greater London area. Planning approval was

given to expand the population to 92,000 within the existing borough boundaries and a 75-ac. trading estate has been developed close to the Swindon-Gloucester railway. In 1947 the first municipally owned arts centre in the country was opened, combining a small concert hall, a stage and an art gallery.

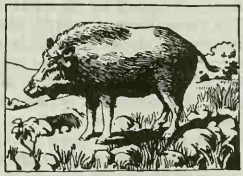
SWINE, a name applied to the domestic pig, but also used to include its wild relatives. The animals constitute the family Suidae (*see* ARTIODACTYLA). The swine are found only in the old world. They are characterized by elongated head and mobile snout, with an expanded, truncate, terminal surface in which the nostrils open; narrow feet with four toes, the outer pair not reaching the ground when walking. The canines in both jaws curve upward, forming large tusks, best developed in boars. The hair is coarse and bristly, while the skin is thick, underlaid by thick fat. The tail is moderately long, generally tufted.



THE DENTITION OF THE BOAR. 12 INCISORS, 4 CANINES, 16 PREMOLARS AND 12 MOLARS: TOTAL 44

The typical genus *Sus* is exemplified by the domesticated pigs (*see* PIG) and the wild boar. (*See* BOAR.) Allied to these are the pygmy hog of Nepal (*Porcula salvania*), a little larger than a hare; and the African river hogs (*Potamochoerus*).

Linking these genera with the wart-hogs is the African *Hylochoerus*, a forest-dwelling form with a thick coat of coarse black hair. The remaining forms are the wart-hogs (*q.v.*) (*Phacochoerus*), and the babirusa (*q.v.*). Extinct swine, some of gigantic size, are know from strata as far back as the Pliocene of Europe and Asia; many of these belonged to the genus *Sus*.



THE WILD BOAR (SUS SCROFA)

SWINE FEVER (HOG CHOLERA) is a devastating acute general infection of swine, caused by a virus, in which high fever and prostration are outstanding symptoms.

Natural infection results from immediate or mediate contagion. The disease may be acquired by healthy herds by the introduction of infected pigs. Exposure of swine in markets in infected areas is a common source of infection; so are vehicles and trucks in which pigs are conveyed from place to place; dealers, castrators who journey from farm to farm and attendants may serve as agencies whereby contagion is conveyed. Outbreaks are likely to recur in premises where the disease has existed previously, as the virus appears to retain its vitality for fairly long periods outside the animal body.

Since the virus may persist in uncooked garbage, Canada and most of the states of the U.S. require that garbage fed to swine be sterilized by cooking.

After an incubation period of four days to three weeks after exposure, the disease is ushered in by fever. Subsequent symptoms vary somewhat with the form the disease assumes. The animal has no appetite, is depressed, ill and separates itself from its companions; the conjunctivas are reddened and congested and may secrete a mucopurulent discharge; vomiting is common, and constipation and diarrhea are both met with; sometimes the feces become dysenteric and the odour is peculiar and disagreeable.

In many cases a skin rash develops, the parts principally involved being the region of the ears, the axilla, groin and the skin covering the abdominal area. The mucous membrane lining the mouth, fauces and pharynx may become inflamed; later ulcers form and become covered with a diphtheritic exudation. As a result the animal is unable to feed and respiration is difficult. The lungs may be the seat of pneumonic changes.

When this is the case the animal then has a short, dry and paroxysmal cough, a nasal discharge, difficulty in respiration and the other symptoms characteristic of acute lung infection. Gastritis and enteritis are common. The animal lies about and if compelled to move does so reluctantly and sometimes with a staggering gait and an arched back; later it is unable to rise and becomes comatose.

As in other acute septicemic diseases, a hyperacute form may develop and death supervene within a few days; or a less acute form may set in, with one or more groups of organs becoming involved.

A chronic type is also encountered, in which life drags on over a considerable period; the animal becomes more debilitated and emaciated and it remains a possible source of infection to other pigs.

Recovery occurs in some cases, the virulence and the mortality varying widely in different outbreaks.

Swine fever has a wide distribution throughout Europe, North America and Africa and is not uncommon in the British Isles. Canada controls the disease by slaughter of all hogs on infected premises, the cleansing and disinfection of premises and the serum treatment of all hogs in the vicinity of an outbreak. In the United States the average annual loss from swine fever, including vaccination costs, was estimated at mid-20th century to be from $20,000,000 to $30,000,000. It is the cause of 90% of deaths in swine.

The great toll that swine fever exacted in the United States stimulated efforts to develop vaccines. The first active vaccination method used, dating from 1907, was the simultaneous method, using active virus simultaneously with hyperimmune serum. By the early 1950s, about 90% of all hog vaccinations in the United States were by the simultaneous method. However, a strong movement developed against the use of virulent virus.

Modified and killed virus vaccines, neither of which could cause the disease, came to replace those made with active virus almost entirely, and eventual eradication of swine fever seemed possible. (A. R. S.; NA. BR.; X.)

SWINOUJSCIE (Ger. SWINEMÜNDE), a fishing port and seaside resort in the province of Koszalin, Poland, at the east end of the island of Usedom, and on the Swine river which connects the Stettiner Haff with the Baltic.

Swinemünde, which was founded in 1748, was fortified and raised to the dignity of a town by Frederick the Great in 1765. It was an important German naval base during World War II. About 70% of the city was destroyed during the war. Its population declined from 30,056 in 1939 to 5,771 in 1946. Population (1960) 16,900. It was transferred to Poland in 1945.

Manufactures include lumber, furniture and shipbuilding.

SWINTON, PHILIP CUNLIFFE-LISTER, 1st EARL OF (1884–), British statesman, was born on May 1, 1884, at Ayton, Yorkshire. The youngest son of Col. Lloyd-Greame, he assumed by royal warrant the name of Cunliffe-Lister in 1924. He was educated at Winchester and University college, Oxford, and was called to the bar in 1908.

During World War I he rose to the rank of major in the king's royal rifle corps. Elected Unionist member for Hendon in 1918, he was chairman of the labour committee of the war cabinet. He was parliamentary secretary to the board of trade and of the overseas trade department in the Lloyd George government; president of the board of trade, 1922–23, 1924–29 and 1931; colonial secretary, 1931–35; secretary of state for air, 1935–38; and minister for civil aviation, 1944–45.

He was created Viscount Swinton in Nov. 1935, and after he had served as secretary of state for commonwealth relations in the Churchill government 1952–55, was, at his retirement, created Baron Masham, of Ellington, and earl of Swinton.

SWINTON, an urban district in the West Riding of Yorkshire, England, 10½ mi. (17 km.) NE of Sheffield by road, at the junction of the Don navigation with the Dearne and Dove navigation. The population of Swinton (1961) was 13,420. Area 2.7 sq.mi. (7.0 sq.km.).

Formerly renowned for its Rockingham ware, it has coal mines, iron foundries, steel and glass works.

SWINTON AND PENDLEBURY, a municipal borough of Lancashire, England, 5 mi. NW of Manchester, contiguous with Salford. Pop. (1961) 40,470. Area 5.3 sq.mi. A great part of the Clifton rural area was added in 1933. Cotton spinning and weaving, coal mining, the manufacture of electric batteries, and

engineering are the chief industries. The town hall was opened in 1938. Anciently parts of Swinton were held by the Knights of the Order of the Hospital of St. John of Jerusalem. Swinton and Pendlebury, incorporated in 1934, forms with Eccles the Eccles parliamentary division of Lancashire, returning one member.

SWISS LITERATURE. From the point of view of language and literature the various nationalities of the Swiss confederation project into those of the neighbouring countries—Germany (Austria), France and Italy. This unusual aspect becomes particularly significant in relation to literary production before the existence or outside the boundaries of the *Ewiger Bund* (the Everlasting league) of 1291. Literatures produced exclusively in Switzerland are those written in the Raeto-Romance language or in one of the numerous varieties of local Swiss dialects. Latin literature, of course, chronologically precedes all the others. In the following survey all these literatures are considered both on their own account and as parts of a Swiss literature in an over-all sense.

Latin Swiss Literature.—The oldest document relating to what is now Switzerland was not produced within its territory, for the first notable description of Helvetia and the Helvetians is to be found in the first book of Caesar's *Commentarii*, commonly referred to as *Bellum Helveticum*. After the humanists rediscovered this text, it gradually rose in the historical consciousness of the Swiss to the rank and dignity of a Great Charter. Among the Latin chronicles of the middle ages the *Casus Sancti Galli*, written by Ekkehart IV in the 11th century, gives an impressive description of the highly developed culture of the abbey of St. Gall, founded by Irish Benedictines. Another chronicle, the *Johannis Vitodurani Chronicon*, contains the world picture as seen by a Franciscan of Winterthur in the first part of the 14th century. The Swiss humanists who expressed themselves in Latin devoted most of their prose and verse to a description of their own country; *e.g.*, Albrecht von Bonstetten, dean of the abbey of Einsiedeln, in his *Superioris Germanie Confoederationis descriptio* (1479) and Heinrich Loriti Glareanus in his *Helvetiae descriptio* (1514). An important position in the history of general linguistics is taken up by the Zürich polyhistorian Konrad Gesner, who in his *Mithridates* (1555) undertook a survey of all languages known to him. For three centuries to come the two volumes entitled *De Republica Helvetiorum* (1576) by the Zürich clergyman Josias Simler remained the standard work in the constitutional law and politics of the older Swiss confederation.

The final phase in Latin Swiss literature belongs to the Age of Reason and is largely connected with natural science. An outstanding example of it is the description of Switzerland given by the city physician of Zürich, J. J. Scheuchzer, and entitled *Uresiphoites helveticus sive itinera alpina tria* (1708). It was printed in London as a publication of the Royal society with the imprimatur of Isaac Newton.

German Swiss Literature.—In bulk, the most important literature of Switzerland is written in the German language. It first flourished in the age of the Minnesingers, of whom more than 30, *i.e.* practically one-fifth of the total number, lived on what is now Swiss territory. The identity of the Codex Manesse with the famous manuscript of that name in Heidelberg has traditionally been accepted. The Codex Manesse is an example of how specimens of prenational literature may be incorporated in a national literature: in the 18th century J. J. Bodmer was the first to have the Codex printed, and in the 19th century Gottfried Keller made it the subject of one of his *Züricher Novellen*. In this way the poetry of the Minnesingers became part of the history of Swiss letters.

The first literary productions with national characteristics were popular ballads and popular chronicles. The popular ballads commemorate especially the great battles fought by the Swiss from the 14th to the 16th century and later on. The most effective and influential of these is the ballad of William Tell, which through its moral and political aspects maintained an important position as a narrative and a play from the late 15th century onward. The popular chronicles, occasionally written on behalf of some local authority and often illustrated, flourished especially in Berne, Lucerne and Zürich. The most famous chronicle, however, was written by Giles (Aegidius) Tschudi, a leading magistrate of Glarus. The contents of his *Chronicon Helveticum* was widely known after its completion in the 16th century, although it was only printed in the years 1732 and 1734 at Basel. In it Johann von Schiller found the material for his drama *Wilhelm Tell* and Walter Scott the text for his translation of the ballad "The Battle of Sempach."

The reformation activities of Zwingli had only an indirect influence on literature. Zwingli himself wrote mainly in Latin. The so-called Zürich Bible was soon replaced by the Luther Bible, which actually made for a closer connection between the writings produced in Switzerland and in Germany. One of the most important results of the work done by Zwingli was that the Protestant majority of German-speaking Switzerland established a permanent connection with the Protestant parts of western Switzerland and with Protestant countries abroad. Thus one of the earliest Swiss accounts of England was produced by the Zürich clergyman and lexicographer Josua Maler (*c.* 1550), and in the 18th century the first German translations of Shakespeare were published in Zürich.

The beginning of 18th-century literature in German-speaking Switzerland is marked by the works of a Bernese poet and the end by those of a historian from Schaffhausen. The poetry of Albrecht von Haller, especially *Die Alpen* (1728), helped to bring about a new awareness in poetic experience in Germany and belongs to the earliest specimens of German poetry appreciated in France. Johannes von Müller's *Geschichten schweizerischer Eidgenossenschaft* (1786–1808), though a fragment, became the most representative piece of historical writing in German classicism and had a far-reaching influence on the style of historical writing in general. Between these two cornerstones appears what is occasionally referred to as the Zürich school of criticism. Two members of the "Carolinum," J. J. Bodmer and J. J. Breitinger, opposing the critical views of French rationalism and classicism, strongly promoted the reception and appreciation of English literature from John Milton to Thomas Percy in Germany. In this connection a revival of religious and national conceptions was prepared which had its special effect in the succeeding generations. Salomon Gessner had a European success with his revival of the idyll; for predominantly moral reasons J. K. Lavater turned to physiognomy and became one of its innovators; J. C. Hirzel surprised the reading public with the description of the village Socrates Kleinjogg. In importance, however, they were surpassed by Johann Heinrich Pestalozzi, who not only wrote the first village story in German, entitled *Lienhard und Gertrud* (1781), but included in it the nucleus of his later meditations, covering the entire field of the nature and destiny of man. It is indicative of the profound impression his ideas made on his contemporaries that his private institute of education at Yverdon was visited by philanthropists from all over Europe including, for instance, Robert Owen.

In the 19th century international fame and critical estimate at home did not always concur. Thus *The Swiss Family Robinson* (1812–27) by Johann David Wyss, and *Heidi* (1881) by Johanna Spyri are examples of a world-wide success though their purely literary merits were never considered very high at home. From the first to the second half of the 19th century the significance of German Swiss literature gradually increased. Berne produced its greatest novelist in the person of Jeremias Gotthelf (pseudonym for Albrecht Bitzius), whose epic power was combined with a profound knowledge of human nature. The series of his great novels, most of which describe village life, opened with the *Bauernspiegel oder Lebensgeschichte des Jeremias Gotthelf* (1837) and ended in the year of his death (1854) with *Erlebnisse eines Schuldenbauers*. Fiction and poetry are the fields in which the Zürich writers Gottfried Keller and Conrad Ferdinand Meyer achieved fame. Both realized their own possibilities and destinies comparatively late, a phenomenon frequently observed in the history of Swiss thought. Gottfried Keller reached the first peak of his creative powers with *Der grüne Heinrich* in 1854 when he was 35 and Meyer with his verse epic *Huttens letzte Tage* in 1871

when he was 46. Meyer drew upon the mountains and the princely courts of the past, and as a writer of fiction he succeeded in giving the Renaissance cult original expression. Keller in some of his best stories devoted himself to subjects corresponding to those of Cervantes and Shakespeare and brought the art of the Novelle to perfection. The love he had for his country appears in all shades from hymns of praise to pungent social criticism. His understanding of the human soul is seen at its height in the description of odd characters of both sexes and of unforgettably charming young women. His prose is both homely and delicate and is matched by the quality of his fundamentally human attitude. Many of his admirers believe that his work is the one real contribution of Switzerland to world literature. The importance of this period is emphasized by the writings of the Basel historian Jakob Burckhardt, whose statements about the past, the present and even the future have since become increasingly impressive and valid. His greatest disciple was Heinrich Wölfflin, who continued and completed that work of his master which was interrelated with the history of fine arts. The last poet of international repute was Carl Spitteler, who was distinguished for epics in verse and prose.

French Swiss Literature.—Though the French-speaking part of the Swiss population is only one-fifth of the total, the literature it has produced carries considerably more than the proportionate weight. This is due partly to the immigration of a number of highly talented Huguenot families, partly, though to a lesser degree, to the fact that occasionally writers born in the German-speaking part of the country would decide to express themselves in the French idiom. This applies especially to the Bernese. The beginnings of French Swiss literature proper coincide with the Reformation. Jean Cauvin from northern France fashioned Geneva into the *"cité de Calvin."* From Geneva he exercised his influence on French literature through his diction and on the Anglo-Saxon world through his theology. In this connection it is significant that one of the first important interpreters of England in France was the Bernese B. L. von Muralt, whose stay in England in 1694 resulted in his *Lettres sur les Anglais et les Français* (1725). His work was admired by J. J. Rousseau, the father of continental romanticism. The formative influences at work on the Genevese philosopher have usually been connected with the comparatively narrow strip of country between the Alps and the Jura. He was, however, not only a *"citoyen d'un état libre,"* as he described himself in the opening chapter of the *Contrat social*, but obviously shared the political and psychological conceptions of all Switzerland to such a degree that one is tempted to consider him its most brilliant publicity agent (*see*, for example, his *Considérations sur le gouvernement de Pologne*). Obviously Swiss subject material was dealt with by H. B. de Saussure, one of the founders of modern geology, in his *Voyages dans les Alpes* (1779–96).

At the beginning of the 19th century, however, French Swiss literature on the whole preferred international themes and was largely connected with the influential circle of Mme. de Staël, who was born in Paris but was of German and Swiss extraction. Of her books setting forth the qualities of the various nations of Europe the one entitled *De l'Allemagne* had the widest influence. Her friend Benjamin Constant, though less prolific, expressed himself all the more pointedly; his adaptation of Schiller's trilogy *Wallenstein* contains the memorable essay *Sur le théâtre allemand*. Another member of the circle of Mme. de Staël, who in her manor house on the Lake of Geneva entertained a great many of the leading intellectuals of western Europe, was J. C. L. S. de Sismondi, who made an important contribution to Italian history with his *Histoires des républiques italiennes du moyen-âge* (1807–18). Further there was C. V. de Bonstetten, fundamentally a European in outlook, especially in his book *L'homme du midi et l'homme du nord* (1824).

To a later development belong the humorous tales and sketches of Rodolphe Toepffer, depicting a more homely world (*La bibliothèque de mon oncle*, 1832), while the poems of Juste Olivier testify to a pronounced patriotism (*Poèmes suisses*, 1830). The criticism of Alexandre Vinet, based on an essentially personal interpretation of Christianity, found an echo both in Switzerland and in France. Toward the end of the 19th century a powerful international influence was exercised by the posthumous publication of the fragments of H. F. Amiel's *Journal intime* (1883), which caused Tolstoy to write a moving preface for its Russian translation. In his historical writings Philippe Monnier, in a manner reminiscent of Jakob Burckhardt, deals with Italy in the 15th and 18th centuries (*Le Quattrocento*, 1901, *Venise au 18. siècle*, 1907). C. F. Ramuz wrote strongly imaginative novels about life amongst vinegrowers and Alpine herdsmen (*Oeuvres complètes*, 1940–). Gonzague de Reynold, with a consistency hardly known before him, demonstrated the inward connection between German and French-speaking Switzerland in the 18th century and depicted the whole of his country in his *Cités et pays suisses*.

Italian Swiss Literature.—Only about 4% of the total population of Switzerland use Italian as their native language. Their home is in the Ticino country and in some of the southern Grisons valleys. It is thus unlikely that the literature of such a small group could exist without a strong and permanent relation to Italian literature proper. Since the 18th century there have been a few nationally distinct features in what is now often referred to as the Swiss Italianità. The themes of these minor writings devoted to poetry and fiction, natural science and history, economics and politics are fairly characteristic. As an exponent of the Age of Reason, Father Francesco Soave of Lugano translated John Locke and Salomon Gessner into Italian and, following the example of the latter, wrote his own *Novelle Morali* (1782), which since then have appeared in hundreds of editions and have also made the story of William Tell popular in Italy. A representative figure in the first half of the 19th century is Stefano Franscini, who in theory and practice helped to bring about the new unity of the Ticino country and to connect it with the spirit of the Swiss confederation. His main work, *La Svizzera Italiana* (1837–38), bears as a motto a quotation from Albrecht von Haller: "Alpibus quidem ad Italiam spectantibus ego plurimum boni spero" (The Alps looking toward Italy are my greatest hope). His *Nuova statistica della Svizzera* (1847) proved to be a particularly effective description of Switzerland in that period.

So far, undeniably the most outstanding poet and novelist of the Ticino country has been Francesco Chiesa. A convinced regionalist he combines a profound love of his country with an admiration for the great Lombard poets and thus creates the high literary standards necessary for the Italian-speaking part of Switzerland. His vast influence on his contemporaries has essentially contributed towards keeping the Italianità of his region intact (*Calliope*, 1907; *I viali d'oro*, 1911; *Tempo di marzo*, 1925). One of his disciples equally bent on exploring and interpreting his home country was Giuseppe Zoppi (*Leggende ticinesi*, 1928, and *Presento il mio Ticino*, 1933).

Raeto-Romance Literature.—The Raeto-Romance language was formerly spoken far beyond the boundaries of Switzerland; today it is only used in a few Grisons valleys. A treatise, published by the Grisons historian and philologist Joseph von Planta in the *Philosophical Transactions of the Royal Society of 1776*, made the existence of the language known to a wider public. Only recently, however, was it officially recognized. On Feb. 20, 1938, following a plebiscite of the whole country, Raeto-Romance, though represented by only 1% of the total population, was declared one of the four national languages of the Swiss confederation. Raeto-Romance literature was in its origins predominantly ecclesiastical and began in the middle ages. The Reformation gave it new impulses. There is a rich variety of popular songs, especially of the political kind. Owing to its geographical distribution Raeto-Romance literature is essentially regional in character. Nevertheless the anthologist Caspar Decurtins and the poet Peider Lansel achieved decidedly more than mere local fame in the 20th century.

Swiss Dialect Literature.—The fortunes of Swiss dialect literature vary according to the language situation in each respective section of the country. In the Ticino country dialect literature is being produced with much devotion, but so far seems to have been of little consequence. In the French-speaking part of the country the local dialects are on the wane and, consequently, so are dialect writings. The important instances of this sort of literature belong to past centuries such as the different Genevese ballads

commemorating the victory of the *Escalade* in 1602. International fame was achieved by the various *Ranz des vaches,* melodies sung or played on the horn by herdsmen and in one case taken down by Rousseau in his *Dictionnaire de musique* (1768).

At present dialect literature flourishes mainly in the German-speaking part of the country, chiefly because there the population, regardless of social rank and education, consistently sticks to the use of the dialect for everyday purposes. The existence of numerous local idioms might even produce an ever increasing variety of dialect writings. This, however, would be at cross purposes with the accepted determination of the writers and their reading public to remain on common ground with German literature as a whole, which explains why some of the best poets express themselves both in high German and in their dialect. Thus Adolf Frey published a volume of poems in the dialect of the Aargau (*Duss und underm Rafe,* 1891 and 1921) and Meinrad Lienert one in the dialect of Schwyz (*'s Heiwili,* 1908). There are vigorous novels in the Bernese dialect by Rudolf von Tavel, who began in 1907 with *Jä gäll, so geit's,* and by Simon Gfeller, who depicted life among the peasants of the Emmenthal beginning with *Heimisbach* (1911).

Patrimonium Helveticum.—Without detriment to their connections with European civilization many Swiss consider the existence of a sixfold national literature with deep satisfaction. They are perfectly aware of the perplexing problems it inevitably involves and even look upon them as an asset. Between the extremes of absolute diversity and absolute unity there is room for all aspects and expressions of thought and imagination. This is what the Swiss spirit of freedom fundamentally wills. A representative cultural centre—such as Florence once was for Italy and Paris still is for France—is inconceivable for Switzerland. Each group and even each individual expresses itself according to its character and region. Such a state of affairs may have its greatness, but it certainly also has its drawbacks and its dangers, which can only be overcome on the level of a vigorous national consciousness. So far this consciousness has proved adequate for the solution of the problems involved. A deteriorating structural change could only occur if the minorities were estranged from the confederate whole or if the majority forces were tempted to exploit their position regardless of tradition.

Although there is at present a trend for a slight shift in the numerical relationship of the different languages one cannot seriously speak of a peril to the well-established balance of the different cultures. On the contrary the fact that the underlying ideals are very much alive may produce further high lights of national integration. A development of over 1,000 years is at the basis of the literary life of Switzerland and will continue to shape what might be called an ever-growing *patrimonium Helveticum.* (F. E.; H. Sn.)

Contemporary Writers.—Swiss contemporary authors have found themselves inescapably confronted with two world wars, totalitarian revolutions, the triumphs of technical civilization and a growing disintegration of the great traditional values.

A considerable minority stick to the old values still current in the village and the small town, praising the good gifts of the earth (Heinrich Federer, 1866–1928; and Alfred Huggenberger). Others furnish antidotes to the conventional order of things and are alert to the stimuli of international modern thought. After 1914 Albert Steffen, Jakob Schaffner (1875–1944), Robert Faesi and others opened broader and deeper vistas of the individual and of nations. Marxist views are expressed by Hans Mühlestein and Jakob Bührer. On the other hand the complicated, highly strung modern individual preoccupies novelists like Monique Saint-Hélier and Kurt Guggenheim.

A wave of irrationalism has set free new forces in Swiss poetry. In some writers, a refined sense of form (Siegfried Lang) and the bold use of intuitive imagery (Albin Zollinger, 1895–1941, Werner Zemp, Silja Walter) are reminders of the nearness of C. G. Jung, the Swiss explorer of the unconscious. Religious thought in this epoch has been greatly influenced by the theology of Karl Barth and Emil Brunner. A promising, provocative playwright is Friedrich Dürrenmatt.

With a few exceptions (*e.g.,* the Zürich dramatist Max Frisch and the young Swiss Italian Felice Filippini) existentialism has had little chance with Swiss writers so far, but many of the best believe that the truth lies in a sort of holism, an interpenetration of both the forces of the native soil and the world at large, *e.g.,* the novelists Gustav Keckeis and Cecile Ines Loos and the poet Urs Martin Strub; whereas the novelist Meinrad Inglin and the playwright Caesar von Arx (1895–1949) stress the native element. A similar sense of many-sided relationships is an essential feature of the historic and literary essays of Gonzague de Reynold, Carl J. Burckhardt, Fritz Ernst and Max Rychner, representatives of a truly European Switzerland. (A. Bx.)

Bibliography.—R. Faesi, *Anthologia Helvetica* (Leipzig, 1921) is a collection in the six languages mentioned. Anthologies based on each of the four national languages are: E. Korrodi, *Geisteserbe der Schweiz,* 2nd ed. (Erlenbach-Zürich, 1943); C. Clerc, *Le génie du lieu* (Neuchâtel, 1929); *Scrittori della Svizzera italiana,* 2 vol. (Bellinzona, 1936); C. Descurtins, *Rätoromanische Chrestomathie,* 12 vol. (Erlangen, 1896–1919).

General histories: G. de Reynold, *Histoire littéraire de la Suisse au 18 siècle* (Lausanne, 1909); E. Jenny and V. Rossel, *Geschichte der schweizerischen Literatur,* 2 vol. (Bern, 1910); C. Clerc *et al., Littératures de la Suisse* (Paris, 1938).

German Swiss literature: J. Baechtold, *Geschichte der deutschen Literatur in der Schweiz* (Frauenfeld, 1892); J. Nadler, *Literaturgeschichte der deutschen Schweiz* (Zürich, 1932); E. Ermatinger, *Dichtung und Geistesleben der deutschen Schweiz* (Munich, 1933); O. von Greyerz, *Die Mundartdichtung der deutschen Schweiz* (Swiss German dialect literature) (Leipzig, 1924).

French Swiss literature: V. Rossel, *Histoire littéraire de la Suisse romande des origines à nos jours,* 2 vol. (Geneva, 1889, 1891); P. Godet, *Histoire littéraire de la Suisse française,* 2nd ed. (Paris, 1895); *see* also C. Clerc, "La Suisse romande" in J. Bédier and P. Hazard, *Histoire de la littérature française,* 2 vol., 2nd ed. (Paris, 1948, 1949).

Italian Swiss literature: G. Zoppis, introduction to *Scrittori della Svizzera Italiana* (*see* above).

Raeto-Romance literature: *Historisch-biographisches Lexikon der Schweiz,* vol. iii, pp. 708 ff. (Neuenburg, 1926).

Contemporary trends: A. Bettex, *Die Literatur der deutschen Schweiz von heute* (Olten, 1950). *See* also F. Baldensperger and W. P. Friederich, *Bibliography of Comparative Literature* (Chapel Hill, 1950); H. Weilenmann, *Die vielsprachige Schweiz* (Basle, 1925); F. Ernst, *Helvetia Mediatrix* (Zürich, 1939) and *European Switzerland Historically Considered* (Zürich, 1951).

SWITHUN (or Swithin), **SAINT** (d. 862), bishop of Winchester and patron saint of Winchester cathedral from the 10th to the 16th century. He is scarcely mentioned in any document of his own time. His death is entered in the *Anglo-Saxon Chronicle* under the year 861; and his signature is appended to several charters in Kemble's *Codex diplomaticus.* Of these charters three belong to 833, 838, 860–862. In the first the saint signs as "Swithunus presbyter regis Egberti," in the second as "Swithunus diaconus," and in the third as "Swithunus episcopus." Hence if the second charter be genuine the first must be spurious, and is so marked in Kemble.

More than 100 years later, when Dunstan and Ethelwold of Winchester were inaugurating their church reform, St. Swithun was adopted as patron of the restored church at Winchester, formerly dedicated to St. Peter and St. Paul. His body was transferred from its almost forgotten grave to Ethelwold's new basilica on July 15, 971, and according to contemporary writers, miracles accompanied the translation.

The revival of St. Swithun's fame gave rise to a mass of legendary literature. The so-called *Vitae Swithuni* of Lantfred and Wulstan, written about A.D. 1000, hardly contain any germ of biographical fact; and the earliest detailed authority is a biography ascribed to Gotzelin (*fl.* 1058–78). From this writer, who has perhaps preserved some fragments of genuine tradition, it is learned that St. Swithun was appointed bishop of Winchester under Aethelwulf. At his request Aethelwulf gave the tenth of his royal lands to the church. He died on July 2, 862, saying that he was not to be buried within the church but outside in "a vile and unworthy place."

William of Malmesbury adds that, as Bishop Alhstan of Sherborne was Aethelwulf's minister for temporal, so St. Swithun was for spiritual matters. The same chronicler uses a remarkable phrase in recording the bishop's prayer that his burial might be "ubi et pedibus praetereuntium et stillicidiis ex alto rorantibus

esset obnoxius." This expression has been taken as indicating that the well-known weather myth contained in the doggrel lines—

> St. Swithin's day if thou dost rain
> For forty days it will remain;
> St. Swithin's day if thou be fair
> For forty days 'twill rain na mair—

had already, in the 12th century, crystallized around the name of St. Swithun; but it is doubtful if the passage lends itself by any straining to this interpretation.

The so-called lives of St. Swithun written by Wulstan, Lantfred, and perhaps others toward the end of the 10th century may be found in the Bollandist *Acta sanctorum* (July), i, 321–327; Mabillon's *Acta SS. O. B.* vi, 70, etc., vii, 628, etc.; and J. Earle's *Life and Times of St. Swithun*, 59, etc. *See* also William of Malmesbury, *Gest. reg.*, i, 150, and *De gest. pont.* 160, 167, 179; Florence of Worcester, i, 168; T. Rudborne ap. Wharton's *Anglia sacra*, i, 287; T. D. Hardy's *Cat. of mss.*, i, 513–517; J. Brand's *Popular Antiquities;* R. Chambers's *Book of Days:* Ethelwulf's *Tithe Charters*, nearly all of which refer to St. Swithun in the body of the text, may be studied in Haddon and Stubbs's *Councils*, iii, 636–645.

SWITZERLAND (Ger. SCHWEIZ; Fr. SUISSE; It. SVIZZERA; Lat. HELVETIA) is a confederation or *Eidgenossenschaft* of 22 cantons (3 of which have half-cantons) with a republican and federal constitution. It is situated in central Europe and bounded north by Germany, east by Austria and the principality of Liechtenstein (included since 1924 in the Swiss customs and monetary union), southeast and south by Italy, and southwest, west and northwest by France.

PHYSICAL GEOGRAPHY

Switzerland extends between 45° 49′ and 47° 48′ 30″ N. lat., and from 5° 58′ to 10° 30′ E. long. It forms an oval-shaped mass of which the greatest length (west-southwest to east-northeast) is 226 mi., and the greatest breadth (north to south) is 137 mi. It has little or no physical unity, consisting as it does of numerous small districts, differing widely in language, religion, ethnology and customs, bound together in a political alliance, made originally for common defense.

The nation consists of three great river valleys (Rhône, Rhine and Aar), lying to the north of the main chain of the Alps and including within their Swiss basins all the region between the Alps and the Jura. In addition, the wedge-shaped canton of Ticino runs south from the St. Gotthard massif and drains to the Po, while in the extreme east a tongue of the Grisons canton drains northeast via the Engadine, into the Danube. The Rhône and Rhine valleys are shut off from that of the Aar by the great northern ridge of the Bernese Oberland and Tödi Alps. The Aar valley is wide and undulating, but the upper Rhine and Rhône have cut deep trenches in the structural depression which separates the great parallel chains. The main chain of the Alps provides the loftiest wholly-Swiss summit (15,200 ft.) in the crowning Dufourspitze of Monte Rosa, though the Dom (14,942 ft.) in the Mischabel range, immediately north of Monte Rosa, is the highest entirely Swiss mountain mass. The highest summit in the northern parallel ridge is the Finsteraarhorn (14,026 ft.) in the Bernese Oberland, while the lowest level within the confederation is on the Lago Maggiore (646 ft.).

The geological build of the Alps has been proved to be exceedingly complex; the contorted, folded, and even overfolded recent rocks have been fractured, exposing old crystalline cores to denudation; the Jura are much less complicated in their folding and in their exposures. Much of the central hummocky plain is covered with undisturbed very recent rocks of Oligocene and Miocene age, many of which are marine deposits formed at a period when an arm of the Mediterranean spread up the present Rhône course along the outer border of the Alps as far east as Austria. These deposits are concealed in many parts of the plain by the most recent glacial and alluvial accumulations.

The present confederation drains into the North sea, Mediterranean sea (direct), Adriatic sea and Black sea which, respectively, receive the Rhine, Rhône, Ticino-Po and Inn-Danube. The Swiss portions of their basins are approximately, in the order named: 11,160; 2,760; 1,360; and 660 sq.mi.

Many of its rivers, with their Alpine origin, tend to seasonal overflows which have necessitated artificial embankment. The most important control scheme was that of Conrad Escher of Zürich (later Conrad von der Linth) who in 1807–27 turned the turbulent Linth into Walensee, from which it emerges as a canalized stream. An earlier (1714) successful work was the diversion of the troublesome Kander into the Lake of Thun, where it is now placidly building an extensive fertile delta.

The lakes of Switzerland are very numerous: the largest, Geneva (southwest) and Constance (northeast) are on the frontiers and are not wholly Swiss. Neuchâtel (83 sq.mi.) is the largest wholly Swiss lake. About 16 sq.mi. at the north end of Lago Maggiore (82 sq.mi.) belongs to Switzerland; next in order of size are Lucerne, Zürich, about half of the Swiss-Italian Lake of Lugano (19 sq.mi.), Thun, Brienz, Morat, the Wallen and Sempach (5.5 sq.mi.); no others exceed 4 sq.mi. Eleven of these lakes are in the Aar basin, two (Maggiore and Lugano) are in the Po basin, and Geneva is the great Rhône filter. The lakes of the Swiss portion of the Inn basin are small; the largest are Sils (1.5 sq.mi.), and the still smaller, slightly lower and adjacent Silvaplana. Many small mountain lakes are of interest, such as the dreary Daubensee (7,264 ft. alt.) near to the Gemmi pass; the extremely beautiful Oeschinensee (5,223 ft. alt.) mirroring the snowy Blümlis Alp, and the remarkable, though sometimes empty Märjelensee (7,766 ft. alt.) on which float miniature icebergs from the Great Aletsch glacier.

Of the countless waterfalls in Switzerland those of the Rhine (near Schaffhausen), 100 ft. high, inclusive of rapids, and 340 ft. wide, are the grandest, but the most beautiful are of less volume and greater height, such as those of the Lauterbrunnen valley and particularly the Staubbach, a mere veil of water hanging in front of a precipice 980 ft. high.

Switzerland contains many more glaciers than the combined total for adjacent lands; the number is estimated at more than 1,000, but no exact computations are possible on account of numerous detached ice masses which may or may not rank as glaciers. Practically all of them are now in retreat, though they will long continue to feed all the important rivers and streams of Switzerland. They probably occupy 700 sq.mi., very unequally distributed; 11 of the cantons possess no glaciers. The greatest area is found in the Valais (more than half of the total area), followed by the Grisons and Berne (about one-sixth each); then by Uri, Glarus and Ticino (the last about 13 sq.mi.); the remaining cantons—Unterwalden, Vaud, St. Gallen, Schwyz and Appenzell have, in aggregate, a glacier area little superior to that of Ticino alone. The longest glacier in the main Alpine chain is the Gorner (9.25 mi.), but it is exceeded by the Great Aletsch (16.5 mi.) and by the Fiescher and the Unteraar (10 mi. each) which run down from the high eastern mountains of the Bernese Oberland.

Climate.—In an area such as Switzerland which extends through less than two degrees of latitude, climate will be influenced more particularly by differences of altitude, aspect and gradient; these show wide and surprisingly sudden variations, *e.g.*, Monte Rosa (15,217 ft. high) is only 30 mi. distant in a "straight" line from Lago Maggiore, with waters in the same latitude, but only 646 ft. in altitude. Great heights are characterized by (1) low barometer readings: Monte Rosa summit records about half sea-level pressure; hence the possibility of "mountain sickness"; (2) usually more brilliant sunshine and dry air, particularly in winter, producing sunburning and ice-dazzle, which in some subjects amounts to snow blindness; (3) much lower temperatures; among the Alps the average fall in temperature is about 1.2° F. on the south side and about 0.9° F. on the north side for each 330 ft. of ascent. The height of the Alpine crest and its main direction also profoundly influence the winds. In addition to local winds of certain higher levels which show a daily change of direction, downhill in the morning and uphill in the evening, there are other characteristic winds which are much less localized in their effects, as the southerly *föhn*, warm, dry and oppressive, which affects considerable areas in east Switzerland, particularly during spring and autumn; and the northerly *bise*, a cold wind experienced especially in Geneva. Switzerland's many climates are important, not only in Europe's playground generally, but also in the "nests of

sanatoria." Hence, climate and weather statistics have been carefully compiled for many stations.

January is the coldest month, the following averages in degrees F. being recorded: Basle (909 ft. alt.) 31.8; Altdorf (1,480 ft. alt.) 32.4; Davos (5,121 ft. alt.) 18.7; St. Gotthard (6,877 ft. alt.) 18.1; Säntis (8,202 ft. alt.) 16.2, while an unusually low average is from Bevers (5,610 ft. alt.) with 14.2. Precipitation (either rain or melted snow) shows wide variations, *e.g.*, Basle has 32.5 in. total annual precipitation; Altdorf, 49.0 in.; Davos, 35.7 in. and Säntis, 95.7 in. At several stations much of this falls as snow. On Säntis the precipitation is as snow from November to April inclusive, and only July and August have more rain than snow. Snow accumulates to a depth of 20 to 25 ft. at Bevers and upward of 45 ft. on Säntis. July is the hottest month, the average July records in degrees F. being: Basle 66.4, Altdorf 64.4, Davos 53.8, St. Gotthard 46.2, Säntis 41.0, Bevers 53.2. The range of 39.0° at Bevers tends to be extreme; Säntis, though cold, is much more equable, 24.8°. The snow line, showing considerable local variations, is about 9,000 ft. high on the western Alps, and about 10,500 ft. in the drier eastern mountains.

Vegetation.—Though Swiss vegetation is mainly central European in character, in some of the warm sheltered valleys Mediterranean types are found. The true alpine flora (plants growing above the tree limit) is akin to that of the Pyrenees rather than to either of the other elements (*see* Alps). Mediterranean species, found locally in the wide valley between the Jura and the Alps, in the Rhône valley in Valais, and on the southern slopes of the mountains running down to the Italian lakes, include vines, walnut trees, camellias, magnolias and palms. The lowlands and foothills are largely cultivated, and wherever possible land is farmed in the mountains. The natural climax vegetation of the mountain region, up to about 4,000 ft. on the northern slopes and 5,000 ft. or more on the southern, consists of deciduous trees such as oak, beech, ash and sycamore; but mixed forests of timber trees such as spruce, silver fir, larch and birch have been planted to a great extent in their place. The subalpine region covers the next 1,000 ft. or so where the pure coniferous forest (spruce, silver fir, larch and Scotch pine) are found. These merge into the scrub of the alpine region which contains dwarf pines and willows, juniper, bilberry, creeping azalea, mezereon, white dryas, rhododendrons, etc. This is also the region of the alpine pastures. Above about 6,500 ft. (where the summer lasts only five or six weeks) the flora is typically alpine, and cushion plants, mosses and lichens flourish among the bare rocks, screes and glaciers. Finally, between 8,000 ft. and 9,500 ft. comes the region of perpetual snow.

The alpine meadows are transformed into brilliant gardens in May, June and July. Among the first to blossom are crocuses, anemones, *Gentiana verna, Primula farinosa* and pheasant's eye narcissus, while on the higher slopes the edelweiss is in bloom. In June come St. Bruno's lily, orchids, campanulas, larkspurs and many others, followed by the martagon lily, oxeye daisy, etc. The alpine roses, *Rhododendron ferrugineum* and *R. hirsutum* (only found on limestone), clothe the sides of the Alps and Jura together with *Saponaria ocymoides, Helianthemum alpestre, Gentiana nivalis,* the alpine lily and many more. As soon as the meadows have been cut for hay the autumn crocus (*Colchicum autumnale*) appears.

Fauna.—In Switzerland the animals are protected and there is only a very short shooting season. Many of them, for example the chamois that lives high up in the mountains and the marmot that is found near the snow line, are alpine animals. Among the inhabitants of the forests are squirrels and foxes and woodpeckers, jays, tits and many other birds. Hares and field mice are common in the meadows. In the lakes and rivers there are trout and other fish, and salmon are found in the Aar, the Rhine and Lake Zürich. Among reptiles the viper is common. Insects are numerous: butterflies and moths abound in the meadows when the flowers are in bloom and ants are plentiful in the woods.

See E. de Martonne, "La Suisse" in part 1 of *Europe centrale,* vol. iv of *Géographie universelle,* ed. by P. Vidal de la Blache and L. Gallois (Paris, 1931). (W. A. B. C.; X.)

HISTORY

For indications of the prehistory and ancient history of what is now Switzerland *see* Alamanni; Burgundy; Celt; Geneva; Helvetii; La Tène; Raetia; Savoy, House of; Vindelicia; etc. The following pages treat of the history of the Swiss confederation from its origin to the present day.

The Swiss confederation forms a nation the patriotism of whose members is universally acknowledged; but Swiss history is very intricate and very local. A firm hold on a few guiding principles is therefore most desirable, and of these there are three which must always be borne in mind. (1) The first is the connection of Swiss history with that of the empire. Swiss history is largely the history of the drawing together of bits of each of the imperial kingdoms (Germany, Italy and Burgundy) for common defense against the Habsburgs; and when this dynasty had secured to itself the permanent possession of the empire, the Swiss league little by little won its independence of the empire, practically in 1499, formally in 1648. (2) The second is the German origin and nature of the confederation. Around a German nucleus (the three Forest districts or, as they are now called, cantons) there gradually gathered other German districts; and it was not until 1803 and 1815 that its French- and Italian-speaking cantons were raised to political equality with the German confederates. (3) The third is that Swiss history is a study in federalism. Based on the defensive alliances of 1291 and 1315 between the three Forest cantons, the confederation was enlarged by the admission of other districts and towns, all leagued with the original three members, but not necessarily with each other. Hence great difficulties are encountered in looking after common interests; the diet was merely an assembly of ambassadors with powers very strictly limited by their instructions, and there was no central executive authority. The whole constitutional history of the confederation is summed up in its transition to a federal state, which, while a single state in its foreign relations, in home matters maintains the more or less absolute independence of its members.

The League of 1291.—On Aug. 1, 1291, the men of the valley of Uri, the free community of the valley of Schwyz, and the association of the men of the lower valley or Nidwalden—Obwalden or the upper valley is not mentioned in the text, though its name was afterward added to the attached seal of Nidwalden—formed the Everlasting league for the purpose of self-defense against all who should attack or trouble them, a league which was expressly stated to be a confirmation of a former one. This league was the foundation of the Swiss confederation.

The legal and political conditions of each district were very different. (1) In 853 Louis the German granted all his lands in the *pagellus Uroniae* to the convent of SS. Felix and Regula in Zürich (the present Fraumünster) and exempted them from all jurisdiction save that of the king (*Reichsfreiheit*), so that though locally within the Zürichgau they were not subject to its count, the king's deputy. The abbey thus became possessed of the greater part of the valley of the Reuss between the present Devil's Bridge and the Lake of Lucerne, for the upper valley (Urseren) belonged at that time to the abbey of Disentis in the Rhine valley and did not become permanently allied with Uri until 1410. The privileged position of the abbey tenants gradually led the other men of the valley to "commend" themselves to the abbey. The important post of protector (*advocatus* or *vogt*) of the abbey was given to one family after another by the emperor; in 1218 the office was granted to the Habsburgs, but in 1231 the German king Henry VII bought it back in order to control the St. Gotthard pass, the valley thenceforward depending immediately on the king. (2) In Schwyz (first mentioned in 972) it is necessary to distinguish between the districts west and east of Steinen. In the former the land was in the hands of many nobles, among whom were the Habsburgs; in the latter there was, at the foot of the Mythen, a self-governing community of free men; both, however, were politically subject to the king's delegates, the counts of the Zürichgau, who after 1173 were the ever-advancing Habsburgs. But in 1240 the free community of Schwyz obtained from the emperor Frederick II a charter which removed them from the jurisdiction of the counts, placing them in immediate dependence

on the king, like the abbey men of Uri. In a few years, however, the Habsburgs contrived to dispense with this charter in practice. (3) In Unterwalden things were very different. The upper valley (Obwalden or Sarnen), like the lower (Nidwalden or Stans), formed part of the Zürichgau, while in both the soil was owned by many lay and ecclesiastical lords, among them being the Habsburgs and the Alsatian abbey of Murbach. Hence in this district there were privileged tenants, but no free community and no centre of unity; this explains why Obwalden and Nidwalden won their way upward so much more slowly than their neighbours in Uri and Schwyz.

The Habsburgs had risen steadily from the position of an unimportant family in the Aargau to that of powerful landed proprietors in Swabia and Alsace and had attained political importance as counts of the Zürichgau and Aargau. In one or both qualities the cadet or Laufenburg line, to which the family estates in the Forest cantons around the Lake of Lucerne had fallen on the division of the inheritance in 1232, seems to have exercised its legal rights in a harsh manner. In 1240 the free men of Schwyz obtained protection from the emperor. It was natural that when, after the excommunication and deposition of Frederick II by Innocent IV in 1245, the head of the cadet line of Habsburg sided with the pope, some of the men of the Forest cantons should rally around the emperor. Schwyz joined Sarnen and Lucerne (though Uri and Obwalden supported the pope); and in 1247 the men of Schwyz, Sarnen and Lucerne were threatened by the pope with excommunication if they persisted in upholding the emperor and defying their hereditary lords the counts of Habsburg. The rapid decline of Frederick's cause soon enabled the Habsburgs to regain their authority in these districts. These incidents are the only foundation in fact of the legendary stories of Habsburg oppression told of and by a later age. After this temporary check the power of the Habsburgs continued to increase rapidly. In 1273 the head of the cadet line sold all his lands and rights in the Forest cantons to the head of the elder or Alsatian line, Rudolph, who a few months later was elected German king. He recognized the privileges of Uri but not those of Schwyz; and as he united in his person the characters of German king, count of the Zürichgau and landowner in the Forest cantons, the union of offices might be expected to result in a confusion of rights. On April 16, 1291, Rudolph bought from the abbey of Murbach in Alsace (of which he was advocate) all its rights over the town of Lucerne and the abbey estates in Unterwalden. It thus seemed probable that the other Forest cantons would be shut off from their natural means of communication with the outer world by way of the lake. Rudolph's death, on July 15 of the same year, cleared the way, and a fortnight later (Aug. 1) the Everlasting league was made between the men of Uri, Schwyz and Nidwalden for the purpose of self-defense against a common foe.

Morgarten and the League of 1315.—In the struggle for the empire, which extended over the years following the conclusion of the league of 1291, the confederates supported without exception the anti-Habsburg candidate. On Oct. 16, 1291, Uri and Schwyz allied themselves with Zürich and joined the general rising in Swabia against Albert, the new head of the house of Habsburg. It soon failed, but hopes revived when in 1292 Adolph of Nassau was chosen German king. In 1297 he confirmed to the free men of Schwyz their charter of 1240 and, strangely enough, confirmed the same charter to Uri, instead of their own of 1231. It is in his reign that the first recorded meeting of the *Landesgemeinde* (or legislative assembly) of Schwyz (1294) occurred. But in 1298 Albert of Habsburg himself was elected German king. His rule was strict and severe, though not oppressive. He did not indeed confirm the charters of Uri or of Schwyz, but he did not attack the ancient rights of the former, and in the latter he exercised his rights as a landowner and did not abuse his political rights as king or as count. In Unterwalden the two valleys were joined together (1304) under a common administrator (the local deputy of the count)—a great step forward to permanent union. The stories of Albert's tyranny in the Forest cantons are not heard of until two centuries later. His successor, Henry of Luxemburg, confirmed, on June 3, 1309, to Uri and Schwyz their charters of 1297 and, for some unknown reason, confirmed to Unterwalden all the liberties granted by his predecessor, though as a matter of fact none had been granted. This charter and the nomination of one royal bailiff to administer the three districts had the effect of placing them all in an identical political position, and that the most privileged yet given to any of them—the freedom of the free community of Schwyz. On June 13, 1311, Henry granted the request of Leopold I of Habsburg, who had accompanied him on his expedition to Rome, that Baron Eberhart of Bürglen and Count Frederick of Toggenburg should inquire into the precise rights of the Habsburgs in Alsace and in the Forest cantons with a view to restoring them. But no steps were taken to draw conclusions from this investigation.

On Henry's death in 1313 the men of Schwyz seized the opportunity for making a wanton attack on the great abbey of Einsiedeln, with which they had a long-standing quarrel as to rights of pasture. The abbot caused them to be excommunicated, and Leopold's brother Frederick of Habsburg (chosen king, as Frederick III, by a minority of the electors), who was the hereditary advocate of the abbey, placed them under the ban of the empire. His rival, Louis of Bavaria, to whom they appealed, removed the ban; whereupon Frederick issued a decree by which he restored to his family all their rights and possessions in *the three valleys* and Urseren and charged Leopold with the execution of this order. The confederates hastily made arrangements with Glarus, Urseren, Arth and Interlaken to protect themselves from attack on every side. Leopold collected a brilliant army at the Austrian town of Zug in order to attack Schwyz, while a body of troops was to take Unterwalden in the rear by way of the Brünig pass. On Nov. 15, 1315, Leopold moved forward along the shore of the Lake of Aegeri, intending to assail the town of Schwyz by climbing the slopes of Morgarten above the southeastern end of the lake. There they were awaited by the valiant band of the confederates, who utterly defeated them. Leopold fled in haste to Winterthur, and the attack by the Brünig was driven back by the men of Unterwalden. On Dec. 9, 1315, representatives of the victorious highlanders met at Brunnen, on the Lake of Lucerne and renewed the Everlasting league of 1291. In their main lines the two documents are very similar, the later being chiefly an expansion of the earlier. That of 1315 is in German (in contrast with that of 1291, which is in Latin) and has one or two striking clauses largely indebted to a decree issued by Zürich on July 24, 1291. None of the three districts or their dependents was to recognize a new lord without the consent and counsel of the rest. Strict obedience in all lawful matters was to be rendered to the rightful lord in each case, unless he attacked or wronged any of the confederates, in which case they were to be free from all obligations. No negotiations, so long as the "Länder" had no lord, were to be entered on with outside powers, save by common agreement of all. Louis solemnly recognized and confirmed the new league in 1316; and in 1318 a truce was concluded between the confederates and the Habsburgs. The lands and rights annexed belonging to the Habsburgs in the Forest cantons were fully recognized as they existed in the days of Henry of Luxemburg, and freedom of commerce was granted. But there was not one word about the political rights of the Habsburgs as counts of the Zürichgau and Aargau.

The League of Eight Members.—The name Switzerland (*Sweicz;* derived from Schwyz, which had always been the leader in the struggle) applied to the three Forest cantons as early as 1320 and extended to the confederation as a whole in 1352. But it was not until after Sempach (1386) that it came into popular use, and it did not form the official name of the confederation until 1803. This is in itself a proof of the great renown which the league won by its victory at Morgarten. Another is that as years passed other members were admitted to the privileges of the original alliance of the three Forest cantons. First to join the league (1332) was the neighbouring town of Lucerne, which had grown up around the monastery of St. Leodegar, Ludger or Leger (a cell of the great house of Murbach in Alsace) and had been sold to the Habsburgs in 1291. Its accession brought a new element into the pastoral association of the Forest cantons, which surrounded the entire Lake of Lucerne. Next came the ancient

town of Zürich in 1351, when it got the help of the Forest cantons and of Lucerne against the threat of an Austrian attack. The league was then advanced from the hilly country to the plains, though the terms of the treaty with Zürich did not bind it so closely to the confederates as in the other cases and hence rendered it possible for Zürich now and again to incline toward Austria in a fashion which did great hurt to its allies. In June 1352 the league was enlarged by the admission of Glarus and Zug. Glarus belonged to the monastery of Säckingen on the Rhine and the Habsburgs as advocates of the abbey claimed many rights over the valley, which, however, refused to admit them and joyfully received the confederates who came to its aid. However Glarus was placed on a lower footing than the other members of the league, being bound to obey their orders. Three weeks later the town and district of Zug, attacked by the league and abandoned by their Habsburg masters, joined the confederation. The immediate occasion of the union of these two districts had been a war begun by the Austrian duke Albert II against Zürich, and the peace of Sept. 1352 restored Glarus and Zug to the Habsburgs, who also regained their rights over Lucerne. Zug was won for good by a bold stroke of the men of Schwyz in 1364, but it was not until the day of Näfels (1388) that Glarus recovered its lost freedom. These temporary losses were, however, far outweighed by the entrance into the league in 1353 of the famous town of Berne. Founded in 1191 by Berchtold V of Zähringen, Berne had become a free imperial city in 1218 on the extinction of the Zähringen dynasty and made a treaty with the Forest cantons as early as 1323. In 1352 it had been forced to take part in the war against Zürich, but soon after the conclusion of peace entered the league as the ally of the three Forest cantons, being thus only indirectly joined to Lucerne and Zürich. The special importance of the accession of Berne was that the league began to spread to the west and so came into contact for the first time with the French-speaking land of Savoy. The league thus by the end of the 14th century numbered eight members, the fruits of Morgarten, and no further members were admitted until 1481, after the Burgundian war. But while each of the five new members was allied with the original nucleus—the three Forest cantons—these five were not directly allied to one another.

Sempach and the Aargau.—Soon afterward, the quarrels with Austria broke out afresh. Lucerne fretted much under the Austrian rule, received many Austrian subjects among its citizens and refused to pay custom duties to the Austrian bailiff at Rothenburg, on the ground that it had the right of free traffic. An attack on the customhouse at Rothenburg and the gift of the privileges of burghership to the discontented inhabitants of the little town of Sempach a short way off, so irritated Leopold III (who then held all the possessions of his house outside Austria) that he collected an army, with the intention of crushing his rebellious town. Lucerne meanwhile had summoned the other members of the league to its aid, and approximately 1,600 men of Uri, Schwyz, Unterwalden and Lucerne opposed the 6,000 which made up the Austrian army. The decisive fight took place on July 9, 1386, near Sempach. Leopold, with a large number of his followers, was slain and the Habsburg power within the borders of the confederation finally broken. Glarus at once rose in arms, but it was not until the expiration of the truce made after Sempach that Leopold's brother, Albert III of Austria, brought an army against Glarus and was defeated at Näfels (not far from Glarus) on April 9, 1388, by a handful of Glarus and Schwyz men.

In 1389 a peace of seven years was made, the confederates being secured in all their conquests; and on July 16, 1394, the peace was prolonged for 20 years (and again in 1412 for 50 years), various stipulations being made by which the long struggle of the league against the Habsburgs was finally crowned with success. By the peace of 1394 Glarus was freed on payment of £200 annually; Zug too was released from Habsburg rule. Schwyz was given the *advocatia* of the great abbey of Einsiedeln; Lucerne got the Entlebuch (finally in 1405), Sempach and Rothenburg; Berne and its ally Solothurn were confirmed in their conquests. Above all, the confederation as a whole was relieved from the overlordship of the Habsburgs, to whom, however, all their rights and dues as landed proprietors were expressly reserved, Berne, Zürich and Solothurn guaranteeing the maintenance of these rights and dues with power in case of need to call on the other confederates to support them by arms. Though the house of Habsburg entertained hopes of recovering its former rights, so that technically the treaties of 1389, 1394 and 1412 were but truces, it was finally to renounce forever all its feudal rights and privileges within the confederation by the Everlasting compact of 1474.

The victory at Sempach enabled the league to extend both its influence and its territory. Both the league and its individual members were able to take the offensive. In the 15th century each member increased and rounded off its territory, though it usually withheld political rights from the men of the newly acquired lands. It was in the same century that Appenzell, St. Gallen and the Upper Valais first became associated with the league, though they did not become full members for a long time—Appenzell in 1513, St. Gallen in 1803, the Valais in 1815. Above all, the 15th century saw the first attempt of the confederation to secure a footing south of the Alps.

In 1412 the treaty of 1394 between the league and the Habsburgs had been renewed for 50 years. When in 1415 the German king and future emperor Sigismund placed Duke Frederick IV of Austria under the ban of the empire, the league hesitated, because of their treaty of 1412, until Sigismund declared that this treaty did not release them from their obligations to the empire. In Sigismund's name, therefore, and by his special command, the different members of the league overran the extensive Habsburg possessions in the Aargau. The chief share fell to Berne, but certain districts (known as the *Freie Ämter* and the county of Baden) were joined together and governed as bailiwicks held in common by all members of the league (save Uri, busied in the south, and Berne, which had already secured the lion's share of the spoil for itself). This is the first case in which the league as a whole took up the position of rulers over districts which, though guaranteed in the enjoyment of their old rights, were nevertheless politically unfree.

First Italian Conquests.—As the natural policy of Berne was to seek to enlarge its borders at the expense of Austria and later of Savoy, so Uri, shut off by physical causes from extension in other directions, as steadily turned its eyes toward the south. In 1410 the valley of Urseren was finally joined to Uri, an acquisition which gave to Uri complete command over the St. Gotthard pass, long commercially important and henceforth useful for purposes of war and conquest. Already in 1403 Uri and Obwalden had occupied the long narrow upper Ticino valley on the south of the pass called the Val Leventina or Livinental; in 1410 the men of the same two lands, exasperated by the insults of the local lords, called on the other members of the league, and all jointly (except Berne) occupied the Val d'Ossola (Eschental) on the south side of the Simplon pass. Having lost this to Savoy in 1414, they obtained in 1416–17 the alliance of the men of the Upper Valais, then fighting for freedom, and thus regained the valley (1416). In 1419 Uri and Obwalden bought from its lord the town and district of Bellinzona. This rapid advance, however, did not meet with the approval of the duke of Milan, Filippo Maria Visconti; the confederates were not at one with regard to these southern conquests; Filippo Maria intrigued with them; and finally in 1426, by a payment of a large sum of money and the grant of certain commercial privileges, the Val Leventina, the Val d'Ossola and Bellinzona were formally restored to him.

The First Civil War.—The original contrasts between the social condition of the different members of the league became more marked when the period of conquest began. Quarrels and ill-feeling arose on the subject of the Aargau and the Italian conquests and a few years later ripened into a civil war, brought about by the dispute as to the succession to the lands of Frederick VII, count of Toggenburg, the last male representative of his house. Count Frederick's predecessors had greatly extended their domains, so that they took in not only the Toggenburg or upper valley of the Thur, but Uznach, Sargans, the Rhine valley between Feldkirch and Sargans, the Prättigau and the Davos valley. He himself, the last great feudal lord on the left bank of the Rhine, had

managed to secure his vast possessions by making treaties with several members of the league. His death (April 30, 1436) was the signal for the outbreak of strife. The Prättigau and Davos valley formed the League of the Ten Jurisdictions in Raetia, while Frederick's widow took sides with Zürich against Schwyz for different portions of the great inheritance which had been promised them. After being twice defeated, Zürich was forced in 1440 to buy peace by certain cessions to Schwyz, the general feeling of the confederates being opposed to Zürich, so that several of them went so far as to send men and arms to Schwyz. Zürich, however, was bitterly disappointed at these defeats and had recourse to the policy which it had adopted in 1356 and 1393—an alliance with Austria (concluded in 1442), which now held the imperial throne in the person of Frederick III (Frederick IV as German king). Though permissible according to the terms of Zürich's joining the league in 1351, this act caused the greatest irritation in the confederation, and civil war broke out. In 1443 the Zürich troops were completely defeated at St. Jakob on the Sihl, close under the walls of the city. Next year the city itself was long besieged. Frederick, unable to get help elsewhere, procured from Charles VII of France the dispatch of a body of Armagnac free lances (the Ecorcheurs), who came, 30,000 strong, under the dauphin Louis, plundering and harrying the land. At the very gates of the free imperial city of Basle the desperate resistance of a small body of confederates (1,200 to 1,500), until cut to pieces, checked the advance of the freebooters, who then returned whence they had come. Several small engagements ensued, Zürich declining to make peace because the confederates required, as the result of a solemn arbitration, the abandonment of the Austrian alliance. At length peace was concluded in 1450, the confederates restoring almost all the lands they had won from Zürich.

(W. A. B. C.; H. Nz.)

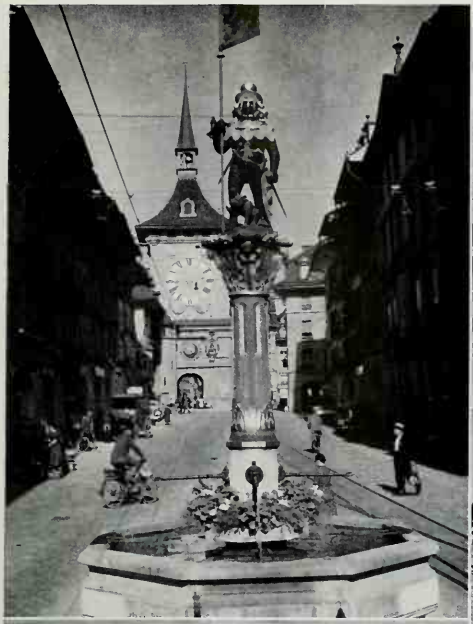
Waldshut and the Burgundian War.—Various communities such as the abbey and the city of St. Gallen (1451), Appenzell (1452), Schaffhausen (1454), Rottweil (1463) and Mulhouse (1466) now became allies or associates of the eight cantons, while a successful war secured Thurgau as a common lordship, administered by all the states, except Berne, in agreement (1461). In 1467 and 1468 Berne used various pretexts to secure control of the Aargau; and war followed in 1468, partly in order to maintain the continued independence of Mulhouse against Sigismund of Tirol, the emperor Frederick III's cousin, who had succeeded to the Habsburg lands on the Rhine and in Alsace on his father's death in 1439 (in the years immediately following Frederick had been acting as his guardian). The war was hurriedly ended in Aug. 1468 by the treaty of Waldshut, by which Sigismund was obliged to pay an indemnity of 10,000 gulden within a year, failing which he was to forfeit Waldshut and his possessions in the Black forest. Sigismund had no ready cash and his efforts to secure a loan led to an even greater war.

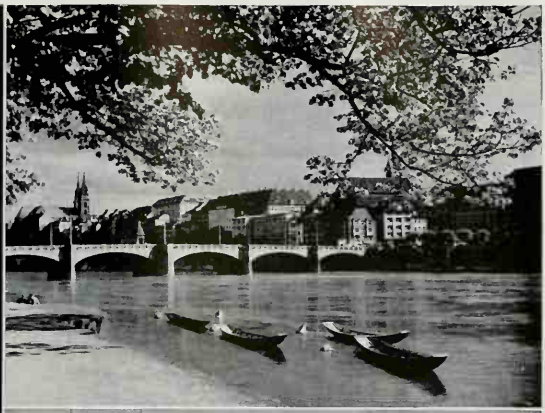
Louis XI of France refused to advance any money; but by the treaty of Saint-Omer (May 1469) Charles the Bold, duke of Burgundy, was willing to offer 50,000 gulden in cash in return for a pledge of the Austrian lands in Upper Alsace, Ferrette, the Black forest and the Rhenish towns of Waldshut, Säckingen, Rheinfelden and Laufenburg. Skilfully interpreting some permissive clauses of the treaty favourably to his own expansionist ambitions, Charles proceeded to occupy the whole of Upper Alsace. He appointed Peter von Hagenbach as his agent with instructions to redeem mortgages, repair fortifications and improve administration in a way that would permanently unite Alsace with the Burgundian territories. This threatened the empire, and Frederick III declared the treaty void; it also greatly perturbed Berne, whose prospects of expansion were greatly reduced. Further, Charles the Bold entered into an alliance with Savoy and with Milan and also threatened to occupy Mulhouse, closely associated with Berne.

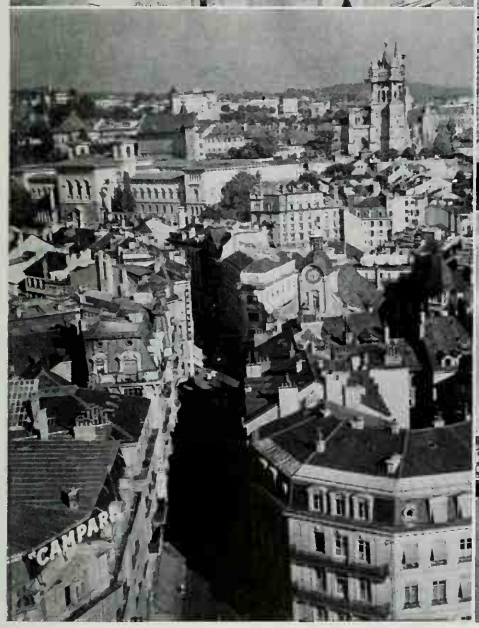
One result was the formation of the Lower union in 1473 by Basle, Colmar, Mulhouse, Sélestat and Strasbourg; another, the Perpetual understanding of 1474 between Sigismund of Tirol and the Swiss, by which the former renounced any further territorial claims in Switzerland in return for promises of money to pay off his debts. Hagenbach was illegally arrested and executed in the spring of 1474—a deliberate act of defiance of Charles the Bold, who seemed to be fully occupied elsewhere. The Lower union, supported by Berne, attacked Héricourt (Nov. 1474) and Blamont (1475), Louis XI promising cash subsidies while the war continued. Inspired by Niklaus von Diesbach, this western and southern movement continued all through 1475, much of the Vaud and part of Savoy, including Grandson, Orbe, Echallens, Jougne, Morat, Avenches, Estavayer and Yverdon being occupied by the end of the year, the inhabitants of the Valais acting in collaboration with Berne to prevent any interference from the side of Milan. Enraged and even a little alarmed, Charles the Bold came south in person in 1476, occupying Grandson at the end of February and hanging its Bernese garrison. Berne had gathered a considerable force from many states of the confederation and from the Lower union for the relief of the castle: too late to effect this, they offered battle on March 2 at Grandson. Charles' faulty generalship in the face of a disciplined force of pikemen, whom he mistakenly despised as ignorant peasants, led to the rout but not to the destruction of the Burgundian army. The duke regrouped his forces at Lausanne and again advanced against Berne in June. His way was blocked by the castle of Morat, held by Adrian von Bubenberg and 2,000 Bernese. The threat to this position was sufficient to bring 20,000 Swiss to its defense; and the battle of Morat (June 22, 1476) ended in a complete Swiss victory. Charles was forced to fall back on Lorraine, whose ruler, René, had turned against him. At Nancy (Jan. 5, 1477) the duke was defeated and killed. The Burgundian lands devolved upon his daughter Mary; Louis XI of France secured the duchy of Burgundy; Savoy regained, with French help, a good deal of the Vaud; and a Swiss incursion into Franche-Comté failed to consolidate itself into a permanent occupation.

Fribourg and Solothurn Admitted.—Such unity as the Swiss had so far enjoyed had been based almost entirely on successful defense against external enemies. They had been, in fact, poor and apparently insignificant communities co-operating with difficulty for self-preservation. Their military prowess, nurtured by necessity, had made them potentially great: the latter part of the 15th century was the golden age of the mercenary soldier, and the Swiss had proved themselves the best available recruits. There was much money to be made by military service and, for the first time, prospects of lucrative employment opened to young men of valour. But the cessation of danger from outside and the lure of foreign gold were almost sufficient to turn friendly states into competitors. There was no constitution to hold them together, no common officers, seal, law courts, coinage or flag; there was a growing divergence of interest between the cities and the mountain-dwellers, as well as between east and west; there were numerous family feuds and endless local jealousies, culminating in disputes about the distribution of the booty taken in the Burgundian war and about the administration of the territories held in common, particularly the Thurgau. Fribourg and Solothurn were anxious to come within the circle of the confederation, but neither Zürich nor Lucerne was very anxious for their support, while the country districts were anxious to maintain their predominance over the cities. By 1481 these differences had become so exacerbated that all common purpose seemed at an end and civil war threatened. It was averted by the intervention at Stans of the mysterious and self-effacing hermit Nicholas Lowenbrügger (Nicholas von der Flüe or Bruder Klaus), who arranged the compromise which held the confederation together. By this *Stanser Verkommnis* the full sovereignty of the component states was maintained and reasserted, but each agreed to abstain from attempts to foment internal discord or separatist activity among the subjects of any other state; the confederate oath of union was to be renewed every five years; loot captured in war was to be divided proportionately among the actual participants in the campaign, but lands occupied were to be shared equally by the states involved. The ancient guarantees of mutual aid and assistance in time of war were renewed, and Fribourg and Solothurn were admitted as full sovereign members of the confederation.

Zürich and the Swabian War.—Zürich had grown steadily in power and influence after its readmission to full confederate







BY COURTESY OF (TOP LEFT, TOP RIGHT, BOTTOM LEFT) SWISS FEDERAL RAILWAYS; PHOTOGRAPHS, (CENTRE RIGHT, BOTTOM RIGHT) J. ALLAN CASH

CITIES OF SWITZERLAND

Top left: Berne, showing the 16th-century clock tower and a statue of a bear in armour. The bear as a symbol of Berne appeared on a city seal of 1224

Top right: Bridge across the Rhine joining north and south Basle

Centre right: Railroad station, Zürich

Bottom left: Business district of Lausanne

Bottom right: City of Zürich on the bank of the river Limmat

SWITZERLAND

BY COURTESY OF (CENTRE RIGHT) SWISS FEDERAL RAILWAYS; PHOTOGRAPHS (TOP LEFT, TOP RIGHT, BOTTOM LEFT, BOTTOM RIGHT) J. ALLAN CASH

SCENES IN THE SWISS MOUNTAINS

Top left: Ski lift at Zermatt, Valais canton. The Matterhorn may be seen in the background
Top right: St. Gotthard pass, Ticino
Centre right: Resort of St. Moritz, Grisons
Bottom left: Horse-drawn sled in Zermatt
Bottom right: Engelberg, Unterwalden. The village is entirely surrounded by mountains, the highest being the Titlis

status in 1446. Its textile manufactures had increased and with them the influence of the guilds; the effective authority within the city was shared by the ancient aristocracy, the *Constafel*. Hans Waldmann (1436–89) was recruited to the one by apprenticeship and to the other by marriage; he took part in a number of campaigns, including the battle of Morat, and was equally successful in war, business, diplomacy and politics. Burgomaster in 1482, he became virtual dictator of the city, upholding the interests of the citizens against the peasants and turning the city itself into a community whose welfare, amusements, food and clothing were subjected to regulation and supervision. In 1489, however, Waldmann had to face an armed peasants' revolt, which led to a constitutional revolution and to his execution. The aristocracy of birth and of labour joined hands, and the possibility of Zürich's dictatorship extending to the confederation was averted. The struggle between the government and the peasants, however, was repeated with local variations in Schwyz, Unterwalden, Zug, Lucerne and St. Gallen, although in no instance did the country people secure all their hopes.

The temporary eclipse of Zürich was followed by the last attack from the north. The emperor Maximilian I, uniting in his own person Habsburg and Burgundian traditions, sought to extend alike the imperial prerogatives and the possessions of his house. In 1495 the imperial diet of Worms had put forth a revised constitution for the empire ensuring internal peace and common taxation and setting up an imperial chamber (*Reichskammergericht*) as a final court of appeal. The Swiss communities were technically imperial subjects and had reason to fear renewed Habsburg aggression. They were therefore anxious to renew the understanding with France—Charles VIII was opposing Maximilian in Italy and was ready to hire Swiss mercenaries—and unwilling to join the South German Swabian league for the maintenance of peace.

South of the ten states of which the confederation was now composed lay the Grisons. In Raetia many small separate communities had united into an Upper league, a League of God's House and a League of the Ten Jurisdictions, and the two former were in alliance with the Swiss confederation. In 1499 a boundary dispute between Tirol and the League of God's House had led to a Tirolese incursion into the Engadine and the Münsterthal and thus to war with Maximilian; the Swiss came to the help of their ally, the emperor took up the implied challenge and in 1499 there was fighting from Basle to Bormio. Battles raged at Hard, Bruderholz, Schwaderloh, Frastenz and, most notably, against Maximilian himself at Dornach. Everywhere Maximilian, as usual, was unsuccessful; and the peace of Basle, Sept. 22, 1499, mediated by Ludovico Sforza (il Moro) of Milan, ended the war. In fact, although not in name, it was the recognition of Swiss independence.

Basle, Schaffhausen and Appenzell Admitted.—Basle as a member of the Lower union had taken part in the Burgundian war with the Swiss and had learned the value of such co-operation: as an imperial city it evaded with difficulty an obligation to support Maximilian in the Swabian war, while the confederates obviously needed security in the northwest. The bishop of Basle was politically much less influential than he had formerly been; and on July 13, 1501, the state was admitted to the confederation. It was a notable addition, for the city was well fortified, strong, relatively wealthy, a home of scholarship with an active university and with an industrious and busy population. Schaffhausen, although smaller, was in much the same position: it had had much difficulty with local knightly landlords, it had sent men to help its Swiss neighbours since 1453 and been accepted by most as an ally; and it had firmly refused to join the Swabian league. It received the reward of full admission to the confederation on Aug. 10, 1501. Finally, on Dec. 17, 1513, a little reluctantly, Appenzell, which had been an ally rather especially closely associated with the confederation, was admitted to full membership and partnership. It was the last addition to the whole body, now 13 independent states, until the French Revolution, and it was accepted only because it helped to redress the balance of the "country" as against the "city" states.

The Milanese Campaigns.—The northern frontiers of the Italian states, especially Milan, were exceedingly ill-defined. The southern route through Uri over the St. Gotthard was of growing importance for trade and much used by the mercenaries hired for service in the Italian wars. The inhabitants of the Val Leventina renounced Milanese allegiance in 1475 and were upheld in 1478 by armed men from Uri and other Swiss states. Bellinzona was besieged but not taken, and at Giornico a small Swiss force defeated a much larger Milanese one. But Berne and the western states were not interested in this southward expansion and refused to support any further advances; and it was with some difficulty that in April 1503 Bellinzona and the Blenio valley were secured by Uri, Schwyz and Nidwalden from Louis XII (as duke of Milan) in return for help against Ludovico il Moro. There was, however, growing resentment felt at the expenditure of Swiss lives in fighting on behalf of strangers, and opinion was markedly divided about the continued value of the French alliance, especially when France controlled Milan. The greatest man in the Valais was Matthäus Schiner, bishop of Sion and a cardinal, who had turned against the French. In 1510 Schiner arranged for Swiss support of the Holy league formed by the masterful pope Julius II for the expulsion of the foreigner from Italy. After the death of Gaston de Foix at Ravenna in 1512 Swiss, papal and Venetian forces in co-operation were able to secure the removal of the French from the plains of Lombardy. Genoa regained a transitory independence, Maximilian Sforza, son of Ludovico il Moro, returned as duke of Milan and acquiesced in the occupation of Locarno, Lugano, Mendrisio and the Val d'Ossola by the confederates, while the Grisons secured the Valtellina, Bormio and Chiavenna. At Novara (1513) a fresh French army under Louis de La Trémoille was almost annihilated in a hard-fought battle, after which the Swiss penetrated into the duchy of Burgundy as far as Dijon but returned home when promised a substantial sum of money. Francis I renewed the French invasion of northern Italy in 1515. The division of interests between the eastern cantons (who were reluctant to leave the St. Gotthard route uncovered) and Berne (traditionally well-disposed to France), together with a growing dislike of mercenary service and an equally strong distrust of Cardinal Schiner, led to a division of forces. Berne, Fribourg, Solothurn and the Valais were bought off at Gallarate; and a truncated Swiss army, with Milanese and papal support, was defeated on Sept. 14, 1515, at Marignano (*q.v.*) with heavy losses. Francis I willingly agreed to easy terms in 1516 retaining the duchy of Milan but acknowledging all that the confederates and the Grisons had gained in the Bellinzona area except Domodossola and paying a considerable sum as well.

The Reformation.—Five dioceses—Basle, Lausanne, Sion, Chur and Constance—provided the spiritual framework for the ministrations of the church to the Swiss people. Neither the institutions nor the men were suited to the changing needs of the 16th century. The bishops were sometimes nonresident, sometimes incompetent, often, as absentee landlords, disliked. The parish priests were frequently ignorant and sometimes lazy or immoral. While there were no territorial princes, as in Germany, to benefit by a redistribution of church property, the considerable monastic lands were coveted by certain cities, and if there was little active opposition to the indulgence system which so enraged Martin Luther, there was little popular support for it either. Humanism could hardly impinge upon the lives and thoughts of peasant mountain dwellers, but contacts with Italy were easy and Basle became the home of Erasmus and Joannes Froben's printing press. All these things affected the career of Huldreich Zwingli (*q.v.*) (1484–1531) who became the most influential figure in the Swiss Reformation. A musical, intelligent boy of good peasant stock from the Toggenburg district he had taught classics at the University of Basle and had been parish priest at Glarus (1506) and at Einsiedeln (1516) and people's priest (*Leutpriest*) in the Zürich minster in 1518. He had been to Rome and with Swiss armies in the field, had studied widely in the Fathers and read the Bible in the original Hebrew and Greek. Experience had convinced him that mercenary service was harmful to his countrymen, and his Bible study and reading of the works of Erasmus had made him critical of the pope and opposed to indulgences.

In the all-important city of Zürich the pulpit, properly used, could be most influential, and on Jan. 2, 1519, Zwingli started a course of sermons on St. Matthew's Gospel such as had not been

heard in the minster before: his teaching was inevitably called Lutheran although it was, in fact, the independent outcome of his own thought and studies. In the same year there was plague in Zürich, which caused a revivalist movement. Next, in 1520, Zwingli was denouncing the compulsory payment of tithe to the canons. The city council then permitted the public preaching of the gospel, and in 1522 Zwingli formally challenged ecclesiastical authority by eating meat (sausage) during Lent. He also denounced clerical celibacy and married Anna Reinhard, setting out his convictions and theological standpoint in 67 conclusions, which were publicly debated in Zürich in 1523. Appealing to the Bible, he convinced his hearers; the Great minster was reformed, monks and nuns began to leave the cloister, the clergy to marry and the communion to be administered in both kinds, the income from church property being diverted to education or poor relief. Images were discountenanced, preaching encouraged, relics hidden or destroyed, monasteries dissolved and finally, in April 1525, the mass was superseded by a simple commemorative service. Zürich had thus by steady stages accepted the Reformation, and Zwingli was left almost as head of the state as well as religious leader. He set out his religious teaching in writing, issued a vernacular translation of the Bible, rejected the teaching of the Anabaptists (who were proscribed) and secured a permanent majority of supporters in the city council.

The Catholic states were by then seriously alarmed and made every effort to isolate the new teaching in Zürich and to prevent its extension even to the common lordships over which Zürich shared jurisdiction. It soon became apparent that religious and economic interests were likely to coincide; the strongly Catholic cantons were, generally, the rural mountainous areas, Uri, Schwyz, Zug, Unterwalden, Lucerne, Solothurn and Fribourg, while the city-dominated communities, Basle, Berne and Schaffhausen in particular, were lost to the old faith. The coherence of the confederation had been difficult to maintain in 1480: it proved almost impossible in 1529. In February of that year the Catholic communities, encouraged by pope and emperor, formed a separatist group known as the Christian union; Zürich then tried to force the pace by an armed campaign in June, but the western states, particularly Berne, were not interested, and the antagonists were singularly loath to kill one another. The first armistice at Kappel (June 24) was followed by an agreement to dissolve the Christian union, to allow freedom of worship in the common lordships and to decline further foreign pensions. But neither side kept the spirit of the agreement. Then, in Oct. 1529, the efforts of Philip of Hesse to unite the Protestants by bringing Luther and Zwingli together at Marburg conspicuously failed. By 1531, when an Italian adventurer attacked the Grisons, the Forest cantons could refuse to come to the help of a non-Catholic state and could complain that the evangelical faith was being forced on the Thurgau. Choosing time and place with great skill, the five Catholic cantons declared war on Zürich at Lucerne on Oct. 4, 1531, and sent a force north to Kappel, where they met a hurriedly raised detachment of volunteers from Zürich led by Jorg Göldli and with Zwingli in person among them. The battle of Kappel (Oct. 11, 1531) resulted in a Catholic victory and in Zwingli's death.

Zürich thus lost the effective leadership of the confederation, which devolved, not very satisfactorily, upon Berne; the Catholics regained an appreciable amount of ground, so that when the Jesuits came they were in a strong position. For almost three centuries after 1531 religion so divided the Swiss people that co-operation was hardly possible. The rare diets of the whole confederation, necessary only for the administration of commonly owned territories, were brief and formal; the Protestant cities entered into cordial relations with England, Scotland, northern Germany and Scandinavia; the Catholic states were the willing agents of France, Spain and Austria on the rare occasions when these powers were found acting together. The wonder is that the confederation survived at all.

Geneva was not to be admitted as anything other than an allied state until 1815. Important strategically, it had entered into a pact of co-citizenship with Berne and Fribourg in 1526 in order to throw off the rule of its bishop and to prevent its absorption into the expanding and ambitious duchy of Savoy. The Protestant faith was brought into the Vaud from France and preached by Guillaume Farel and Pierre Viret under the protection of Berne, which had already taken various parts of the Vaud into a kind of protective custody. The Reformers were successful in 1535, after the usual public disputation, in inducing the independent city council of Geneva to forbid the public celebration of the mass. Threatened by their former bishop and by Savoy, the Genevans appealed to Berne, and a Bernese force under Hans Franz Nägeli overran the Vaud, occupied Lausanne and brought temporary security to Geneva. At this juncture John Calvin, who in 1536 had published at Basle his definitive *Christianae Religionis Institutio* came to Geneva and, not without some setbacks, soon made that city the centre of an active and independent religious teaching. Calvinist doctrine was explicit and logical, the duties of the Christian household were clearly prescribed, attendance at church, particularly at sermons, was enforced, sumptuary laws were enacted by the city, an excellent system of schools established and ministers of religion were chosen with the greatest care for their character and capacity. By the agreement of Zürich (*Consensus Tigurinus*), made with Heinrich Bullinger in 1549, Calvinism became the creed of the whole of Protestant Switzerland. The Catholic cantons would not agree to the inclusion of Geneva in the confederation so that the city had to rely upon the alliance with Berne, upon the desire of the communities of the Valais that Savoyard aggression should be restrained, and upon the obvious interest of France in a friendly neutral at such a strategically important point. Savoy was to make a last attempt to capture Geneva by surprise in 1602: this failed, and the city was never again in serious danger until it was annexed to France in 1798.

The Counter-Reformation and the Valtellina.—Charles Borromeo (*q.v.*) became archbishop of Milan in 1565, and much of southern Switzerland lay within his province. He was determined to stamp out the new faith wherever possible even if this meant straining the law and peace-agreements; he set a personal example of piety and of activity, visiting Einsiedeln and encouraging active Catholic conversion wherever possible. From the Swiss college at Milan, which he founded, and from the Jesuit houses missionaries came north; but the Protestants expelled from Locarno found a place of refuge in Zürich, and the Inquisition found little support even in the Catholic cantons. However, the nomination of a permanent papal nuncio was welcomed, and Lucerne and Fribourg gladly received Jesuit colleges. The historian Giles Tschudi, by a well-managed war, the so-called Tschudikrieg (1559–64), and through French and papal mediation, got the decrees of the Council of Trent enforced in Glarus, with the ultimate elimination of the Zwinglians; and Appenzell between 1579 and 1597 was subjected to an intensive process of reconversion which was so successful that the canton split into the two halves, Ausser Rhoden and Inner Rhoden (it thus remained politically impotent until 1848, and the division is still operative). Lucerne not only became the centre of Catholic Switzerland but it was also represented in France, during the Wars of Religion, by Ludwig Pfyffer (d. 1594) and it was the chief component of the Golden league which supported Philip II of Spain. Melchior Lussy (d. 1606) of Nidwalden not only advanced this cause by his skilful diplomacy, making himself very wealthy in the process, but he also secured the confidence of the pope and averted the threatened entire dissolution of the confederation.

The loosely united communities (Grey league, League of the Ten Jurisdictions, League of God's House) known as the Grisons, two-thirds Reformed, one-third Catholic, controlled the important eastern passes, which were of interest alike to Milan and to Venice. Milan was under Spanish control and jealous of its ancient commercial and territorial rival Venice, which in 1603 had secured the friendship of the Grisons partly through the powerful Planta family. The Valtellina (*q.v.*) was governed by the Grisons and inhabited by Catholics; its roads formed the most convenient link between Milan and Austria. Habsburg allies whose co-operation was unwelcome to France. A revolt in the Valtellina in 1620 led to Spanish support for the inhabitants against their northern masters and to successful Austrian intervention. The treaty of Mon-

zon (1626) called in the pope as a third party and led to further disputes, fomented by the local patriot, pastor, politician and pervert Georg Jenatsch (*q.v.*) and ended by a return to the *status quo* in 1639, with the Valtellina restored to Grisons sovereignty but with the Catholic rights markedly reinforced.

The Thirty Years' War.—All Switzerland was interested in the Bohemian war of 1618, which grew to a conflict that involved every great power on the continent, but the religious division in the confederation effectually prevented any corporate action. This was advantageous, for a Habsburg victory might have meant a renewal of Habsburg claims or an enforcement of the Edict of Restitution (whereby the Catholics recovered lost ecclesiastical lands), while a decisive Swedish triumph must have altered the balance of power in the confederation decisively. Many Swiss soldiers were hired by either side, and much material was sold to the combatants at a profit, while the necessities of frontier defense caused all the 13 states to come to a rare agreement in the Defensionale of Wil (1647). The frontiers would be defended against any aggressor and a mixed council of war of Protestant and Catholics would direct a force recruited from all parts of the confederation and its subjects and allies. The Defensionale was to be revised in 1668 and certain details were challenged later, but the acceptance of the principle of defense of neutrality was significant. The negotiation of the peace of Westphalia (1648) owed much to the indefatigable labours of Johann Rudolf Wettstein, burgomaster of Basle, and it was appropriate that this European instrument should take notice of the explicit renunciation by the emperor Ferdinand III of all imperial claims over Basle and the former Habsburg territories in Switzerland, an act followed by the abrogation of any claim by the *Reichskammergericht* to exercise appellate jurisdiction there.

The Villmergen Wars and Louis XIV.—The soil of Switzerland, however carefully cultivated, could not support its growing population, and the temporary prosperity caused by the wars gave way to a depression in the latter part of the 17th century. Exports were necessary, and for this an industry was required which should use as little heat as possible and make the maximum use of human ingenuity and activity. Home industries were popular, partly because conditions in the valleys restricted outdoor work, particularly in the winter. Hence wood carving, spinning, frame-work knitting and embroidery were popular; water power made textile manufacture (silk, cotton and linen) profitable, particularly around the lake of Zürich and in St. Gallen; and in western Switzerland the construction of clocks and watches and of musical instruments, at first luxury trades allied to the art of the goldsmith and the jeweller, centred in Geneva, spread to Neuchâtel, Olten, La Chaux-de-Fonds and Le Locle. Basle was a commercial city, where stockings and silk ribbon and the preparation of tobacco and snuff occupied many; Berne remained primarily the capital of a great agricultural state, where the real or imaginary virtues of the peasant received an exaggerated recognition which was later to become a kind of intellectual cult (popularized by Hans Caspar Hirzel in his *Wirthschaft eines philosophischen Bauers*, 1761). Mercenary service, particularly with the French army, was a popular form of employment, and returning soldiers brought news of foreign markets and fashions that was often valuable.

The government of a number of states, Berne, Lucerne, Fribourg and Solothurn in particular, was in the hands of a patrician aristocracy, exclusive and splendid. Difficult economic conditions after the Thirty Years' War led to a rising in 1653 of the peasants of the Entlebuch against Lucerne. With the help of neighbouring governments it was suppressed, but it was followed by similar trouble in the Bernese Oberland and in the rural territories of Basle. The peasants seemed united, and the cities in alarm cooperated for the defense of their privileges, irrespective of religious differences. A battle at Herzogenbuchsee established their superiority, and harsh punishments were meted out (followed by some belated reforms). Zürich tried to make the confederation more of a reality, but the attempt broke on the rock of religious prejudice. The Catholic states reinstituted the Golden league in 1655, and the expulsion of some Protestant families from Arth led to a split between Schwyz and Zürich which developed into civil war between the five inner states and Zürich allied with Berne. At Villmergen (Jan. 24, 1656) the Catholics under Christopher Pfyffer defeated their opponents and reasserted to the full the favourable position secured at Kappel and the complete individual sovereignty of the cantons.

In the age of Louis XIV the Swiss could hardly escape the all-pervading influence of Versailles. If the Protestant states were not very eager to help the persecutor of the Huguenots, and the Catholic states looked to Spain for help, both could be influenced by the appeal of Jean de La Barde, the French ambassador, to renew the Perpetual understanding with France. In 1663 an embassy confirmed this renewal by a spectacular public oath in Notre Dame at Paris and for the next half-century French influence was paramount in such external relations as the confederation had. Louis XIV's aggressive policy was in many respects at the expense of ancient Swiss interests. At the end of the 15th century Swiss influence had been almost paramount in the Franche-Comté and in Alsace, and these were exactly the areas against which Louis proceeded most successfully. In 1667–68 the Habsburgs succeeded in retaining possession of the Franche-Comté, but the French threat to it alarmed the confederates. Its final annexation to France in 1678 brought France to the Bernese frontier and caused fears for the future of the French-speaking Vaud, while the French occupation of Strasbourg meant that Mulhouse and Rottweil could not hope to remain permanently linked with the confederation. Basle, too, recognized how serious French pressure could be when a French battery at Hüningen confronted its defenses. The confederation remained discreetly neutral during the War of the Spanish Succession, but politically its independence was the safer because of the victories of the duke of Marlborough and Prince Eugene. One other setback to France occurred in Neuchâtel: on the extinction of the ruling French family of Orléans-Longueville in 1707, a struggle ensued between the French candidate François Louis de Bourbon, prince of Conti, and the Prussian king Frederick I, in which the latter won the principality for the Hohenzollerns.

The religious question in the common lordships, particularly the Thurgau and the Rheintal, was a delicate one, as the populations of both creeds were ruled by nominally co-operating states of similarly divergent faiths. Questions of mixed marriages and the education of children were urgent ones and roused strong feelings, which were accentuated by the attempt of the prince-abbot of St. Gallen to dictate to his mainly Protestant subjects in the Toggenburg district. A road entirely under Catholic control was being constructed from Schwyz into imperial territory, and the abbot Leodegar Bürgisser was determined that the Toggenburgers should contribute labour and materials to its construction. They demurred and appealed successfully for support to Berne and to Zürich, and in 1712 an armed conflict broke out. The abbot had the support of the emperor Charles VI, which made France willing to mediate; moreover, the French representative, Charles François de Vintimille, comte du Luc, was well-disposed to the Protestant cause. Berne and Zürich co-operated well, while Lucerne was unable to co-ordinate effectively the forces of the Forest cantons. The abbot fled, the Protestants won a victory at Bremgarten, besieged Baden and blockaded their enemies elsewhere. A final battle (July 25, 1712) at Villmergen settled the issue in their favour; and the peace of Aarau (Aug. 11, 1712) established real equality of treatment for both sides in the common lordships, increased decisively the influence of Berne and Zürich in the confederation and entirely superseded the second peace of Kappel.

The Eighteenth Century.—Every form of government existed side by side: a patrician aristocracy, proud, exclusive and competent, ruled in Berne, Fribourg, Lucerne and Solothurn; Zürich, Basle and Schaffhausen were governed by their craft guilds and trading corporations; in the oldest states, Zug, Uri, Unterwalden, Schwyz, Glarus and Appenzell, the primitive democracy of the *Landesgemeinde* assured the participation in the government of every adult male citizen; in Neuchâtel and St. Gallen autocratic princes were in power. Every state, individually or collectively, ruled subject territories; there was little sense of nationality and no common language, coinage, system of weights and measures,

army or law courts. There was a general air of contentment, prosperity and industry and determination not to change. A vigorous and varied intellectual life was evidenced by the work of Johann Jakob Bodmer, Johann Caspar Lavater, the Bernouilli family at Basle, Albrecht von Haller and Johann Heinrich Pestalozzi (*qq.v.*), as well as the philosopher Isaac Iselin (1728–82). The country was increasingly visited by foreigners: Edward Gibbon wrote much of the *History of the Decline and Fall of the Roman Empire* in Lausanne; Voltaire was to be found on the frontier at Ferney. Jean Jacques Rousseau was proud to describe himself as a citizen of Geneva, and his thought was profoundly influenced by Swiss conditions. The general complacency was disturbed by several manifestations of discontent. John Daniel Davel attempted to free Lausanne and part of the Vaud from the domination of Berne in 1723; he was quite unsuccessful and was executed, but not forgotten. There was trouble in Appenzell raised by Lorenz Zellweger in 1733. Geneva unwillingly acquiesced in a greater degree of democratic government in 1738. A serious conspiracy against Berne headed by Samuel Henzi was decisively dealt with in 1749. There were demonstrations against the authority of Uri in the Val Leventina in 1755 and armed opposition to the patricians of Fribourg, led by Nicolas Chenaux, in 1781; but neither these, nor somewhat similar explosions in Schwyz, Lucerne and Geneva, presented any real challenge to constituted authority. All the time a consciousness that something more than mere defense and self-interest held the Swiss states together was growing, and this found public expression in the formation of the Swiss society (*Helvetische Gesellschaft*) in 1761.

The French Revolution and Napoleon.—The ideas of "enlightenment," rationalism and general improvement which became widespread in 18th-century Europe were eagerly absorbed in Switzerland. The writings of Montesquieu, Voltaire, Denis Diderot, Jean Le Rond d'Alembert and the Encyclopédistes, as well as the almost home-bred solvent thought of Rousseau, powerfully influenced the intellectual circles; many talked of liberty and equality in societies in which there was little practical application of either. Some of the more advanced thinkers had found life more congenial in France than at home in 1789, and in 1790 a "club of Swiss patriots" was founded in Paris by Jean Castellaz and supported by a number of Vaudois refugees, by Henri Monod, by Jean Jacques Cart and, most notably, by Frédéric César La Harpe who had been employed by Catherine the Great in Russia to teach her grandson Alexander. In 1791 F. C. La Harpe was urging the French people to support the malcontents in the Vaud and the Valais while, as in Great Britain, the conservatives in Switzerland were in strong opposition to Jacobinism. Feeling was roused to violent heights by the massacre of the Swiss guards of Louis XVI in the Tuileries (Aug. 6, 1792); but neither this event nor the invasion of the lands of the bishop of Basle by a French army nor even the formation of the First Coalition led to war. In Paris Robespierre urged correctly that the interests of France called for a friendship with the Swiss people; and the war party in Switzerland, led by Niklaus von Steiger, was able to secure only an agreement that the allies might recruit Swiss mercenaries. There was, however, a good deal of anti-French activity in Switzerland, including the collection of war material for use against France and a spy service in the pay of William Pitt's representative, Sir William Wickham.

The successful invasion of Italy by Napoleon Bonaparte in 1796 and the peace of Campo Formio (1797) was soon followed by the annexation to the Cisalpine Republic of the three provinces ruled by the Grisons, the Valtellina, Chiavenna and Bormio, and the Swiss could do nothing but accept the accomplished fact. In Paris, La Harpe and Peter Ochs of Basle were urging a French invasion of Switzerland in the name of liberty, using the separatist movement in the Vaud as an excuse for intervention. There were inspired riots in the Liestal near Basle and elsewhere, culminating in the proclamation of a "Republique Lémanique" in Jan. 1798. The French Directory at once sent soldiers to the help of the insurgents: Berne resisted, but received no help from the diet of the confederation, which debated fruitlessly at Aarau. A French contingent was driven back at Neuenegg (March 5) but at the same time the main Bernese force capitulated at Grauholz and the city of Berne was occupied. Great quantities of supplies and bullion were carried off to France by Bonaparte's agent Guillaume Brune, while Peter Ochs drafted a new constitution for a unitary Helvetian republic. The common lordships and the associated districts were abolished; the number of cantons was enlarged from 13 to 23; a bicameral legislature (great council and senate), chosen indirectly by manhood suffrage, was to be responsible for law-making; and a directory of five members was constituted the executive. The rights of the individual citizen were written into the constitution, including freedom of worship, of meeting, of petition and of the press. A national militia was embodied, and local government was controlled by centrally appointed prefects.

The new constitution was not imposed without resistance. In Nidwalden a revolt was savagely put down by Gen. Balthasar de Schauenbourg, but discontent continued to manifest itself in the east, especially when the country was marched over by opposing armies. A compulsory alliance with France gave that country the right to use Swiss roads for military purposes, and the French guarantee of the Helvetian republic offered easy opportunities for intervention. The allies regarded the republic as a satellite of France, and in May 1799 Austrian and Russian forces under Suvorov met a French army under André Masséna near Zürich before making a classic retirement into the Vorarlberg. The campaign had devastated the countryside; the price of provisions rose alarmingly; and a bad harvest made the winter of 1800 one that was long to be remembered for misery and unemployment. The cost of the new centralized government was relatively high, although the burden was more equally distributed, direct taxation being introduced for the first time and feudal dues, corvées, tithes and ground rents abolished without compensation. Some money was spent on education, and an elaborate scheme of free or cheap public instruction (culminating in a national university as in France) was adumbrated, but little could be done in a short time. J. H. Pestalozzi, however, was given a free hand to proceed with his innovations, which attracted much notice, although his adventures in journalism, as editor of the *Helvetisches Volksblatt*, were less satisfactory. The first constitution barely outlived the Directory in France. Then, after violent struggles between the protagonists of greater centralization and those who stood for more cantonal authority, the Malmaison constitution was established in May 1801, and Aloys Reding was chosen as the first *Landammann* or head of the state of 15 cantons. He had to meet political and financial difficulties, culminating in a rebellion in the Vaud and a kind of counter-revolutionary movement in Aargau and Berne.

Napoleon promptly intervened to dictate the Act of Mediation from Saint-Cloud (Sept. 30, 1802; amplified on Feb. 19, 1803). The cantons were again defined with boundaries substantially as they have remained ever since. Berne lost most, Aargau and Vaud becoming separate entities at its expense; the Grisons became part of the confederation, with Ticino, St. Gallen and Thurgau. There were now 19 cantons: the new ones, where representative democratic institutions were accepted; the old *Landesgemeinde* cantons, where popular open-air voting continued but was limited to men of 20 years or more; and the city cantons Zürich, Berne, Fribourg, Solothurn, Basle and Schaffhausen, where the old aristocracy regained some of its ancient influence under new forms. With Lucerne substituted for Schaffhausen, these latter became the directing cantons, having seniority in turn, the chief magistrate for the year having custody of the seal of the republic and presiding over the assembly. The sovereign rights of the cantons, the issue of money, control over customs, education, monasteries and the posts, were restored, but freedom before the law, of settlement and of occupation was guaranteed to all citizens.

Napoleon was at first a relatively popular protector until the war with England obliged him to demand the 16,000 men promised him by the military capitulation. British recruiting of mercenaries was stopped and Switzerland forced to participate in the continental system with very great harm to its own commerce and industry, particularly in cotton. The Swiss frontiers were violated by armed forces. Thousands of Swiss took part in the Russian campaign, distinguishing themselves in the crossing of the Beresina

during the retreat but losing most of the effectives. The Valais, which had remained aloof from the Swiss republic was united with France as the *département* of the Simplon in 1810, nominally to put an end to anarchy; Ticino was occupied by French-commanded Italian troops in order to stop smuggling; and Neuchâtel was handed over to Marshal P. A. Berthier as a kind of feudal fief. When therefore it became apparent in 1813 that the Napoleonic empire was crumbling after the battle of Leipzig, the *Landammann* for the year, Hans Reinhard, called a special assembly at Zürich which voted for a policy of armed neutrality. Prince Metternich soon saw this to be on the whole to the advantage of France and encouraged certain discontented elements, who desired a restoration of prerevolutionary conditions, to apply for military help. A force marched through northern Switzerland on its way into France at the beginning of 1814, while the Napoleonic Act of Mediation was formally denounced.

The Restoration.—The congress of Vienna devoted comparatively little time to the Swiss question, but certain of its decisions were none the less of far-reaching importance. Swiss affairs were left largely to the competent advice of Baron Heinrich vom Stein, Baron Wilhelm von Humboldt, Count Capo d'Istria and Stratford Canning, who made as few boundary alterations as possible. The most serious loss was to Austria, which annexed from the Grisons the three districts of Chiavenna, Bormio and the Valtellina. On the other hand, partly as a result of the efforts of Charles Pictet de Rochemont, Geneva was at last united to the confederation with reasonable boundaries, as too was Valais. Neuchâtel, part of the dominions of the king of Prussia, also came in, and the independent existence of the Vaud as a separate canton was accepted. Berne received the city of Biel and the former prince-bishopric of Basle. The northern part of Savoy, on the southern shore of the lake of Geneva, was declared neutral, as was, permanently, the whole Swiss state. This declaration and a guarantee of Swiss integrity was reinforced by the second peace of Paris (Nov. 20, 1815). The return of Napoleon from Elba made no substantial difference; Swiss soldiers in the service of France did not join him, a Swiss contingent marched with the allies, who were allowed to cross Swiss territory, into Upper Burgundy. There was a small modification of the frontier of the Vaud and the creation of a free zone in Gex and its neighbourhood under the treaty of Turin (March 16, 1816). Mulhouse remained in French hands and the Swiss confederation now consisted of 22 contiguous cantons. Its constitution, after a conservative patrician restoration in Berne and some rioting elsewhere, had been settled at the "Long diet" of Zürich in 1814. Its keynote was the full restoration of the sovereignty of the cantons, almost the only federal institution being the army, for the defense of the frontiers, and for the upkeep of which a special fund was voted. Each canton was equally represented in the diet, Lucerne, Zürich and Berne acting as chief city for two years in turn. The possessions of the monasteries were placed under federal protection, but the earlier general freedom of worship, as of settlement and occupation, were not retained. Each canton continued to collect its own customs, coin its own money and manage its own affairs by the customary machinery. The franchise was everywhere limited so as to favour the old ruling families and the property owners, and any alteration of the federal *status quo* was made constitutionally impossible.

Military service with foreign powers again became possible, Louis XVIII gladly enlisting a number of men. Reserving its neutrality, the confederation declared its adhesion to the Holy alliance in 1817, thereby pleasing the tsar. Refugees in some numbers (including the future Napoleon III) obtained asylum within its borders, so that Metternich complained that the country was a focus of revolutionary plots and in a state of anarchy. The diet even agreed to place limitations on the freedom of the press and on the reception of aliens. Economically the country suffered from the rigid tariff restrictions of its neighbours, France and Austria in particular, from the decline in European purchasing power after the Napoleonic wars, from successful British competition in world textile markets and from its own constitutional inability to have any common policy on imports. Apart from some slight manifestations of embryonic nationalism and some rather more substantial signs of a religious revival, the Swiss states showed little activity in the 1820s.

News of the July revolution in Paris in 1830 caused a good deal of excitement among the politically minded young men of certain Swiss cities. Clubs, societies, even athletic associations began to demand larger measures of cantonal democracy, more popular education and some protection of women and children against industrial exploitation. A number of mass demonstrations led to the introduction of more liberal legislation in 11 of the larger cantons in 1831, implying the acceptance of the principle of popular sovereignty and the expansion of the franchise. More schools and training colleges for teachers were opened and universities were founded at Zürich and Berne. There were, however, exceptions and the surge forward was followed by a reaction. In Schwyz the opposition of reformers and conservatives was so violent that only the intervention of federal troops saved the state from civil war or self-partition. In Basle long-standing differences between the relatively wealthy city, with great external interests, and the predominantly agricultural countryside came to a head in 1831. The adoption of a democratic constitution for the state would give political preponderance to the country people and this the citizens were determined to resist. After violent demonstrations in 1831 the canton was divided in 1833 into two halves, Basel-Stadt and Basel-Land, which have remained permanent. Frederick William III of Prussia in 1831 most reluctantly abandoned a small part of his absolute rights in his principality of Neuchâtel (*e.g.*, that of nominating the whole of the cantonal council); but there were clashes between the republicans and the monarchists, neither of which was long satisfied with the compromise agreed to by the monarch. One result of these differences was the formation in 1832 of the Siebnerkonkordat (Zürich, Berne, Lucerne, Solothurn, St. Gallen, Aargau and Thurgau) opposed by the conservatives united in the Sarnenbund (Basel-Stadt, Uri, Unterwalden, Schwyz and Neuchâtel). None of these internal changes affected the traditional willingness of the country to shelter political refugees: Poles were welcomed in considerable numbers in the north, while the harbouring of Prince Louis Napoleon (particularly after his attempted coup at Strasbourg) and the notorious use of Swiss bases by Giuseppe Mazzini almost brought about war.

The Sonderbund War.—The 1830s were years of considerable literary activity and of much thought about the future of the confederation. Strictly constitutional revision was difficult, if not impossible, but the changing needs of the times and the general modification of opinion toward more liberal forms of government clearly suggested change. With a growth of national feeling went a growing conviction that there were certain minimum individual rights of the citizens that ought to be secured to all, and some were prepared to enforce this point of view; others were impressed by the expense and dislocation caused by each canton's having its own tariff and customs barriers. There were several local manifestations of discontent but they remained localized and ineffective. The bond of union was found in religion. How was the principle of religious freedom to be enforced? The Articles of Baden (1834), condemned by Pope Gregory XVI but upheld by the liberal cantons, called for a general diminution of ecclesiastical privileges and exclusive claims; the mass demonstration against the appointment of D. F. Strauss to a chair of theology at Zürich in 1839 indicated the depth of passion aroused. Matters came to a head in Aargau. There in 1835 the monks of Muri had objected to taking an oath of loyalty, and in 1841 the Catholic minority demonstrated in arms against a liberal revision of the constitution. The government then dissolved the Aargau monasteries, which were held to have fomented the revolt. The federal pact of 1815, however, had expressly guaranteed the property of the monasteries against secularization, and the Aargau Catholics appealed to their friends in the south. The chief Catholic canton was Lucerne, and there Joseph Leu had taken advantage of the situation to secure the recall of the Jesuits, invited by Eutych Kopp, to undertake the teaching of theology and the control of the clerical seminary. This action was lawfully done and lay within the proper province of the canton, but was highly provocative. A demonstration against it in Lucerne itself was severely punished, and a general popular de-

mand for the expulsion of the Jesuits from the confederation was ignored. The Radicals passed from argument to violence, Joseph Leu was found shot, and from most of the Protestant cantons bands of volunteers (*Freischärler*) marched on Lucerne under the command of Ulrich Ochsenbein, a Bernese staff officer. The expedition was mismanaged, and in a fight at Malters more than 100 men were killed and more than 1,800 made prisoner (April 1, 1845).

The battle was followed by the union of the Catholic cantons, Lucerne, Uri, Schwyz, Unterwalden, Fribourg and Valais into a separatist confederation, the Sonderbund. The primary purpose was to prevent any alteration of the weak federal act of 1815, even if this were the declared will of the majority of the people, and foreign alliances were welcomed to this end. Through 1846 opposition grew. Berne, Solothurn, Zürich, Vaud and finally Geneva (under James Fazy) elected radical governments, determined to declare the Sonderbund an unlawful pact. In May 1847 St. Gallen joined the liberal cantons, and 12 votes out of 22 were certain. The diet met in Berne in July 1847 under the presidency of Ochsenbein and by a majority declared the Sonderbund to be dissolved. A commission was appointed to draft a new federal pact, and the expulsion of the Jesuits from Lucerne, Fribourg, Schwyz and Valais was ordered, together with their exclusion from the whole confederation. In the name of religion and of state rights the minority determined to resist by force. Their leader, Constantin Siegwart-Müller, purchased arms, collected an army under the command of Johann Ulrich von Salis-Soglio and looked for support from Louis Philippe and Metternich. This, partly because Lord Palmerston openly sympathized with the opposite side, was slow in coming. The Sonderbund War was soon over. While at Lucerne there were divided counsels and uncertainty of purpose, the Radical cantons entrusted their men and fortunes unreservedly to Guillaume Henri Dufour of Geneva, a somewhat conservative professional soldier, simple, upright and unselfish. Dufour had more men than his opponents, and these better trained and equipped, with better guns and superior economic advantages. All these he used to decisive effect. Fribourg capitulated on Nov. 14, 1847, almost without fighting, and the ring closed relentlessly in on Lucerne. Zug surrendered on Nov. 21, and two days later the decisive battles were fought at Meierskappel and Gislikon. Lucerne was entered, officers, Jesuits and nuns fleeing by steamer to Fluelen and then over the St. Gotthard to exile. Valais was the last canton to submit, on Nov. 28. Somewhat more than 100 men on both sides had been killed. The costs of the war were charged to the losers, and Appenzell Inner Rhoden and Neuchâtel had to pay fines for their neutrality. A collective note from the powers, unsupported by Palmerston (who sent Stratford Canning with sympathetic instructions to Berne), could be and was ignored until the greater troubles of 1848 left Switzerland free to work out its own destiny.

The New Federal State, From 1848.—The revision commission had a new constitution ready for submission to the cantons in June 1848, and it was adopted with little change by 15½ cantons (1,897,887 votes) against 6½ (293,371 votes). The central authority of the confederation was strengthened, particularly in regard to defense, law and order and the promotion of the common welfare. There was to be a cabinet or executive (*Bundesrat*) of seven, a supreme federal court (*Bundesgericht*), a senate or council of states (*Ständerat*) consisting of two representatives from each canton and a national council (*Nationalrat*) representing the county as a whole. Foreign affairs became the concern of the federation only, although separate cantons might, with federal approval, make agreements not of a political nature with one another and with neighbouring countries. The posts, currency, weights and measures and customs duties became the province of the federation, which thus acquired an independent revenue, cantonal expenses being met by local levies, taxes on salt, an excise on wine and stamp duties. Universal military service in a federal force was accepted; very small cantonal armies were approved, and the cantons controlled infantry training but not the cavalry or artillery. Service with foreign armies, military capitulations and the acceptance of foreign pensions or decorations were prohibited. Freedom of the press, of worship, of association and of settlement and equality before the law were secured to all, but these were somewhat narrowly interpreted in some cantons—the Jews remained under special disabilities and the Jesuits were excluded. Berne became the seat of the government, Lausanne of the high court and Zürich of the federal institute of technology. It had been the hope of many that a great federal university at Zürich would unite the higher education of the country, the ordinary teaching of children being left to the cantons; but the existing academies at Geneva, Lausanne and Neuchâtel and the universities of Berne and Basle were not prepared to accept this measure of academic unification, and in the end only technology was thus centralized. Similarly it was only after considerable differences of opinion that the French franc and the metric system of weights and measures were fully adopted.

Foreign Affairs.—Neuchâtel, part of the dominions of Frederick William IV of Prussia, had in 1848 adopted a republican form of constitution, which had been accepted as that of a federal canton. This was partly the work of the industrial region of Le Locle and La Chaux-de-Fonds and was resisted by the older ruling families, who proclaimed their loyalty to their monarch. The last of several attempts at counterrevolution took place in Sept. 1856, when the loyalists under Comte Frédéric de Pourtalès-Steiger gained possession of the castle of Neuchâtel and had to be ejected by force. The king of Prussia threatened to intervene on behalf of his "subjects" and mobilized his men; the federal army was placed on a war footing and sent to the northern frontier. British good offices and diplomatic pressure on Prussia by Lord Palmerston were followed by the mediation of Napoleon III, and in 1857 the king gave way, renounced all sovereign rights and secured an amnesty for his supporters.

More serious was the question of refugees. Switzerland, like Great Britain, was a home for political exiles from many countries, and Giuseppe Garibaldi and Mazzini now found safety from the Austrian police there. Mazzini continued to plot the violent overthrow of the lawful rulers of Lombardy and Venetia and attempted to gain recruits for his secret associations until he left the country. In 1850 some measures were taken to prevent Switzerland from becoming too obvious a base for revolutionary propaganda but the law on this subject was never very strictly enforced. In the Ticino the sympathy of the population for Italian nationalism combined with provocative anticlerical actions led to Austrian reprisals in the form of the expulsion of Ticinese from Lombardy and of an economic boycott which caused much suffering. Perhaps only Austrian preoccupation elsewhere prevented matters from being pushed to extremities. During the war of 1859 federal troops prevented Swiss sympathizers from crossing the frontier to join the insurgents and secured the internment of belligerents who fled north. The war was followed in 1860 by the cession of Nice and Savoy to France and the reopening of the question of the neutral zone of Chablais, Faucigny and the Genevois. A federal councillor, Jacob Stämpfli, threatened to readjust matters by a Swiss occupation of this area, a threat impossible to carry into effect as the easy repulse of a Genevan force that attacked Thonon in March 1860 clearly showed. In the end the neutral zone was enlarged a little, arrangements made about the collection of customs and the issue shelved until 1919 (when full French sovereignty over the area was to be asserted and accepted). The war had further been responsible, indirectly, for the establishment of the International Red Cross organization, a widely read pamphlet by Henri Dunant having drawn attention to the unnecessary sufferings of the wounded at the battle of Solferino (1859). In 1864, after much local activity, the Convention of Geneva regulated the care of wounded in war and arranged for the neutralization of the medical and ambulance services, distinguished by the federal cross in reverse colour.

Economic Development.—The Swiss people did not at first welcome the advent of the railway: it was believed to be aesthetically and technically unsuited to mountainous country and it threatened the prosperity of the roadside hotels. The interests of Basle and Zürich conflicted and cantonal state-rights were a serious obstacle. It was therefore not until 1847 that an experimental line

was opened between Baden and Zürich. After a report from Robert Stephenson the federal council accepted the necessity for planned railway construction, and a conflict between Alfred Escher on the side of private enterprise and Jacob Stämpfli on that of state ownership and control resulted in the victory of the former. There was soon a mania for railway speculation, and lines were built with little or no co-ordination, sometimes in conscious rivalry. In 1853 the North Eastern railway linked Zürich with the Rnine and the lake of Constance, and in the same year the route Morges-Lausanne-Yverdon was opened; by 1859 the Central railway joined Basle with Olten and thus to Lucerne on the east and to Solothurn, Berne and Thun on the south. Finance, politics, personalities, cantonal rivalries and strategic considerations soon caused difficulties. The construction of a railway through the St. Gotthard tunnel was not completed until 1882, after much negotiating with Baden and Württemberg and with the Italian government. Escher was chairman of the company responsible, the confederation refusing to accept any financial obligations. The estimates proved inaccurate, and Escher was forced to resign, some part of the excess cost being borne by public funds. There were difficulties about gauge, locomotives and differential rates, and in 1883 a suggestion of unification by public ownership was rejected by a small majority. A similar suggestion was rejected by a referendum in 1891; but in 1898 nationalization of the main lines was accepted by a popular vote, and the principle was extended in 1903 and 1909.

The existence of ample water power and an industrious population gave Swiss industry a good start, but the absence of coal made further developments difficult until the progress of hydroelectric schemes supplied an alternative source of heat and energy. Swiss textile manufactures, including cotton goods, largely held their own in a predominantly free-trade world but declined in face of growing protection and large-scale competition. Embrodiery of good quality, however, was to maintain a wide market until about 1920, as did the Basle silk-ribbon industry. Dyeing led to a considerable development of the chemical industry; and the traditional localized watchmaking and clockmaking was supplemented by large-scale factory production, which met with considerable success and led to the making of a wide range of machines and locomotives. The home production of grain decreased in the 1880s, when it became much cheaper to import foreign wheat; the farmers turned to stock breeding and dairy farming. Cheese making, the export of milk products of all kinds and the chocolate industry became widespread. Better communications with the outside heightened the flow of the valued tourist traffic.

Constitutional Revision and Internal Politics, 1860–1914.—The German *Kulturkampf* had its repercussions in Switzerland. The issue of the *Syllabus errorum* by Pius IX in 1864 attracted a good deal of attention and helped to make the Radicals also to some degree anticlerical, and the definition of papal infallibility in 1870 and the emergence of the Old Catholics in Germany, with many sympathizers in Switzerland, created difficulties. The Catholics of the diocese of Basle, in particular, refused to follow their bishop, Eugène Lachat, in accepting papal infallibility and declared him deposed. In this they were upheld by the cantonal governments, particularly Berne, but a majority of the clergy stood by their bishop, refused to accept his removal and protested against the secularization of his property. Berne now required that parish priests should be chosen by their parishioners from among those who were approved by the government and had undergone a state examination. In Geneva the action of the pope in appointing Gaspard Mermillod to an apostolic vicariate there without any prior consultations led to Mermillod's being expelled from Switzerland (1873), to the recall of the papal nuncio and to the prohibition of the activities of religious associations in schools.

The feeling thus aroused helped forward the growing demand for a general revision of the constitution of 1848. The introduction of the "initiative" (*i.e.*, the right of the people to propose legislation) into the canton of Zürich in 1865 was a victory of the left-wing democrats over Escher and his followers. The freedom of the press was extended, the principle of payments to councillors was accepted, and the example of Zürich was largely copied by Berne, Lucerne and Solothurn. In 1866 the Jews, who had been denied freedom of settlement on commercial grounds, were admitted to full civic rights. In 1874, partly as a result of the Franco-German War, a revision of the conditions of military service was generally demanded. The division between those demanding greater centralization and increased federal authority and those upholding the rights of the cantons was focused on the rivalry of Emil Welti of Aargau and Louis Ruchonnet of the Vaud. The Radicals were only partially successful, but a revised federal constitution was accepted by popular vote in 1874. Direct popular control over the government was further extended by the principle of the referendum, by which a national verdict on an issue could be sought if properly demanded by 30,000 votes or eight cantons. In 1891 it was agreed that a demand by 50,000 votes (the initiative) could bring similar machinery directly into operation. The army was more directly centralized: discipline and military education, the provision of arms and the collection of exemption-payments were to be federal matters, the force being based on the cantons (which also promoted some of the officers). Freedom of trade and choice of occupation within any canton was secured to all, and each canton was obliged to make provision for full primary education. Teaching was to be secular, civil marriages were made compulsory and the foundation of new monasteries was forbidden. Certain other anticlerical regulations were made. The authority and appellate jurisdiction of the federal court at Lausanne was upheld and extended, and federal laws dealing with factories, bank notes and the railways were made more comprehensive. The money needed to meet the additional expenditure was secured by the extension and raising of duties on imports.

The opening of the St. Gotthard route attracted some attention to the Ticino; an increasing number of immigrants heightened party rivalries, Radicals and Conservatives being almost precisely equally balanced. Religious feeling made matters worse; the Conservatives were accused of manipulating the elections so that they would remain permanently in power; and rioting in 1890 ended in the formation of a provisional government. Armed federal intervention became necessary, a revised electoral law introduced proportional representation, and an increased electorate was followed by a Radical and peaceable administration. Growing industrialization and an increase in the size of the urban populations were accompanied by the spread of socialism. The Social Democratic party formed in 1870 avowedly existed to forward the claims of the manual workers and, under Otto Lang, began to demand social reforms and positive opposition to large-scale capitalist enterprise. A notable piece of practical socialism was the slow nationalization of the railways. Some action was necessary to prevent important sections of the railways from passing under foreign control, and in 1891 the federal council negotiated the purchase of the Central railway; but the proposal was negated by a popular referendum. Seven years of discussion followed, accompanied by an occasional strike. Then in 1898 agreement was secured for the public purchase of certain main lines. The principle was extended in 1903 and 1909 and was to make much easier the later electrification of the federal lines.

Political refugees, their numbers increased by German anti-socialist legislation, continued to be a source of difficulty. In 1889 a German police inspector, August Wohlgemuth, arrested as a spy, was expelled from the country only after an acrid exchange of notes between the Swiss government, represented by Numa Droz, and Bismarck. Anarchists were more difficult to control or eliminate. In 1898 the Austrian empress Elizabeth was assassinated in Geneva; and in 1902, after the murder of King Humbert of Italy, the Italian government demanded that action should be taken against Giulio Silvistrelli for provocative newspaper articles. A temporary breaking-off of diplomatic relations between Switzerland and Italy followed, but normal intercourse was soon restored. The principle of permanent neutrality had become firmly accepted by the people when World War I began in Aug. 1914.

World War I.—Attention in 1914 in Berne was soon transferred from the Swiss national exhibition to the needs of defense. The wisdom and the value of the policy of neutrality, imposed

alike by public instruments, by national interest and by tradition, was never in doubt. The citizen-army was rapidly mobilized and sent to guard the frontiers. There was a momentary controversy over the choice of a supreme commander, the parliamentary representatives voting for Theophil Sprecher von Bernegg, the federal council (supported by a decisive popular vote) choosing Ulrich Wille. Fortunately a neutral Switzerland had certain advantages both for France and for Germany, and no attack upon its frontiers was ever seriously considered. Universal military service, the rigours of campaign conditions without an enemy to fight and the use of soldiers for public works were less resented than the economic conditions imposed on a country which could neither entirely feed itself nor supply all its own raw materials. The tourist traffic, which the 20th-century enthusiasm for mountaineering and allied pursuits had rendered especially active, came to an end; a run on the banks was with difficulty averted; and the large number of resident foreigners added certain difficulties. But it was soon apparent that Swiss services could be of value alike to both belligerent groups and a certain amount of trade became possible. For certain purposes Cette became an open port for Switzerland. Heavy taxation and a good deal of borrowing were necessary; there was a violent controversy when two officers, Col. K. Egli and Col. M. von Wattenwyl, were discovered to have given military information to the Germans (Feb. 1916); sympathy with the Allied war aims was most prominently expressed by the French-speaking Swiss. The factory workers in the cities were discontented at the rapidly rising prices and at political and industrial conditions unfavourable to them; and the Social Democratic party became a vehicle for the teaching of Lenin, resident in Zürich until his transference to Russia by the Germans. The constitution, however, made political action slow and ineffective, and in Nov. 1918 a general strike was called. The workers demanded a 48-hr. week, old-age pensions, national health insurance, female suffrage, the election of a new national council by proportional representation, conscription of labour and a capital levy. Prompt action by the administration, the hurried expulsion of the Soviet legation, the arrest of strike leaders and the use of some show of military force ended the general strike quickly, after which a number of the demanded reforms were enacted. Concurrently with the general strike the serious influenza epidemic that affected so much of Europe wrought considerable havoc.

Swiss agencies, governmental and private, including the Red Cross, were much used by the belligerents, but the expression of Swiss opinion had little effect on the course of the war or on the peace negotiations in Paris. The collapse and partition of the Habsburg empire decisively altered conditions on the Swiss eastern frontier, where the inhabitants of the Vorarlberg wished to form themselves into a new canton of the confederation. The decision about this rested with the Allies, who declined to detach the territory in question from the new Austrian state and ignored the principle of self-determination, so that the treaty of Saint-Germain-en-Laye left the frontier there unchanged. Switzerland was affected in several ways by the treaty of Versailles: the international control of shipping on the Rhine thereby became very much a concern of France; the agreement with Germany about the regulation of the St. Gotthard line was revised; and the neutralized "free" zone of Gex and Upper Savoy became an undifferentiated part of the French republic.

The League of Nations.—The Covenant of the League of Nations, in spite of the flattering choice of Geneva as the headquarters of the organization, was not regarded with satisfaction in Switzerland. The guarantees for the maintenance of peace in Europe were less adequate than the Swiss project had suggested, and it was obviously difficult for a permanently neutral state to be associated with the enforcement of collective security and with active resistance to aggression. But Swiss idealism welcomed the general implications of the League, and outstanding difficulties were solved when, by the Declaration of London (Feb. 13, 1920), the council of the League recognized the special position of Switzerland as permanently neutral and agreed that no military assistance, direct or indirect, could be required from it in the event of a breach of the covenant, although "economic sanctions" would become applicable. The issue was long debated; the federal council advised entry into the League, which was accepted by the *Nationalrat* by 115 votes to 50, by the *Ständerat* by 30 to 6. The necessary referendum on May 16, 1920, revealed even greater caution: 11½ states were for, 10½ against, and the final figures revealed a favourable majority of 90,000 in a total vote of 700,000. On the social, economic, intellectual, industrial and medical committees of the League of Nations Swiss help was prominent. The country readily accepted the Kellogg pact abolishing war as an instrument of policy, and it advocated the early admission of Germany into the League. The passage of foreign armed forces across Switzerland, even on the most direct and manifest League missions, was strenuously and successfully resisted. Relations with the U.S.S.R., embittered by the murder of the Soviet representative V. V. Vorovsky at Lausanne in 1923, remained unsatisfactory, and Switzerland opposed the admission of that country into the League in 1934. In 1935 trade relations with Italy were not broken off when Italian forces invaded Abyssinia, although an embargo was placed on the export of arms to both sides. The Sino-Japanese War, the failure of the Disarmament conference, the war in Spain and the successful *Anschluss* between Germany and Austria caused a national reaction, and in May 1938 Switzerland publicly refused to participate in future sanctions.

World War II and After.—Switzerland had foreseen the outbreak of a general conflict and was militarily prepared for World War II. On Aug. 30, 1939, Henri Guisan was elected general (commander in chief) by a large majority. The army was mobilized, the frontiers guarded and conscription of labour enacted. The tourist traffic and with it the hotel industry, already affected by European conditions, suffered greatly, and both imports and exports were exposed to severe and immediate limitations. There are many expensive diplomatic and Red Cross activities, and with the fall of France in 1940 a large number of French and Polish troops crossed the frontier and were interned. Belligerent aeroplanes flew high over Swiss territory and were seldom shot down; certain towns, notably Schaffhausen, were accidentally bombed. Defense measures were deliberately concentrated upon the mountainous interior, and these were not put to the test. War materials were supplied to both sides and conveyed by rail between Germany and Italy. Home agriculture, corn-growing in particular, received all possible encouragement, and the country emerged from the war less impoverished than might have been expected. A careful censorship was exercised over the press; incitement was avoided, as dislike for German National Socialist activities increased and the Communist party was proscribed.

Diplomatic relations with the U.S.S.R. were reopened in 1945, and every effort was made to co-operate with international organizations that was consistent with the maintenance of the increasingly difficult principle of permanent neutrality.

Swiss currency remained remarkably stable in the 1950s but rising world prices had their effect: a number of trade agreements were negotiated and every effort made to extend exports in the face of increasing competition and tariff barriers. A country solidly opposed to Communism and to spending increased amounts on defense, it upheld its tradition of neutrality. Switzerland did not accept membership of the United Nations organization or make any direct contribution to western defense or join the Western European union.

BIBLIOGRAPHY.—J. Dierauer, *Geschichte der schweizerischen Eidgenossenshaft*, 5 vol. (Gotha, 1920–21), vol. 6 (Gotha, 1931); E. Gagliardi, *Geschichte der Schweiz*, 3 vol. (Zürich, 1934–37); E. Dürr, R. Feller, L. von Muralt and H. Nabholz, *Geschichte der Schweiz* (Zürich, 1930); W. Oechsli, *History of Switzerland, 1499–1914* (Cambridge, 1922); G. Soloveytchik, *Switzerland in Perspective* (London, New York, 1954); E. Bonjour, H. S. Offler and G. R. Potter, *Short History of Switzerland*, 2nd ed. (Oxford, 1955); Jacqueline Belin, *La Suisse et les Nations Unies* (New York, 1956); H. Kohn, *Nationalism and Liberty: the Swiss Example* (London, New York, 1956).

(G. R. P.)

POPULATION AND SOCIAL CONDITIONS

Switzerland has an area of 41,287 sq.km. or 15,941 sq.mi. of which 22.6% is "unproductive" (*i.e.*, bare mountains, plateau

slopes, lakes and rivers), 24.8% is occupied by forests, 24.3% consists of alpine pastures and only 28.3% is agricultural land.

Well-organized census returns were made at ten-year intervals from 1850 on, except that censuses were taken in 1888 and not in 1890, and in 1941 and not in 1940. The total population by the census of Dec. 1, 1960, was 5,429,061. This represented an increase of 714,069 or 15.1% over the figure for 1950 and was more than double that for 1850 (2,340,740). One-half of the increase since 1950 was due to natural causes and the remainder to immigration. The cantons registering the greatest gains were the industrial ones of Zürich, Zug, Basle and Geneva. (*See* Table I.) There was a steady increase in density of population from 150 per square mile in 1850 through 207 per square mile in 1900 and 255 per square mile in 1930 to 341 per square mile in 1960. The least densely populated cantons are the Alpine ones; *e.g.*, Grisons (53 per square mile) and Uri (77). The Jura cantons are much better populated: Vaud (346) and Neuchâtel (479). The densest are the cantons of Geneva (2,378), Basle Land–Basle-Stadt (2,089), and Zürich (1,428).

TABLE I.—*Area and Population of Switzerland by Cantons, 1941–60*

Cantons*	Area in sq. mi.	Population		Increase		Pop. per sq. mi. 1960
		1950	1960	Number	Per cent	
Zürich (Zurich), 1351	667	777,002	952,304	175,302	22.6	1,427.7
Bern (Berne), 1353	2,659	801,943	889,523	87,580	10.9	334.5
Luzern (Lucerne), 1332	577	223,249	253,446	30,197	13.5	439.2
Uri, 1291	415	28,556	32,021	3,465	12.1	77.2
Schwyz, 1291	350	71,082	78,048	6,966	9.8	223.0
Obwalden (Unterwald-le-Haut), 1291†	190	22,125	23,135	1,010	4.6	121.8
Nidwalden (Unterwald-le-Bas), 1291†	106	19,389	22,188	2,799	14.4	209.3
Glarus (Glaris), 1352	264	37,663	40,148	2,485	6.6	152.1
Zug (Zoug), 1352	92	42,239	52,489	10,250	24.3	570.5
Fribourg (Freiburg), 1481	645	158,695	159,194	499	0.3	246.8
Solothurn (Soleure), 1481	305	170,508	200,816	30,308	17.8	658.4
Basle-Stadt (Bâle-Ville), 1501†	14	196,498	225,588	29,090	14.8	16,113.4
Basle Land (Bâle-Campagne), 1501†	165	107,549	148,282	40,733	37.9	898.7
Schaffhausen (Schaffhouse), 1501	115	57,515	65,981	8,466	14.7	573.7
Appenzell Ausser Rhoden (Appenzell Rhodes Extérieures), 1513†	94	47,938	48,920	982	2.0	520.4
Appenzell Inner Rhoden (Appenzell Rhodes Intérieures), 1513†	67	13,427	12,943	−484	−3.6	193.2
St. Gallen (Saint-Gall), 1803	778	309,106	339,489	30,383	9.8	436.4
Graubünden (Grisons), 1803	2,745	137,100	147,458	10,358	7.6	53.7
Aargau (Argovie), 1803	542	300,782	360,940	60,158	20.0	665.9
Thurgau (Thurgovie), 1803	389	149,738	166,420	16,682	11.1	427.8
Ticino (Tessin), 1803	1,085	175,055	195,566	20,511	11.7	180.2
Vaud (Waadt), 1803	1,240	377,585	429,512	51,927	13.8	346.4
Valais (Wallis), 1815	2,020	159,178	177,783	16,605	11.7	88.0
Neuchâtel (Neuenburg), 1815	308	128,152	147,633	19,481	15.2	479.3
Genève (Genf, Geneva), 1815	109	202,918	259,234	56,316	27.8	2,378.3
Total	15,941	4,714,992	5,429,061	714,069	15.1	340.6

*The cantons are given in the official order with the year of the entrance of each into the confederation. †Half canton.
Source: *Statistisches Jahrbuch der Schweiz 1956* (Basle, 1957).

The non-Swiss element of the population increased from 3% in 1850 to 11.6% in 1900, but fell to 8.8% in 1930; its total increased from 71,570 in 1850 to 355,522 in 1930. By 1941 there were only 223,554 aliens in the country (5.2%), the most numerous being Italians (96,018); next in order came Germans (68,622), French (24,396) and Austrians (9,652). In 1950 there were 283,367 foreigners. From 1800 to 1940 about 683,000 Swiss emigrated, mainly beyond the seas. The annual average of emigrants was 6,800 in the decade 1871–80, reached 11,500 in the years 1881–88, fell to 4,400 in the years 1889–1900, reached the peak of 11,800 in the years 1910–20 and was about 6,000 in the decade 1931–40. The number of Swiss-born residents in the United States rose steadily from 13,358 in 1850 to 124,834 in 1910; in 1940 the number was 88,293 and in 1950 71,515.

Language.—By the federal constitution of 1874, German, French and Italian are recognized as national languages for the purposes of debates in the federal parliament and for the public notification of federal laws and decrees. The recognition on July 8, 1937, of Romansh (*see* RAETO-ROMANCE DIALECTS) as a fourth national language did not affect the constitution. German predominates in 16 cantons, including even the Grisons (where, however, about one-third of the people speak Romansh in the form known as Ladin and about one-tenth Italian); French prevails in Vaud, Neuchâtel, Geneva, Fribourg and Valais; Ticino is Italian-speaking. Detailed census returns as to language had revealed a small amount of shifting. (*See* Table II.)

TABLE II.—*Distribution of Population by Language*

Year	German	French	Italian	Romansh	Others
1880	2,030,792 (71.3%)	608,007 (21.4%)	161,923 (5.7%)	38,705 (1.4%)	6,675 (0.2%)
1910	2,594,186 (69.1%)	793,264 (21.1%)	302,578 (8.1%)	40,234 (1.1%)	23,031 (0.6%)
1950	3,399,636 (72.1%)	956,889 (20.3%)	278,651 (5.9%)	48,862 (1.0%)	30,954 (0.7%)

Religion.—The constitution of 1874, while recognizing no established Swiss church, guarantees full religious liberty and freedom of worship, as well as exemption from any compulsory church rates. It repeats the constitution of 1848 in forbidding the settlement of Jesuits and their affiliated societies in Switzerland; the interdiction can be extended to any other religious orders that may endanger the safety of the state or interfere with the peace of other creeds. Protestants outnumber the Catholics in 12 cantons, of which 9 are German-speaking and 3 (Geneva, Vaud and Neuchâtel) French-speaking. The Catholics outnumber the Protestants in 10 cantons, including 7 German-speaking, 2 French-speaking (Fribourg and Valais) and the Italian-speaking Ticino.

The Roman Catholic priests are more numerous than Protestant clergy and comprise more than 6,000 priests under five diocesan bishops: viz., those of Basle and Lugano (resident at Solothurn), of Lausanne and Geneva (resident at Fribourg), of St. Gallen, of Sion (Sitten) and of Coire (Chur). (*See* Table III.)

TABLE III.—*Distribution of Population by Religion*

Year	Protestant	Catholic	Jewish	Others
1880	1,667,100 (58.6%)	1,160,782 (40.8%)	7,373 (0.2%)	10,838 (0.4%)
1910	2,107,814 (56.2%)	1,503,538 (42.4%)	15,462 (0.5%)	33,470 (0.9%)
1950	2,655,375 (56.3%)	1,959,046 (41.6%)	19,048 (0.4%)	81,523*(1.7%)

*Including 28,568 "Christian Catholics."

Cities.—In 1960, 13 towns had a population exceeding 30,000.

TABLE IV.—*Cities With Population of More Than 30,000 in 1960*

City	1850	1900	1960	City	1850	1900	1960
Zürich	41,585	168,021	440,170	Lucerne (Luzern)	10,068	29,255	67,433
Basle	27,844	109,161	206,746	Bienne (Biel)	5,609	29,557	59,216
Berne	29,670	67,550	163,172	La Chaux-de-Fonds	12,638	35,968	38,906
Geneva	37,724	97,359	176,183	Neuchâtel	7,901	21,195	33,430
Lausanne	17,108	46,732	126,328	Fribourg	9,065	15,794	32,583
St. Gallen	17,858	53,796	76,279	Schaffhausen	8,064	15,840	30,904
Winterthur	13,651	40,961	80,352				

Education.—Under the 1848 constitution each canton or half-canton is sovereign in education. The constitution stipulates that elementary education is compulsory and free; that all public schools must be state directed; and that religious freedom be guaranteed. The confederation financially supports public elementary schools and vocational training. By regulating examinations for the professions, also influences the curriculum of the higher middle schools. The only institution of higher education maintained by the confederation is the Federal Institute of Technology at Zürich (*Eidgenössische Technische Hochschule*). The ministry of the interior in Berne is in charge of education; there is an educational department in every canton.

Children enter the primary school at the age of 6 or 7 and leave it at 14 or 15. Secondary and lower middle schools form a link between primary and higher or vocational schools. Pupils enter between the ages of 11 and 13, remaining for two to five years; they are grouped in language-history and mathematics-science branches. The technical or vocational schools serve pupils under-

going an apprenticeship between the ages of 15 and 19. The higher middle schools either take their pupils from the senior classes of a primary school or themselves provide the preparatory schooling; in the latter case the course lasts 12½ years. They prepare their pupils for the higher school certificate in three types of schools: classical (with compulsory Latin and Greek), modern language and science; there are boys', girls' and coeducational institutions.

In addition to the Federal Institute of Technology at Zürich (founded 1855), there are seven cantonal universities; viz., Basle (1460), Lausanne (1537), Geneva (1559), Zürich (1833), Berne (1834), Neuchâtel (1838) and Fribourg (1889). The commercial college of St. Gallen (1898) ranks also as an institution of higher education. At Lausanne, Geneva and Neuchâtel instruction is given only in French; at Basle, Zürich and St. Gallen only in German. Fribourg and Neuchâtel have no medical faculties. Swiss universities are lay (*laïques*), but all, except Fribourg, are in cantons of Protestant tradition. At Basle, Geneva, Lausanne, Neuchâtel and Zürich the theological faculties are Protestant; at Berne there are Protestant and Catholic faculties; at Fribourg only a Catholic one. Of the 16,501 students in 1950–51 nearly a quarter were foreigners; at Lausanne one in three students is a foreigner, at Geneva one in two. (*See* Table V.) (K. Sm.)

TABLE V.—*Education*

Type of school	Year	Pupils			Teaching staff
		Boys	Girls	Total	
Primary schools	1953–54	263,281	254,769	518,050	15,204
Secondary schools	1953–54	29,902	29,442	59,344	2,462
Lower middle schools	1953–54	14,982	11,151	26,133	980
Higher middle schools	1953–54				
Classical and modern		6,262	2,540	8,802	1,034
Mathematics-sciences		2,581	157	2,738	
Vocational training schools	1947–48	85,744	75,719	161,463	—
Commercial colleges	1947–48	2,172	1,549	3,721	157
Teachers' training colleges	1947–48	1,143	1,245	2,388	257
Universities	1956–57	10,537	2,326	12,863	1,569
Fed. Inst. of Technology	1956–57	2,866	122	2,988	437
St. Gallen Comm. college	1956–57	596	18	614	89

GOVERNMENT

The Cantons.—The political divisions are the communes (of which there are now about 3,000), districts (198), and cantons (22). Of the cantons, 19 are undivided. In 1831 the rural districts of Basle revolted against the undemocratic rule of the urban trade guild; the latter were twice defeated in the field before they agreed (in 1833) to the subdivision of the canton into Basle-Stadt and Basle Land. Unterwalden, from before 1291, was divided into Obwalden and Nidwalden, representing two distinct physical divisions in different river valleys. In 1291, Nidwalden accepted alliance with Uri and Schwyz; Obwalden did not join until later. The Reformation led to a division (1597) of Appenzell into the Inner Rhoden, mainly Roman Catholic and pastoral, and Ausser Rhoden, largely Protestant and industrial.

Each of the 22 cantons has its own legislature, executive and judiciary. The older cantons have in some instances (Unterwalden, Appenzell and Glarus) preserved their ancient democratic assemblies (*Landesgemeinden*), in which all male citizens of full age meet, usually annually during April or May, in the open air for the purpose of legislation and the selection of an annual administrative council. In the remaining cantons the legislature (*Grosser Rat* or *grand conseil*) is composed of representatives chosen by universal male suffrage and usually by proportional representation. These councils exercise all the functions of the *Landesgemeinden*. The executive also in most of the newer cantons is elected by a popular vote and not as hitherto by the cantonal legislature. All the cantons have the referendum and the popular initiative, the application of which varies in the different cantons.

The cantons are built up of two main types of communes, or *Gemeinden*. These are the real units. The burgher communes (*communes bourgeoises* or *Bürgergemeinden*) are now principally of historical interest, having for the most part been gradually absorbed by the second class of political communes. Initially burgher communes were groupings of the people to manage the "lands subject to common user" (mainly summer pastures and forests). Later, by purchase or otherwise, the burghers became possessors of manorial rights. But when the federal diet imposed the care of the poor on the several communes (1551) these naturally aided only their own burgher members. Thus, all nonburgher settlers were excluded from any share in the enjoyment of the common lands, or in their management, though they paid local rates. Increased communication facilities, together with a shifting industrial population, made such restrictions invidious and unfair, particularly after the introduction, under the Helvetic republic (1798–1801), of a federal citizenship superior to cantonal citizenship. Later, the increase of communal public duties meant financial obligations exceeding the sums produced by the common lands. To avoid some of these inconveniences, political communes (*municipalités* or *Einwohnergemeinden*) were set up. The duties of these communes were largely increased after the liberal movement of 1830; the care of the highways, the police, the schools, the administration of the poor law being successively handed over to them. Swiss citizens belonging to cantons other than those in which they resided were, by the federal constitution of 1848, given rights of voting there in cantonal and federal matters, but not in those relating exclusively to the commune itself, but the federal constitution of 1874 gave to permanent Swiss settlers communal voting rights, subject to three months' residence.

In most cantons a number of communes are grouped together to form a district (*Amtsbezirk*) under a prefect who represents the cantonal government. Minor communes are controlled by a small council. In the larger communes an executive council is chosen by a general assembly of all male Swiss citizens over 20 years of age, of good conduct and with three months' communal residence. A *maire* (*Gemeindepräsident*) is elected by the larger body to preside over the council which has the management of all local affairs, including the carrying out of cantonal and federal laws or decrees, save and except matters relating to the pastures and forests held in common.

The Confederation.—On the cessation of the federal pact in 1848 a new constitution was accepted by general consent. This is fundamentally the constitution in operation today, but sundry revisions were made in 1874 and 1931, and the amended constitution is the one now in force. Any subsequent revision can be made by normal federal legislation with a compulsory referendum on the demand of 50,000 franchised citizens, or by popular initiative. Changes require a favourable majority in total votes and total number of cantons.

Federal sovereignty is exercised by the legislative federal assembly (*Bundesversammlung*) and by the executive federal council (*Bundesrat*), both meeting at Berne. The assembly consists of two chambers: the national council (*Nationalrat*) and the council of the states (*Ständerat*).

Every whole canton is represented in the national council, on the basis (from 1950) of one member for every 24,000 people in the canton, with an additional deputy for any remainder exceeding 12,000. On the census returns of 1950 there were thus 196 representatives varying from 33 for Berne and 32 for Zürich to 1 for Uri. The deputies receive attendance pay and travelling expenses from federal funds; neither allowance is lavish. (*See* Table VI.)

TABLE VI.—*Elections to the National Council*

Party	1922	1925	1928	1931	1935	1939	1943	1947	1951	1955
Liberal Democratic	10	7	6	6	6	6	8	7	5	5
Catholic Conservative	44	42	46	44	42	43	43	44	48	47
Peasants, Artisans and Middle Class	34	30	31	30	21	22	22	21	23	22
Radical Liberal	60	60	58	52	48	50	47	52	51	50
Independents'	—	—	—	—	7	9	5	9	10	10
Democratic	4	5	3	2	4	4	6	5	4	4
Social Democratic	43	49	50	49	50	45	56	48	49	53
Labour (Communists)	2	3	2	2	2	4	—	7	5	4
Others	1	2	2	2	7	4	7	1	1	1
Total	198	198	198	187	187	187	194	194	196	196

The council of the states consists of two members from each canton or one from each half-canton; *i.e.* 44 in all. The mode

of their election, term of membership and rate of pay are entirely cantonal matters and vary accordingly. Their average financial allowance is less than that for a deputy to the national council.

A general election of representatives to both chambers takes place every four years; voting is by proportional representation with universal male franchise for persons of 20 years of age or more, and any voter, other than a clergyman, may be elected a deputy. Laws passed by the federal assembly, though this represents both people and canton, can be vetoed as a result of a majority-supported referendum based on a popular initiative demand by 30,000 enfranchised citizens.

The seven members of the federal council or government are elected for four years by the federal assembly and are responsible for the following departments: foreign affairs; justice and pelice; interior; defense; finance and customs; public economy; and railways and post office. No canton may have more than one representative in the federal council; but according to custom it should always include a Bernese, a Zurichese and a Vaudois. Usually four members are German-speaking, two French-speaking and one Italian-speaking.

The president of the federal council and the vice-president of the federal council (who is one of its seven members) are elected for one year only (Jan. 1–Dec. 31) by the federal assembly and cannot be immediately re-elected. Federal councilors cannot hold a plurality of office, either federal or cantonal, and must not engage in any profession or trade. The federal government (assembly and council) is being increasingly entrusted with matters formerly considered to be purely cantonal.

Justice.—Each canton elects, controls and pays (often merely a nominal sum, even if the posts are not entirely honorary) its own magistracy for ordinary civil and criminal trials. There are cantonal variations in the penal code, *e.g.*, capital punishment exists only in nine cantons and one half-canton.

The federal supreme court (*Bundesgericht*) sits at Lausanne. There are 26–28 full members (plus 9 supplementary judges), elected by the federal assembly to hold office for six years with eligibility for re-election. The federal assembly also elects every two years the president and vice-president who, as such, cannot be re-elected. The original and final jurisdiction of the *Bundesgericht* extends to disputes between the confederation, the cantons, corporations, and private individuals, so far as these differences refer to federal matters and involve more than 4,000 Fr. in financial suits. It is a court of appeal against cantonal authorities in the application of federal laws and also against decisions of other federal departments. It is a court of trial for persons accused of treason, or other offenses against the confederation. The four courts of the *Bundesgericht* are: the court of accusation, the criminal court (with paid jurors elected by the people), the federal penal court and the court of cassation. (W. E. Wh.; X.)

DEFENSE

The Militia System.—Being composed of militia, the Swiss army contains no forces maintained permanently with the colours except a corps of 327 instructors: a system which is unique in Europe. The total number of men physically fit and trained for military duty was approximately 700,000 in 1952.

All male citizens are liable to military service from the beginning of the year in which their 20th birthday falls until the year of their 48th birthday. Service may be rendered either personally or by the payment of a military tax. Volunteering before the legal age is permitted. The military tax is payable up to the age of 48. Enlistment takes place in the year of the 19th birthday. Personal service includes attendance for training, active service at home or abroad in defense of the country, and maintenance of public order and security. Men passed fit for auxiliary service pay a tax in lieu of military training for the years in which it would otherwise be performed. After physical training for boys, provided by the cantons under the supervision of the federal government, the training is carried out under the military department.

The army consists of a first line (*Auszug* or *élite*) for ages from 20 to 32 inclusive, *Landwehr* for ages from 33 to 42 inclusive, and *Landsturm* for ages from 43 to 48 inclusive. After a recruit course varying from 62 to 118 days according to the arm of the service envisaged, members of the *élite* undergo seven 20-day annual trainings. Men in the *Landwehr* and the *Landsturm* are called up for a few repetition trainings (*Wiederholungskurse*) not exceeding 60 days in all. Officers are trained in cadet schools, courses lasting from 62 to 104 days, according to the arm of the service envisaged.

As the militia keep their rifles in their own houses, the whole army can be mobilized in 48 hr. In 1939 about 500,000 were mobilized, but after the collapse of France in 1940 the number under arms was reduced to 250,000.

Order of Battle.—The country is divided into nine divisional districts. There are also four independent mountain brigades and three armoured brigades. There is a staff organization for four army corps.

The 37 infantry regiments are armed with special infantry guns and mortars and anti-tank guns. The 17 artillery regiments are armed with 105-mm. guns and 105-mm. and 150-mm. howitzers. There is a small air force, which by 1953 had 170 jet fighter aircraft in the first line.

In 1950 Switzerland started an extraordinary rearmament program, the total cost of which amounted to about 1,400,000,000 Fr. In 1950 the total military expenditure amounted to 480,000,000 Fr., four times as much as in 1938. The military estimates for 1956 amounted to 673,000,000 Fr.

Higher Command.—In peacetime the Swiss army has no general, colonel being the highest rank. One of the seven federal councillors is head of the military department, the central authority in peacetime for national defense. If troops are mobilized on a large scale the federal assembly appoints a commander in chief, as it did in 1939 when Gen. Henri Guisan was appointed. There is a general staff under the military department and a national defense committee, with the head of the military department as chairman and the chief of the general staff and with four army corps commanders and the officer commanding infantry as members. The committee ceases to function when a commander in chief is appointed.

ECONOMIC CONDITIONS

According to the census of 1950, the economically active population was employed as follows: 46.6% in the manufacturing industry, handicrafts, building, mining and quarrying; 16.5% in agriculture and forestry; 11.8% in commerce, banking and insurance; 7.6% in public administration and the professions; 4.2% in the hotel industry; 4.5% in transport and communications; and the rest in domestic and other services. As these proportions clearly show, Switzerland is preponderantly an industrial country, depending for its existence and prosperity on the export of machinery, apparatus and instruments. This is all the more surprising as the country has practically no useful mineral resources except salt and is entirely landlocked, without direct access to overseas markets. Apparently condemned by nature to a meagre pastoral economy, the Swiss by hard work, inventiveness and perseverance have succeeded in developing other industries but their hold is always precarious. (*See* Table VII).

TABLE VII.—*Economically Active Population*

Category	1910	1930	1950
Agriculture, forestry, fishing	477,118	413,366	355,427
Mining and quarrying	6,009	7,531	6,263
Manufacturing, handicrafts, building	805,148	852,371	908,041
Commerce, banking, insurance	117,500	190,520	252,481
Hotels	78,159	94,321	90,791
Transport and communications	84,734	84,998	98,014
Public administration and professions	80,900	116,714	164,252
Domestic service	105,935	133,902	111,723
Day labourers	8,640	6,448	4,881
Institutions with resident staff	12,902	42,495	64,831
Unemployed			8,952
Total	1,783,195	1,942,626	2,155,056

Agriculture.—The mountainous character of the large part of Switzerland, the high average altitude, the excessive rainfall and the prevailing system of small holdings all combine to prevent the country from being preponderantly agricultural. From the middle of the 19th century the rural population has declined: in 1888 the proportion of the economically active population engaged in agriculture was 37% as compared with 20.8% in 1941. At the same time the country has become increasingly dependent on imports for its food supply, especially for its bread grain. By 1850 the grain production was sufficient to cover about two-thirds of domestic requirements; in 1934–48 this proportion was reduced to one-third. Switzerland is predestined rather to be a cattle-raising country. Milk production per cow is high and reaches about 2,900 kg. or 740 gal. annually; *cf.* an average of about 500 gal. in the United States. (*See* Tables VIII, IX and X.)

Industry.—Practically all raw materials, be they minerals or vegetable or animal fibres, are imported; and hydraulic power is the only source of energy to be found in the country. Having no appreciable home market and being not only stifled by the protectionist policies of all their customers abroad but also burdened with very

TABLE VIII.—*Agricultural Production*

(In 000 metric tons)

Crop	1934–38	1948–52	1955	1956	1957
Wheat*	160	260†	321	232	335
Barley	8	55	63	95	60
Rye	33	34†	41	35	40
Oats	22	68	58	75	58
Potatoes	736	1,030	1,100	1,500	1,300
Sugar beet	74	207	210	224	230

*Includes estimates of spelt. †1948–50.

TABLE IX.—*Livestock*
(In 000 head; date of estimate: April)

Type	1947–52	1953	1954	1955	1956	1957
Cattle	1,544	1,635	1,593	1,583	1,645	1,643
Pigs	892	1,017	950	1,038	1,150	1,160
Sheep	186	185	195	195	201	
Horses	135	128	125	120	117	113

heavy transport charges, Swiss manufacturers cannot indulge in mass production. They can thrive only by producing wares which will find purchasers even if expensive and which present a very high value in a reduced volume. That is the case with intricate producers' goods such

TABLE X.—*Swiss Production and Net Import of Major Foods*
(In 000 metric tons; 1934–38 annual average)

Food	Production	Net Import	Total
Meat	190.3	7.6 (4%)	197.9
Fish and poultry	5.8	9.7 (63%)	15.5
Eggs (000,000 units)	423.2	238.2 (36%)	661.4
Fluid cow milk	1,016.8	9.2 (1%)	1,026.0
Cheese	50.7	16.6*	34.1
Butter	26.3	0.9 (3%)	27.2
Other animals fats	11.7	3.5 (23%)	15.2
Vegetable fats and oils	—	28.1 (100%)	28.1
Sugar	9.4	148.0 (94%)	157.4
Bread grains	203.9	421.1 (67%)	625.0
Potatoes	334.4	31.6 (9%)	366.0
Citrus fruit, etc.	—	45.2 (100%)	45.2
Wine (000,000 l.)	65.0	92.0 (59%)	157.9

*Net export.
SOURCE: J. Rosen, *Wartime Food Developments in Switzerland* (1947).

as complicated machinery, durable consumers' goods such as watches or hoots and shoes, and luxury products of the textile industry.

Switzerland is richly endowed by nature in hydroelectric potentialities. Its total installed capacity rose between 1929 and 1955 from 1,520.000 kw. to 3,720,000 kw. (incl. 3,520,000 kw. from hydroelectric power stations). The production of electric power rose between 1937 and 1956 from 6,842,000,000 kw.hr. to 14,892,000,000 kw.hr. Switzerland has the fifth highest output of electricity per capita in the world (after Norway, Canada, Sweden and the United States). The production of cement in 1956 was 2,379,000 metric tons (as compared with 995,000 tons in 1947).

Foreign Trade.—The importance of foreign trade for Switzerland is reflected in the annual value per head of the population of imports and exports together. In 1938 that figure, in Swiss francs, was 464; *cf.* 564 for Belgium, 523 for the Netherlands and, on the other hand, 430 for Great Britain. Imports, consisting of such necessities as raw

TABLE XI.—*Foreign Trade*
(In millions of francs)

Type	1938	1949	1950	1951	1952	1953	1954	1955	1956
Imports	1,607	3,791	4,536	5,916	5,206	5,071	5,592	6,401	7,597
Exports	1,317	3,457	3,911	4,691	4,749	5,165	5,271	5,622	6,203
Trade balance	−290	−334	−625	−1,225	−457	+94	−321	−779	−1,394

materials, fuel and food, surpass exports in value, but the resulting deficit is generally covered by the tourist trade. (*See* Table XI.)

Watches, clocks and parts were both before and after World War II the most important Swiss exports, amounting to 20% of the total in 1956; machinery represented in that year 21% and textiles 14%. Main imports in 1952 were: machinery and vehicles 12%; steel mill products 10%; coal and petroleum 10%; grain 6%; textile fibres 5%.

Whereas in 1938 Germany had been Switzerland's most important customer (16% of the total) with the United Kingdom in the second place (11%) and France in the third (9%), by 1954 the United States and Canada took most of the Swiss exports (14%), followed by Germany (12%) and then by France (7%). In 1938 Germany was also the main source of Swiss imports (23%), with France in the second place (14%) and the United States in the third (8%); by 1954 Germany recovered its first place (22%), with the United States and Canada second (16%) and France third (13%).

Transport and Communications.—In 1956 Switzerland had 49,000 km. of roads, 308,679 motor cars, 2,594 buses and 59,189 trucks. There were 2,976 km. (route length) of railways, of which 2,810 km. were electrified. (*See* also ALPS for principal passes and tunnels.) In 1956 the state railway goods traffic amounted to 3,264,000,000 ton-km. (as compared with 1,710,000,000 ton-km. in 1938). That of passenger traffic amounted in 1954 to 6,954,000,000 pass.-km. (as compared with 3,297,000,000 pass.-km. in 1938). The traffic on the national airlines (Swissair) in 1956 amounted to 603,600,000 pass.-km. (as compared with 13,575,000 pass.-km. in 1938) and for cargo, including mail, to 15,276,000 ton-km. (as compared with 268,000 ton-km. in 1938). In 1956 there were 1,293,743 telephones, 1,267,824 licensed radio sets and 19,971 television sets.

Federal Finances.—The principal expenditure of the confederation is under the heads of defense, social insurance, contributions toward the promotion of economy and administration. Until World War I the main revenue of the confederation was from indirect taxes (customs duties and stamp duty), while direct taxes were levied only by cantons and communes. In 1915 federal taxes on property and income were introduced. (*See* Table XII.)

TABLE XII.—*Federal Finance*
(In millions of francs)

	1913	1938	1952	1953	1954	1955	1956
Revenue	101	731	1,949.9	1,974.8	2,320.4	2,245.3	2,610.6
Expenditure	106	772	2,161.5	1,884.1	1,959.2	1,948.7	1,963.6

The unit of currency is the franc of 100 Rappen or centimes. The exchange rate of the franc on Dec. 31, 1955, was 4.284 Fr. per U.S. dollar (as compared with 4.338 Fr. on Dec. 31, 1938). The Swiss National bank, founded in 1907, is the central bank of issue. On Dec. 31, 1956, it had gold reserves of U.S. $1,676,000,000 and foreign exchange to the amount of U.S. $238,000,000, compared with a note issue of 6,240,000,000 Fr. and deposit money of 8,360,000,000 Fr.

BIBLIOGRAPHY.—Kurt B. Mayer, *The Population of Switzerland* (1952); G. Sauser-Halt, *The Political Institutions of Switzerland* (Zürich, 1946); William E. Rappard, *La Constitution fédérale de la Suisse* (Neuchâtel, 1948); Max Barthell, *La Défense de la Suisse* (Paris, 1939); J. Rosen, *Wartime Food Developments in Switzerland* (1947); Jean Pavillon (ed.), *L'Agriculture et l'industrie dans l'économie nationale* (Neuchâtel, 1951); André Siegfried, *Switzerland: A Democratic Way of Life* (London, 1950); also the annual *Statistisches Jahrbuch der Schweiz* (Basle). (K. SM.)

SWOLD, BATTLE OF: *see* OLAF I TRYGGVESSÖN.

SWORD, a hand weapon with a sharp metal blade that is longer than that of a dagger, though the distinction is somewhat arbitrary. "Sword" is a general term that includes many special types of weapons, such as sabres, rapiers, scimitars, etc.

Early History.—It is possible that the earliest swords derived from the flint dagger of the Neolithic period, which may have been the model for the first hand weapons made of copper. Daggers with copper blades found in Mycenae have hilts attached to the blades by rivets. An important step in the construction of both sword and dagger was taken when the blade was continued in an extension, or tang, which became the core of the handle. The blade and hilt could thus not easily be broken apart by a blow. Most swords of the Bronze Age had leaf-shaped blades and small hilts. They have been excavated in large numbers throughout Europe. Swords of this kind (*ziphos; akinakes*) were used in the Homeric period, and are shown in paintings on Greek vases of the Hellenic period. Alternatively to the straight-edged, leaf-shaped sword was the so-called Bronze Age rapier, a sword with a long narrow blade for thrusting only. Its opposite was the *machaera* or *kopis*, whose blade had a double curve and a single cutting edge.

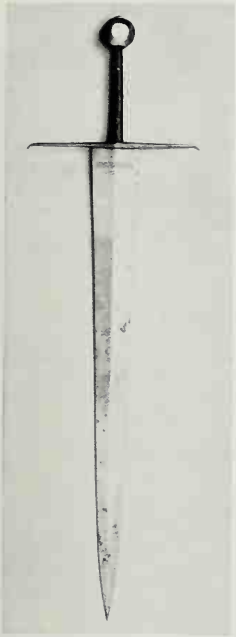
During the Hallstatt period when iron was used in addition to bronze, for the first time iron blades were fitted with hilts of bronze or other material.

In classical times the sword as a weapon was subsidiary to the spear. The Roman soldier's sword (*gladius*) was generally short, straight and broad, though a longer type of sword, known as the spatha, was also used. The word ensis was generally applied to swords of tribes outside the empire.

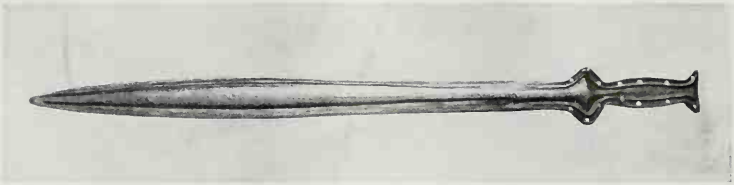
The Franks of the late Iron Age continued to use the long spatha, but as it was made of soft iron, the blade was easily bent. Some protection for the hands was given in early swords by the form of the hilt, but it was not until the early middle ages that the crossbar, or *quillons*, became fully developed for protective purposes.

The swords of the 5th century, found in the bog at Nydam, Schleswig, show waisted grips. By the 6th century Viking swords began to take the form which lasted for a thousand years. The sword had then acquired its four principal parts: (1) the pommel, a knob at the anterior end, which acted as a counterpoise to the blade; (2) the grip, covering the tang and grasped by the hands; (3) the crossbar or *quillons*, which protected the user's hand, and (4) the blade. (*See* figure.)

The Middle Ages.—It was the Vikings from Scandinavia who introduced the sword with the blade of carbonized iron. There has always been a mystery about the making of sword blades, hence the legends of Weyland the Smith, Siegfried's sword "Nothung" and King Arthur's "Excalibur." Modern science



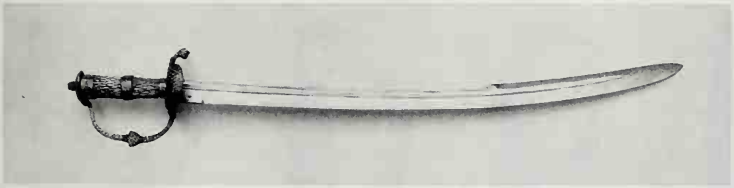
Medieval: Italian, 15th century



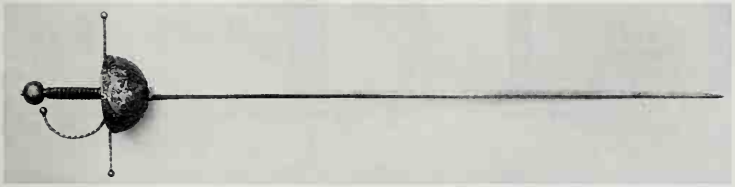
Bronze age: Ireland, 1000–500 B.C.

Carolingian: so-called sabre of Charles the Great (Charlemagne), 8th century

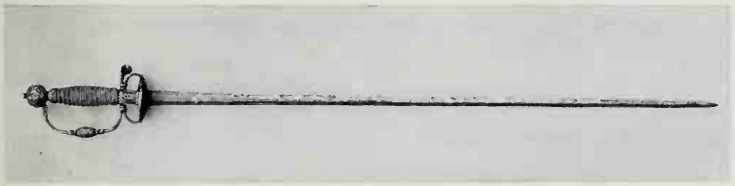
Viking sword of the 10th century



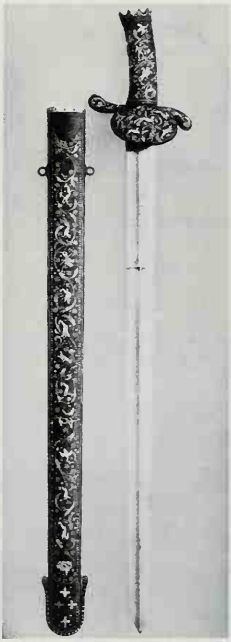
English cutlass (hanger), early 17th century



French cup hilt rapier, middle of the 17th century



German decorated smallsword, middle of 18th century



French hunting sword, 17th century

DEVELOPMENT OF THE SWORD IN EUROPE

BY COURTESY OF (ALL EXCEPT SECOND ROW RIGHT) THE METROPOLITAN MUSEUM OF ART, NEW YORK CITY, (SECOND ROW RIGHT) KUNSTHISTORISCHE MUSEUM

SWORD

English sabre and scabbard dated 1856. Presented by Queen Victoria to a U.S. navy captain

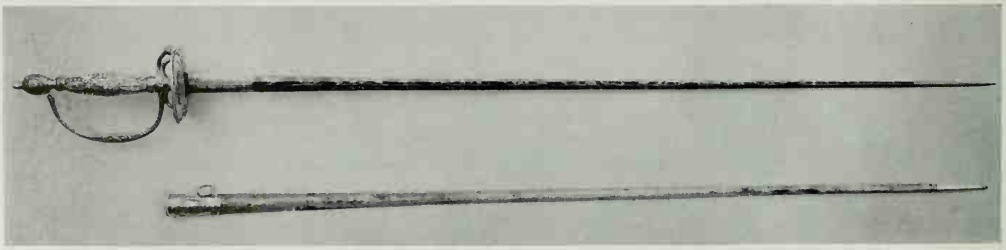
American naval cutlass of Revolutionary War; copied from a British pattern

American-made sabre of 1750–80. H. W. Williams collection

U.S. naval officer's sword, 1820–35

U.S. light cavalry sabre, model 1860

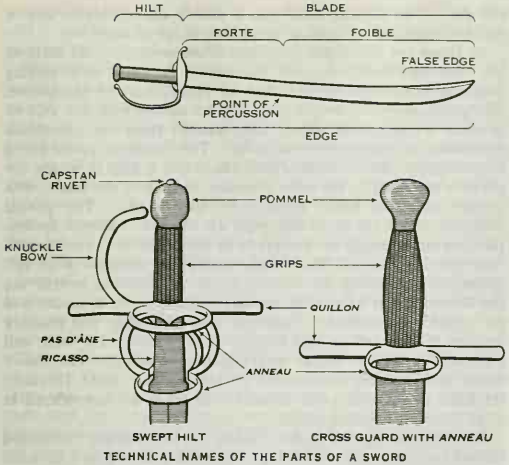
U.S. cavalry sword, model 1913, last sword to be carried by the U.S. cavalry



French smallsword presented by U.S. congress to Col. Marinus Willett for his services during the American Revolution. Dated 1777

PRESENTATION SWORDS AND U.S. MILITARY SWORDS

BY COURTESY OF (TOP, BOTTOM) THE METROPOLITAN MUSEUM OF ART NEW YORK CITY, (CENTRE) HAROLD L. PETERSON



TECHNICAL NAMES OF THE PARTS OF A SWORD

has discovered that some of these blades, which have patterns or "watering" inherent in them, were produced by what is called strip welding. Bundles of iron strips were bent over, hammered together, cut and bent again repeatedly, and the carbonization of metal, probably discovered accidentally, improved their strength and temper.

The Viking sword, and its successor, the medieval knightly sword, usually had a straight blade, double-edged and pointed, with a shallow hollow down the centre, giving lightness with rigidity. The medieval sword was purely an offensive weapon; the user relied for bodily protection on his shield and defensive armour. The user's hand was protected by rigid *quillons,* either straight or with the ends curved toward the blade. The pommel was usually rounded, either as a sphere or a disk (so-called wheel pommel) and later fig-shaped.

In the course of years certain European centres of the craft were established, notably at Passau on the Danube, and later at Solingen in western Germany, where the craft was still active in the mid-20th century. Blades were exported from Solingen to all parts of Europe, often bearing the mark of a running wolf. Swords bearing distinctive names, denoting a common origin, were widely distributed in Europe, such as those inscribed Ulfbehrt and Ingelred. Other centres of the sword cutler's craft were places where the making of arms and armour in general had been established, notably in Milan in north Italy, and in the south German towns of Nürnberg and Augsburg. In Spain great skill was shown by the swordsmiths of Toledo. The names of many master craftsmen are known and are found engraved upon the blades themselves, which also carry the marks of the maker and often the city guild where they were proved.

In the late middle ages the knightly sword was sometimes lengthened, both in blade and hilt, and the latter so shaped that it could be grasped by both hands if necessary. In the 15th century the blades of straight-bladed swords were more usually double-edged and flat, sometimes stiffened by a sharp ridge down the centre. This was particularly the case in Italy. A purely thrusting sword, with a rigid blade, was known as an *estoc.* At the turn of the century the big two-handed sword had a short vogue, but the space required for wielding it limited its advantages.

In order to have better control of the weapon, a custom developed of passing the right forefinger over the *quillons.* To protect it, a metal hook or ring was placed in front of the *quillons.* The next step was to place a corresponding half ring on the other side of the blade for the sake of symmetry. This development, nicknamed the *pas d'âne* from its resemblance to a cloven hoof mark, apparently developed first in Spain and then passed to Italy before the end of the 15th century. It is important as foreshadowing the development of the hilt whereby the hand was protected by a cagelike series of rings and bars forming part of the sword, instead of by a steel glove.

The Later Sword.—The sword underwent rapid changes in Europe during the first half of the 16th century, and proliferated in a variety of forms. These changes correspond with the period when defensive armour was beginning to fall into disuse, and special corps of soldiery were becoming components of national armies. The protection of the hand was carried a step farther by the provision of a knuckle bow; *i.e.,* a semicircular bar extending from the *quillons* toward the pommel. The next move was to add additional rings to guard the whole hand, both in the vertical and the horizontal planes. The old arming sword, which was wielded with a gauntleted hand, had been strictly cruciform. The 16th century saw the arrival of the rapier, with a long, narrow blade, and with the hand protected by a series of guards and counterguards.

The habit of carrying a sword with civilian dress had not been universal in the middle ages but it later became the custom for every gentleman to carry a rapier in civil life, and in consequence dueling increased. Previously, dueling had taken the form of judicial combats, conducted with strict formality. Dueling with a rapier rapidly encouraged the art of fencing. Schools conducted by masters of fence were set up in which gentlemen learned their skill at arms. The earliest known fencing book is a treatise, *Flos Duellatorum* or *Il Fior di Battaglia,* by Maestro Fiore dei Liberi (compiled in Italy in 1410). It demonstrated in a series of drawings combats with swords, sword and buckler, wrestling, etc. It was followed in the 15th century by a picture book by H. Talhofer, of which several manuscript versions exist. In the 16th century, books on fencing came from the printing presses, generally written by one of the masters of the schools of fencing, such as A. Marozzo and C. Agrippa in Italy, H. de St. Didier in France and George Silver, who was a stout defender of the rough-and-ready English fashion against his rival, "Signior" Rocco, an Italian who had set up his own school in London. The weapons used in dueling could be rapier and dagger, the latter held in the left hand and used for parrying, rapier and cloak, or two rapiers, but eventually fencing with the rapier or smallsword alone prevailed. At first both edge and point were used, but with the increase of manual dexterity it was found that the point, which allowed a long reach, was more effective by itself. The swordsman was liable to lay himself open when he used striking blows, so thrusting, parrying and feinting within a narrow compass became a practised art. (*See* also FENCING.)

The hilts of rapiers took many forms, often being richly decorated with gold and silver, damascening, enameling and even incrustation with jewels. Fine specimens are to be seen in the national armouries of Vienna, Madrid and Paris. Notable swordmakers were A. Piccinino of Milan, L. Palumbo of Naples, the Wundes family and Clemens Horn, among many, at Solingen, and Juan Martinez, Tomas de Ayala and Sebastian Hernandez at Toledo. The blades of most rapiers bear the signature of the maker, his town, and the stamp of the city guild, which is usually placed upon the *ricasso* (the base of the blade), which usually has a rectangular section. Many of these long, narrow, thrusting blades made in Germany were stamped with imitations of the ♀ stamp of Toledo. Blades of all kinds were made in Germany, inscribed with the names of swordsmiths who may be regarded as having specialized in blades of a certain kind, such as A. Sahagun and Andrea Ferrara. The last name is quite commonly found on the blades of broadswords of the 17th century and is no indication of the actual maker. The blade is classified in two parts, the forte which is the anterior part, and the foible which is the forward part farthest from the hilt. In Spain a particularly elegant form of rapier hilt was developed in which the principal protection of the hand was a steel cup, often beautifully pierced with decorations. This developed at Toledo, but soon spread throughout much of western Europe. Generally speaking, those made in the south had deep cups while those produced in the Low Countries and England had shallower "dish" guards.

In the third quarter of the 17th century the rapier was reduced in size and simplified to what became known as the smallsword. This too was principally a civilian weapon, and the subject of much delicate and elaborate ornamentation, following the trends of fashion through the 18th century. The blade of the smallsword was purely for thrusting, and usually given a hollow, triangular section. In other cases the blades were hexagonal, diamond-shaped or even elliptical in section.

New weapons produced their own writers on swordsmanship, notably the Scotsman, Sir William Hope, who wrote a number of manuals in the late 17th and early 18th centuries. H. Angelo's school of fencing in London is well known from his books illustrated by George Rowlandson.

Swords for specialized use and local types are legion, such as the straight broadsword with S-shaped *quillons* used by the German *Landsknechte* (professional infantry) in Germany and Switzerland in the 16th century. In the highlands of Scotland the two-handed sword lasted to the end of the 17th century in the form of the big *claidheammor* (great sword; whence "claymore") with drooping *quillons* ending in quatrefoils, and Scotland produced a characteristic variety of the basket-hilted broadsword.

The medieval arming sword had been replaced for war by the sword rapier, a heavier version of the civilian sword. Its successor was the cavalry broadsword, a simple and handy weapon on which the knuckles were protected by a single or basketlike guard, and the blade was flat and broad, sometimes double-edged and sometimes single, when it was called a backsword. Many broadswords of this type have come down from the time of the Civil War in England. Sometimes the solid guards of the hilts are chased with the heads of Charles I and his queen, and these are popularly called "mortuary" swords, referring to the execution of the king.

The military form of the smallsword during the 18th century was more heavily built, and was furnished with a single knuckle bow and a pair of oval-shaped plaques in front of the *pas d'ane*. The creation of standing armies with regiments uniformly equipped resulted in swords of regulation patterns.

The Curved Sword and the East.—The sabre is essentially a cutting weapon, with the blade sharpened on the convex edge only. It cuts well because only a small part of the arc first reaches the objective, and then bites into it.

Whereas the straight sword was the favourite in the west, the curved sword predominated in eastern Europe and Asia, although curved swords were not unknown in the west. In the middle ages there was the falchion, a sword with a cleaverlike blade, broadening toward the point and with a single curved cutting edge. That it was regarded as an oriental weapon is shown by the fact that in medieval manuscripts the sleeping guards at the Sepulchre are shown with falchions. The common people of England practised on foot as a sport with sword and buckler, when those more wealthy jousted with lances on horseback. In the 17th century light, curved riding swords or hangers were much used in Europe, and from this followed the cutlass. Hunting swords with appropriate decoration, used for deer and boar, form a class by themselves.

The Turkish invasion of Europe in the 16th century brought the west into close contact with the curved weapons of the east. The scimitar, which has a keen arc-shaped edge, derived from the Persian shamsheer. It was brought into Hungary and Poland, and western cavalry regiments in the late 18th and 19th centuries were armed with the curved sabre which derived directly from the east. Turkish and Persian scimitars were often made with finely watered blades. These blades are often described as damascened, but Damascus was never a centre of the craft, only an entrepôt. Fine swords were made at Ispahan in Persia, at Constantinople, at Cairo under the Mamelukes and at Fez in north Africa. The north and middle European scimitar generally has a pistol-shape grip, which continues the curve of the blade in the opposite direction.

In India the scimitar has always had a permanent place, and is called a tulwar. Among other Indian swords are the Maratha khanda and the gauntlet sword known as a pata, and in Burma the *dha*. The Ghurka kukri has a doubly curved blade, and in skillful hands can be used to decapitate an ox at one blow.

In Japan the sword has long been taken seriously. As early as the 13th–14th centuries great masters of the craft were making blades of remarkable quality, such as Yoshimitsu and Masamune. The possession of a sword is closely connected with the cult of ancestor worship, and swords were handed from one generation to another as objects of veneration. The Japanese sword has a long, slightly curved, single-edged blade and a grip in which the curve is extended in the same direction as that of the blade, with ample room for being gripped by two hands. The guards (tsubas), which are small and solid, are objects of special decoration; an owner might have as many as sixty different sword guards with which to furnish his sword, ringing the changes as he felt appropriate. Among the various kinds of Japanese swords are the *daisho*, a pair of swords, the larger being called a *katana* and the smaller a *wakizashi*. Japanese swordsmanship was possibly an even greater cult in that country than it was in the west, and expressed itself in formal and elaborate dueling. The sword which the Japanese commander in chief, Count Juichi Terauchi, tendered to Admiral Lord Mountbatten when he surrendered in 1945 is now at Windsor castle.

In north Africa one finds a liking for the straight two-edged blade of western type. This type of sword is known as a *kaskara*. It is often said that these swords derived from the crusades, but a glance will show that in construction they have one important difference. The heavy pommel of the European medieval sword is missing from the *kaskara*, which has only a light disk.

The Modern Sword.—Since the middle of the 18th century the sabre has been predominant. Indeed, if one accepts the definition of a sabre as a sword with a single edge designed primarily for cutting, regardless of whether the blade is curved or straight, it becomes apparent that more than 80% of all European and American swords of the period fall within this category. The civilian small sword had virtually disappeared by 1800. Diplomats and some staff officers continued to wear small swords of one type or another, but their numbers were few. Among the military sabres, those used by the cavalry were most important as weapons. The first sabres to be used by horsemen were usually straight or very slightly curved. Early in the 19th century the curve increased. Then, with the discovery that the point was used more often than the edge, the straight blade returned to favour in almost every European nation and in America. Another form of sabre, the naval cutlass, achieved its distinction during this era. It was a short weapon with a straight or slightly curved blade and a heavy guard suitable for fighting in close quarters and easily carried while climbing. It maintained its basic design without change until the long-range naval actions of the twentieth century rendered it obsolete.

Officers and ranking noncommissioned officers of almost all arms and branches of service carried swords at some time during the 19th century. Indeed, the sword was used to designate the rank and branch of service of the wearer as well as to provide for his defense. For this reason the swords of the last two hundred years are found in a greater variety of styles and forms than for any other comparable period.

As long as firearms were single-shot weapons and firepower was relatively low, the sword remained a vital weapon. During the American Revolution, cavalry leaders considered the sword the only true weapon of a trooper and scorned the use of firearms except for skirmish or picket duty. Infantry officers participated in charges with sword in hand or stood by with that weapon to help repel an enemy attack.

The adoption of repeating arms ended the sword's importance. As late as the U.S. Civil War cavalry still charged with sabre in hand, but the concept of the trooper as a mounted infantryman was developing. His revolver and his repeating carbine carried more weight than his sword. The last traditional cavalry charge by U.S. troops occurred on Luzon in 1906. There were some isolated actions by European cavalrymen during World War I and again on the eastern front in World War II, but these were anachronisms. The era of the sabre as a weapon had ended. As

a symbol of rank and of arm of service, however, it is still in use, and it has been found to serve an important function from the standpoint of morale and the sense of tradition.

Presentation and State Swords.—A special category of swords comprises those used in state ceremonies and those presented to individuals as a mark of esteem or honour. The ceremonial or state swords have been used for many centuries at coronations, inaugurations and other solemn public occasions. Among them are the sword of justice, the sword of mercy and the sword of mourning. Usually they are huge two-handed swords richly mounted in precious metals, often set with gems, and frequently bearing the arms of the country, state or city to which they belong. Such swords have figured in the regalia of almost every European monarchy, and some are still in use. Special presentation swords or swords of honour have a much shorter history. The practice of giving a sword as a mark of esteem or as recognition for a heroic deed is as old as the sword itself. The manufacture of special swords for this purpose, however, did not achieve widespread popularity until the 18th century. From then until the third quarter of the 19th century huge quantities of them were made and presented by nations, states, cities, organizations and individuals. Some were even awarded by "vote" in lotteries during the U.S. Civil War. After 1875 the practice rapidly declined in popularity, although it has not yet died out completely. A few were presented to leading generals at the close of World War II. Usually these presentation swords were patterned after the principal swords of the day but distinguished from them by the richness of their mountings. In the 18th century they were primarily small swords; in the 19th century they were sabres. Most recently, however, there has been a tendency to revert to the medieval arming sword. Most of those presented at the close of World War II, in fact, were of this traditional form.

The Headsman's Sword.—One last class of sword may also be mentioned—the "heading" sword or executioner's sword. In many countries, both in Europe and Asia, it was often the practice to decapitate condemned criminals with the sword instead of the ax. Queen Anne Boleyn was beheaded with a sword at the Tower of London in 1536, for instance, and even in the United States at least one such execution is recorded in Massachusetts in 1644. In Asia the practice continued and was still being used as late as the end of World War II. The European "heading" sword traditionally had a long straight blade with a square "point." The balance was forward; the hilt was cruciform with grips long enough to permit the use of both hands. Asian swords, while also normally possessing the two-handed grips, tended to have slightly curved blades.

BIBLIOGRAPHY.—*Prehistoric, Greek, Roman and Dark Ages:* J. Naue, *Die Vörromischen Schwerter aus Kupfer, Bronze und Eisen* (1903); A. E. Remouchamps, *Griechische Dolch-und Schwertformen* (1926); W. P. Brewis, "The Bronze Sword in England," *Archaeologia,* lxxiii (1924); P. Coussin, *Les Armes Romaines* (1926); W. Ginters, *Das Schwert der Scythen und Sarmaten* (1928); M. Jahn, *Die Bewaffnung der Germanen in der alteren Eisenzeit* (1916).

Middle Ages and Renaissance: J. Petersen, *De Norske Viking Sverd* (1919); G. F. Laking, *Record of European Armour and Arms through Seven Centuries,* 5 vol. (1920–23); E. de Beaumont, *La Fleur des Belles Epées* (1885); R. Forrer, *Die Schwerten und Schwertknaufe der Sammlung C. von Schwerzenbach* (1905); J. Schwietering, *Geschichte von Speer und Schwert im 12 Jahrhundert* (1912); A. Bruhn-Hoffmeyer, *Middelalderens Tvaeggede Svaerd,* i and ii (1954); R. E. Oakeshott, *The Archaeology of Weapons* (1960).

17th–19th Centuries: C. Welch, *History of the Cutlers' Company of London,* 2 vol. (1916, 1923); B. Dean, *Catalogue of European Court Swords and Hunting Swords* (1929); J. Drummond, *Ancient Scottish Weapons* (1881); J. D. Aylward, *The Smallsword in England* (1905); M. Bottet, *L'Arme Blanche de Guerre Française aux XVIIIe Siècle* (1910), *L'Arme Blanche des Armes Françaises 1789–1870, La Manufacture des Armes à Versailles* (1903); Z. Hartleb, *Szabla Polska* (1926); H. Seitz Svärdet och Värjän (1955); C. J. ffoulkes, *Sword, Lance and Bayonet* (1938); T. A. Bosanquet, *The Naval Officer's Sword* (1955); War Office, *Text Book on Small Arms* (various dates); H. L. Peterson, *The American Sword, 1775–1945* (1954); H. L. Peterson, *American Silver Mounted Swords 1700–1815* (1955).

Oriental: Lord Egerton of Tatton, *Indian and Oriental Armour,* 2nd ed. (1896); P. Holstein, *Armes Orientales* (1931); G. C. Stone, *A Glossary of . . . Arms and Armour* (1934); G. B. Gardner, *Keris and Other Malay Weapons* (1936); H. Joly and I. Hogitaro, *The Sword Book and the Same* (1913); Jungi Homma, *The Japanese Sword* (trans. 1939); Inami Hakusi, *Nippon Tó, The Japanese Sword* (1948); B. W. Robinson, *Primer of Japanese Sword Blades* (1956).

Information can be obtained from illustrated catalogues of public museums; *e.g.,* the Tower of London (swords in preparation); Wallace Collection (3 parts, 1925–45); Victoria and Albert Museum; Metropolitan Museum, New York; Musée de l'Armée, Paris; Real Armeria, Madrid; Dresden (E. Haenel, *Kostbare Waffen*); Swiss National Museum, Zurich; Historical Museum, Berne; Waffensammlung, Vienna. *See* also articles on swords in *Zeitschrift für Historisches Waffenkunde* (1896–1945), and useful information in catalogues of auction sales of arms (full bibliography in the *Wallace Collection Catalogue of European Arms and Armour,* part iii, London, 1945).

(J. G. MN.; HD. L. P.)

SWORDFISH (*Xiphias gladius*), a large fish, widely distributed in warm seas, dark purplish in colour, and with the snout prolonged to form a long, flat "sword." As in the mackerels, to which it is related, the caudal peduncle is slender and the caudal fin is widely forked. It reaches a weight of about 1,000 lb. It swims swiftly at the surface and feeds on other fishes, especially mackerel, and is said to dash into a school, killing a number with slashes of its long sword. The flesh of the swordfish is valued as food; fisheries exist in the Mediterranean and off the Atlantic coast of America, the usual method being harpooning. The record swordfish taken with rod and reel was 13 ft. 9 in. long and weighed 860 lb.

Related to the swordfish are the spearfishes (*Tetrapturus*), marlins (*Makaira*) and sailfishes (*Istiophorus*): in these the rostrum is not flattened, but is spearlike, rounded in transverse section. Each genus has two or three species, all large fishes of warm seas. *Istiophorus americanus* of the Atlantic is a beautiful fish, silvery and blue-backed, with the very large saillike dorsal fin deep blue, with numerous round black spots. The members of this group are said to attack whales, and whereas this is doubtful, it is certain that they frequently strike ships, possibly mistaking them for whales. The sword or spear may be driven into a wooden ship with such force that the fish, in order to get free, has to break it off short. In the Natural History museum in London is a piece of timber, less than a foot square, containing the broken ends of three spears.

SWYNFORD, CATHERINE (*c.* 1350–1403), wife of John of Gaunt, duke of Lancaster, was a daughter of Sir Paon Roet (Sir Payne Roelt), a knight who came to England from Hainault in the train of Edward III's queen, Philippa. About 1367 she married Sir Hugh Swynford (1340–1372), a Lincolnshire man, by whom she had a son, Thomas (*c.* 1368–1433), who was a friend and companion of Henry IV.

Soon after her husband's death in 1372 Catherine became the mistress of John of Gaunt, and in 1396 she was married to him at Lincoln. She died at Lincoln on May 10, 1403. By John of Gaunt Catherine had four children, all of whom were born before their marriage. They were declared legitimate in 1397 and took the name of Beaufort from one of their father's castles in Anjou. *See* BEAUFORT.

SYAGRIUS (d. 487), the last of the independent Roman administrators of Gaul, was the son of Aegidius, who had seized Gaul while Ricimer was master of Italy. Defeated in 486 by Clovis, king of the Salian Franks, at the battle of Soissons, Syagrius fled, leaving his land at the mercy of the Franks. He sought refuge with Alaric II, king of the Visigoths, at Toulouse, but Alaric imprisoned him instead of granting him refuge, and delivered him up to Clovis. He was executed in 487, secretly and by the sword, according to Gregory of Tours.

SYBARIS, a city of Magna Graecia, on the Gulf of Tarentum, between the rivers Crathis (Crati) and Sybaris (Coscile), which meet 3 mi. from the sea, but in ancient times had independent mouths, was the oldest Greek colony in this region. It was an Achaean colony founded by Isus of Helice (about 720 B.C.), but had among its settlers many Troezenians, who were ultimately expelled. Placed in a very fertile, though unhealthful, region, and following a liberal policy in the admission of citizens from all quarters, the city became great and opulent, with a vast subject territory and daughter colonies even on the Tyrrhenian sea (Posidonia, Laus, Scidrus). For magnificence and luxury the Sybarites were pro-

verbial throughout Greece, and in the 6th century no Hellenic city could compare with its wealth and splendour. At length contests between the democrats and oligarchs in which many of the latter were expelled and took refuge at Crotona, led to a war with that city, and the Crotoniats, with inferior forces, were victorious. They razed Sybaris to the ground and turned the waters of the Crathis over its ruins (510 B.C.). Explorations have so far failed to lead to a precise knowledge of the site. The only discoveries have been (1) that of an extensive necropolis, about 8 mi. to the west of the confluence of the two rivers, of the end of the first Iron Age, known as that of Torre Mordillo, the contents of which are preserved at Potenza; (2) that of a necropolis of about 400 B.C.—the period of the greatest prosperity of Thurii (*q.v.*)—consisting of tombs covered by tumuli, in some of which were found fine gold plates with mystic inscriptions in Greek characters.

SYBEL, HEINRICH VON (1817–1895), German historian, was born on Dec. 2, 1817, at Düsseldorf. Sybel was educated at the gymnasium of his native town, and then at the University of Berlin, where he came under the influence of Savigny and of Ranke, whose most distinguished pupil he was to become. In 1841 he settled at Bonn as *Privatdozent* in history. He was already known by his *Geschichte des ersten Kreuzzuges* (Düsseldorf, 1841; new ed., Leipzig, 1881; Eng. trans. by Lady Duff-Gordon, 1861). This was followed by a study on the growth of German kingship (*Die Entstehung des deutschen Königtums*, Frankfurt, 1844, and again 1881), after which he was appointed professor. In the same year (1844) Sybel came forward as an opponent of the Ultramontane party. In 1846 he was appointed professor at Marburg, while a seat in the Hessian *landtag* gave him his first experience of political affairs. In 1848 he was present at Frankfurt, but did not sit in the national assembly. His opposition to the revolutionary party made him unpopular with the mob, who broke his windows, as his liberalism made him suspected at court. He sat in the Erfurt parliament of 1850, and was attached to the Gotha party, which hoped for the regeneration of Germany through the ascendancy of Prussia. During the years of reaction that followed all political activity was impossible, but he was fully occupied with his great work, *Geschichte der Revolutionzeit 1789–1800*, for which he had made prolonged studies in the archives at Paris and elsewhere. The later editions of the earlier volumes are much enlarged and altered, and a new edition was published at Stuttgart in 1882. The first three volumes have been translated into English by W. C. Perry (1867–69). In this work he for the first time showed the connection between the internal and external history of France.

In 1856 Sybel accepted the post of professor at Munich. There he founded the *Historische Zeitschrift*, the original and model of numerous technical historical publications. In 1861 he accepted a professorship at Bonn, which he held till 1875. He was at once elected a member of the Prussian lower house, and during the next three years was an active member, opposing the policy of Bismarck, not only on financial but also on the Polish and Danish affairs. In 1864 he did not stand for re-election. He had a seat in the constituent assembly of 1867, and while he joined the National Liberals he distinguished himself by his opposition to the introduction of universal suffrage, the effects of which he, as did many other Liberals, much distrusted. In 1874 he again accepted a seat in the Prussian parliament. In 1880 he retired.

In 1875 he had been appointed by Bismarck director of the Prussian archives. Under his superintendence was begun the great series of publications, besides that of the correspondence of Frederick the Great, in the editing of which he himself took part. His last years were occupied on his great work, *Die Begründung des deutschen Reiches durch Wilhelm I* (7 vol., Munich, 1889–94; Eng. trans., 1890–91). He died at Marburg Aug. 1, 1895.

SYCAMORE, the name applied to American species of plane trees of the genus *Platanus*, which also embraces the European and Asiatic planes and in Europe to a species of maple (*q.v.*). The distinctive eastern sycamore (*Platanus occidentalis*), also known in certain localities as buttonwood or American plane, is probably the largest hardwood species in the eastern United States. Best development is attained on rich alluvial bottomlands, where it often reaches a height of 150 ft. and a diameter of 8 to 12 ft. It is occasionally found in pure groves, but is much more common as an occasional tree in association with other bottomland hardwood species such as river birch, red and silver maples, American elm, cottonwood and willows. Sycamore features large, simple, shallowly three- to seven-palmately-lobed leaves, each with a long petiole or leafstalk swollen at the base and enclosing next year's bud. A pair of large leafy outgrowths or stipules at each node is another conspicuous feature.

The flowers are unisexual. Both sexes are borne in separate, solitary heads that appear in advance of the leaves. The fruit is a multiple head of obovoid achenes, each furnished at the base with a fringe of brown silken hairs.

Winter characters are also distinctive. The stout, brown, conspicuously zigzag twigs feature alternately disposed conical buds that are capped with a single scale and nearly surrounded by a leaf scar. Above each leaf scar is a prominent stipule scar which encircles the twig. The bark on small branches is creamy white; that on larger limbs and the trunk (except near the base of large trees) is characterized by the annual exfoliation of the brown outer

RUTHERFORD PLATT (TOP), J. HORACE MCFARLAND CO. (BOTTOM LEFT), ROCHE (BOTTOM RIGHT)

SYCAMORE (PLATANUS OCCIDENTALIS) SHOWING THE BROWN, OUTER LAYERS OF BARK WITH THE WHITE LAYERS BENEATH THEM (TOP), A LEAF (BOTTOM LEFT) AND STAMINATE FLOWERS (BOTTOM RIGHT)

layers to expose the lighter yellowish to white layers immediately below.

The Arizona sycamore (*P. wrightii*) is one of the most abundant hardwoods of southwestern United States. The tree features deeply five- to seven-lobed leaves and the multiple heads of achenes are borne in racemose clusters of from two to four. The California sycamore, *P. racemosa*, a medium-sized tree of the coast

ranges and lower Sierra Nevada mountains of California, features three- to five-lobed leaves, and fruit in racemose clusters of two to seven.

See also PLANE. (E. S. HR.; X.)

SYDENHAM, THOMAS (1624–1689), one of the greatest English physicians of the 17th century, was born on Sept. 10, 1624, in Wynford Eagle, Dorset. His medical studies at Oxford were interrupted by the Civil War, in which he served for four years in Oliver Cromwell's army. After graduation from Oxford, he studied at Montpellier, Fr., and then set up practice in London. His greatest contributions to contemporary medicine were his careful observations and records of diseases and his revival of the Hippocratic outlook at a time when superstition and mysticism had crept into medicine. His simple and classic descriptions of disease are contained in various publications such as *Medical Observations* in 1676 and *A Treatise on Gout* in 1683. He insisted that in treatment, observations should have precedent over theory. He treated fevers by cooling, introduced "Sydenham's laudanum," a tincture of opium, and popularized the use of Peruvian (cinchona) bark in treating malaria. He died in London on Dec. 29, 1689.

See C. E. A. Winslow, *The Conquest of Epidemic Disease* (Princeton, N.J., 1943).

SYDENHAM, a residential area almost wholly in Lewisham (*q.v.*), but also extending into Camberwell (*q.v.*), London, Eng. Its development dates from the re-erection there (1854) of the Crystal palace (burned down 1936). Sydenham Wells park contains the site of medicinal springs popular in the 17th and 18th centuries. Former residents included the actress Sarah Siddons, the poet Thomas Campbell and the colour printer George Baxter.

SYDNEY, a city situated in the Sydney basin (County Cumberland) is the capital of New South Wales, Austr., and lies (lat. 33° 52′ S., long. 151° 12′ E.) mainly on the south shore of Port Jackson somewhat south of the central east of the state. The oval-shaped area which extends from about Broken Bay (Hawkesbury river estuary) on the north to a little beyond Port Hacking on the south, and from the coast east of Sydney about 40 mi. inland (near Glenbrook) forms the lower part of a geological and tectonic basin which is disposed fairly symmetrically about the line Botany bay-Penrith.

The Sydney basin is enclosed on three sides by flattish highlands which are deeply dissected and seamed with gorges by streams which now, for the most part, form part of the Hawkesbury system but which formerly had different (westward and northwestward) courses and were dislocated by the late (Tertiary?) uplift which raised the plateaus. On the seaward side, also, recent earth movements have resulted in the drowning of river valleys and created spacious inlets—Port Hacking, Botany bay, Port Jackson, Broken bay—but there also the basin-edge structure is apparent in the upstanding cliffs which form an outer rampart along the coast and give rise, at the entrance of Port Jackson, to the famous Sydney Heads, great sandstone cliffs which admit between them a deep (80 ft. low water) sea passage of a minimum breadth of $\frac{3}{4}$ mi.

The Sydney basin is one of the largest coastal lowlands of the state, but its lack of easy land communications makes the presence and growth of Sydney seem somewhat surprising. The reasons are partly physical but more historical: the superiority of the deep and sheltered harbour, defensible yet easily accessible from the sea, gave it an advantage over the shallower and more open Botany bay and the semisilted Port Hacking, and when Gov. Arthur Phillip, in 1788, moved the early settlement from Botany bay, he chose the small deep-water cove at the head of which stands Circular Quay, but where was then a fresh-running stream (Tank stream) later obliterated.

Port Jackson, sheltered from the southerly gales, covers an area of 21 sq.mi. and has a highly embayed coastline—a succession of romantic low bluffs and coves, still often densely wooded, giving a total shore line of 152 mi. It opens a way into the heart of the lowlands and upon its southern and lower shores (elevation 100 ft.) the original settlement grew. For some time it was, from the point of view of its founders, not only "the state," but "all Australia," and from it emanated, and upon it centred, much of the energy and initiative which built up the growing colony. Hence its triumph over physical obstacles, for the roads and the railways have had to cross the highlands at one of their most difficult sections, and, outside the basin, and apart from a few special areas (mining centres, and the tourist resorts at Katoomba) attracted by elevation and scenery—the barren and rugged highlands at a distance of 50 mi. are still almost unoccupied. The city expanded rapidly (pop. 1861, 95,789; 1901, 481,830; 1933, 1,235,267) until in 1961 the total metropolitan area contained 2,183,388 inhabitants. The metropolitan area includes the city proper (172,202 persons in 1961) and a group of suburbs organized into a total of 31 additional municipalities and parts of 4 others. In 1961, it covered an area of 344.7 sq.mi. and had a density of 6,334.2 persons per square mile. But this concentration of a large portion of the citizens of a vast state into what is virtually the Sydney basin is characteristic. The remarkable growth of large industrial cities, which in so many countries drained the countryside during the last half century, has characterized Australian growth from almost the beginning and without the typical agrarian preliminary stage.

However, as the city grew and industrial and commercial occupation advanced, residential areas were pushed farther out. Thus areas adjacent to but outside the existing boundaries of the metropolitan area have shown a considerable increase in population—from 124,000 persons in 1933 to 320,000 persons in 1952. The main industrial and commercial areas are situated in or in close proximity to the city of Sydney proper and in the western and southern suburbs. The northern and eastern suburbs are mainly residential, but there is also a considerable residential population in the southern and western suburbs. The climate is generally healthful and invigorating, cool and bracing with westerly winds in winter (May to September inclusive), hot and sometimes unpleasantly humid in summer. The prevailing winds from October through April are from the northeast. Mean temperatures range from 71.3° F. in summer to 55.2° F. in winter, with extremes 113.6° F. and 35.7° F. Sydney is a city of sunshine with 2,465 hr. per annum average. Rainfall is distributed throughout the year but shows an autumn and winter maximum (February-July) and averages 45 in. Sydney, besides its splendid site, has had the advantage of fine and easily procured building stone, the sandstones of Pyrmont and other local quarries. The old town had winding, narrow and relatively enclosed streets. In spite of much rebuilding, some of these persist and give the city an old-world flavour. The public buildings are, in many cases, notably fine, the two cathedrals (Anglican and Roman Catholic), the university, post office, several of the government offices, and some of the public galleries, being worthy of note. The main streets also offer excellent specimens of modern layout and substantial architecture, while many other buildings including government house, have historical interest. The city is well endowed with parks, recreation grounds and pleasure resorts generally. The 699 ac. of parks within the city boundaries include Moore park, the fine Botanic gardens, the Domain, Hyde park which commands many splendid views; in addition the Sydney Cricket ground and the Royal Agricultural society's ground where is held annually the most famous agricultural show in the state. There are 14,000 ac. of similar reserves in the municipalities, and two large national reserves: National park (34,400 ac.) 16 mi. south of Sydney around Port Hacking, and the Ku-ring-gai Chase (38,300 ac.) 15 mi. north on Broken bay —both with magnificent scenery.

The harbour of Sydney, with its coves, creeks and bays, is a favourite pleasure resort, as are also such beaches as Manly for surf bathing and water sports. Sydney has an excellent and assured water supply. In 1952 the storage reservoirs had a capacity of 124,892,000,000 gal., and the catchment area for these reservoirs covered 3,759 sq.mi. Electric trams, motor omnibuses, electric and steam railways serve the city and suburbs, and ferries ply in the harbour. The great steel-arch Sydney bridge, completed in 1932, which spans the harbour from Dawes Point (south) to Milson's Point (north), has a total length of 3,770 ft. (main arch span, 1,650 ft.; clearance for vessels passing underneath, 170 ft.), and carries a roadway 57 ft. wide, two footways, and four lines of railway. It is one of the largest steel-arch bridges in the world. The

city is, in respect of its social, educational and artistic life a leading city of the southern hemisphere. There are about 12,000 factories in the Sydney basin area (County Cumberland), most of which are in the metropolitan area. Clothing, metal, machinery (including railway engineering) industries are the most important, with food and drink preparation second. Sydney is also the chief commercial, banking, warehousing and shipping centre of the state. Upon it are centralized the lines, not only from its own basin, but from a very large part of the state. Wheat, meat, butter, fruit and many other primary products are brought to the city, from inland or by coasting steamers, for consumption or for export, but most important and most characteristic is wool. Sydney is by far the largest wool-selling centre in Australia and its sales, which begin in September and last for eight to nine months are world famous. Port Jackson offers unrivalled opportunities for harbour development.

The Maritime Services board has pursued a vigorous policy of development and improvement and has largely transformed the earlier unsatisfactory state of much of the water front. The principal wharves lie four to five miles from the Heads and close to the business centre of the city. The board controls about 60,000 ft. of commercial wharfage, about half of which is leased to shipping companies. The principal wharves are located at Woolloomooloo bay, Sydney Cove, Glebe Island, Walsh bay and Pyrmont, where the main oversea passenger terminal is situated. Glebe Island is the wheat-exporting harbour with modern appliances including an elevator of 7,500,000 bu. capacity. There are also wool, timber, coal, and other special cargo wharves and warehouses. All these are on the southern shores and do not include ferry wharves or private jetties. The islands (9) in Port Jackson are either pleasure resorts or used for special purposes, such as naval and fire-stations. There are five large graving docks in Sydney harbour, one of which is capable of accommodating the largest vessels afloat.

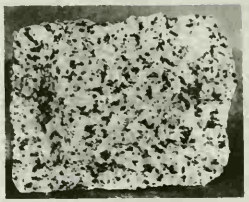
BIBLIOGRAPHY.—Griffith Taylor, "The Warped Littoral Around Sydney" in *Proc. Royal Soc. N.S.W.*, vol. lvii (Sydney, 1923), and in *Commonwealth Year Book*, no. 20 (Canberra, 1927, *et seq.*); K. Ussher, *The Cities of Australia* (London, Toronto, 1928); *Official Year Book of N.S.W.* (Sydney, 1950–51, *et seq.*).

SYDNEY, a city of Nova Scotia, Can., is the commercial and industrial centre of Cape Breton county. Founded by Joseph Frederic DesBarres in 1783 as a haven for United Empire Loyalists, and named after the colonial secretary, it was the capital of Cape Breton Island until 1820 when the island was united with Nova Scotia. The population was increased during the early 19th century by the immigration of Scottish highlanders. Sydney is a divisional terminus of the Canadian National railways and on the south arm of the harbour are excellent facilities for ocean-going vessels. The Dominion Steel and Coal corporation is the city's chief industry. It is the centre of a large coal-mining district with an urban population exceeding 100,000. Glace Bay (*q.v.*), 14 mi. E., New Waterford, 14 mi. N.E., and Sydney Mines on the north side of the harbour are the major mining centres. North Sydney, on the northwest arm of the harbour, is the terminus of the ferry service to Port aux Basques, Nfd. Xavier Junior college (1952), an affiliate of St. Francis Xavier university, Antigonish, is located in Sydney. Across the south arm of the harbour is the Point Edward naval base. Pop. (1961) 33,617. (C. W. RD.)

SYENITE, in petrology, the group name for a class of plutonic rocks composed essentially of an alkali-feldspar and a ferromagnesian mineral. A special group of alkali-syenites is characterized by the presence of a feldspathoid mineral (nephelite, or nepheline, leucite, analcime or sodalite).

The name was first used by Pliny for the rock occurring at Syene (Assuan) on the Nile river in upper Egypt. This rock, extensively worked in ancient times for monumental structures, is a hornblende-granite with abundant quartz. A. G. Werner subsequently adopted the name for the class of rocks defined above, of which the type example is the hornblende-syenite of the Plauen'scher Grund, near Dresden, Ger. As essential constituents there occur soda orthoclase and green hornblende with subordinate amounts of oligoclase feldspar.

The texture of syenites is like that of the granites, hypidiomorphic-granular. Mineralogically these rocks differ from granites only by the absence or scarcity of quartz. The alkali-feldspars include orthoclase, perthite or albite, more rarely microcline, while the ferromagnesian mineral may be biotite, hornblende or pyroxene. In the alkali-syenites, the amphiboles or pyroxenes are frequently soda-bearing varieties. The more normal syenites are divisible into augite-, hornblende- and biotite-syenites according to their prevalent dark-coloured mineral, but syenites are also divisible, as in the case of granites, into potash and soda syenites, according to the type of alkali-feldspar. The accessory constituents include sphene, apatite, zircon, magnetite and pyrites. Syenites are by no means common and are not of equal importance with granites and diorites from a geological standpoint.



B. M. SHAUB

HORNBLENDE-SYENITE FROM CUTTINGSVILLE, VT. THE BLACK GRAINS ARE HORNBLENDE, LIGHT ONES ARE FELDSPAR

Among the potash syenites are classed the original hornblende-syenite (Plauen) and the hornblende-syenite of Biella, Piedmont, Italy, of which analyses are given below.

The rocks known as nordmarkite and pulaskite are classed among the soda syenites. Nordmarkite originally described from the Oslo region (Norway) is built up of pink microperthite and subordinate aegirine and arfvedsonite together with some quartz, while the pulaskite syenites, first described from the Fourche mountains, Ark., consist of soda-orthoclase and alkali-hornblende together with some biotite. Rocks transitional between syenites and diorites are known as monzonites (*q.v.*).

The following analyses show the chemical composition of a few of the principal types of syenite. They are characterized by a moderate amount of silica, relatively high alkalis and alumina, while lime and magnesia are more variable, but never in great amount.

	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O
1	60.52	16.65	2.97	2.15	2.32	4.73	4.43	4.39
2	59.37	17.92	6.77	2.02	1.83	4.16	1.24	6.68
3	59.88	17.87	2.67	1.50	1.04	2.01	7.96	5.69
4	58.97	16.33	2.72	4.14	2.72	4.29	4.10	4.54

1. Hornblende-syenite, Plauen'scher Grund, Dresden, Ger.
2. Hornblende-syenite, Biella, Piedmont, Italy.
3. Nordmarkite, Oslo, Nor.
4. Syenite (mean of 12 analyses). (C. E. T.)

SYKES, SIR MARK, 6TH BARONET (1879–1919), English traveller and politician, was born on March 16, 1879, in London, the only son of Sir Tatton Sykes, 5th baronet. He was educated at Beaumont college, Windsor, at Monaco, at Brussels and at Jesus college, Cambridge. He served with the Yorkshire militia in the South African War (1902). On his return he travelled for some time in Syria, Mesopotamia and southern Kurdistan, and, after a short period (1904–05) as secretary to George Wyndham in Ireland, was appointed honorary attaché to the British embassy in Constantinople.

At the outbreak of World War I he raised a battalion of the Yorkshire regiment, but did not accompany it to France. He was sent by the government on important special missions to Bulgaria, Serbia, Russia, Mesopotamia and Syria. His special knowledge and qualifications led, in 1915, to his participation in the Anglo-French conversations in London on the Syrian question. He was an enthusiastic advocate of Arab independence. The settlement arrived at is known as the Sykes-Picot agreement. Sykes then became attached to the foreign office and was on several occasions sent on missions to Egypt. He was on friendly terms with the Arabs and designed the flag under which they marched to Damascus. In 1918, hoping to reconcile the aims of the French and the Arabs, he established himself at Aleppo, and

his diplomacy in a difficult position rendered valuable service to the government. His efforts, however, told on his health and he died in Paris of influenza on Feb. 16, 1919.

His publications include: *Through Five Turkish Provinces* (1900); *Dar-ul-Islam* (1904); *Five Mansions of the House of Othman* (1909); *The Caliph's Last Heritage* (1915).

SYLHET, a municipality and district in the Chittagong division of East Pakistan. The town is on the right bank of the Surma river, on rising ground, embowered in groves. It is on a branch of the metre gauge section of the Eastern Bengal railway, 123 mi. N.E. of Dacca. Pop. (1961) 37,740.

SYLHET DISTRICT has an area of 4,785 sq.mi. The present district comprises five-sixths of the former British Indian district of the same name, which was part of Assam until 1947. It consists of the lower valley of the Surma (Barak) river, and for the most part is a uniform plain, broken only by scattered clusters of sandy hillocks called *tilas,* and intersected by a network of rivers and drainage channels. It is a broad and densely cultivated plain. The rainfall is heavy, reaching an annual average of 159 in., of which more than 100 in. fall between June and October.

In 1961 the population was 3,489,589. The normal cultivated area is 2,400,000 ac., giving an average of nearly an acre a head but much of the district is low lying and subject to floods. Tea cultivation is a flourishing industry in the southern hills. Lime is extensively quarried; under the Mogul government Sylhet lime was a monopoly of the nawabs of Bengal. Other industries are boatbuilding and the manufacture of fine reed mats, buttons from fresh-water shells, ironwork inlaid with brass, and perfume. The perfume is exported to Turkey, Arabia and China. There are undergraduate colleges connected with Dacca university at Sylhet, Habiganj and Sunamganj.

SYLLOGISM, a two-premiss argument of any of several rather simple types distinguished by Aristotle or by those who followed in his tradition. Syllogisms are usually classified as categorical, hypothetical or disjunctive, depending upon the nature of their premisses. A categorical syllogism consists of three categorical statements (*i.e.*, statements of the forms "All *S* is *P*," "No *S* is *P*," "Some *S* is *P*," or "Some *S* is not *P*") which contain a total of at most three terms among them, each term occurring in at least two of the statements. A typical example would be:

No matadors are pirates.
Some Spaniards are matadors.
Therefore, some Spaniards are not pirates.

which is of the form (or mood) called *Ferio*:

No *M* is *P*.
Some *S* is *M*.
Therefore, some *S* is not *P*.

In the hypothetical syllogism one or both of the premisses are hypothetical in form (*i.e.*, have the form "If *p* then *q*"). Thus

If *p* then *q*.
If *q* then *r*.
Therefore, if *p* then *r*.

is one of the valid moods of the hypothetical syllogism. Similarly, a syllogism is disjunctive if one of its premisses is a disjunction. For example, every argument of the form:

Either *p* or *q*.
Not *p*.
Therefore, *q*.

is a valid disjunctive syllogism.

The term *syllogismos* ("syllogism") was used by Plato and occasionally by Aristotle simply with the sense of "argument." To this effect the latter gave his famous definition: "A syllogism is a discourse in which, certain things being posited, something else follows from them by necessity." In most cases, however, Aristotle used the term technically for the categorical syllogism. Later the meaning was extended by the Peripatetics to include hypothetical and disjunctive syllogisms. In Stoic usage the adjective *syllogisticos* ("syllogistic") was applied to any arguments of their five basic types.

See LOGIC and LOGIC, HISTORY OF. (B. MS.)

SYLT, the largest German island in the North sea, being about 36 sq.mi. in area and nearly 23 mi. long. It is, however, very narrow, being generally about half a mile in width, except in the middle, where it extends as a peninsula to the east 7 mi. across. It lies from 7 to 12 mi. from the Schleswig coast. The chief places are Keitum, Tinnum, Morsum, Rantum and Westerland. The latter is a favourite pleasure resort. During the Danish War of 1864 the island was in the hands of the Danes but was occupied by the Prussians later in that year. *See* FRISIAN ISLANDS.

SYLVESTER, JAMES JOSEPH (1814–1897), English mathematician who with his contemporary, Arthur Cayley (*q.v.*), founded the theory of algebraic invariants, was born in London on Sept. 3, 1814. In 1831 he entered St. John's college, Cambridge, where he gained high honours. In 1838 he was appointed professor of natural philosophy at University college, London. In 1841 he accepted the professorship of mathematics at the University of Virginia, but resigned after several months. Sylvester returned to London in 1845 where he obtained employment in actuarial work and took private pupils, one of whom was Florence Nightingale. In 1846 he entered the Inner Temple and in 1850 was called to the bar. About this time Sylvester and Cayley began an enthusiastic and profitable mathematical partnership. From 1855 to 1870 he was professor of mathematics at the Royal Military academy, Woolwich. It was not until 1876 that Sylvester got the first opportunity of his life commensurate with his talents, when he was appointed professor of mathematics at Johns Hopkins university, Baltimore, Md. In 1883 he returned to England to succeed Henry J. S. Smith as professor of geometry at Oxford. He died in London, March 15, 1897.

Sylvester was primarily an algebraist. With Cayley, he created the theory of algebraic forms. In the theory of numbers he did brilliant work, particularly in partitions and diophantine analysis. He worked by inspiration, and frequently it is difficult to detect a proof in what he confidently asserted. All his work is characterized by powerful imagination and inventiveness. While at Johns Hopkins, Sylvester founded (1878) the *American Journal of Mathematics,* thus contributing to the development of mathematical research in the U.S.

See Eric T. Bell, *Men of Mathematics* (1937). (E. T. B.; X.)

SYLVESTER, JOSUAH (1563–1618), English poet-translator, best known as a translator of a popular biblical epic, was born in Kent and died on Sept. 28, 1618, at Middleburg in the Low Countries, where as secretary to the Merchant Adventurers he wound up a varied career. His chief production, the *Divine Weeks and Works,* translated from the French Protestant poet Guillaume du Bartas (1544–90), appeared in sections in 1592 and complete in 1608. This epic on the Creation, the fall of man and other early parts of Genesis had high prestige in Europe generally and, in Sylvester's translation, was extremely popular in England in the first half of the 17th century. Its appeal was in its biblical matter, its reassuringly orthodox treatment of religion, nature as God's handiwork, science and cosmology, and in the grandiose, picturesque vigour of the writing—a quality only heightened in Sylvester's frequent extravagance of phrase, which Dryden, a youthful admirer, came to regard as "abominable fustian." Sylvester had some small influence on Milton and also had a share in the early development of the heroic couplet and the "poetic diction" which the Augustans were to refine.

The *Complete Works of Josuah Sylvester* were edited by A. B. Grosart in 2 vol. (1880).

BIBLIOGRAPHY.—P. Weller, *Joshuah Sylvesters englische Übersetzungen der religiösen Epen des Du Bartas* (1902); H. Ashton, *Du Bartas en Angleterre* (1908); A. H. Upham, *The French Influence in English Literature* (1908); G. C. Taylor, *Milton's Use of Du Bartas* (1934); G. Tillotson, *On the Poetry of Pope* (1938); J. Arthos, *The Language of Natural Description in Eighteenth-Century Poetry* (1949); V. L. Simonsen, "Josuah Sylvester's English Translation of Du Bartas' *La Première Semaine,*" *Orbis Littcrarum,* viii, pp. 259–285 (1950).
(D. BH.)

SYLVIIDAE: *see* WARBLER.

SYLVITE, a mineral consisting of potassium chloride and the chief commercial source of potassium compounds, was first observed in 1823 as an incrustation on Vesuvius lava. Well-formed crystals were subsequently found in the salt deposits of Stassfurt, Ger., and Kaluszyn, Pol. It also occurs in the southern

Harz region of Germany, in the nitrate deposits of Chile and Peru and in the Permian salt basin of southeastern New Mexico and adjoining parts of Texas. Sylvite crystallizes in the cubic system with the form of cubes and cubo-octahedra and possesses perfect cleavages parallel to the faces of the cube. The composition is KCl and the crystal structure is similar to that of rock salt, NaCl. Crystals are colourless (sometimes bright blue) and transparent; the hardness is 2; specific gravity 1.99. Like salt, it is highly transparent to heat waves. The name sylvite is from the old pharmaceutical name, *sal digestivus sylvii*. (Cl. F.; X.)

SYLVIUS, FRANCISCUS (Franz de le Boë, or du Bois) (1614–1672), German physician and anatomist, called the true founder of the iatrochemical school (*see* Iatrochemistry) was born in Hanau on March 15, 1614. He studied at various universities, acquired a degree from Basel and became professor at the University of Leyden (1658), where his enthusiasm and original ideas attracted numerous students. An instance of pedagogical novelty, although not originality, was Sylvius' introduction of ward instruction in medical education.

Sylvius based his medical system upon such then recent developments in anatomy and physiology as the circulation of the blood and the recognition of the lymphatics, although still within the general framework of Galen's humoural physiology. At the same time, like his predecessor, J. B. van Helmont (*q.v.*), but omitting the mystical element so strong in the latter, he emphasized the chemical aspects of physiology. His study of salts led him to the belief that they were the result of the interaction of acids and bases, and he recognized their existence in living matter in his *Praxeos medicae idea nova* (1671–74).

According to Sylvius, digestion is a process of fermentation in which the saliva and pancreatic juice play the chief roles, the former determining digestion in the stomach and the latter directing the course of the digested food into blood or feces. The most important processes of normal and pathological life take place in the blood, and diseases are explained and treated chemically, the indications for treatment being based upon the presumed relationship between disease and drug.

In other aspects of medicine Sylvius distinguished between conglomerate and conglobate glands, and it is his name that is associated with the Sylvian aqueduct and Sylvian fissure, described in his *De spirituum animalium in cerebro cerebelloque confectione* (1660), although these structures appear to have been noted earlier by others. He also made some contribution to pediatrics through his posthumous *De morbis infantium* (1674), which, translated into English in 1682, gained favour perhaps more through its author's academic stature than through any singular importance of the contents.

Sylvius died in Leiden, Holland, Nov. 15, 1672.

Bibliography.—M. Foster, *Lectures on the History of Physiology During the Sixteenth, Seventeenth and Eighteenth Centuries* (1924); S. E. Jelliffe, "Franciscus Sylvius," *Proceedings of the Charaka Club*, vol. iii (1910); K. E. Rothschuh, *Geschichte der Physiologie* (1953).

(C. D. O'M.)

SYMBIOSIS is the general term for the partnerships of dissimilar organisms; it denotes simply the living together of different kinds of organisms. In this discussion three major types of symbiosis are recognized: (1) commensalism, in which food or space may be shared without evident benefit or harm to either organism or in which the benefit relation is one-sided, but without harm to the host; (2) mutualism, in which both host and guest (or the two partners) are obviously or apparently benefited and (3) parasitism, in which the host is subject to varying degrees of injury while the parasite benefits. The term symbiosis has been widely used in the restricted sense of mutualism. The broader definition corresponds both to the original definition of symbiosis by Heinrich de Bary (1879) and to modern usage. Because of the great importance of parasitism to human life and economy, this subject is dealt with separately (*see* Parasitology).

Commensalism.—Pierre van Beneden (1876) referred to a commensal as an organism that "requires from his neighbour a simple place on board his vessel, and does not partake of his provisions. The messmate does not live at the expense of his host; all he desires is a home or his friend's superfluities." The arrangement represents a level of relationship a degree removed from that in aggregations of animals about a common food supply.

The commensal relation is often between individual larger host and individual smaller guest; the host organism is unmodified, whereas the commensal guest may show great structural adaptation consonant with its habits, as in the remoras (*q.v.*) that ride attached to sharks and other fishes, or no modification other than the habit of dependence on its host. The main types of dependence shown in commensalism are for shelter, support, locomotion (*i.e.*, distribution and spread) and food supply. These may be single or variously combined. Some of the more familiar household insects associated with man illustrate several of these relations—the house cricket, the common cockroach, etc. Only when the host-guest relation is recognizably specific, a particular guest species being regularly commensal with a particular host species, is commensalism sharply defined. The shelter relation is seen in the inhabitants of the tubes of marine worms or in the insect inhabitants of field mouse nests. The support relation may involve plants sessile upon animals, as algae on turtle shell, or animals on plants, as bryozoans and hydroids on floating sargassum weed (*see* Sargasso Sea). Locomotion and food supply are combined in the remoras. Food supply is the main factor in the host of bacteria and protozoa that inhabit the gut of larger animals without apparent contribution to the digestive process of the host.

Mutualism.—Relationships in which benefits to both partners may be discerned and in which both may be radically modified in adjustment to the mutualistic relation are widespread and present diverse aspects. Mutualism may exist between plant and plant, between plant and animal and between animal and animal. The adjustments are both operational and evolutionary, and the mutualistic relation is one of the major types of co-operation in nature.

Plant and Plant.—In apparently mutualistic relations between plant and plant the degree of benefit to one or the other partner may be obscure. One of the best known of such partnerships is that between fungi and algae to form lichens (*q.v.*), the discovery of which led to the coining of the term symbiosis. The support and protection supplied by the fungus, combined with the carbohydrate synthesis of the alga, makes possible the invasion of inhospitable rock surfaces and other adverse environments by lichens.

The relation of certain fungi with the roots of higher plants (mycorhiza; *q.v.*) is extremely widespread and not, indeed, confined to the higher plants. Fungus mycelium (of various species) may form an enclosing feltlike mat around the roots of trees and other woody plants. In this case the benefit to the higher plant appears to lie in improved absorption of the food-bearing solution at the root surface. Even more intimate relations appear in the mycorhiza of orchids, in which the fungus mycelium may be essential to normal germination and development. The fungus mat may be entirely external, or the fungus may send filaments throughout the host plant. In the latter case the filaments of mycelium on the roots replace the normal root hairs.

The combination of nitrogen-fixing bacteria with leguminous plants recalls the mycorhizal relation in the obscurity of benefit to the bacteria, though in the legumes gall-like root nodules harbour the bacteria. The existence of the fixed association is the presumptive evidence of mutual benefit.

Plant and Animal.—The combination of green algae with various simple animals is made conspicuous by the resultant colouring of the animals. Various single-celled green algae, zoochlorellae, together with superficially similar yellow or brown flagellates, zooxanthellae, are regularly associated with protozoa, sponges, coelenterates, flatworms and certain other types. Zoochlorellae and zooxanthellae are regularly present in the polyps of reef corals thus, in vast total numbers and mass. In its life cycle the free-living marine flatworm *Convoluta roscoffensis* begins existence entirely free from zoochlorellae, acquires a dense intracellular population of the algae as it develops, whereupon it ceases active feeding, and finally loses the algae during senescence. In the similar association of algae with the fresh-water *Chlorohydra*

viridissimus the supply of oxygen and carbohydrate food to the host animal and of carbon dioxide and nitrogenous food to the plant was experimentally examined and largely substantiated. In *Chlorohydra*, the transmission of the zoochlorellae to the egg cell takes place directly, without an alga-free juvenile stage.

The habit of growing various fungi for food developed independently among beetles, termites and ants. The fungi grown by various species of beetles of the families Ipidae and Platypodidae in burrows in tree trunks appear to be species-specific. The leaf-cutting ants commonly extend their destructive activities to horticultural plants and have thus been the focus of much study. The fungus-growing termites of the old world quite obviously developed this habit independently of both ants and beetles. For all these insect gardeners some means of transmitting the fungus to a new nest is essential, and remarkable special apparatus exists in both beetles and ants for this purpose. The means of transmission remains to be discovered in the fungus-growing termites (*see* Social Insects).

An important mutualistic relation exists between larger animals (*i.e.*, many ruminants such as cows, goats, sheep, camels, etc.) and a series of bacteria that inhabit their alimentary tracts, for the bacteria aid in the digestion of cellulose or other refractory food materials and synthesize certain vitamins. Enormous numbers of such bacteria may be harboured in special structures such as caeca or elaborate divisions of the stomach.

Among the most interesting of the mutualistic plant-animal symbioses are those between insects and certain bacteria and fungi. In some instances the microorganisms (or symbiotes) live freely in the lumen of the alimentary tracts or in the associated caeca of their insect hosts. In other instances the symbiotes live within the tissue cells of the insect. In both instances the microorganisms appear to be highly specific and are transmitted from generation to generation by rather complicated mechanisms, usually via the eggs of the insect. In most instances the beneficial role played by the symbiotes is unknown, or only assumed; however, in many cases it has been shown that the insect cannot survive without the symbiotes and that the latter may produce certain essential vitamins or hormonelike substances. In aphids, for example, the symbiotes are believed by some to fix atmospheric nitrogen. The insect-symbiote relationship is an intimate one since usually every individual of an insect species harbours the microorganism. The intracellular symbiotes live in cells called mycetocytes which are frequently grouped into an organ or structure known as a mycetome. Some of the extracellular symbiotes occurring in the alimentary tracts of insects are known to take an active part in the digestive processes of the host (such as the bacteria found in the "fermentation pouches" of larvae of lamellicorn beetles). Certain bacteria are found in large numbers in the "gastric caeca" of higher Hemiptera. The arthropod hosts of the intracellular symbiotes occur in most of the principal orders of insects and numerous families of ticks and mites. Insects of both medical and agricultural importance are concerned, including such groups as cockroaches, lice, bedbugs, reduviid bugs, aphids, scale insects, beetles, ants and many others.

A vast series of plant-animal mutual benefit associations is supplied by the phenomenon of insect fertilization of flowers. These range from the most casual visitation of an open type of flower by a wide variety of insects, in which there may be little or no modification for the accomplishment of pollination, to the most strict relation of a single species of flower with a single species of insect, in which both are modified for their mutual relations. Elaborate and complex flower mechanisms, in which exactness of fit of flower parts to a specific insect, and correspondingly elaborate insect structures that function solely in cross-fertilization of the appropriate plant, have been demonstrated (*see* Pollination).

Animal and Animal.—The mutualistic relations between animal and animal may begin at the simple level of the association of birds with large herbivorous mammals, as of the North American cowbird with domestic cattle, the various tickbirds of other parts of the world and the little white cattle heron (*Bubulcus*), which associates itself with elephants and rhinoceroses as well as with domestic cattle throughout its range from North Africa to the East Indies.

A more intimate level of such mutualism between diverse types of animals is found among marine animals. Hermit crabs of various species may beset their snail-shell house with hydroids (*e.g.*, *Eupagurus constans* with *Hydractinia sodalis*) or with sea anemones (*Eupagurus prideauxi* with *Adamsia* species). Other crabs may regularly bear small sea anemones in their claws. Still others regularly have their shells grown over with various species of sponges. The advantage to the sessile animals is evidently that of supplying locomotion, the advantage to the crustacean lies in protection or camouflage. That the relation is fully mutualistic is shown especially by the hydroid *Hydractinia sodalis*, in which the basal layer that bears the polyps grows forward from the mouth of the snail shell inhabited by the hermit crab, thus making it necessary for the crab to undergo the hazard of changing its snail-shell house as it grows. The foot of the sea anemone *Adamsia* performs the same function for its host.

The cellulose-digesting mutualism of vertebrates and insects with bacteria is paralleled at the animal-animal level by a series of protozoa that accompany the bacteria in the alimentary tracts of various animals, ranging from harmless commensals to effective agents for cellulose digestion and to parasitism. The most sharply defined relation of this kind is represented by the association of primitive termites with characteristic and highly modified protozoans. It has been shown that the faunules of protozoans are characteristic of the species of termite inhabited. It is illuminating that the higher termites have accomplished the physiological step in their evolution that enables them to digest cellulose without the aid of the protozoans.

The fungus gardens of the ants are matched by the relation between ants and aphids, and certain other insects. The ants give elaborate care to the aphids from which they obtain the secreted honeydew, protecting the aphids from adverse weather and moving them from one pasture plant to another. Ants also protect aphids from certain of their natural enemies.

The series of nest inhabitants found regularly in the nests of both termites and ants exhibit the range from commensalism to a pseudomutualism of an elaborate and special form. Various families of insects, especially beetles, are associated with greatly modified ant and termite guests not found elsewhere. These forms produce special secretions, so much valued by the hosts that they treat them as normal and even favoured nest inhabitants.

The interrelations of commensalism, mutualism and parasitism are extremely complex. It is evident that parasitism may evolve from either external or internal commensalism and that mutualism may commonly develop from the various types of commensalism. In the opposite direction, it is quite possible that mutualism may arise by complete adjustment of host and guest in an originally parasitic relation. Thus symbiosis in its manifold physiological and social relations, appears as a biological phenomenon of fundamental importance.

Bibliography.—A. S. Pearse, *Animal Ecology*, 2nd ed. (1939); Sir Frederick Keeble, *Plant-Animals* (1910); Paul Buchner, *Endosymbiose der Tiere mit Pflanzlichen Mikroorganismen* (1953); W. M. Wheeler, *Social Life among the Insects* (1923); E. A. Steinhaus, *Insect Microbiology* (1946) and *Principles of Insect Pathology* (1949); Richard Hesse and Franz Doflein, *Tierbau und Tierleben* (1943); Maurice Caullery, *Parasitism and Symbiosis*, Eng. trans. by A. M. Lysaght (1952).
(W. C. A.; K. P. S.; Ed. A. St.)

SYMBOL, the term given to a visible object representing to the mind the semblance of something which is not shown but realized by association with it. One of the first symbols of the Saviour, the fish, was derived from an acrostic of the Greek word ἰχθύς, the component letters of which were the initials of the five words Ἰησοῦς Χριστός, Θεοῦ Υἱός, Σωτήρ, Jesus Christ, Son of God, Saviour. The ship, another early symbol, represented the Church, in which the faithful are carried over the sea of life. Other symbols are those which were represented by animals, real or fabulous, and were derived from Scripture: thus the lamb typified Christ from St. John's Gospel (i. 29 and 36), and the lion from the Book of Revelation, where (v. 5) Christ is called the "Lion of the tribe of Juda." The peacock stood for immortality;

the phoenix for the Resurrection; the dragon or the serpent for Satan; the stag for the soul thirsting for baptism. The sacred monogram Chi Rho, ☧ supposed to have been the celestial sign seen by the emperor Constantine on the eve of the defeat of Maxentius, represents the two first letters of the Greek word Χριστός which Constantine figured on his labarum, or standard, and is found on early Christian coins bearing also the favourite decoration of the Byzantine sarcophagi.

SYMBOLISTS, THE, the name given to a large but loosely connected group of French poets (many of them of foreign extraction) writing in the 1880's. The term *décadent* was often used as an alternative, especially of the extreme manifestations of the movement. In 1888 the critic Brunetière suggested that the history of French poetry from the 17th century could be plotted in three phases, the architectural, the pictorial and the musical. Bearing this programme in mind and taking as examples from his own literature Dryden's *Ode to Mistress Anne Killigrew,* Keats's *Ode on a Grecian Urn* and Tennyson's *Claribel,* the English reader will best understand the scope and extension of the Symbolist movement. In its origin the movement was a revolt against Naturalism as being too concrete, and against Parnassianism as being too clear-cut, to satisfy young minds who had caught from E. A. Poe and Baudelaire, from Wagner, and to some extent from the Pre-Raphaelites, the "sense of the ineffable" and were seeking a medium for its expression. Their object was not to declaim, not to depict, and least of all to transcribe, but to suggest: to communicate in their lost subtlety the most intimate and evanescent tones of experience. Baudelaire himself, Banville, Gérard de Nerval and Villiers de l'Isle Adam had already been engaged in the same quest. The common stock of poetic symbolism has been accumulated to serve common uses: a private symbolism, unless it is forced into currency by a great poet, is bound to be largely unintelligible and will often be suspected at least to be nonsense. It is not surprising therefore that the French public was at first half inclined to regard the new poetry as a hoax, and on the other hand to take seriously a volume of parodies *Les Déliquescences d'Adoré Floupette* (1885) which first made it popularly familiar. But in the meantime Verlaine had entered the circle; he gave it a watchword: "Pas de couleur, rien que la nuance," and in his book *Trois Poètes Maudits* introduced the young decadents to their true masters, Rimbaud, Corbière, Mallarmé. The year 1885 may be taken as the centre of the movement; besides the three already mentioned, Rodenbach, Verhaeren and Jean Moréas had then already appeared: Laforgue, H. de Régnier and Vielé Griffin published their first volumes in 1885, and they were followed by Rémy de Gourmont in 1886, Maeterlinck in 1889, Claudel in 1890, Robert de Montesquieu in 1893. With Samain's first volume in this last year the movement may be said to rejoin the main stream. Its principal representative, in prose, had been J. K. Huysmans: as poets, Rimbaud, Verlaine and Mallarmé, still after 40 years dominant influences in French literature. It was Mallarmé who provided the symbolists with their aesthetic theory and the most accomplished models of metric style, and in the whole course of the movement his weekly conversations were probably the most effective influence at work—an influence not unfelt in England, where it coincided with the aesthetic movement and the Celtic renascence.

BIBLIOGRAPHY.—A. Symons, *The Symbolist Movement in Literature* (1899); P. Berthelot, "Symbolisme," in *La Grande Encyclopédie* (edit. A. Berthelot, etc., 31 vols., 1887–1903); Catulle Mendès, *Rapport sur le mouvement poëtique français de 1867 à 1900* (1903); all with bibliographies; also R. de Gourmont, *Livre des Masques* (1896); E. Vigié-Lecocq, *La poésie contemporaine* (1896).

SYMEON, "the Great," Bulgarian tsar (? –927) succeeded his father Boris (*q.v.*) in 893, after the short intervening reign of his dissolute elder brother, Vladimir. Educated in Constantinople, Symeon was probably destined by his father for the archbishopric of Bulgaria, and at one time took monastic orders.

His dominant ambition, himself to ascend the imperial throne of Byzantium—an aspiration afterwards shared by many Bulgarian tsars—led Symeon to embark on many wars with the empire (894, 896, 913, 917, 923). He never took Constantinople, although it was more than once at his mercy; but he took for himself the title of "Tsar of all the Bulgarians" (918) and raised the archbishop of Bulgaria to the rank of a patriarch. In the Balkans he extended the power of Bulgaria over south Macedonia, south Albania and Serbia, which became his vassal; but Bulgaria's shadowy dominion north of the Danube was probably lost during his reign. Symeon was twice married, and left four sons, of whom the second, Peter, succeeded him.

SYMEON METAPHRASTES, the most renowned of the Byzantine hagiographers. The menology of Metaphrastes, who is thought to have lived in the second half of the 10th century, is a collection of lives of saints for the twelve months of the year, consisting of about 150 distinct pieces, some of which are taken bodily from older collections, while others have been subjected to a new recension (μετάφρασις). Among other works attributed (though with some uncertainty) to Symeon are a *Chronicle,* a canonical collection, some letters and poems, and other writings of less importance. The Greeks honour him as a saint on Nov. 28, and an office has been composed in his honour.

See A. Ehrhard, *Die Legendensammlung des Symeon Metaphrastes* (Rome, 1897).

SYMMACHUS, pope from 498 to 514, a native of Sardinia and apparently a convert from paganism, was in deacon's orders when he was elected as successor to Anastasius II and promptly consecrated, in the Lateran, on Nov. 22, 498. The election, however, was not unanimous: another candidate, Laurentius, having the support of a strong Byzantine party, was also consecrated, in Sta. Maria Maggiore, on the same day. A decision was obtained in favour of Symmachus from the Gothic king Theodoric; but peace was not established until 505 or 506, when Theodoric ordered the Laurentian party to surrender the churches of which they had taken possession. An important incident in the protracted controversy was the decision of the "Palmary synod" of 502, in favour of Symmachus. The dispute gave rise to a body of literature which was drawn on by later exponents of the doctrine *quod prima sedes non judicatur a quoquam* (*i.e.,* that no-one can pass judgment on the pope). Symmachus died on July 19, 514.

SYMMACHUS, QUINTUS AURELIUS (*c.* 345–410), was a brilliant representative of 4th-century paganism in Rome. He was educated in Gaul, and in 373 was pro-consul of Africa (for his official career see *C.I.L.* vi. 1699). His public dignities, which included that of pontifex maximus, his great wealth and high character, added to his reputation for eloquence, marked him out as the champion of the pagan senate against the Christian emperors. In 382 he was banished from Rome by Gratian for his protest against the removal of the statue and altar of Victory from the senate-house, and in 384, when he was prefect of the city, he addressed to Valentinian II. a letter praying for the restoration of these symbols. After this Symmachus was involved in the rebellion of Maximus, but obtained his pardon from Theodosius, and appears to have continued in public life up to his death. In 391 he was Consul ordinarius. His honesty, both in public and in private affairs, and his amiability made him very popular. As his letters do not extend beyond the year 402, he probably died soon after that date.

Of his writings we possess: (1) *Panegyrics,* written in his youth, two on Valentinian I and one on the youthful Gratian; (2) nine books of Epistles; (3) parts of *Complimentary Orations,* five from a palimpsest (also containing the *Panegyrics*), of which part is at Milan and part in the Vatican, discovered by Mai, who published the Milan fragments in 1815, the Roman in his *Scriptorum veterum nova collectio,* vol. i (1825), and the whole in 1846; (4) the *Relationes,* which contain an interesting account of public life in Rome, composed for the emperor. All earlier editions are now superseded by that of O. Seeck in *Monumenta Germaniae historica. Auctores antiquissimi* (1883), vi, 1, with introductions of his life, works and chronology.

SYMMETRIC FUNCTIONS. A symmetric function is a rational function of $x_1, x_2, \ldots, x_n$, which is unchanged when any of the $x_1, x_2, \ldots, x_n$ are interchanged. If $x_1, x_2, \ldots, x_n$ are roots of an equation $x^n+a_1 x^{n-1}+\ldots+a_n=0$, then the coefficients a_1 represent certain symmetric functions of the roots, for example, $x_1+x_2+\ldots+x_n=-a_1$, $(-1)^n x_1 x_2 \ldots x_n=a_n$. This type of function occurs in algebra (*see* EQUATIONS, THEORY OF), group theory and statistics. (R. G. SR.)

SYMONDS, JOHN ADDINGTON (1840–1893), British

essayist and historian of the Italian Renaissance, was born at Bristol on Oct. 5, 1840. He was educated at Harrow and at Balliol college, Oxford, and was elected a fellow of Magdalen college. He died in Rome on April 19, 1893.

Of acute sensibility, he developed early symptoms of pulmonary disease and, after many journeys in search of health, finally (1877) settled at Davos Platz in the Swiss highlands, where most of his work was done, though with frequent journeys to Italy. Symonds found in literary work and in the society of his Swiss friends and neighbours relief from torturing metaphysical doubts, and his output was very large. His chief work, *The Renaissance in Italy* (7 vol., 1875–86), is rather a series of extended essays than a systematic history but has passages of both insight and eloquence. He also translated the *Sonnets* of Michelangelo and Campanella (1878), the *Autobiography* of Benvenuto Cellini (1887), and the *Memoirs* of Carlo Gozzi (1890). His *Life of Michelangelo Buonarrotti* (1892) is a work of great industry and research, but shows signs of failing powers.

Earlier he had published two series of *Studies of the Greek Poets* (1873–76), several volumes of travel sketches, and monographs on Shelley (1878), Sir Philip Sidney (1878) and Ben Jonson (1886), and on the day of his death appeared his study of Walt Whitman, with whom he had corresponded and of whom he was one of the first European admirers. His verse is of less account.

BIBLIOGRAPHY.—H. R. F. Brown, *John Addington Symonds; a Biography,* 2nd ed. (1903); H. R. F. Brown, ed., *Letters and Papers of J. A. S.* (1923); Van Wyck Brooks, *J. A. S.* (London, 1914); R. L. Stevenson, *Memories and Portraits* (1887), in which Symonds figures as Opalstein in "Talks and Talkers"; P. L. Babington, *Bibliography of the Writings of J. A. S.* (1925). (J. Ps.)

SYMOND'S YAT, one of the most famous view points on the River Wye, England. It stands on the River Wye 9 mi. above Monmouth and 12 mi. below Ross, on a deeply incised meander nearly 5 mi. long and 600 yd. across the neck. The peninsula forms Huntsham hill, in which there are caverns.

SYMONS, ARTHUR (1865–1945), English poet and critic, was born in Wales of Cornish parents, on Feb. 28, 1865. He was educated privately, spending much of his time in France and Italy. He was a contributor to the *Yellow Book* and in 1896 became with Aubrey Beardsley editor of the *Savoy.* Symons was an admirer of Stéphane Mallarmé, Charles Baudelaire and Paul Verlaine, and he did much to spread the ideas of the French symbolists in England. He was an acute critic and covered a wide field. His poetry belongs mainly to the *fin de siècle* in feeling. His works include: verse—*Silhouettes* (1892), *London Nights* (1895), *Amoris Victima* (1897), *Images of Good and Evil* (1900), *The Fool of the World* (1906); prose—*The Symbolist Movement in Literature* (1899), *Studies in Prose and Verse* (1904), *Spiritual Adventures* (1905), *Studies in Seven Arts* (1906), *Studies in Elizabethan Drama* (1920), *Charles Baudelaire* (1920), *Confessions,* an autobiography (1930). He also contributed a number of articles to the *Encyclopædia Britannica.*

BIBLIOGRAPHY.—*Collected Works of Arthur Symons* (London, 1924, etc.); T. E. Welby, *Arthur Symons* (London, 1925).

SYMPATHETIC SYSTEM: *see* NERVOUS SYSTEM: *Autonomic Nervous System.*

SYMPHONIC POEM (TONE POEM, SYMPHONISCHE DICHTUNG, TONDICHTUNG, POÈME SYMPHONIQUE). A type of composition for orchestra of 19th-century origin, it is generally in one movement in contrast to the several movements of a symphony; it is a kind of program music (*q.v.*) based on a literary, pictorial or any extramusical idea which may affect both its content and form. The term program symphony denotes a symphony based on extramusical ideas—Beethoven's *Pastoral* symphony (No. 6) for example. Whether the listener should be acquainted with the composer's program or simply listen to the music as music is a debatable question; opinion generally favours a knowledge of the program by the listener.

The symphonic poem proper began with Franz Liszt. (However, some overtures of Beethoven, the "Leonore No. 2" for example, are near ancestors.) Liszt first applied the term to a series of 12 compositions, the first of which was called "Ce qu'on entend sur la montagne" ("Heard on the Mountain") (1848–49), based on a poem by Victor Hugo. Also from this series are the popular "Les Préludes" (1850) and "Hamlet" (1858). Liszt was influential in the development of new forms. A contribution to composition technique was his system of theme metamorphosis by which a single theme, through altering the rhythm, changing the tempo, etc., was made to serve the several sections of an entire composition, thus providing musical unity where the program determined the form. "Les Préludes" illustrates the technique. Liszt emphasized the poetic qualities of the theme although at times he inclined toward bombast.

The spirit of nationalism found expression in such works as Friedrich Smetana's set of six compositions, *My Country* (1874–79), of which "The Moldau" is the best known; Jean Sibelius' *Finlandia* (1899); Nicolas Rimsky-Korsakov's *Scheherazade* (1888), its four movements linked by the recurring themes associated with Scheherazade and the sultan. Descriptive ability and character delineation are illustrated in Peter Ilich Tschaikovsky's *Romeo and Juliet* (1869); poetic treatment of subject matter is found in César Franck's *Psyché* (1886–88). Camille Saint-Saëns' *Danse macabre* (1874) is descriptive but not musically involved. Richard Strauss's masterful technique of orchestration and his descriptive and narrative ability are demonstrated in such works as *Don Juan* (1888) and *Till Eulenspiegel* (1895). Extreme realism is often encountered in the Strauss works.

The first quarter of the 20th century witnessed such imaginative, expressive and varied compositions as Claude Debussy's *La Mer* (1903–05); Frederick Delius' *In a Summer Garden* (1908); Arnold Schönberg's *Pelléas and Mélisande* (1902); Charles Griffes' *The Pleasure Dome of Kubla Khan* (1919); and many more. During the second quarter of the 20th century composers generally turned their attention to more specifically musical matters.

See R. W. S. Mendl, "The Art of the Symphonic Poem," *Musical Quarterly* (July 1932); H. Ulrich, *Symphonic Music* (New York, 1952). (P. A. P.)

SYMPHONY, in music. (1) The term *συμφωνία* was used by the Greeks, first, to denote concord, in general, whether in successive or simultaneous sounds; second, in the special sense of concordant pairs of successive sounds (*i.e.*, the "perfect" interval of modern music; the 4th, 5th and octave); and third, as dealing with *ἀντιφωνον,* the concord of the octave, thus meaning the art of singing in octaves, or *magadizing,* as opposed to *ὁμοφωνία,* or singing and playing in unison. In Roman times the word appears in the general sense, which still survives in poetry, viz., as a harmonious concourse of voices and instruments. It also appears to mean a concert. In St. Luke xv, 25, it is distinguished from *χόροι* and translated as signifying "music and dancing." Polybius and others seem to use it as the name of a musical instrument.

(2) In the 17th century the term is used, like "concerto," for certain vocal compositions accompanied by instruments, *e.g.*, the *Symphoniae sacrae* of Heinrich Schütz. The modern use of the word symphony for the instrumental ritornello of a song is also found in Schütz's *Kleine geistliche Concerte.*

(3) The principal modern meaning of the word is a sonata for orchestra (*see* SONATA FORMS). The orchestral symphony originated in the operatic overture (*q.v.*), which in the middle of the 18th century began to assimilate the essentials of the sonata style. Mozart's overture to his early opera, *La Finta Giardiniera,* marks the breaking point between three-movement symphony and operatic overture, since it contains the usual first movement and slow movement, and the curtain rises with what sounds like the beginning of its third movement.

Though the sonata style is dramatic, the stiffness of its early forms did not help C. W. Gluck toward his ideal of an overture that should prepare the listener for the drama. Hence the overtures of Gluck are based on the contrast of loosely knit passages of various textures in vague forms which he learned from San Martini. These are no less evident in the symphonies of Philipp Emanuel Bach.

The differentiation between symphony and overture raised the dignity of the symphony; but the style was more essential than the form; and in Mozart's and Haydn's mature works we find the

sonata form as firmly established in the overture as in the symphony, while the styles are quite distinct. Mozart's most elaborate overture, that of *Die Zauberflöte*, could not possibly be the first movement of one of his later symphonies; nor could the finale of his *Jupiter* symphony be taken for a prelude to an opera.

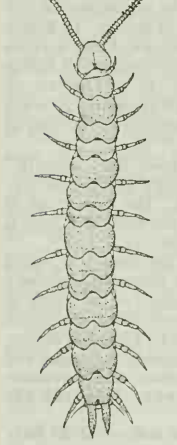
See also MUSIC; SONATA FORMS; ORCHESTRATION; OVERTURE; SCHERZO; VARIATIONS. (D. F. T.)

SYMPHORICARPOS, a genus of the honeysuckle family comprising about 15 species of deciduous shrubs, native to North America except for a single species found in China. The leaves are opposite, simple, usually entire, and without stipules. The flowers are borne on short stalks either clustered or solitary and relatively inconspicuous. The fruit is a berrylike drupe with two seeds, and constitutes the chief decorative value of the plants. Perhaps the best known species are snowberry or waxberry (*S. albus*); wolfberry, also called buckbrush (*S. occidentalis*); and Indian currant or coralberry (*S. orbiculatus*). These, as well as a few other species, are excellent ornamental shrubs for either partly shaded or open places. (J. M. BL.)

SYMPHOSIUS (SYMPOSIUS), Latin writer of riddles, who provided a model for this type of literature in the middle ages, probably lived in north Africa in the 4th or 5th century A.D. Most of his 100 riddles, which consist of three hexameters each, are about animals or objects of everyday life. The prologue states that they were composed for the Saturnalia.

There is an edition by E. Baehrens in the Teubner series, *Poetae Latini minores*, 4:364–385 (1882).

SYMPHYLA, a class of small, many-legged tracheate arthropods (myriapods), the largest representatives being slightly less than one centimetre in length. The skin of these animals is thin, soft and colourless. They live under logs, leaves and stones, where their white bodies contrast sharply with the darker soil. Although symphylids usually subsist upon decaying vegetable matter and delicate rootlets and tissues of living plants. One species in particular, *Scutigerella immaculata*, is known to do extensive damage to such crops as celery, lettuce, asparagus and beets. This form is often spoken of as the "garden centipede," although the absence of poison jaws and other basic differences widely separate it from true centipedes (*q.v.*).



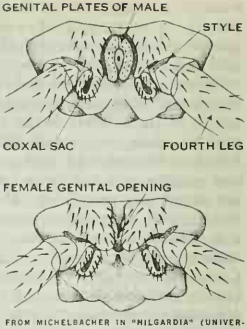
FROM MICHELBACHER IN "HILGARDIA" (UNIVERSITY OF CALIFORNIA)

FIG. 1.—ADULT SYMPHYLID (SCUTIGERELLA IMMACULATA)

As in other myriapods, so-called, the head is distinct. The body is composed of 14 segments, bearing 15, or rarely 22 tergites. Twelve pairs of usually five-jointed legs are present, one pair to each of the first twelve segments. The head, which bears a Y-shaped dorsal suture like that of insects, lacks eyes and bears a pair of long, many-segmented antennae. The mouthparts include a pair of mandibles, a pair of maxillulae (paragnatha) and two pairs of maxillae of which the second pair combine to form a labium as in insects and centipedes. Only one pair of respiratory spiracles are present; these open on the head near the bases of the antennae. The paired reproductive organs open by a single pore lying between the third and fourth pairs of legs, a characteristic placing them with millipedes and pauropods in the now widely recognized group Progoneata. The females lay their eggs in clusters in the ground. As with the millipedes, the young hatch out with three pairs of legs, the rest being developed later.

The symphalids have a world-wide distribution. Because of neglect by most collectors only about 100 species have so far been recognized. Formerly, they were placed in three families, two of which have been placed with the centipedes, the family Scutigerellidae remaining. The name Symphyla refers to the intermediate position of these animals, all of which show obviously connecting characters with other groups of tracheate arthropods, especially insects.

BIBLIOGRAPHY.—H. J. Hansen, "Genera and Species of the Symphyla," *Quart. J. Micr. Sci.*, vol. 53 (1901); C. G. Attems, in Kukenthal and Krumbach's *Lehrbuch der Zoologie* (1926); Verhoeff in Bronn's *Klassen und Ordnungen des Tierreichs*, vol. 5 (1913–14, 1926). (R. V. C.)



FROM MICHELBACHER IN "HILGARDIA" (UNIVERSITY OF CALIFORNIA)

FIG. 2.—SCUTIGERELLA IMMACULATA, FOURTH SEGMENT AS SEEN FROM BELOW, SHOWING DIFFERENCE BETWEEN MALE AND FEMALE

SYNAGOGUE, a Jewish house of worship. In Hellenistic Greek, *synagoge* was the term for assembly and later for the place where the assembly gathered, corresponding to the Hebrew *Keneset* and *Bet ha-keneset*. The phrase "meeting places of God" in Ps. lxxiv, 8 is generally assumed to be a reference to synagogues (and was so translated in the Authorized Version). Talmudic literature also uses the terms "house of the people" (based on Jer. xxxix, 8) and "small sanctuary" (based on Ezekiel xi, 16) while Greek sources speak of "house of prayer" (*proseuche* or *proseukterion*) and occasionally "sabbath house." In modern times "temple" and *shul* (Yiddish, from the German *schule*) are common. United States usage often distinguishes between a temple, meaning a Reform Jewish synagogue, and a *shul*, an Orthodox Jewish house of worship.

The beginnings of the synagogue are obscure. It most probably originated during the Babylonian Exile as an institution for public worship and instruction soon after the destruction of the Temple in 586 B.C. After the return from the Babylonian Captivity, especially under Ezra and his successors, the synagogue developed side by side with the cult of the Temple in Jerusalem. Some scholars trace the origin of regular worship in the synagogue to the institution of Maamadoth, which formed part of the Temple worship; along with each of the 24 priestly divisions, certain towns would arrange for a two-week assembly of prayer and worship. The oldest dated evidence of a synagogue is the dedication of a house of worship to Ptolemy III (3rd century B.C.) inscribed on a marble slab discovered near Alexandria, Egy. Approximately 50 ancient synagogues were excavated in Palestine and the surrounding area during the 19th and 20th centuries, but they are all from the 2nd to the 6th centuries A.D. Philo and Josephus, as well as the New Testament and the early rabbinic sources,

FOTO REISEWITZ, NEUSTADT AN DER WEINSTRASSE, FRIEDRICHSTRASSE 25, GERMANY

INTERIOR OF RASHI CHAPEL, AT WORMS. THE CHAPEL, ATTACHED TO THE WORMS SYNAGOGUE, WAS RESTORED AFTER ITS DESTRUCTION IN 1938

speak of the synagogue as an institution that had been in existence from early antiquity. The literature of the 1st century also refers to numerous synagogues not only in Palestine but in Rome and Greece as well as in various cities in Egypt and Babylonia and throughout Asia Minor. Toward the end of the second commonwealth (middle of 1st century A.D.) there were synagogues wherever there was a Jewish settlement of sufficient size to support organized congregational activity. The destruction of the Temple by Titus (*c.* A.D. 70) had no adverse effect on the synagogue, which had long been a well-established and independent institution. During the centuries following the fall of Jerusalem the synagogue gained in importance, becoming the centre of the corporate life of the Jewish community.

The character of the synagogue, often termed "the child of the dispersion," represented from its inception a radical departure from that of earlier places of worship. It was not restricted to any geographic locale that presumably possessed inherent sanctity; it had no sacrificial or sacramental ritual; it did not require the intermediacy of a special priesthood. Any group of Jews, regardless of place, may establish a synagogue and share the responsibility for the conduct of its affairs. Essentially democratic, the synagogue represents a fellowship of worshipers seeking God through prayer and study.

WERNER BRAUN

PULPIT AND WOMEN'S GALLERY OF CONEGLIANO VENETO SYNAGOGUE, 1701, MOVED FROM VENICE TO JERUSALEM IN 1940

Certainly from the 1st century A.D., and probably much earlier, morning, afternoon and evening services were conducted in the synagogue daily and on the Sabbaths and festivals. These worship services consisted of the recitation of Psalms and of set prayers of praise, petition, thanksgiving and confession as well as readings from the Pentateuch, the Prophets and selections from rabbinic literature. On special occasions, particularly on the Sabbath, the service included a sermonic exposition of the scriptural lesson as well as public lectures on various aspects of the law. The facilities of the synagogue, particularly the rooms adjoining the main prayer hall, were often employed for a variety of other purposes throughout the ages. They served as the meeting place for the leaders of the community and for popular assemblies, for publication of legal notices and decrees of the secular government (during the medieval period), for funerals of its leading members and rabbis, for weddings, circumcisions and, at times, even for lodging of travelers. Above all, the facilities of the synagogue were used for the education of children. The modern synagogue provides regular public worship, a program of education for children, youth and adults and a variety of social, cultural and philanthropic activities.

There is no typical synagogue architecture; the houses of worship were and are built in the prevailing architectural style. Numerous synagogues built after World War II are among the finest examples of advanced architectural design. The only prescriptions found in the Talmud are that the synagogue be built on the heights of a city and that it have windows ("light dwells with Him," Dan. ii, 22; and of Daniel praying, "he had windows in his upper chamber open," Dan. vi, 10), but even these few requirements are only theoretical. The modern synagogue usually contains an ark where the Scrolls of the Law are kept; a light above the ark called the eternal light or perpetual lamp; two candelabra; pews for the worshipers; and a raised platform, *bimah*, from which the scriptural lesson is read and, often, services conducted. In the medieval synagogue, the *bimah* was usually located in the centre of the auditorium and a gallery was built in the back or on the sides of the auditorium for women worshipers. Also, a *mikvah*, a ritual bath, occasionally was included in the synagogue premises. Many Orthodox synagogues today retain all or several of these features.

BIBLIOGRAPHY.—For synagogue in general, *see* various articles under "Synagogue," "Liturgy," etc., in *Jewish Encyclopedia, Universal Jewish Encyclopedia,* Hastings' *Encyclopaedia of Religion and Ethics. See* also G. F. Moore, *Judaism,* vol. i, pp. 281–308 (1927); S. Krauss, *Synagogale Altertümer* (1922); S. Baron, *The Jewish Community* (1945) and *A Social and Religious History of the Jews,* 8 vol. (1952–58). For architecture, *see* E. L. Sukenik, *Ancient Synagogues in Palestine and Greece* (1934); W. G. Tachau, "The Architecture of the Synagogue," *American Jewish Year Book,* vol. xxviii, pp. 155–234 (1926); R. Wischnitzer, *Synagogue Architecture in the United States* (1955). (E. MI.)

SYNAGOGUE ORGANIZATIONS. Each Jewish congregation is an independent entity, governed in most cases by an elected board of directors. Nearly all describe themselves as Reform, Conservative or Orthodox (*see* JUDAISM). No precise figures on the relative strength of the three are available, but it is believed that in the United States their numbers are approximately equal.

United States.—In the past, Orthodoxy was strongest in New York city and in the Middle Atlantic and New England areas. In smaller cities of the midwest, the south and far west the Reform temple was frequently the most prominent Jewish religious institution. Conservatism centred in some of the largest Jewish communities, though it was not especially prominent in New York city. The evolution of American Jewry, however, changed this picture considerably. Thus, the movement of the Jews of New York city to the suburbs added strength to Reform and Conservatism; the growth of Los Angeles as a centre of Jewish population meant the strengthening of Conservatism and Orthodoxy in that area.

In regard to the constituency of each group, the most notable trend is the general decline in social-class differences. Many present-day Reform congregations, in contrast to those of an earlier period, include few if any persons who are members of the upper class. The constituency of Reform is now also overwhelmingly composed of persons descended from immigrants who originated in eastern Europe rather than Germany. While the Orthodox group is extremely variegated in composition, it still possesses the strongest links with the immigrant past. However, Orthodoxy increasingly utilizes American methods of organization, and in many Orthodox congregations English has replaced Yiddish as the language of common discourse.

National synagogue federations exist in the United States for each of the three groups: (1) the Union of American Hebrew Congregations for the Reform group; (2) the United Synagogue of America for the Conservative group; and (3) the Union of Orthodox Jewish Congregations of America for the Orthodox group. The Reform federation, founded in 1873, is the oldest and the best organized of the three. The Conservative federation, founded in 1913, was relatively inactive until the 1940s, becoming after that period, however, an important agency. The Union of Orthodox Jewish Congregations, although founded in 1898, has remained a loosely organized group. Most Orthodox congregations

are small in size and do not have ties with any national body. The synagogues affiliated with the Young Israel movement, however, have achieved a measure of co-ordination, and the Yeshiva University Synagogue council also has become a force in the Orthodox group.

During the 19th century some thought that all synagogues might affiliate with a single federation—hence names such as "Union of American Hebrew Congregations," or the "United Synagogue." Today, however, even those who foresee a single "American Judaism" conceive of this as a far-off development. The Synagogue Council of America, an over-all agency, is composed of both rabbinical and lay representatives from each of the three groups. It has confined its efforts to representing the common interests of Reform, Conservatism and Orthodoxy before interreligious and governmental bodies.

American Jewish congregations have typically been organized by local residents who join together, hold services, start a religious school and raise funds for a synagogue building. Initially at least it is the members themselves who shape the exact religious complexion of the congregation. They decide whether the group should affiliate with one of the national synagogue federations. However, after World War II the federations themselves became somewhat active in promoting the establishment of congregations. Substantial variations exist in each movement, and they tend to overlap. The national bodies have been generally hesitant to promulgate detailed codes of doctrine, though they publish textbooks, magazines and diverse pronouncements for the guidance of their member congregations.

In many cases the synagogue has added an extensive list of activities to its traditional function of providing a place of worship and a centre for study by male adults of the sacred Jewish texts. Congregations increasingly become "synagogue centres" and sponsor club programs, athletic activities and public lectures. They make their facilities available for a variety of Jewish and general communal purposes. Many congregations possess affiliated organizations such as sisterhoods, men's clubs and youth groups, which provide a varied program, generally with an emphasis on social activities. Sisterhoods, men's clubs and youth groups are affiliated with federations sponsored by the national bodies.

The task of providing elementary Jewish education also has become a congregational rather than a communal responsibility. An important motivation for congregational affiliation is the desire to enroll children in the religious school of the institution.

The religious life of Canadian Jewry is essentially the same as that of the United States. Canadian synagogues belong to any one of the three U.S. congregational bodies.

Great Britain.—In Great Britain, greater centralization is provided for. The largest religious body is the United Synagogue, founded in 1870 and operating in greater London (no connection with the United Synagogue of America). In contrast to U.S. practice, the property of affiliated congregations belongs to the United Synagogue. Another distinguishing feature is the office of the chief rabbi, who functions as the religious head both of the United Synagogue and of the other Ashkenazic congregations outside the United Synagogue in London and the provinces.

Since the United Synagogue and the chief rabbinate have stood for Orthodoxy, Reform and Conservative tendencies could not be accommodated within this particular framework. Non-Orthodox congregations have formed themselves into two groups, the Reform Synagogues in Great Britain and the Union of Liberal and Progressive Synagogues. Both these are congregational unions in the American sense: their constituent bodies constitute independent entities.

The provincial congregations are independent, and the non-Orthodox groups in all parts of the country are outside the traditional ecclesiastical framework; in addition, there are Orthodox congregations in the London area that do not belong to the United Synagogue. Two congregational unions serve such bodies, many of which are ultra-Orthodox congregations, or have a more recognizable link with the immigrant past than the average United Synagogue congregation. Furthermore, as in the United States, a few Sephardic congregations exist, and these stand outside the United Synagogue and the other groups, all of which follow the Ashkenazic tradition. Reform and Conservative movements have much less strength in Great Britain than they do in the United States, and English congregations carry on a much more restricted program of activity than do American congregations.

(M. Sk.)

SYNAPSE, the functional conjunction between two neurons, making conduction of nervous impulses continuous from one to the other. Synapses occur between the axon terminations of one neuron and the cell body of an adjacent cell. This type of synapse is called axosomatic. Synapses also occur between axon terminations of one neuron and the dendrites of the next, and are then termed axodendritic. Adjacent neurons may be connected by both types of synapse. The term synaptic junction is used to describe this functional and anatomical relation. (*See* Axon; Nerve; Nerve Conduction; Nervous System.) The term is also used in embryology to denote the pairing and union of the homologous chromosomes of the male and female pronuclei. This process is also named syndesis. (W. M. M.; X.)

SYNCRETISM, the act or system of blending, combining or reconciling inharmonious elements. The term is used technically in politics, as by Plutarch, of those who agree to forget dissensions and to unite in the face of common danger, as the Cretans were said to have done; in philosophy, of the efforts of Cardinal Bessarion and others in the 16th century to reconcile the philosophies of Plato and Aristotle; and in theology, of a plan to harmonize the hostile factions of the church in the 17th century, advocated by Georg Calixtus, a Lutheran professor of theology at Helmstedt. Its most frequent use, however, is in connection with the religious development of antiquity, when it denotes the tendency, especially prominent from the 2nd to the 4th centuries of the Christian era, to simplify and unify the various pagan religons. During this period, as a result of the intimate knowledge of the world's religions made possible by the gathering of every known cult of importance into the religious system of the Roman empire, belief in the identity of many deities which resembled each other, and indeed in the essential identity of all, received a special impulse. Not only were various forms of the same deity, such as, for example, Jupiter Capitolinus and Jupiter Latiaris, recognized as being really the same under different aspects, but even the gods of different nations were seen to be manifestations of a single great being. Roman Jupiter, Greek Zeus, Persian Mithras and Phrygian Attis were one. The Great Mother, Isis, Ceres, Demeter, Ops, Rhea, Tellus, were the same great mother deity under different masks (*see* Great Mother of the Gods). Venus and Cupid, Aphrodite and Adonis, the Great Mother and Attis, Astarte and Baal, Demeter and Dionysus, Isis and Serapis, were essentially the same pair. Syncretism even went so far as to blend the deities of paganism and Christianity. Christ was compared with Attis and Mithras, Isis with the Virgin Mary, etc. Isis, perhaps more than any other deity, came to be regarded as the great maternal goddess of the universe whose essence was worshiped under many different names.

Naturally, the influence of Greek philosophy was very pronounced in the growth of syncretism. Plutarch and Maximus of Tyre affirmed that the gods of the different nations were only different aspects of the same deity, a supreme intelligence and providence which ruled the world. The Neoplatonists, however, were the first school to formulate the underlying philosophy of syncretism: "There is only one real God, the divine, and the subordinate deities are nothing else than abstractions personified, or celestial bodies with spirits; the traditional gods are only demons, that is, beings intermediate between God and man. . . . All, like every other created being, are emanations from the absolute God" (Jean Réville, *La Religion à Rome sous les Sévères,* Paris, 1886). Care must be taken, however, not to place too much emphasis upon syncretism as a conscious system. The movement which it represented was not new in the 2nd century A.D. The identification of Latin with Etruscan gods in the earliest days of Rome, and then of Greek with Italian, and finally of Oriental with the Graeco-Roman, were all alike syncretistic movements, though not all con-

scious and reasoned. The ideal of the common people, who were unreflecting, as well as of philosophers who reflected, was "to grasp the religious verity, one and constant, under the multiplex forms with which legend and tradition had enveloped it" (Réville). The advent of Greek philosophy only hastened the movement by conscious and systematic effort.

(*See* PHILOSOPHY, HISTORY OF.)

Syncretism, being a movement toward monotheism, was the converse of the tendency, so prominent in the early history of Rome, to increase the number of deities by worshiping the same god under special aspects according to special activities. In the hands of the Neoplatonists it was instrumental in retarding somewhat the fall of paganism for the time, but in the end contributed to the success of Christianity by familiarizing men with the belief in one supreme deity.

See also GNOSTICISM.

See J. Réville, *La Religion à Rome sous les Sévères* (1886), especially pages 104–127, 159–174, 284–295. (G. S.)

SYNDICALISM is an anticapitalist doctrine stressing the autonomy of working groups and advocating direct industrial action to overthrow the capitalist order. Syndicalists thus differ from socialists, who advocate capturing the political state, whether by parliamentary or by revolutionary means, as a crucial step in enacting their programs. The syndicalists, like the anarchists, by whom they were influenced, conceive the state as inherently oppressive. For its control, they would substitute that of organized workers based on production units. The structure and administration of this proposed syndicalist community remain vague and utopian.

The syndicalist doctrine was first worked out by certain leaders of the French trade-union movement toward the end of the 19th century. The word *syndicalisme* means simply trade-unionism; what is known in English as syndicalism the French call *syndicalisme révolutionnaire*. The movement flourished in France chiefly between 1900 and 1914 and had a considerable influence on other countries, especially Italy, Spain and Central and South American nations.

It had also certain close resemblances to the direct action advocated in the United States at about the same time by the Industrial Workers of the World (*q.v.*). Like the French syndicalists, the I.W.W. repudiated parliamentary action; but the I.W.W. differed in favouring strong trade-union centralization rather than emphasizing the typically small working units of the French syndicalists. There was also an analogous movement in Great Britain in the early decades of the 20th century, but this movement, called guild socialism (*q.v.*), was greatly modified both by the traditional parliamentarism of the British working class and by the nonrevolutionary character of the movement's intellectual proponents. Nevertheless certain British trade-unions were permeated by syndicalist ideas between 1906 and 1914.

After World War I, syndicalism largely ceased to be a dynamic force, even in France. Its followers tended to be attracted either by revolutionary communism, especially as represented by the U.S.S.R., or by the increasing possibilities in western democracies for socioeconomic improvement through ordinary trade-union activity as well as parliamentary action. However, syndicalism remained an important residual force, at least during the interwar years, in many European countries. Left-wing trade-union agitations, usually called Trotskyite by the Communists, often resulted from old syndicalist influence. In Spain, in particular, an anarchist opposition to Communist party discipline existed in the 1930s. During this same period, the Italian fascist dictatorship of Mussolini sought to use syndicalist sentiment, in perverted form, in recruiting support for his corporate state (*q.v.*). Syndicalist overtones were also discerned in some of the support, but not in the leadership, of the British general strike of 1926. The failure of this strike, which incidentally was hardly revolutionary in its conduct, did much, especially in Britain, to turn working-class efforts toward parliamentary action instead of direct industrial action to achieve political ends.

Syndicalists often claimed that theirs was the most proletarian of all radical social theories and owed the least to middle-class theorists. This was true of its origin but hardly of its later development by Georges Sorel, whose exposition *Reflections on Violence* (1908) is the best-known syndicalist work. Sorel introduced into syndicalism the notion of social myths by which the workers could be irrationally stirred to revolutionary action. The general strike, advocacy of which the syndicalists had taken over from the anarchists, became in Sorel's hands a myth of this kind.

The syndicalist started from the assumptions common to most schools of socialist thought. He affirmed the inherent injustice of the wage system and the fundamental immorality of capitalist society, based, in his belief, on the exploitation of labour. He accepted and pushed to its logical conclusion the Marxian dogma of the class war; he therefore affirmed that solidarity of interests does not, and cannot, exist between capitalists and wage earners. From these premises he drew the usual socialist conclusion: that individual ownership of the instruments of production must be abolished and communal ownership and control substituted for it. But at this point syndicalism and socialism parted company. Whereas the socialist demanded ownership by the state and its dependent organs such as the municipalities, the syndicalist demand was typically for direct ownership and control by the workers, acting through the organizations of their own creation—the trade-unions and federations of unions.

State organization and control of industry are, in the syndicalist view, incompatible with true working-class emancipation. The state is, and must be, an instrument of class domination; it is, indeed, "the executive committee of the capitalist class." It exists to defend the interests of that class and is consequently as much the enemy of labour as capital itself. To extend its powers would be to twine the bonds of wage slavery ever more firmly about the workers' limbs.

The state is, moreover, hopelessly wedded to an uncreative bureaucracy, incapable of initiative and ignorant of industrial technique. Its control, even if it were benevolent (which the syndicalist denies it could be), would necessarily be despotic and inefficient; the spirit of routine would combine with inexperience to crush out the possibility of economic progress. Here the syndicalist endorsed the ordinary individualistic criticism of state socialism. Workers' control, exercised through the *syndicats*, would combine freedom with efficiency.

The form of social organization in which this ideal could be realized was ordinarily conceived somewhat as follows: The unit of organization would be the local *syndicat*. This would be brought into touch with other groups by means of the local *bourse du travail* ("labour exchange"), the established function of which was to act at once as an employment agency and a general centre for trade-union activities. When all the producers were thus linked together by the *bourse*, the administration of the latter would be able to estimate the economic capacities and necessities of the region, could co-ordinate production and, being in touch through other *bourses* with the industrial system as a whole, could arrange for the necessary transfer of materials and commodities, inward and outward. A species of economic federation would thus replace the structure of capitalist industry, along with which the political and administrative machinery of the state would necessarily disappear.

Two features of this utopia need to be emphasized: consumers as such were excluded from any share in industrial control, and a localized system of industry was envisaged. This latter feature was a direct reflection of economic circumstances in France; both industry and trade-unionism were much more local in range than in other and more highly developed countries. But the movement toward large-scale organization which so profoundly affected every aspect of modern economic life produced a corresponding modification in later syndicalist ideals.

Syndicalist theory starts from the idea of a class war which must be waged relentlessly till a complete social transformation has been accomplished. The essential weapon in this struggle is the power of the organized workers. As the cause of the conflict is economic it must necessarily be fought out in the economic sphere. Syndicalist congresses thus persistently repudiated political action and pinned their faith to a general strike as the grand instrument

of social revolution. This reliance upon industrial or direct methods of action flows necessarily from the fundamental notions of syndicalism as to the nature of the state, and also from strictly practical considerations.

Outside the mine or factory, working men hold divergent religious or political opinions which make effective mass action difficult, if not impossible. Inside, the nature of their employment gives them a sense of solidarity which overrides other differences and bands them together in the *syndicat* for common defense. To persuade them to pass from the defensive to the offensive is the syndicalist's task, and in the accomplishment of this, political labels and controversies would be a hindrance. The strike, therefore, is the characteristic syndicalist weapon. However limited in its scope and object, it is an educative experience. Successful, it inspires the workers with a sense of power; unsuccessful, it impresses upon them the servility of their lot and the necessity for better organization and wider aims. Thus, every strike is a preparation for the revolutionary day when the workers, or a fighting minority of them (for syndicalism repudiates as bourgeois the dogma of the sacredness of majority rule), shall seize the instruments of production by an expropriatory general strike. In the meantime, they are working out from day to day, in the ordinary course of their employment, the ethics and the jurisprudence of the new social order.

The syndicalist idea, as understood in France, may be said to have developed out of the strong anarchist and antiparliamentary traditions of the French working class, and to have been much influenced by the teachings of P. J. Proudhon and, on another side, by Louis Auguste Blanqui. A French delegate to the congress of the International Working Men's association (known as the first international) held at Basel in 1869, for instance, prophesied that "the grouping of different trades in the city will form the commune of the future" when "government will be replaced by federated councils of *syndicats* and by a committee of their respective delegates regulating the relations of labour—this taking the place of politics."

These tendencies manifested themselves with increasing strength during the 1890s in the two main labour organizations of the period—the Confédération Générale du Travail (or C.G.T.) and the Fédération des Bourses du Travail. The secretary of this latter organization, Fernand Pelloutier, did more perhaps than any other individual in his day to work out the characteristic doctrines of syndicalism and to spread them among his fellow workers. When these two bodies joined forces in 1902, trade-unionism in general and syndicalism in particular received an immense accession of strength; and the syndicalist doctrine subsequently remained —in spite of the efforts of political socialists to capture the *syndicats* for their own purposes—the characteristic expression of French revolutionary idealism, at least until 1914. *See* also Anarchism; Communism; General Strike; Guild Socialism; Industrial Workers of the World, The; Sabotage; and Socialism.

Bibliography.—H. Lagardelle, *Le Socialisme ouvrier* (1911); G. Sorel, *Réflexions sur la violence* (1908; Eng. trans. by T. Hulme, *Reflections on Violence*, 1914); J. G. Brooks, *American Syndicalism* (1913); L. Levine, *Syndicalism in France* (1914); G. D. H. Cole, *Self-Government in Industry*, 3rd ed. with appendix on syndicalism (1918), *Guild Socialism Restated* (1920); F. W. Coker, *Recent Political Thought*, ch. viii (1934); P. F. Brissenden, *The I. W. W.*, 2nd ed., 2nd ptg. (1957). (G. D. H. C.; J. M. Re.; S. H.; L. D. E.)

SYNECDOCHE (Gr. *synekdoche*, "a receiving together"), a figure of speech in which a part stands for the whole, or the reverse. Examples: "bench," law court; "hand," a worker, as in "hired hand"; "ABC's," alphabet. *See* Figures of Speech: *Figures of Relationship.* (G. W. A.)

SYNESIUS (*c.* 370 or earlier–*c.* 414), bishop of Ptolemais in the Libyan Pentapolis, was born at Cyrene of wealthy parents who claimed descent from Spartan kings.

His extant works, written in Greek, are: (1) a speech before Arcadius, *De regno;* (2) *Dio, sive de suo ipsius instituto;* (3) *Encomium calvitii*, a comic defense of baldness; (4) *De providentia;* (5) *De insomniis* (on dreams); (6) *157 Epistolae;* (7) *12 Hymni;* and several homilies and occasional speeches.

Bibliography.—For a modern edition of the texts *see* N. Terzaghi (ed.), *Hymni* (Rome, 1939) and *Opuscula* (1944); for an English translation, A. Fitz-Gerald, *The Essays and Hymns of Synesius of Cyrene*, 2 vol. (London, 1930), and *The Letters of Synesius of Cyrene* (1926). *See* also C. Lacombrade, *Synésios de Cyrène* (Paris, 1951).

SYNGE, JOHN MILLINGTON (1871–1909), Irish dramatist, was born at Rathfarnham near Dublin, April 16, 1871. Educated at Trinity college, Dublin, he traveled for some years on the continent, spending much of his time in Paris. He returned in 1898 to the Aran Isles. There he wrote a number of sketches dealing with the life of the islanders, which were later collected in *The Aran Islands* (1907). In these and other sketches of the same period he had not quite shaken off the obsession of stylism, and still had a wish "to do for the west of Ireland what Pierre Loti had done for the Bretons." Gradually, however, Ireland took hold of him, and turning to the dramatization of incidents in the life he now knew intimately, he began to elaborate, partly from his notebooks and partly from the writings of Lady Gregory and Douglas Hyde, that richly imaginative Anglo-Irish dialect which he used with such complete success. When, in 1904, he became a director of the newly opened Abbey theatre, Dublin, he had already produced two one-act plays, *The Shadow of the Glen* (1903) and *Riders to the Sea* (1904), published in one volume (1905). *The Well of the Saints,* a beautiful three-act play produced in 1905, was regarded by some as an affront to Irish morals, and when *The Playboy of the Western World* appeared in Jan. 1907, it was interrupted every night for a week by an organized disturbance. When the real merits of the play became known, it contributed largely to the fame of the Abbey theatre. Time has shown how deeply Synge penetrated into the soul of the Irish peasant. The richness of the dialogue and the suggestion which it conveys of a permanent human enigma combine to make it a masterpiece. Synge's latest volumes were a collection of his poetical works entitled *Poems and Translations* (1909), and another play, *Deirdre of the Sorrows* (1910) which he all but completed before his death. Synge's greatest work is probably *The Playboy*, though some critics give pride of place to *Deirdre.* Synge died in Dublin March 24, 1909.

Bibliography.—W. B. Yeats, *Synge and the Ireland of His Time* (1911); M. Bourgeois, *John Millington Synge and the Irish Theatre* (1912); P. P. Howe, *J. M. Synge: A Critical Study* (1912); J. Masefield, *John M. Synge: A Few Personal Recollections, with Biographical Notes* (1915); J. Thorning, *J. M. Synge* (1921).

SYNGE, RICHARD LAURENCE MILLINGTON (1914–), British biochemist, was awarded the 1952 Nobel prize for chemistry jointly with Archer John Porter Martin (*q.v.*) for their method of identifying and separating chemical elements by chromatography. He was born on Oct. 28, 1914, at Liverpool and educated at Winchester college and Trinity college, Cambridge. At Cambridge and in the Wool Industries Research association's laboratories at Leeds he collaborated in chemical research with Martin. In 1943 he was appointed to the Lister Institute of Preventive Medicine, London, and from 1948 he worked at the Rowett Research institute, Bucksburn, Aberdeenshire. He was elected fellow of the Royal society in 1950. Partition chromatography, devised by Martin and Synge, effected the partition of a substance between two liquids by very simple means instead of by a series of intricate chemical operations, and it became so widely used in chemical, biological and medical research as to constitute almost a new scientific tool. In 1951 Synge, with four other scientists, formed a "science for peace" committee, and supported the "People's Congress for Peace" (Vienna, 1952). (D. McK.)

SYNOPTIC GOSPELS: *see* Gospels.

SYNOPTIC WEATHER CHART, a term used peculiarly in meteorology for maps of any large or small portion of the globe, on which are plotted weather observations taken at various points over the area represented and at approximately the same time. In addition the data on the map are usually analyzed by means of conventional symbols, isolines and shadings (*see* Weather Forecasting). The expression weather map or weather chart is essentially synonymous with synoptic chart. The adjective synoptic implies a brief or condensed means of presenting a combined or general view of something; but in meteorological parlance the word came to have the connotation of synchronous because of the

fact that according to most schools of meteorology only synchronous weather observations lend themselves to rational analysis of maps that are to be used for the purpose of preparing forecasts. The notable exception is in the U.S.S.R. where the official weather maps are not based on entirely synchronous observations but on observations made at the same local sun time.

Synoptic charts are constructed in various forms according to their purpose and the kind of data plotted on them. The basic chart for the use of forecasters is that which shows the synchronous observations at sea level or surface stations over a more or less extended area, generally at least a large part of a continent and often a whole hemisphere. For each weather station whose observations are taken at internationally standardized synoptic hours—one to four times a day—and received by radio or telegraph at the forecast offices, there are plotted in a model grouping around the circle representing the station on the map the values of or symbolic indications of a number of weather elements. For all stations at least the barometric pressure (usually reduced to sea level), the air temperature, the present weather, the wind direction and speed or force, and the sky cover will be reported. In some countries many stations also give visibility, dew point or relative humidity, forms, amounts and movements of clouds at different levels, net amount of pressure change during the last three hours, the characteristics or the tendency in the pressure during the last three hours, and the amount of precipitation after the last observation. In case of ship reports the water temperature, state of sea, position and motion of ship are also reported.

An innumerable variety of charts could be constructed from such reports, but in practice only a few are made in addition to the standard charts containing the plotting and analysis of the complete surface reports; *e.g.*, charts to show the change in pressure and the change in temperature since the previous chart or in the last 12 hr. or 24 hr. are often made. The analysis of the basic surface chart consists usually of isobars (*q.v.*), fronts (*q.v.*), and air mass (*q.v.*) designations, and shadings or symbols to show where precipitation and fog are present.

Such maps are two-dimensional. To obtain a three-dimensional synchronous view of the atmosphere, it is customary, in regions where upper-air soundings are regularly made, to construct synoptic charts for several levels in the free atmosphere. Plotted on them, at the appropriate points where soundings are reported, are the values of the elements given by the soundings at that level, usually only pressure, temperature, humidity and wind. The levels selected for upper-air charts may be surfaces of constant height, constant pressure or constant entropy. Constant-height surfaces of 5,000 ft., 10,000 ft. and 20,000 ft. mean sea level have been much used, but during the 1930s and the World War II years, after some experimentation with isentropic charts (*q.v.*), nearly all weather services gradually came to prefer constant-pressure charts; by 1945 they became practically internationally standard.

The World Meteorological organization recommended that data for the following surfaces be evaluated from the soundings and broadcast: 1,000, 850, 700, 500, 300 and 200 millibars; and that analysis of the 700- and 500-millibar surfaces be made as a minimum. After World War II transmittal of sea-level and upper-air synoptic charts by means of facsimile (both telegraphic and radio) was developed, in the United States particularly.

With the advent of jet aircraft, analysis of much higher levels in the atmosphere was required. In this connection some forecasters draw lines of equal wind speed (isotachs, isokinetics) on the constant-pressure charts for 500-, 300-, 200-millibar and even higher surfaces, and charts showing the topography of the tropopause are prepared.

For tropical regions isobaric and constant-pressure charts have not proved very helpful for forecasting, and experiments indicated that streamline charts would be more successful, though the technique is difficult and weather reporting stations were still too sparse over most of the tropics in the 1950s.

In addition to the various types of charts already mentioned, which might be called the basic ones, it is customary in the larger forecast centres to prepare many auxiliary charts such as those showing the changes in height of a pressure surface with time, and those for the thickness and temperature of the layers between a lower and a higher constant-pressure surface. Some of the systems of analysis and forecasting under trial in the 1950s involved the preparation of many charts of an intermediate character representing the step-by-step computation of derived rather than observed factors; *e.g.*, in vorticity analysis and in some methods of numerical prediction.

Finally, there are the prognostic charts which represent the forecast patterns of pressure, height of pressure surfaces, or other elements (temperature, wind, etc.), for some future time.

Thus the total number of charts constructed in a large forecast central may amount to hundreds per day, and consume the services of a large staff working in shifts around the clock.

In connection with forecasting for longer periods of time than one or two days, it is desirable or necessary to construct mean synoptic charts which show averages of the elements over five days, a week, a month or more.

Over much of the northern hemisphere synoptic observations collected immediately by wire and radio permit maps to be made one, two, three or four times a day as desired; in some countries maps can even be made for every hour. The maps can be plotted and analyzed within two to six hours after the observations are made. Upper-air data are available only once or twice a day, however. Large areas of the oceans, and particularly of the polar regions and southern hemisphere, have few observing stations or ships and hence cannot be synoptically charted in the detail required by weather forecasters. Methods of extrapolation of soundings or map patterns to heights above those reached by many of the soundings have been found expedient under certain conditions.

Synoptic charts are published daily by the leading national weather services. A remarkable series of historical weather maps for the northern hemisphere, one a day from Jan. 1, 1899, to June 30, 1939, compiled and analyzed by a uniform technique, were published by the U.S. government (weather bureau) during World War II and were resumed in 1946. The earliest weather charts were made well before 1835 by collecting synchronous weather reports by mail. The first telegraphic collection of synoptic reports and mapping thereof for forecasting was accomplished by Urbain J. J. Leverrier following the Crimean War. From about 1870 the leading nations have published charts each day along with the official forecasts. Simplified sea-level weather charts are published daily in many newspapers, posted in public places and discussed on television programs.

For the principles of analysis and use of synoptic charts in forecasting *see* AIR MASS; FRONTS; ISALLOBAR; ISENTROPIC CHART; ISOBAR; WEATHER FORECASTING.

BIBLIOGRAPHY.—S. Petterssen, *Weather Analysis and Forecasting* (1940); V. P. Starr, *Basic Principles of Weather Forecasting* (1942); H. C. Willett, *Descriptive Meteorology* (1944). (R. G. SE.)

SYRA (SYROS), a Greek island in Cyclades *nomos* (department). Pop. (1961) 19,570; area 34 sq.mi. Its prehistoric importance is attested by Aegean antiquities. Once remarkably fertile, the destruction of its forests left it a brown, barren rock, with scanty aromatic scrub, pastured by sheep and goats. In the 19th century Syra became the commercial centre of the archipelago.

Hermopolis (Hermoupolis), the chief town (pop. [1961] 14,402), on the east coast, is the residence of the nomarch of the Cyclades and the seat of the central law courts. Old Syra, on a hill behind the port town, was built by the Capuchins, who in the middle ages chose Syra as a mission headquarters. On the outbreak of the War of Greek Independence about 40,000 refugees from Chios (Khios) rallied at Syra under the protection of the French until Greek freedom was achieved.

SYRACUSE, a city of Sicily (Gr. Συράκουσαι; Lat. *Syracusae,* Ital. *Siracusa*), the capital of a province of the same name, is a port on the Ionian sea. The city is situated on the east coast of the island, 54 mi. by rail S.E. of Catania, and about 32 mi. direct.

The estimated population (1957) was 78,488 (commune).

Syracuse was the chief Greek city of ancient Sicily, and one

of the earliest Greek settlements in the island. It was founded from Corinth, and Thucydides, who gives the date as the year after the foundation of Naxos (*i.e.*, 734 B.C.), mentions that the leader Archias expelled the Sicel inhabitants from Ortygia. (*See* under SICILY.) Their presence there was definitely proved by the discovery of traces of huts and of a rock-cut tomb of the beginning of the second Sicel period on Ortygia, while similar tombs may be seen both on the north and south edges of the terrace of Epipolae, and on the peninsula of Plemmyrium. (For this and other prehistoric sites in the neighbourhood of Syracuse *see* SICILY, *Archaeology*.)

Till the beginning of the 5th century B.C. our notices of Syracuse history are fragmentary, and we do not even know when the mole connecting the island with the mainland was constructed. In its external development Syracuse differed somewhat from other Sicilian cities. Although it lagged in early times behind both Gela and Acragas (Agrigentum), it very soon began to aim at a combination of land and sea power. In 663 it founded the settlement of Acrae, in 643 Casmenae, and in 598 Camarina, of which the first was unusually far inland. The three together secured for Syracuse a continuous dominion to the south-east coast. They were not strictly colonies but outposts; Camarina indeed was destroyed after a revolt against the ruling city.

Gelon and Hieron.—Hippocrates, tyrant of Gela (492–485), threatened the independence of Syracuse as well as of other cities, and it was saved only by the joint intervention of Corinth and Corcyra and by the cession of the vacant territory of Camarina. In 485 the *Gamoroi*, or landowners (*i.e.*, the descendants of the original settlers, who formed an aristocratic body), who had been expelled by the Demos and the Sicel serfs and had taken refuge at Casmenae, craved help of Gelon, the successor of Hippocrates, who took possession of Syracuse, without opposition, and made it the seat of his power. He gave citizenship both to mercenaries and to settlers from Greece and added to the population the inhabitants of other cities conquered by him, so that Syracuse became a city of mixed population in which the new citizens had the advantage. He then extended the city by including within the fortifications the low ground (or at any rate the western portion of the low ground) between Upper Achradina and the island and making the agora there; at the same time (probably) he was able to shift the position of the crossing to the island by making a new isthmus in the position of the present one, the old mole being broken through so as to afford an outlet from the Little Harbour on the east. The island thus became the inner city, the stronghold of the ruler. Gelon's general rule was mild, and he won fame as the champion of Hellas by his great victory over the Carthaginians at Himera. He is said to have been greeted as king; but he does not seem to have taken the title in any formal way.

Gelon's brother and successor, Hieron (478–467), kept up the power of the city and won himself a name by his encouragement of poets, especially Aeschylus and Simonides, and philosophers, while his Pythian and Olympian victories made him the special subject of the songs of Pindar and Bacchylides. But in his internal government he was suspicious, greedy and cruel. After family disputes the power passed to his brother Thrasybulus, who was driven out next year by a general rising. Syracuse thus became a democratic commonwealth. Renewed freedom was celebrated by a colossal statue of Zeus Eleutherius and by a yearly feast in his honour. But when the mercenaries and other new settlers were shut out from office new struggles arose. The mercenaries, as in the last rising, held Ortygia and Achradina. The people now walled in the suburb of Tyche to the west of Achradina. The mercenaries were at last got rid of in 461. Although we hear of attempts to seize the tyranny and of an institution called *petalism*, like the Athenian ostracism, designed to guard against such dangers, popular government was not seriously threatened for more than 50 years. The part of Syracuse in general Sicilian affairs has been traced in the article SICILY (*q.v.*). We hear of a naval expedition to the Etruscan coast and Corsica about 453 B.C. and of the great military and naval preparations of Syracuse in 439.

The Athenian Siege.—The Athenian siege (415–413) is of the deepest importance for the topography of Syracuse. (The chief authorities for the siege are Thucydides [bks. vi. and vii.], Diodorus [bk. xiii.] and Plutarch, *Nicias*.) Through the whole siege there was a treasonable party within the city, which kept up a correspondence with the besiegers. When the Athenian fleet under Nicias, Alcibiades and Lamachus was at Rhegium in Italy, Athens had the prize within her grasp, and she lost it wholly through the persistent dilatoriness and blundering of Nicias. It was at his advice that the summer and autumn of 415 were frittered away and the siege not begun till the spring of 414. By that time the Syracusans had built a wall from north to south across Epipolae, taking in Tyche and Temenites, so as to screen them from attack on the side of Epipolae on the north-west. They did not, however, occupy Euryelus, at the western extremity of the high ground of Epipolae, and this omission allowed the Athenians to obtain possession of the whole plateau and to begin the investment of the city. The Syracusans then carried a counterwork up the slope of Epipolae, which, if completed, would cut in two the Athenian lines and frustrate the blockade. At this point Nicias showed considerable military skill. The Syracusans' work was destroyed by a prompt and well-executed attack; and a second counterwork carried across marshy ground some distance to the south of Epipolae and near to the Great Harbour was also demolished after a sharp action, in which Lamachus fell, an irretrievable loss. However, the blockade on the land side was now almost complete, and the Athenian fleet had at the same time entered the Great Harbour. The citizens began to think of surrender, and Nicias was so confident that he neglected to push his advantages. He left a gap to the north of the circular fort which formed the centre of the Athenian lines, the point where Epipolae slopes down to the sea, and he omitted to occupy Euryelus. He made a final and irretrievable blunder in letting the Spartan Gylippus first land in Sicily and then march at the head of a small army, across the island, and enter Syracuse by way of Epipolae, past Euryelus. Just before his arrival a few ships from Corinth had made their way into the harbour with the news that a great fleet was already on its way to the relief of the city. The tables were now completely turned. The military skill of Gylippus enabled the Syracusan militia to meet the Athenian troops on equal terms, to wrest from them their fortified position on Plemmyrium, which Nicias had occupied as a naval station shortly after Gylippus's arrival, and thus to drive them to keep their ships on the low beach between their double walls, to take Labdalum, an Athenian fort on the northern edge of Epipolae, and make a third counterwork right along Epipolae in a westerly direction, to the north of the circular fort. The Athenians were thus reduced to such a plight that, as Nicias said in his despatch towards the close of 414, they were themselves besieged rather than besieging. The naval preparations of the Syracusans, under the advice of Hermocrates, had led them, too, to confidence in their powers of giving battle to the Athenian fleet. In the first sea-fight, which took place simultaneously with the capture of Plemmyrium, they had been unsuccessful; but in the spring of 413 they actually won a victory over the Athenians in their own element.

On the very next day, however, a second Athenian fleet arrived under Demosthenes and Eurymedon, with 73 ships of war and a large force of heavy infantry and light troops. Demosthenes decided at once to make a grand attack on Epipolae, with a view to recovering the Athenian blockading lines and driving the Syracusans back within the city walls. The assault was made by night by way of Euryelus under the uncertain light of the moon, and this circumstance turned what was very nearly a successful surprise into a ruinous defeat. The army was now thoroughly out of heart and Demosthenes was for at once breaking up the camp, embarking the troops, and sailing back to Athens. But Nicias could not bring himself to face the Athenian people at home, nor could he be prevailed on to retire promptly to some position on the coast, such as Catania or Thapsus. He dallied till the end of August, many weeks after the defeat, when the coming of Syracusan reinforcements decided him to depart; but

on Aug. 27 was an eclipse of the moon, on the strength of which he insisted on a delay of almost another month. The Syracusans now blocked the mouth of the Great Harbour, and the Athenian fleet, after a frantic effort to break out and a desperate conflict, was utterly defeated and half destroyed. The broken and demoralized army, its ranks thinned by fever and sickness, at last began its hopeless retreat, attempting to reach Catania by a circuitous route; but, harassed by the numerous Syracusan cavalry and darters, after a few days of dreadful suffering it was forced to lay down its arms. The Syracusans sullied the glory of their triumph by putting Nicias and Demosthenes to death.

Dionysius I.—Her great deliverance and victory naturally stirred up the energies of Syracuse at home and abroad. Syracusan and Selinuntine ships under Hermocrates now play a distinguished part in the warfare between Sparta and Athens on the coast of Asia. Under the influence of Diocles the constitution became a still more confirmed democracy, some at least of the magistracies being filled by lot. Diocles appears also as the author of a code of laws of great strictness. Under these influences Hermocrates was banished in 409. In 407 he was slain in an attempt to enter the city, and with him was wounded Dionysius, son of another Hermocrates and an adherent of the aristocratic party, but soon afterwards a demagogue, though supported by some men of rank, among them the historian Philistus. By accusing the generals engaged at Acragas in the war against Carthage, by obtaining the restoration of exiles, by high-handed proceedings at Gela, he secured his own election as sole general with special powers. He next procured from a military assembly at Leontini a vote of a bodyguard; he hired mercenaries and in 406–405 came back to Syracuse as tyrant of the city. Dionysius kept his power till his death 38 years later (367). But it was well-nigh overthrown before he had fully grasped it. After his defeat before Gela (*see* SICILY) his enemies in the army reached Syracuse before him, plundered his house, and horribly maltreated his wife. He came and took his vengeance, slaying and driving out his enemies, who established themselves at Aetna. In 397 Syracuse had to stand a siege from the Carthaginians under Himilco, who took up his quarters at the Olympieum, but his troops in the marshes below suffered from pestilence, and a masterly combined attack by land and sea by Dionysius ended in his utter defeat. Dionysius, however, allowed him to depart without further pressing his advantage. This revolution and the peace with the Carthaginians confirmed Dionysius in the possession of Syracuse, and left it the one great Hellenic city of Sicily, which, however enslaved at home, was at least independent of the barbarian. Dionysius was able, like Gelon, though with less success and less honour, to take up the rôle of the champion of Hellas.

During the long tyranny of Dionysius the city grew greatly in size, population and grandeur. Syracuse absorbed the population of Gela, Camarina and Acragas, and received large accessions from some of the Greek cities of southern Italy, from Hipponium on its west and Caulonia on its east coast, both of which Dionysius captured in 389 B.C. There had also been an influx of free citizens from Rhegium. Dionysius largely extended the fortifications of Syracuse. The island (Ortygia) had been provided with its own defences, converted, in fact, into a separate stronghold. Dionysius, to make himself perfectly safe, drove out a number of the old inhabitants and turned the place into a barracks, he himself living in the citadel. Profiting by the experience gained during the Athenian siege, he included in his new lines the whole plateau of Epipolae, with a strong fortress at Euryelus, its apex on the west; the length of the outer lines (excluding the fortifications of the island) is about 12m. Syracuse was now the most splendid and the best fortified of all Greek cities. Its naval power, too, was vastly increased; the docks were enlarged, and 200 new warships were built. The fleet of Dionysius was the most powerful in the Mediterranean. It was doubtless fear and hatred of Carthage, from which city the Greeks of Sicily had suffered so much, that urged the Syracusans to acquiesce in the enormous expenditure which they must have incurred under the rule of Dionysius. Much, too, was done for the beauty of the city as well as for its strength and defence. Several new temples were built, and gymnasia erected outside the walls.

Dionysius' Successors.—Under his son, the younger Dionysius—an easy, good-natured, unpractical man—a reaction set in amongst the restless citizens of Syracuse. His uncle, Dion, was for a time the trusted political adviser of his nephew Dionysius II. Dion's idea seems to have been to make Dionysius something like a constitutional sovereign, and with this view he brought him into contact with Plato. But Dionysius had Philistus and others about him, who were opposed to any kind of liberal reform, and the result was the banishment of Dion from Syracuse as a dangerous innovator. Ten years afterwards, in 357, the exile entered Achradina a victor. A siege and blockade followed, till Dion made himself finally master of the mainland city. Ortygia, provisions failing, was also soon surrendered. Dion's rule lasted only three years, for he perished in 354 by the hand of an assassin.

Of what took place in Syracuse during the next ten years we know but little. The younger Dionysius came back and from his island fortress again oppressed the citizens. Sicily, too, was again menaced by Carthage. Syracuse, in its extremity, asked help from the mother-city, Corinth; and Timoleon (*q.v.*) delivered Syracuse from the younger Dionysius and from Hicetas, who held the rest of Syracuse, and won a decisive triumph over Carthage and the safe possession of Sicily west of the river Halycus, the largest portion of the island. From 343 to 337 he was supreme at Syracuse, with the hearty good will of the citizens. Syracuse, with an influx of a multitude of new colonists from Greece and from towns of Sicily and Italy, once more became a prosperous city. Timoleon, having accomplished his work, accepted the position of a private citizen. After his death (337) a splendid monument, with porticoes and gymnasia surrounding it, known as the Timoleonteum, was raised to his honour.

In the interval of 20 years between the death of Timoleon and the rise of Agathocles (*q.v.*) to power another revolution at Syracuse transferred the government to an oligarchy of 600 leading citizens. It was shortly after this revolution, in 317, that Agathocles, with a body of mercenaries from Campania and a host of exiles from the Greek cities, backed up by the Carthaginian Hamilcar, became tyrant of the city, and the senate and the heads of the oligarchical party were massacred wholesale, while a unanimous vote of the people gave him absolute control over the fortunes of Syracuse. His wars in Sicily and Africa left him time to do something for the relief of the poorer citizens at the expense of the rich, as well as to erect new fortifications and public buildings; and under his strong government Syracuse seems to have been at least quiet and orderly. After his death in 289 comes another period of revolution and despotism; and but for the brief intervention of Pyrrhus in 278 Syracuse, and indeed all Sicily, would have fallen a prey to the Carthaginians.

A better time began under Hieron II., who had fought under Pyrrhus and who rose from the rank of general of the Syracusan army to be tyrant—king, as he came to be soon styled—about 270. During his reign of more than 50 years Syracuse enjoyed tranquillity and seems to have grown greatly in wealth and population. Hieron's rule was kindly and enlightened, combining good order with a fair share of liberty and self-government. His financial legislation was careful and considerate; his laws as to the customs and the corn tithes were accepted and maintained under the Roman government, and one of the many bad acts of Verres was to set them aside. It was a time, too, for great public works—works for defence at the entrance of the Lesser Harbour between the island and Achradina, and temples and gymnasia. Hieron through his long reign was the stanch friend and ally of Rome in her struggles with Carthage.

Capture by the Romans.—Hieronymus, the grandson of Hieron, thought fit to ally himself with Carthage, but he soon fell in a conspiracy. There was a fierce popular outbreak and more bloodshed; the conspirators were put to death and Hieron's family was murdered; whilst the Carthaginian faction, under the pretence of delivering the city from its tyrants, got the upper hand and drew the citizens into open defiance of Rome. M. Claudius Marcellus was then in command of the Roman army in Sicily, and he threatened the Syracusans with attack unless

they would get rid of Epicydes and Hippocrates, the heads of the anti-Roman faction. Marcellus struck his first blow at Leontini, which was quickly stormed; and its sack roused the feeling of the Syracusans against any negotiations with the Roman general, and, putting themselves under Epicydes and Hippocrates, they closed their gates on him. Marcellus, after an unsuccessful attempt to negotiate, began the siege in regular form (214 B.C.) by both land and sea, establishing a camp on Polichne, where stood the old temple of Olympian Zeus; but he made his chief assault on the northern side and on the defences of Tyche. The city was defended by a numerous soldiery and by the ingenious contrivances of Archimedes, whose engines dealt havoc among the Roman ships and frustrated the attack on the fortifications on the northern slopes of Epipolae. Marcellus had recourse to a blockade, and at last, in 212, information was given him that the Syracusans were celebrating a great festival to Artemis. Making use of this opportunity, he forced the *Hexapylon* entrance by night and established himself in Tyche and on the heights of Epipolae. The strong fortress of Euryelus held out for a time, but, being now isolated, it soon had to surrender. The "outer" and the "inner" city of Thucydides still held out, whilst a Carthaginian fleet was moored off Achradina and Carthaginian troops were encamped on the spot. But a pestilence broke out in the autumn of 212, which swept them clean away and thinned the Roman ranks. The ships sailed away to Carthage; on their way back to Syracuse with supplies they could not get beyond Cape Pachynus owing to adverse winds, and they were confronted by a Roman fleet. All hope for the city being now at an end, the Syracusans threw themselves on the mercy of Marcellus; but Achradina and the island still held out for a brief space under the Syracusan mercenaries, till one of their officers, a Spaniard, betrayed the latter position to the enemy, and at the same time Achradina was carried and taken. Marcellus gave the city up to plunder, and the art treasures in which it was so rich—many of the choicest of them, no doubt—were conveyed to Rome. Archimedes perished in the confusion of the sack while he pursued his studies.

Syracuse was now simply one of the provincial cities of Rome's empire. It retained much of its Greek character and many of its finest public buildings. Its importance and historic associations naturally marked it out as the residence of the Roman praetor or governor of Sicily. Cicero often speaks of it as a particularly splendid and beautiful city, as still in his own day the seat of art and culture. Caligula restored its decayed walls and some of its famous temples. It was plundered by Frankish pirates about A.D. 286. In the 4th century it is named by the poet Ausonius, chiefly, perhaps, on the strength of its historic memories; and there was a purple factory here under an imperial procurator. In 665 Heraclius Constans fixed his capital here, but owing to his oppressive government was assassinated in 668. Syracuse has been a place of comparatively little importance since the year 878, when it was destroyed by the Saracens.

Archaeology.—The mediaeval and modern town of Syracuse (with the exception of a new quarter which has sprung up since the construction of the railway between the station and the island) is confined to the island. This contains the remains of two Doric temples. The older, belonging probably to the beginning of the 6th century B.C., appears, from an inscription on the uppermost step, to have been dedicated to Apollo. It was a peripteral hexastyle of 42 columns; the portion excavated shows that its total width is 74½ft., the width of the cella 38½ft., the lower diameter of the columns 6¼ft. The other temple, into which the cathedral was built in A.D. 640, is to be dated 474–460 B.C. It was a peripteral hexastyle of 36 columns, with a total length of 180ft. and a total breadth of 72ft.; the columns have a lower diameter of 5¾ft., and the intercolumniation is 13½ft. It is almost certainly the temple of Athena. The site was previously occupied by buildings of the 7th and 6th centuries B.C., including an archaic temple and a great sacrificial altar, which were destroyed to make room for it. Their terra-cotta decorations were, however, carefully buried under the pavement of the space which surrounded it; and they are of the highest importance in the history of design, showing that even if the method was Corinthian in origin it was in Sicily that it reached its highest development.

Near the west coast of the island is the famous fountain of Arethusa. According to the legend, the nymph Arethusa was changed into the fountain by Artemis to deliver her from the pursuit of the river-god Alpheus (*q.v.*); and the spring, which was fresh until an earthquake broke the barrier and let in the salt water, was supposed to be actually connected with the river. There are interesting remains of mediaeval architecture in the closely built town with its narrow streets; the beautiful 14th-century windows of the Palazzo Montalto may be especially noticed, and also the 13th-century Castello Maniace at the southern extremity of the island. The town also contains the large and well arranged archaeological museum which is the creation of Prof. Orsi, and is of the highest importance for both pre-Hellenic and Hellenic archaeology. The discoveries of recent years in the south-eastern portion of Sicily, including especially the objects found in Sicel and Greek cemeteries, may be studied there. The isthmus connecting the island with the mainland, which was defended by strong fortifications erected by Charles V. and Philip II. (now demolished), does not occupy the site of the mole erected in the 6th or 7th century B.C., running due north from the north point of the island. On the landward side of the new isthmus was the agora, in which remains of a colonnade of the Roman period have been found. To the west are the remains of an extensive building of the Roman period, probably a palaestra with a small Odeum attached. To the west-north-west is an extensive necropolis, mostly of the 7th and 6th centuries B.C. (For the cemeteries see *The Year's Work in Classical Studies*, p. 123 [1924–25].) This necropolis was included within the defensive wall of Dionysius, a portion of which, no less than 18½ft. thick, was found running diagonally across the new cemetery, and later an outwork in front of it was discovered. East of this point it probably followed the edge of the low terrace above the marsh (the ancient Lysimeleia), while in the other direction it ran north-north-west, making straight for the western edge of the gorge known as the Portella del Fusco, which was thus included within the fortifications, as it would otherwise have afforded a means of access to the enemy. Here the wall gained the top of the cliffs which mark the southern edge of the plateau of Epipolae, which from this point onwards it followed as far as Euryelus. The south wall of Epipolae, considerable remains of which exist, shows traces of different periods in its construction, and was probably often restored. It is built of rectangular blocks of limestone generally quarried on the spot. Euryelus, the point where the terrace of Epipolae narrows down to a ridge about 60yds. wide, which is its only link with the hills to the west, had thrice proved during the Athenian siege to be the key to Syracuse. It now bears the ruins of a mighty fortress, finer than that which defends the entrance to the acropolis of Selinus—the most imposing, indeed, that has come down to us from the Greek period—which there is no doubt is the work of Dionysius. The total length of the works is about 440yds. In front of the castle proper are three ditches, the innermost of which can be reached from the interior of the castle by a complicated system of underground passages. The front of the castle is formed by five massive towers; behind it are two walled courtyards, to the north of the easternmost of which is the well-guarded main entrance to the plateau of Epipolae (narrower minor entrances are to be seen on both the north and the south sides) communicating by a long underground passage with the inner ditch in front of the castle proper. On the north side of Epipolae the cliffs are somewhat more abrupt; here the wall, of a similar construction to that on the south, is also traceable, but here it is apparently all of one period. It is, indeed, recorded by Diodorus that Dionysius built the north wall from Euryelus to the *Hexapylon* in 20 days for a length of 2¾m., employing 60,000 peasants and 6,000 yoke of oxen for the transport of the blocks.

The most important buildings of which we have any remains are to be found in the lower part of Achradina and in Neapolis, a quarter of which we hear first in the time of Dionysius, and which at first was confined to the lower ground below Temenites; but in Roman times included it and the theatre also, though it

did not extend beyond the theatre to the uppermost part of the plateau. In lower Achradina remains of Roman private houses have been found, and it is in this district that the early Christians constructed their catacombs. (St. Paul tarried at Syracuse three days on his way to Rome [Acts xxviii. 12].) Those which are entered from near the 12th-century church of S. Giovanni, situated near an ancient temple, are extensive and important, and include the ancient crypt of S. Marcianus, and the type is different from that of the Roman catacombs, the galleries being far larger (partly owing to the hardness of the limestone in which they are excavated), and having circular chambers at the points of junction. In Neapolis, on the other hand, public buildings predominate. The temple of Apollo Temenites has entirely disappeared, but the theatre, entirely hewn in the rock, is still to be seen. It is the largest in Sicily, being 138½yds. in diameter, and having 59 rows of seats; the 11 lower tiers were originally covered with marble. Each of the nine *cunei* or divisions bore a name; the inscriptions of five of them, still preserved on the rock, are in honour of Zeus, Heracles, King Hieron II., his wife Philistis, and his daughter-in-law Nereis. It was restored after 238, after her marriage to Gelon. Of the stage nothing but cuttings in the rock and foundations are visible. (*See* G. E. Rizzo, *Il Teatro Greco di Siracusa* [1923] for a most careful study.) The situation is well chosen, commanding a splendid view over the Great Harbour, and it is often used for classical performances. Not far off to the southeast is the amphitheatre, probably erected by Augustus when he founded a colony at Syracuse; it is partly cut in the rock and partly built. It is inferior in size only to the Colosseum and the amphitheatres of Capua and Verona, measuring about 153 by 130 yd. over all: the arena is 76 by 43 yd. To the west of the amphitheatre is the foundation of the great altar erected by Hieron II., 217 yd. long by 24 wide, and about 6 yd. in height. To the northwest of the theatre a winding road ascends through the rock, with comparatively late tomb chambers on each side of it. In this district are seen hundreds of small niches cut in the rock, as a rule about 2 ft. square and a few inches deep, which served for containing inscriptions or reliefs, sometimes of a sepulchral character, but sometimes relating to the cult of a divinity. Both the districts just described also contain huge quarries, the famous Lautumiae (from Gr. λᾶας, stone, and τεμεῖν, to cut; hence λατομία, quarry) of Syracuse, over 100 ft. deep and of great extent (though through the collapse of the pillars supporting the undermined rock they have become still larger than they were in ancient times).

We have already seen that immediately outside Lower Neapolis on the south the marshes of Lysimeleia begin, which proved fatal to more than one besieging force. They are traversed by the Anapus, with its tributary the Cyane, the latter famous for the papyrus planted by the Arabs, which here alone in Europe grows wild in the stream. To the south of the Anapus is the hill of Polichne, on which stood the Olympieum, attributed on stylistic grounds to 581 B.C. Its monolithic columns, of which two are still standing, are about 21 ft. in height and 6 ft. in lower diameter: its length is estimated at 197 ft., its breadth at 66⅔ ft. This and the temple of Apollo are the earliest known examples of peripteral temples in stone. The hill was frequently occupied in attacks on Syracuse by the besieging force. The hill of Dascon is a trifle to the southeast. From this point southward the shore of the Great Harbour begins to rise, until the rocky promontory of Plemmyrium is reached.

As an important seaport, Syracuse was bombed by the Allies a number of times during World War II. With the rest of Sicily it was occupied by Anglo-American forces in the summer of 1943.

(E. A. F.; T. A.)

SYRACUSE, a city of New York, U.S., and the seat of Onondaga county, is midway between Albany and Buffalo, at the south end of Lake Onondaga (5 mi. long). It is the central city of the Syracuse standard metropolitan statistical area, which encompasses Madison, Onondaga and Oswego counties. Syracuse itself has lost population to the metropolitan area; in 1960 the population of the city was 216,038, as compared with 220,583 in 1950. The metropolitan area by contrast had a population of 563,781 in 1960, a 21.2 increase since 1950. The growth is spreading north to Liverpool, west to Auburn and east to Chittenango, reaching toward Oneida, Rome and Utica to form a single great central New York urbanized region. (For comparative population figures *see* table in NEW YORK: *Population.*)

History.—The site of Syracuse was the home of the Onondagas and the capital of the Iroquois federation called the Five (later Six) Nations, whose founding father was Hiawatha (himself a Mohawk), the legendary chief of the Onondagas. Known also as "The Long House" or "People of Many Fires," they formed a chain of nations, beginning with the Mohawks at the easternmost point and linking the Oneidas, Onondagas, Cayugas and Senecas on the west (*see* also IROQUOIS). The site was visited by the explorers Samuel de Champlain in 1615 and Pierre Esprit, sieur de Radisson, in 1651 (while a captive of the Mohawks) and by Simon Le Moyne, a Jesuit missionary, in 1654. At the site was a spring from which the Indians would not drink because they believed a demon lived in it, giving the water an evil smell; Le Moyne discovered that it was a fountain of salt brine. In 1655 a mission was founded by Fathers Joseph Chaumonot and Claude Dablon, and the next year a military post was founded near the present village of Liverpool (5 mi. N.W. of Syracuse); both were abandoned in 1658 because of the hostility of the Mohawks, but the mission was resumed in 1668 and continued through the century.

The first white settler at Syracuse was Ephraim Webster, who established a trading post there in 1786. In 1788 Asa Danforth, who came to be called "the father of Onondaga county," built a sawmill and a gristmill and organized the militia. In 1788 the state undertook by treaty with the Onondagas to manage the salt springs; in 1795 it acquired title to the springs and 10 sq.mi. of surrounding land in return for $1,000 in cash and promise of an annual rental of $700 and 150 bu. of salt. In 1797 the state began leasing the salt lands on a royalty basis. Three villages soon sprang up around the southern end of the lake: Webster's Landing and Salina in 1797 and Geddes (founded by James Geddes, the first manufacturer of salt on a large scale) in 1803. The landing was known by several names in succession until 1820, when a post office was established and the name of Syracuse was adopted, after the ancient Greek city in Sicily. Syracuse was incorporated as a village in 1825, became the county seat in 1827 and, absorbing Salina, was chartered as a city in 1847. From time to time various annexations of contiguous territory were made, including the village of Geddes in 1886. The Onondaga Indian reservation is located 6 mi. S. of the city.

Government.—Syracuse has a mayor and council form of government. Because of the development of the metropolitan area, especially Onondaga county, local government is tending more toward a county or metropolitan form; for example, by 1950 it was evident that the supply of water then available was insufficient for further industrial development, and in that year the Onondaga County Water authority was created to attack the problem from an area standpoint. The city and county per capita costs of government are average for the cities and counties of upstate New York, as are the property tax rates. The city has a sales tax and is virtually debt free.

Commerce, Industry and Transportation.—The salt industry provided the original economic foundation of the city, as well as the source of its first private fortunes; Syracuse supplied most of the salt for the United States until 1870. Since then, however, the metropolitan area has developed into a manufacturing centre of considerable importance. The principal products are machinery, metal products, food, paper products and chinaware. It is the wholesale trade centre for the central New York region, rivaled only by Buffalo at one side of the state and Albany at the other. Syracuse is situated at the crossroads of central New York and is well served by highways and railroads. Hancock airport is supplemented by numerous other public and private airports. The State Barge canal is within easy reach; four oil pipelines complete the transportation system.

Education and Cultural Activities.—Syracuse is an outstanding educational centre, led by Syracuse university, a private

coeducational institution founded in 1870 by the Methodist Church. Its colleges and schools offer facilities for graduate, professional and adult education; among them are the Maxwell Graduate School of Citizenship and Public Affairs, the college of forestry (a unit of the State University of New York) and a branch college in Utica. Le Moyne, a Roman Catholic coeducational college, was founded in 1946. Nearby are two state university colleges of education, at Cortland and Oswego, and the Auburn Community college, a two-year institute. There are numerous musical and dramatic groups, including summer theatres at nearby Skaneateles and Fayetteville, the Syracuse Museum of Fine Arts, the Lowe Art centre and the Barrows Art gallery (Skaneateles). The Salt museum near Syracuse and Rail City near Oswego feature the history of those industries. The fair grounds, where state fairs have been held since 1890, are at the western edge of the city.

Recreation.—The recreation opportunities of the city and area include many fine places for hunting, camping, boating, swimming and fishing; eight state parks in the Finger lakes region; Onondaga lake; Lake Ontario and Oneida lake; and the famous Adirondack park. (V. C. C.)

SYR-DARYA (Gr. and Lat. *Jaxartes;* Arab. *Shash* or *Sihun*), (1) a river of Asiatic Russia, flowing into the Sea of Aral, and having a length of 1,370 mi. and a drainage area of about 178,378 sq.miles. Its headstream is the Naryn, rising in the Tianshan south of Lake Issyk-kul, on the south slope (12,000 ft.) of the Terskei Ala-tau. After union with the Barskaun, it flows W.S.W. at 11,000 to 10,000 ft. above the sea, in a barren longitudinal valley between the Terskei Ala-tau and the foothills of the Kokshal-tau. As the Naryn it flows through a wild gorge in the south-west continuation of the Terskei Ala-tau, falling 4,000 feet. Fort Narynsk, 20 m. below the confluence of the Great and the Little Naryn, is only 6,800 ft. above the sea. Here the river enters a broad valley—formerly the bottom of an alpine lake—and flows past the ruins of Fort Kurtka, for 90 m. westward, as a stream some 50 yd. wide and from 3 to 11 ft. deep, used by the Kirghiz for irrigating their cornfields. The At-bash, a large mountain stream, joins the Naryn at the head of this valley and the Alabuga or Arpa at its lower end, both from the left. Before reaching the lowlands the Naryn cuts its way through three ridges which separate the valley of Kurtka from that of Ferghana, and does so by a series of wild gorges and open valleys (170 m.), representing the bottoms of old lakes; the valleys of the Toguztorau, 2,000 ft. lower than Kurtka, and the Ketmen-tube are both cultivated by the Kirghiz. Sweeping north, the river enters Ferghana—also the bottom of an immense lake—where, after receiving the Kara-darya (Black River) near Namangan, it assumes the name of Syr-darya. The Kara-darya is a large stream rising on the northern spurs of the Alai mountains. As it deflects the Naryn towards the west, the natives look upon it as the chief branch of the Syr-darya, but its volume is less. At the confluence the Syr is 1,440 ft. above sea-level.

On issuing from this gorge the Syr enters the Aral depression, and flows for 850 m. in a north-westerly and northerly direction before reaching the Sea of Aral. On this section it is navigated by steamers. Between the Irjar rapids and Baildyr-turgai (where it bends north) the river flows along the base of the subsidiary ranges which flank the Chotkal mountains on the north-west, and receives from the longitudinal valleys of these alpine tracts a series of tributaries (the Angren, the Chirchik, the Keles).

Some 50 m. below Chinaz (770 ft. above sea-level) the Syr bends northwards, but resumes its north-westerly course 150 m. farther down, following with remarkable persistency the edge of the loess. Its low banks, overgrown with reeds and rendered uninhabitable in summer by clouds of mosquitoes, are inundated for 20 m. on both sides when the snows begin to melt. These inundations prevent the moving sands of the Kyzyl-kum desert from approaching the Syr; below Perovsk (Kzyl-Orda, or Ak Mechet) the steppe conquers. Down to Perovsk the river rolls its muddy yellow waters, at the rate of 3 to 5 m. an hour, in a channel 300 to 600 yd. wide and 3 to 5 fathoms deep; at Perovsk its vertical section is 8,220 sq.ft., and 312,500 cu.ft. of water are discharged per second. The Arys and the Bugun are the only tributaries worthy of notice along this part of its course; the other streams which descend from the Kara-tau fail to reach the river. The Kirghiz rear numerous herds of cattle and sheep in the valley of the Arys, while lower down, as far as Julek, the Iginchis carry on agriculture. All this applies of course only to the right bank; on the left the moisture is absorbed by the hot winds which cross the Kyzyl-kum sands towards the river. The dryness of the atmosphere has a marked effect upon the Syr when it gets below Julek, the Kara-kum sands being then on its right. Ten miles below Perovsk the river traverses a marshy depression (the bottom of a lake not yet fully dried up), where it divides into two branches—the Jaman-darya and the Kara-uzyak. The latter spreads out into marshes and ponds, from which it again issues to join the former at Karamakchi, after a course of 80 miles. The main arm, owing to its shallowness and sinuosity, is very difficult to navigate, and the difficulty is increased by the rapidity of the current and the want of fuel. Between Kazalinsk and the Sea of Aral (158 ft.) navigation becomes somewhat easier, except for the last 10 m., where the river divides into three shallow branches before entering the "Blue sea." All three have at their mouths sandy bars with only 3 ft. of water.

Two former right-hand tributaries of the Syr—the Chu and the Sary-su—now disappear in the sands some 60 m. before reaching it. The Chu, which is 600 m. in length, rises in the Tian-shan south-west of Lake Issyk-kul, and as the Kashkar flows towards Lake Issyk-kul, but a few miles before reaching that lake turns suddenly to the north-west, enters under the name of Chu the narrow gorge of Buam, and, piercing the snow-clad Kunghei Ala-tau, emerges on its northern slope, having descended from 5,500 ft. to less than 2,000 in a distance of not more than 50 miles. In this part of its course it receives from the right the Kebin, whose high valley equals in size that of the upper Rhone. It then flows north-westwards through the valley of Pishpek, and, avoiding the Muyun-kum sands, describes a wide curve to the north before finally taking a western direction. Numberless streams flow towards it from the snow-clad Alexander mountains, but they are for the most part lost in the sands before reaching it. The Talas, 170 m. long, formerly an affluent of the Chu, which rises in the highest parts of that range, pierces the Cha-archa mountains, and, flowing past Aulie-ata on the south border of the Muyun-kum, enters the salt lake of Kara-kul 60 m. from the Chu. The Chu terminates in the Saumal-kul group of lakes, 60 m. from the Syr. Another elongated group of lakes—the Uzun-kul—near the above, receives the Sary-su, which has a length of nearly 570 m. and flows rapidly in a narrow channel along the western edge of the northern Famine Steppe (Bekpak-dala).

The delta of the Syr begins at Perovsk, whence it sends a branch to the south-west, the Jany-darya (New River), which formerly reached the south-eastern corner of the Sea of Aral, very near to the mouth of the Amu-darya. The Kirghiz affirm that a canal dug for irrigation by the Kara-kalpaks gave origin to this river. It had, however, but a temporary existence. A dam erected by the people of Khokand at Ak-mechet (Perovsk) caused its disappearance, and the Russians found nothing but a dry bed in 1820. When the dam was removed the Jany-darya again reappeared, but it failed to reach the Sea of Aral; in 1853 it terminated in Lake Kuchka-denghiz, after a course of 250 m.; all traces of its bed were then lost in the sand. Five centuries ago, in the time of Timur, the Mongol prince of Samarkand, the Jany-darya brought the waters of the Syr to the Daukara lakes, close by the present mouth of the Amu. The series of old river-beds in the Kyzyl-kum, which are still seen above Perovsk, indicate that the Syr had a constant tendency to seek a channel to the south-west, and that its present delta is but a vestige of what it was once. At a still more remote period this delta probably comprised all the space between the Kara-tau and the Nura-tau in Samarkand; and the series of elongated lakes at the base of the Nura-tau—the Tuz-kaneh and Bogdan-ata lakes—represent an old branch of the delta of the Syr which probably joined the Zarafshan before reaching the Amu. The cause of this immense change is simply the rapid desiccation of all northern and central Asia. The extension of the Caspian sea as far as the Sary-kamysh lakes

during the post-Pliocene period and the extension of the Sea of Aral at least 100 mi. to the east of its present position are both proved by the existence of post-Pliocene marine deposits.

(P. A. K.; J. T. Be.)

See W. R. Rickmers, *The Duab of Turkestan* (1913); A. Boutquin, *L'Asie Centrale, la Question du dessèchement du globe* (1910).

(2) A former province of Asiatic Russia, now included in the Kazakstan A.S.S.R. and Uzbek A.S.S.R. (*qq.v.*).

SYRIA, a geographical name applied at different epochs to different areas. In ancient times Syria was the fertile strip between the eastern coast of the Mediterranean and the Arabian desert, from the Gulf of Alexandretta to Sinai. It included therefore the territories of the modern states of Lebanon, Israel and Jordan. The name Greater Syria, used in the 20th century by Arab nationalists, vaguely corresponds to this definition.

Lying as it does where three continents converge, Syria was usually part of larger empires; Hittites, Egyptians, Babylonians and Persians successively dominated the country. Alexander of Macedon conquered it. In the 1st century B.C. it became a Roman province. As part of the Eastern Roman empire it was again invaded by the Persians and later by the Arabs, becoming in the 7th century part of the caliphate. Turkish, Mongol and Mameluke invasions, as well as crusades fill the period from 11th to the 16th century when Syria was incorporated in the Ottoman empire.

After World War I it came under the French mandate, but without Palestine and Transjordan.

After World War II the independent states of Syria and Lebanon (*q.v.*) were created from the French mandated territory. This independent Syria is bordered on the northwest and on the north by Turkey, on the east and on the southeast by Iraq, on the south by Jordan and Israel and on the west by Lebanon and the Mediterranean.

Physiography.—The whole area of the states of Syria and Lebanon (*q.v.*) may be regarded as a single geographical unit extending from the eastern shore of the Mediterranean to the Syrian desert (which is crossed diagonally from northwest to southeast by the Euphrates) and the Jazirah east of that river. It is bounded on the north by the foothills of the Taurus ranges and terminates south of the Anti-Lebanon and Mount Hermon with the volcanic region of the Hauran and the Jabal Druse. The coastal strip is a narrow, discontinuous plain measuring about 200 mi. in a direct line from Kara Duran on the northern frontier to Ras an-Nakurah (Ras en Naqura) in the south. There are many varieties of shore line, from the cliffs and headlands north of Latakia and the rugged coast between Tripoli and Beirut, to the sandy shore near Tyre. Behind, from north to south, lie the Ansariyah and Lebanon ranges divided by the valley of the southern Nahr al-Kabir (the ancient Eleutherus). The Ansariyah averages 18 to 20 mi. in width and is of fairly uniform, rounded outline intersected by deep valleys. The highest peak, Nebi Yunis (5,194 ft.), is toward the northern end of the range. The Lebanon rises steeply from the Bika (Bekaa) plain to a tableland more than 8,000 ft. high from which rise several peaks, the highest being Kurnat as-Sauda (Qurnet es Sauda) (10,131 ft.). The main pass, which carries the Beirut-Damascus road, is the Dahr al-Baidar (5,057 ft.). South of this pass the main ridge is narrower and falls away southeast to the plain of Merj Ayun, south to the Litani gorge and southwest to the coast near Sidon. The upper slopes of the Lebanon are bleak, stony and desolate; the middle slopes, below 5,500 ft., are the most densely populated region and cultivation extends up to 5,200 ft. East of the coast ranges is a long depression. At the northern end are the Orontes valley and the Ghab marshes which extend for 38 mi. from north to south with a maximum width of about 9 mi. South of the Ghab the Orontes follows a tortuous course to Lake Homs, south of which is the Bika valley. The northern end of this valley, through which the Orontes flows is, for the most part, poor soil strewn with basalt boulders mixed with limestone and flint. South of the Baalbek ridge, the watershed between the Orontes and the Litani, the soil is rich and well-cultivated. The Litani river falls very gradually through it from northeast to southwest until it reaches the barrier of Jabal Gharbi, a low ridge between the Lebanon and Mount Hermon. The Bika valley is dominated on the east by the Anti-Lebanon range which runs from northeast to southwest nearly parallel with the Lebanon. The main ridge (8,300 ft.) rises steeply about 17 mi. south of Homs and extends for 65 mi. to the Zebdani depression through which pass the railway and road between Beirut and Damascus. Mount Hermon (9,232 ft.) rises south of this gap. Northeast of Damascus are the three Kalamun ridges, the easternmost of which stretches for more than 130 mi. to Palmyra. Northwest of Palmyra lies a semiarid steppe sloping eastward to the Euphrates. The Syrian desert, the Hamad, to the east of the Anti-Lebanon, comes up in places almost to the foot of the range. Where water is available (as in the Ghuta around Damascus which owes its fertility to the Barada [the Abana of the Old Testament]), or rainfall is abundant (as in the Hauran farther south), cultivation pushes out toward the desert. East of the Hauran lies the Jabal Druse (5,900 ft.), much of which is fertile though large areas of the district are lava-covered. To the northeast of the Euphrates is the Jazirah, roughly triangular in its Syrian portion, with a projection in the northeast (the "Duck's Bill") to the Tigris between Jazirat-ibn-Omar (in Turkish Cizre) and Faish Khabur. Except in the neighbourhood of the Euphrates and its tributaries, the Balikh and the Khabur, which cross the Jazirah from north to south, it is a district of stony hills and low ridges.

Climate.—Syria has a climate varying from the Mediterranean type in the west through the steppe type to the continental or extreme type, with a wide range of temperature and absolute aridity for six months, in the inland region, which is open to the Arabian desert on the south. The two gaps in the mountain ranges screening the interior from the sea cause an extension of the maritime influence inland to the region south of Homs and to the Hauran and Jabal Druse south of Damascus. The rainy season is from November to March, though abundant rainfall may occur in April or May. The average of rainfall varies from 20–40 in. in the coastal belt (rising to 50 in. or more in the mountains), through 10–20 in. between Aleppo and Damascus, to 5–10 in. in the crescent from Damascus to the northern frontier and to Dair az-Zor (Deir) and to as little as 0.5 in. in the desert from Palmyra to the southeast. Snow falls on the mountains and may lie even on Mount Hermon until late in the summer. Normally Syria has a hot summer and a mild winter. Temperatures increase from west to east and also, though to a lesser extent, from north to south. Summer temperatures, taking an average over a number of years, vary from mean monthly maxima and minima (August) of 106° F. and 68° F. at Aleppo to 110° and 65° at Palmyra, with absolute maxima at Aleppo (115° F.), Damascus (113°), Palmyra (120°) and Dair az-Zor (115°). Mean monthly maxima and minima for January, which is usually the coldest month, are: Aleppo 58° F. and 22° F., Damascus 62° F. and 28° F., Palmyra 62° F. and 25° F.

(K. C. B.)

Vegetation.—The coastal strip of Syria is intensely cultivated and planted with groves of mulberries, olives, etc., and there is little space for natural flora. Rocky patches bear brushwood with *Calycotome villosa*, buckthorn, tamarisk, etc.; in swampy places reed-grasses grow; and among the crops brilliant flowers of the Mediterranean element such as *Anemone coronaria*, *Ranunculus asiaticus*, *Papaver syriacum*, etc., spring up wherever possible. There is high forest of Aleppo pine in the north of the Ansariyah range with sessile and Syrian oaks (*Quercus syriaca*); in the centre are stunted oak woods; in the south mulberry, valonia oak (*Q. aegilops*), etc. The shrub constituent includes rock roses, tamarisk, terebinth, oleander, etc. On the Lebanon mountains there are forests of oaks, Aleppo pine, cypress, juniper, cedar and fir. Near Beirut there has been considerable re-afforestation with stone pine, and large areas of the lower slopes, up to 5,000 ft., are planted with olive, fig and mulberry (for silkworms). Wild plants in this zone include valerian and danewort, and above, where wheat and barley are cultivated (up to 6,000 ft.) is the montane region of sea holly and rhododendrons. The alpine zone supports peculiar plants such as *Acantholimon libanoticum*, a spiny milk vetch which grows in clumps, and sainfoin. The Bika plain is almost treeless while composites, labiates and milk vetches are characteristic. On the foothills of the steppes terebinth predominates; on the plains

wormwood; and in the Jabal Druse maquis with forests of kermes oak. The Hamad supports only halophytes and xerophytes except along the Euphrates and other water courses where poplars, tamarisk and a few others flourish.

Fauna.—In the upper regions of the Lebanon bear is still occasionally found—the largest remaining wild animal in Syria. Antelope and gazelle are common in the Hamad. Other mammals include deer, polecat, wild cat, ichneumon, porcupine, hedgehog, squirrel, marten, ermine, otter, dormouse, rats, hares and the hyrax. Many migrating birds call at Syria on their way north and south. Flamingoes, pelicans, cormorants, ducks, herons, snipe, etc., frequent the fresh waters and marshes; bustards and ostriches the desert; kingfishers, cuckoos, woodpeckers, etc., the valley; eagles, buzzards, kites and falcons the mountains. Horned and other vipers occur in the desert, and lizards and chameleons are common. Of innumerable insects, the mosquito is the most harmful but locusts and grasshoppers also do damage. In the lakes and rivers mullet, eels and bass are among the fish. (X.)

HISTORY

Flint implements of palaeolithic types have been found in caves in Syria and in Palestine, associated with bones of the rhinoceros and of the hippopotamus and, in the district of Mount Carmel, with skeletal remains of a manlike creature (these dating to the Mousterian period). The remains and industries of man, who already made use of fire, continue through the late Palaeolithic into the Mesolithic age, which is well represented in the so-called Natufian culture found chiefly near Jerusalem. This was the work of men who appear to have been smaller than their predecessors and might be racially similar to the Hamites and Semites. At first purely hunters, they began, before the end of the Mesolithic period, to cultivate the wheat and barley which, in their wild state, were indigenous to the country; and thus they made the decisive step from a nomadic to a sedentary life in which the arts of civilization could be born. Rough pottery is found in a very early age (perhaps *c.* 4000 B.C.) at Jericho, and not long afterward there was centred in northern Syria and in northern Mesopotamia a prehistoric manufacture of the finely made and beautifully painted ware named after Tell Halaf, the place of its first discovery. At about the same time began the use of copper for tools and ornaments. Oxen, sheep, goats and pigs were domesticated, and even irrigation was practised. Population was mixed, the heavier Armenoids being mingled with a slighter race, which may already have spoken a Semitic language.

EARLY HISTORY

History begins with the invention of writing, which took place in southern Babylonia perhaps *c.* 3500 B.C., the script being an original picture-character which developed later into cuneiform. An early Sumerian ruler of the city of Erech, named Lugal-zaggisi, recorded his triumphal progress from the lower sea to the upper (*sc.* the Mediterranean), and this may be taken as the first mention of the land of Syria, though not of the name, in history. Immediately after this mention, Syria appeared as the provenance of a great migration of Semitic invaders into Babylonia by way of the Euphrates, the first of several such movements which may all have begun from northern Arabia, but are never traceable farther back than Syria. These first intruders have been given the name of Accadians, from their leader, Sargon of Agade or Akkad, with whom is connected a legend that he was born obscurely on the upper Euphrates, launched on the river in a reed basket and adopted in Babylonia, where he found favour with a king, supplanted him and became the greatest conqueror and the most famous name in Babylonian history. From his new city of Agade he marched back up the Euphrates to the "cedar mountain" (the Amanus) and far beyond, if the various sources may be trusted. It is probable that he and his equally famous grandson Naram-Sin (*c.* 2300 B.C.) ruled effectively over the whole of Syria, to which they doubtless introduced the Sumerian writing that they had learned and adapted to their Semitic language in Babylonia. But their successors were feebler men, and the dynasty of Agade was soon overthrown at its centre and superseded by the dynasty of Ur. Nothing certain is known about the authority (if any) that the kings of Ur exercised in Syria, so far away from their capital. The end of their dynasty, however, was brought about chiefly by the pressure of a new Semitic migration from Syria, this time of the Amorites (*i.e.*, the westerners), as they were called in Babylonia. Their invasion began before 2000 B.C. and by the 18th century B.C. they had covered both Syria and Mesopotamia with a multitude of small principalities and cities, mostly governed by rulers bearing some name characteristic of the Semitic dialect that the Amorites spoke. The period of Amorite ascendancy is vividly mirrored in the Mari letters, a great archive of royal correspondence found at the site of Mari, near the modern frontier with Iraq. They treat of the wars and general affairs of a multitude of states extending from northern Palestine to the upper Tigris. Among the principal figures in them are the celebrated lawgiver Hammurabi of Babylon (himself an Amorite) and a king of Aleppo, reckoned the most powerful of all, but of whom little is yet known. Part of his kingdom was the city of Alalakh, on the Orontes near what was later Antioch. This has yielded to excavation records both of the Amorite period and of a later age.

At about the same time (later in the 18th century B.C.) there began from Syria a movement of people in the opposite direction. This resulted in the Hyksos invasion of Egypt, which was subject to this foreign domination for about 150 years. The mixed multitude of the Hyksos certainly included Khurrians (Hurrians, Horites), who, not being Aryans themselves, were under the rule and influence of Aryans and learned from them the use of light chariots and horses in warfare, which they introduced both into Egypt and into Syria and Mesopotamia. These Khurrians established the kingdom of Mitanni, with its centre east of the Euphrates, and this was for long the dominant power in Syria, reaching its height in the 15th century B.C. But other nations were growing at the same time, and in the 14th century Syria was the arena in which at least four great competitors contended. The Khurrians were first in possession, and these maintained friendly relations with Egypt, which, after expelling the Hyksos, had established a vast sphere of influence (it was scarcely an "empire") in Palestine and Syria, under the kings of the 18th Dynasty. Third of the powers disputing Syria in the 14th century were the Hittites, who finally, under their greatest warrior Shuppiluliumash (*c.* 1350 B.C.) not only defeated the kingdom of Mitanni but established a firm dominion of their own in northern Syria with its principal centres at Aleppo and Carchemish. Fourth was the rising kingdom of Assyria which became a serious contender in the reign of Ashuruballit I. Indeed a fifth actor also appeared on the scene, the king of Babylon, but he was too distant and impotent to affect the issue. This was the period of the Amarna letters which vividly illustrate the decline of Egyptian influence in Syria, especially under Amenhotep IV (Ikhnaton), the distress or duplicity of local governors and the rivalry of the aforesaid powers, contending with Egypt and with one another in a strife that was not to leave any of them with a final victory. Egyptians and Hittites continued their struggle into the 13th century, but even the famous battle of Kadesh, which Rameses II so boastfully claimed to have won (*c.* 1285 B.C.) led to no more than a treaty maintaining the equal balance. Assyria had already swept away the remains of Mitanni but soon fell into decline, and the Hittites were not long afterward driven from their centre in Asia Minor into a chain of minor states on the northwestern fringes of Syria by the migration of "peoples of the sea," western invaders from the isles of the Aegean and from Europe. These immigrants then marched down the coast of the Levant and threatened Egypt, but were repulsed by Rameses III *c.* 1190 B.C.

While all the neighbouring states had failed to establish a lasting ascendancy in Syria, the actual possession was assumed gradually by a people without sufficient cohesion to appear as a power at all. Various documents of the 14th century mention the Akhlamu, who were forerunners of another vast movement of Semitic tribes called generically Aramaeans. By the end of the 13th century these had covered with their small and loose principalities the whole of central and northern Syria and a great part of Mesopotamia. The Assyrians, however, were able to guard their homeland from this penetration: henceforth much of the warfare of Assyrian kings

was to be a mixture of aggression and defense aimed at the Aramaean states of Syria and at the borderland "Hittite" principalities. At about the same time as the Aramaean invasion there was proceeding the exodus of Israelite tribes from Egypt; and as these, toward the end of the 11th century established a kingdom centred upon Jerusalem, so did the Aramaeans set up their principal kingdom at Damascus, and the wars between kings of Judah or of Israel and kings of Aram make up much of Old-Testament history. But the most formidable enemies of the Aramaeans and often of the Hebrews as their allies, were the great military kings of the Assyrians. After a premature Assyrian incursion under Tiglath-pileser I early in the 11th century B.C., the real contest was reserved for the 9th and the 8th, during which the Assyrian empire was established over the west, though destined to be short-lived. At the battle of Karkar in 853 B.C. Shalmaneser III of Assyria was opposed by Ben-Hadad of Damascus, Ahab of Israel and by 12 vassal monarchs. In 732 Damascus, the Syrian capital, was at length captured by Tiglath-pileser III. But campaigns against the Aramaeans and Anatolians of northern Syria (called "Hatti-land") had to be undertaken by the Assyrians until almost the end of the Assyrian empire. Culturally, the most important achievement of the Aramaeans was the bringing of the alphabet into general use for public and private business. It had been invented along the Syrian coast before the middle of the 2nd millennium B.C. This was perhaps the greatest single contribution ever made to human civilization.

Before the close of the 8th century B.C. began a massive southward movement of people, partly of Aryan stock, from the north and west. Pressure of these upon the Assyrian dominions and homeland became ever more severe, and they deeply affected Syria also; and in the 7th century came the invasion of the Cimmerians, followed by the Scythians. To these and to the Medes, another Aryan people, Assyria finally succumbed with the fall of Nineveh in 612 B.C. A remnant supported itself in Harran for a year, but that city was evacuated before the advance of the Medes, now reinforced by Nebuchadrezzar, king of Babylon. Nebuchadrezzar finally defeated the attempted rescue of Assyria by Necho, king of Egypt, and annihilated his army at Carchemish in 605 B.C. In 586 B.C. he captured Jerusalem and carried its people into exile. Thereafter Syria was, for half a century, under the rule of Nebuchadrezzar's successors on the throne of Babylon. But another and greater Aryan power, the Persians, then came to the fore. Under the leadership of Cyrus they subverted the rule of the Medes and assumed their empire, extending their conquests into Asia Minor, where Cyrus defeated and captured Croesus of Lydia. No longer able to brook a rival, he thus came to a final collision with Babylon, which he occupied in 539 B.C. He returned all the gods of Babylonia to their own places, and among the measures of restoration commanded by him and continued by his successors was a project of sending back the exiled Jewish community to Jerusalem and encouraging them to rebuild the temple. In Darius I's great organization of the Persian dominions more than 20 satrapies were constituted; Syria, with Palestine and Cyprus, was the fifth, bearing the name of "Across the River" (*i.e.*, the Euphrates), and its tribute was fixed at 350 talents of silver. Damascus and the Phoenician cities were still the chief centres of Syria under the Persians, and in Sidon was the core of the Phoenician revolt against Artaxerxes III, which ended with the destruction of that city in 351 B.C. But by this time the end of the Persian domination was at hand, and the Macedonians under Alexander the Great were about to bring the whole near east under Greek rule and influence.

Alexander invaded Asia Minor in 334 B.C., and his first decisive victory over the Persians was won on the fringe of Syria at Issus in 333; it was followed by the capture and enslavement of Tyre and Gaza. With the battle of Gaugamela and the destruction of Persepolis the downfall of Persia was completed.

See further AMORITES; HITTITES; PHOENICIA; DAMASCUS; SIDON; TYRE. (C. J. G.)

HELLENISTIC AND ROMAN SYRIA

After Alexander's death in 323 B.C. his marshals contended for control of the country, until after Ipsus (301) Seleucus I Nicator gained the northern part and Ptolemy I Lagus gained the southern (Coele Syria). This partition between the Seleucids and Ptolemies was maintained for 100 yr. despite intermittent Syrian wars during the 3rd century (*see* SELEUCID DYNASTY). Their administrative methods varied. In the south the Ptolemies respected the existing autonomous cities, imposed a bureaucratic system on the rest of the country and established no colonies, while the Seleucids divided their part into four satrapies which embraced the cities, and founded many cities and military colonies (including Antioch, Seleucia Pieria, Apamea and Laodicea [*qq.v.*]), drawing on settlers of European blood. Republics replaced kings in the Phoenician coastal cities of Tyre (274 B.C.), Sidon, Byblus and Aradus. Further political and cultural changes followed. Antiochus III the Great defeated Ptolemy at Panium (200 B.C. or perhaps as late as 198) and thus secured control of southern Syria, where he introduced the satrapal system. His subsequent defeat by the Romans at Magnesia (Dec. or Jan., 190–189), however, diminished both his territory in Asia Minor and his prestige, thereby fundamentally weakening the Seleucid empire, which ceased to be a Mediterranean power. Antiochus Epiphanes (175–163) stimulated the spread of Greek culture in Syria, with which went Greek political ideas, by a policy of urbanization: increased city organization and municipal autonomy involved greater decentralization of his kingdom; his attempted hellenization of the Jews is well-known. Under the later Seleucid kings, with rival claimants to the throne and constant civil war, Syria disintegrated. In the north the Seleucids controlled little more than the areas of Antioch and Damascus (Coele Syria): Commagene asserted its freedom, many of the cities claimed independence and numerous petty chiefs established little principalities. Southern Syria was partitioned by three tribal dynasties, the Ituraeans, the Jews and the Nabataeans. The country was seized later by Tigranes of Armenia (83) until his defeat by Pompey who ended years of anarchy by making Syria a Roman province (64–63).

Pompey in the main accepted the status quo, but he re-established a number of cities and reduced the kingdom of Judaea; 10 cities of the interior formed a league, the Decapolis. The Roman province of Syria therefore included the native kingdoms of Commagene, Ituraea, Judaea and Nabataea. A Parthian invasion was thrown back in 39 B.C., while Antony's extensive territorial gifts to Cleopatra (Ituraea, Damascus, Coele Syria, etc.) involved only temporary adjustments. Under the principate Syria, which stretched northeast to the upper Euphrates and, until A.D. 73, included eastern Cilicia, became one of the most important provinces. Its governor, a consular legate, generally commanded four legions until A.D. 70. Administrative changes followed; Rome gradually annexed the client kingdoms. Ituraea was incorporated (*i.e.*, its territories were assigned to neighbouring cities) partly in 34 B.C., partly *c.* A.D. 93. Judaea became a separate province in A.D. 6, governed by procurators (apart from Herod I Agrippa's shortlived control, A.D. 37–44), until the destruction of Jerusalem in 70. Then the governor was a praetorian legate in command of a legion; next, under Hadrian he was a consular with two legions, and the province was named Syria Palaestina. Commagene was annexed temporarily from A.D. 17 to 38 and permanently by Vespasian in 72. The caravan city of Palmyra came under Roman control, possibly during Tiberius' reign. Finally, Nabataea was made the province of Arabia in 105, governed by a praetorian legate with one legion. Syria itself was later divided by Septimius Severus into two provinces—Syria Coele in the north with two legions, and Syria Phoenice with one. By the beginning of the 5th century it was subdivided into at least five provinces. The frontiers of Syria were guarded by a fortified *limes* system, which was thoroughly reorganized by Diocletian and his successors (particularly against cavalry attacks) and endured until the Arab conquest; much knowledge of this system of "defense in depth" has been obtained with the aid of air-photography.

Syria's economic prosperity depended on its natural products (*e.g.*, wine, olives, vegetables, fruits and nuts), on its industries (*e.g.*, purple-dyeing, glassmaking at Sidon, linen and wool weaving and metalwork) and on its control and organization of trade passing by caravan from the east to the Mediterranean through such

centres as Palmyra, Damascus, Bostra and Petra. Further, Syrian merchants spread westward throughout the Latin provinces of the Roman empire. The population of the country must have been larger than that of today. Some cities were founded under the empire, but these were not numerous and Syria remained essentially rural. The urban upper and middle classes might be hellenized, but the lower classes still spoke Aramaic and other Semitic dialects in town as well as country. Roman influences were naturally weaker than Greek, though the army at first helped the spread of romanization, while Latin, not Greek, was the language of instruction in the famous law school at Berytus until *c.* A.D. 400. The splendour of Syrian culture is seen in the magnificence of the cities (Antioch, ranking among the greatest cities of the empire, was the residence of the governor and later of the *comes Orientis*), and in the luxury, gaiety and versatility of the Syrians. This splendour is also evident in their schools of rhetoric, law and medicine, in their art (the pre-eminence of Italian or Syrian elements in Syrian art is debated), in their literature and philosophy and in the variety of their religions, both pagan and Christian.

During the three centuries Syria was administered from Constantinople, its cultural and economic life remained active, as names like Libanius, Ammianus Marcellinus, Procopius and John Malalas bear witness. Government became more bureaucratic but it was efficient. From the middle of the 5th century, by which time paganism had died out and monasticism was spreading, religious controversy took a political turn with the growth of the Monophysites, in whom anti-imperial, nationalist and secessionist sentiment found a rallying point. The repercussions of this heresy seriously weakened Syria and the Eastern Roman empire. Nor were external dangers lacking. In the 4th century, during the campaigns of Constantius and Julian against Persia, Syria had again become a base of operations and at times endured Persian invasion; and the Huns had appeared before the end of that century. The Persian menace died down during the 5th century, but it blazed up again in the 6th, when Arabs also added to the danger. The Persian Chosroes I captured Antioch itself (540); and though he was outwitted by Belisarius and though Justinian's Persian wars ended in a "50 years' peace" (562), yet ten years later the Persians were back again. The invasion of Chosroes II, which began in 606, was later rolled back by the victories of Heraclius, but the peace of 628 brought no tranquillity to Syria. The Arabs advanced, Damascus and Antioch fell, and Heraclius was forced to evacuate Syria: Islam was master of the land which seven centuries earlier Pompey had formally added to the Roman empire.

See also PALESTINE; MACCABEES; NABATAEANS; ANTIOCH, SYRIAN ORTHODOX PATRIARCHATE OF; also ANTIOCH; BAALBEK; PALMYRA; SELEUCIA.

BIBLIOGRAPHY.—G. A. Harrer, *Studies in the History of the Roman Province of Syria* (1915); E. S. Bouchier, *Syria as a Roman Province* (1916); M. Rostovtzeff, *Social and Economic History of the Hellenistic World* (1941) and *Soc. and Econ. Hist. of the Roman Empire* (1926); A. H. M. Jones, *Cities of the Eastern Roman Provinces* (1937); A. Poidebard, *La Trace de Rome dans le désert de Syrie* (1934); R. Mouterde and A. Poidebard, *Le Limes de Chalcis* (1945); F. M. Heichelheim, *Roman Syria* (1938). (H. H. Sd.)

MEDIAEVAL SYRIA

In 633 the Mohammedan Arabs invaded Syria from the south. After their decisive victory on the Yarmuk river in 636 they occupied the whole country piecemeal and made it a base for further conquests north and west. When civil war broke out in 656, the Arabs of Syria stood firm behind their governor, Mu'awiya, of the Meccan house of the Omayyads, who, after warding off at Siffin an attack by the Iraqi Arabs, extended his authority over the neighbouring provinces and was recognized as caliph in 661 (*see* CALIPHATE).

During the Omayyad caliphate (661–750) Damascus was the capital of an Arab empire which, despite internal outbreaks, extended by 715 from Spain and Morocco to Turkistan and Sind and held the mastery of the Mediterranean. With the wealth drawn from the dependent provinces the caliphs 'Abd ul-Malik and Walid I endowed Jerusalem and Damascus with magnificent religious monuments, the Dome of the Rock and the Omayyad mosque. The first serious setback was the failure of the third and final Arab assault on Constantinople in 719. During the following decades unrest and weakness in the outer provinces involved the wide dispersion of Syrian troops on expeditions and garrison duty; this, combined with factional feuds among the Arabs, undermined the military power of the Omayyads, who were finally overthrown by the Khurasanian army of the new Abbasid caliphate in Iraq in 750.

Under the Abbasids Syria reverted to the status of a dependent province. It had little share in the economic and cultural development of the Abbasid empire and was increasingly abandoned to its sedentary Arab population and to the nomads on the desert frontiers. As the empire disintegrated, Syria was annexed, in whole or in part, by the Egyptian dynasties (Tulunids, Ikhshidis, Fatimites), except for the brief principality of the Arab Hamdanids of Aleppo founded by Saif ud-Daula (944–967), whose early successes against the Byzantine empire were followed by the Greek recovery of Antioch (969).

During the 11th century the fragmentation of Syria was momentarily arrested by its unification under the Seljuk Tutush (1078–95) and attachment to the empire of his brother, the sultan Malikshah (*see* SELJUKS). On his death the Seljuk kingdom fell to pieces, and Antioch, Jerusalem, the coastal regions and Kerak (Trans-Jordan) were occupied by the crusaders against feeble local opposition between 1098 and 1124 (*see* CRUSADES). The Mohammedan counterattack began with the annexation of Aleppo by the Turkish atabegs of Mosul (1125), with the recovery of Edessa (1144), and with the union of inner Syria under Nureddin (1146–74), who also annexed Egypt in 1169. After Nureddin's death his principality was rebuilt and enlarged by his Kurdish viceroy in Egypt, Saladin (*q.v.*) (1174–93), who recovered all of Palestine and most of the inland crusader strongholds after his victory at Hattin (1187) but lost the Palestinian coast again as a result of the third crusade. Concurrent developments of the 12th century were the rise to power of the sect of the Assassins (*q.v.*) and the rapid proselytism conducted by the Druzes (*q.v.*).

The crusades, paradoxically, did much to rehabilitate inner Syria, both by the economic development resulting from the establishment of Italian trading centres on the coast and by the impetus given to its cultural life by Nureddin and Saladin. Although it was again split up after Saladin's death between the members of his family, the Ayyubites, who established separate principalities in Aleppo, Hama, Homs, Damascus, Baalbek and Trans-Jordan and disturbed by intermittent hostilities between their princes or against the crusaders, Syria as a whole regained during the Ayyubite period (1193–1260) a level of prosperity such as it had not enjoyed since the 3rd and 4th centuries. A severe blow was dealt by the first Mongol invasion and sack of Aleppo in 1260, but the Mongols were thrown back by the Mameluke sultans of Egypt, who re-organized the Ayyubite principalities as Mameluke governorships and completed the expulsion of the crusaders by 1291. After the defeat of a fourth Mongol invasion in 1304 Syria remained relatively peaceful and prosperous as a centre of culture, learning and industry under Mameluke rule until the invasion of Timur (*q.v.*) in 1400. From this fresh blow its economy never really recovered, and it continued to decay slowly under the later Mamelukes until its annexation in 1516 by the Ottoman sultan Selim I, who divided it into three Turkish pashaliks. Its incorporation in the Ottoman empire brought some revival of economic prosperity to Aleppo and Damascus through trade and the pilgrimage to Mecca but did not arrest the decline of agriculture and general security under the pressure of nomad tribesmen. (H. A. R. G.)

THE 19TH AND 20TH CENTURIES

For nearly three hundred years, then, the history of Syria was that of a province of the Ottoman empire (*see* TURKEY); its internecine troubles of the 17th and 18th centuries are outlined in the article DRUSES. It was the French invasion of Palestine in Feb. 1799 that brought the country once more under the full light of international concern. Napoleon Bonaparte sought to secure Syria as a bulwark in defense of his newly conquered Egypt and also as a base from which to threaten both Constantinople and the British power in India. The campaign (*see* FRENCH REVOLUTIONARY WARS), though abortive, was one that Napoleonic propaganda could represent as glorious: the memory of it made a lasting and fatal impression on the mind of the French public.

The French were deeply incensed at the convention of London (1840), when the other interested European powers agreed to Turkey's recovering the lost territories. Syrian rebellion and disputes of Syrian factions,

one of which was encouraged by the French, and the massacres of 1860 culminated in French intervention, and to the constitution of Lebanon as an autonomous province in 1864.

The Turkish revolution of 1908 gave impetus to the nationalist movement in the Arabic-speaking provinces of the empire, especially in Syria. The political demands of the Arabs, whose representatives met in June 1913 in Paris, involved a decentralization of the Turkish empire with administrative and cultural autonomy for the Arab provinces. The Syrian nationalists co-operated with similar movements in Mesopotamia and Egypt. During World War I the nationalist leaders entered into close contact with Faisal (*see* FAISAL I). The Turkish administration persecuted them: several were executed and regarded by their followers as martyrs of the movement of national liberation.

The French Mandate.—Faisal had meanwhile formed an Arab army, which fought together with the British under Lord Allenby. While the British conquered Palestine, the Arabs moved through Transjordan northward and by the end of 1918 had reached Damascus, where, according to the promises given to the Arabs during the war, an independent Arab government was set up under Faisal. But France, already holding Beirut and the coastal zone, desired a mandate over the whole of Syria, in spite of the fact that on Nov. 7, 1918, the British and French governments had jointly declared their intention of establishing in Syria and Mesopotamia "national governments drawing their authority from the initiative and free choice of the native populations." A Syrian national congress, however, elected Faisal king of Syria on March 11, 1920, appointed Hashim el-Atasi (1875–) as prime minister, and adopted on July 3 a democratic constitution for the whole of Syria, including Palestine and Transjordan; and in the summer of 1919 a United States commission found that the population rejected the idea of a French mandate. Nevertheless the French mandate was decided upon at San Remo on April 25, 1920, and Gen. Henri Gouraud, the French high commissioner, sent an ultimatum to Faisal demanding recognition of it in July. The Syrians rejected the demand, but their army under Yusuf al-Azmi was defeated at Khan Maisalun, the French entered Damascus, and Faisal left the country. The terms of the mandate were approved by the League of Nations on July 24, 1922.

The French established on Sept. 1, 1920, a state of the Lebanon. Its nucleus was the formerly autonomous sanjak with its predominantly Christian population; but the French incorporated into it districts with a predominantly Moslem population who protested against separation from the rest of Syria. Syria itself was further divided into four "states": the state of Damascus, the state of Aleppo, the state of the Jabal Druse and the state of the Alawis with the capital at Latakia. The sanjak of Alexandretta was made autonomous. Damascus and Aleppo, however, had soon to be united again in a state of Syria. Beirut was chosen as the seat of the French high commissioner. Paper constitutions were given to the different states, but real power remained in French hands. Discontent found expression in violent demonstrations.

The struggle for independence assumed a new form with the success of a local revolt in the Jabal Druse under the leadership of Sultan el-Atrash which broke out in July 1925. On Aug. 24 the Druses approached Damascus itself, where the nationalists quickly made common cause with them. A revolutionary government was established in the Jabal Druse with the help of the nationalist leader Abd ur-Rahman Shahbandar. The struggle lasted for many months, especially concentrated around Damascus but spreading to all parts of Syria, and subsided only in the summer of 1927. During this war Damascus was invaded by the nationalists on Oct. 18, 1925, whereupon the French bombarded the city for two days by artillery and from the air. The French promised eventually to grant Syria independence.

In 1928 the French abolished martial law, reintroduced certain civil liberties, asked Taj ud-Din el-Hasani to form a cabinet and, on March 10, announced elections for a constituent assembly. The elections in April brought an overwhelming victory to the nationalists, and on June 9 the assembly was opened in Damascus. Hashim el-Atasi was elected president of the assembly, and by August the draft of the constitution was ready. The assembly was finally dissolved in May 1930; but the Syrian constitution, with a transitional clause for the duration of the mandate, was proclaimed on May 14, 1930. It provided for a chamber of deputies elected for four years and for the election of a president of the Syrian republic for five years; every male citizen over 20 years of age had a vote. In the next six years the constitutional life of Syria went through several periods of storm and stress, caused by disagreements with the high commissioner, which led to a temporary suspension of parliamentary life. But in 1936 the growing tension in the Mediterranean and the example set by Great Britain in Egypt forced the French to adopt a more liberal attitude. A delegation under Hashim el-Atasi was invited to Paris, where on Sept. 9 a treaty was signed providing for a perpetual friendship between France and Syria. France was to be represented by an ambassador in Damascus, Syria by a minister plenipotentiary in Paris.

This treaty fulfilled many Syrian demands, except for that of union with Lebanon, with which the French concluded a separate treaty on Nov. 13, 1936. The Syrian parliament ratified the Franco-Syrian treaty on Dec. 26, 1936, but the French postponed ratification. Under an agreement between France and Turkey in Jan. 1937, the sanjak of Alexandretta (Iskenderun), which had a largely Turkish population, was to enjoy autonomy under the suzerainty of Syria; on Sept. 5, 1938, the legislative assembly of the sanjak assumed for the territory the Turkish name of Hatay; and on June 23, 1939, under Syrian protest, the territory was incorporated into Turkey.

Faced with strong opposition in parliament the French government announced on Dec. 14, 1938, that they did not intend for the present to ask parliament to ratify the Syrian treaty. The Syrian government and president resigned, and on July 10, 1939, the French high commissioner, Gabriel Puaux, suspended the constitution of 1930 and set up a directorate under his own control.

Gen. Eugène Mittelhauser, commander in chief in Syria, and Puaux declared on June 28, 1940, that they would cease hostilities against Germany and Italy and put themselves under the Vichy government. The Arab leaders first believed that Great Britain had been decisively defeated and that its empire would be inherited in the middle east by the Germans and the Italians. They sounded the intentions of the German armistice commission; but the successful resistance of the British and the defeat of the Italians in Africa deterred them from committing themselves prematurely. Abd ur-Rahman Shahbandar had been assassinated, yet the idea of a unified Arabia with Syria as its nucleus gained in strength.

The unsettled conditions, the scarcity and hoarding of food and the fast-rising prices for all commodities brought the Arab unrest into the open in a wave of widespread and often violent strikes which paralyzed communications. On March 20, 1941, the Syrian National party under Shukri el-Kuwatli (1891–) issued a manifesto in which it demanded the independence of the country. The revolt grew so violent that, after French troops had been used several times to quell riots, the French were forced again to try to meet the native demands halfway. In April 1941 they appointed Khalid el-Azam head of the Syrian government.

In May the Vichy government allowed German aircraft to land and refuel in Syria en route for Iraq (where hostilities had broken out between the British and a revolutionary pro-German government) and also supplied arms and ammunition to Iraq. At the end of the month the British foreign secretary, Anthony Eden, announced Great Britain's sympathy with Syrian independence and Arab unification. On June 8, British forces and Free French troops under Lieut. Gen. Sir Henry Maitland Wilson started military operations for the liberation of the Levant. It took them about a month to overcome the resistance of the French troops under Gen. Henri Dentz, high commissioner and commander in chief, appointed by the Vichy government in Dec. 1940. Damascus was occupied on June 21, and hostilities ceased on July 14. The food situation remained difficult until the 1942 harvest, but subsequently Syria became once more an exporter of grain, at highly inflated prices.

British troops were in Syria only for the purpose of winning the war, and the British were willing that France should enjoy a predominant position in an independent Syria. Gen. Georges Catroux, representative of the Free French in Syria, declared on Sept. 16, 1941, that "Free France, acting in agreement with her ally Great Britain, has undertaken to terminate the mandate and grant Syria the status of an independent sovereign state and to guarantee the new state by treaty." Pro-French Taj ud-Din el-Hasani was appointed president of the republic, but the nationalist leaders refused to take part in the government. In 1943 the French were persuaded by the British to allow elections to be held in the Levant states, and in Syria government of the National party was overwhelmingly returned under Shukri el-Kuwatli who, on Aug. 17, was elected president of the republic. In Feb. 1945 Winston Churchill met el-Kuwatli in Cairo and persuaded him to open negotiations with the French for a treaty comparable with the Anglo-Iraqi treaty of 1930. But in May Jean Helleu, the French high commissioner, returned with his government's proposals and his return coincided with the disembarkation of two small contingents of French troops. The Syrian and Lebanese governments felt that they would be negotiating under duress; and in any case the French proposals (based on the abortive treaty of 1936) were no longer compatible with their insistence on total independence. Disorders in the principal Syrian towns culminated in the French bombardment of Damascus by aircraft and heavy guns on May 29–30, causing about 500 Syrian deaths. The British government ordered their commander in chief in the middle east to intervene, the French troops were compelled to return to barracks, and French influence came to an end.

Independence.—Meanwhile in March 1945 the Syrian and Lebanese governments were invited to the United Nations conference at San Francisco, signifying their recognition as sovereign states. They appealed to the Security council to order the immediate withdrawal of French and British troops from their territories in Feb. 1946, and the evacuation of Syria was completed by April 15. The National party government was confirmed in power by elections in 1947; but discontent at its inefficient and allegedly corrupt administration was aggravated by the poor performance of the Syrian army in the Palestinian war of 1948, which gave rise to further political scandals. In Dec. 1948 student demonstrations forced the replacement of the government by a nonparty administration led by Khalid el-Azam. On March 30, 1949, this was overthrown and el-Kuwatli deposed by a bloodless coup d'état led by the chief of staff, Husni ez-Zaim, a

Kurd who for a few weeks seemed as if he might prove the long-hoped-for "saviour with a sword." As the political leaders were reluctant to commit themselves by serving under him, however, he sought support from the still-hated French and from Egypt and, rapidly quarrelling with all the neighbouring Arab governments, degenerated into an arbitrary despot. On Aug. 14 he was arrested and summarily shot by another military junta under Col. Sami Hinnawi. A general election in November returned a government of the People's party, opposed to the National party and favouring a closer union with Iraq; but this project was forestalled by a third military coup on Dec. 19, which removed Sami Hinnawi from power. New difficulties arose after Syria's denunciation of the customs union with Lebanon in March 1950. Internal disorders were prolonged by the assassination of the commander in chief of the air force, Mohammed Nasser, in July 1950. A new constitution was adopted on Sept. 6, 1950, and Hashim el-Atasi became president.

A new crisis arose when Great Britain, France, Turkey and the United States submitted their scheme for the middle east defense organization to the countries of the Arab league in Oct. 1951. Hasan Hakim, who had become prime minister in August, wanted to accept the plan and resigned in November because of cabinet opposition to it. After three weeks of crisis, Maruf ed-Dawalibi formed a cabinet, but was overthrown nine hours later by a coup d'état (Nov. 29) engineered by the chief of the general staff, Col. Adib esh-Shishakli. President Atasi resigned on Dec. 1, and the higher military council under Shishakli appointed Col. Fawzi Silo chief of state and prime minister.

Shishakli, as deputy premier and chief of staff, controlled the government. On June 21, 1953, he announced the end of the provisional regime under which Syria had been governed since Dec. 1951. A new constitution was accepted by a referendum held on July 10. On the same day Shishakli (now a brigadier general) was elected president of the republic, having received 86.6% of the votes cast. Shishakli's regime did not last long. In Jan. 1954 it was faced with political disturbances allegedly fomented by the Iraqi army leaders. Sultan el-Atrash, the old Druse leader, revolted in the Hauran and Syrian army dissidents joined the movement. On Feb. 25 Shishakli resigned and left the country, while the 79-year-old Atasi was restored as president of the republic, and the Shishakli constitution replaced by that of 1950. On Aug. 18, 1955, Shukri el-Kuwatli, who in 1949 had taken refuge in Saudi Arabia, was elected president for the third time.

In 1955 Syria was faced with the clash between Egypt and Iraq over the Baghdad defense treaty. It accepted an Egyptian invitation to participate in a new Arab defense treaty, to replace the Arab league collective security treaty of June 17, 1950, which, as the Egyptian government had declared, no longer existed. On Oct. 20, 1955, Syria signed a military treaty with Egypt which established a joint command, while Egypt undertook to finance 60% of Syria's additional military expenditure. (G. E. K.; X.)

Syria United With Egypt.—At the beginning of Nov. 1956 Kuwatli paid a state visit to Moscow. The principal result of this visit was the delivery to Syria of large consignments of Soviet arms and the arrival in Damascus of several hundred Soviet instructors and technicians. The Syrian leaders declared that this did not imply acceptance of Soviet domination or creed. Syria rejected, however, the Eisenhower doctrine for the middle east and declared that Israel, not communism, was the threat to peace and security in the area. In Aug. 1957 three officials of the U.S. embassy were accused of plotting to overthrow the Syrian government and had to leave the country. A month later an international crisis developed when Syria and the U.S.S.R., alleging that Turkey was massing troops near the Syrian frontier, accused the U.S. of plotting a revolution in Syria with Turkey and Shishakli.

On Feb. 1, 1958, Kuwatli and Gamal Abd-el-Nasser, president of Egypt, formally proclaimed the union of Egypt and Syria in a United Arab Republic (U.A.R.). Four days later the parliaments of both countries voted the provisional constitution of the U.A.R. On Feb. 21 this constitution was approved by plebiscites in both countries and, on the same day, Nasser was elected president of the U.A.R.

The Union Ended.—On Sept. 28, 1961, following an army coup, Syria seceded from the United Arab Republic. An independent government was established with Mamun Kuzbari as prime minister, and this government was recognized shortly afterward by leading powers. Diplomatic missions were re-established in Damascus. The Syrian Arab Republic rejoined the United Nations on Oct. 13 without any objection being raised. (X.)

POPULATION AND GOVERNMENT

In 1960 Syria covered an area of 71,227 sq.mi. (181,337 sq.km.). Most of the population is of Arab stock, the main national minorities being the Kurds and the Armenians. The Kurds, estimated in 1954 at 140,000, are grouped mainly in the north, along the Turkish frontier. The 120,000 Armenians (1954 est.) are settled in the towns, about 60,000 of them in Aleppo alone. The Alawis, a Moslem sect, form the majority in the province of Latakia; they are probably the indigenous inhabitants of the country. The most densely populated areas are situated around Damascus, Aleppo and Latakia. The desert, which lies east of the Aleppo-Damascus line is sparsely populated.

The country is divided into 12 provinces (*muhafaza*) (1960); Damascus (346,977), Damascus (city; 653,472), Homs (399,714), Hama (325,076), Latakia (524,246), Aleppo (956,561), Hazekeh (351,661), Dayr az Zawr (221,290), As Suwayda (99,512), Dar'a (167,049), Idlib (331,745), and Rashid (177,964), total: 4,555,267.

TABLE I.—*Distribution of the Population by Religion*

Religion	1943	1953
Moslems		
Sunnis	1,971,053	2,578,810
Shi'is	12,742	14,887
Druses	87,184	113,318
Alawis	325,311	398,445
Ismailis	28,527	36,745
	2,424,817	3,142,205
Christians		
Orthodox	136,957	168,747
Roman Catholic*	103,834	123,881
Armenian (Gregorian)	101,747	110,594
Syriac (Jacobite)	40,135	51,363
Chaldaean (Nestorian)	9,176	11,167
Protestant (all denominations)	11,187	13,209
	403,036	478,961
Jews	29,770	31,647
Others (mainly Yezidis)	2,788	3,091
Total	2,860,411	3,655,904

*Roman Catholic subdivides into six rites. The figures for 1943 (1953) were: Greek Melkite 46,733 (55,880); Syriac 16,247 (19,738); Armenian 16,790 (19,492); Maronite 13,349 (16,580); Latin 5,996 (6,749); Chaldean 4,719 (5,492).

The population of major towns (1960) was: Damascus (cap.) 529,579; Aleppo 425,029; Homs 136,474; Hama 110,809; Latakia 67,799.

Constitution.—The Syrian constitution of 1930, as revised in Sept. 1950, specifies that the president of the republic shall be a Moslem, elected by a majority in the chamber of deputies for a period of five years. The president should sign legislation passed by the chamber within three days or else refer it to the state council constitutionally established. Members of the chamber are elected by direct suffrage, and in the election to the constituent assembly in 1949, women voted for the first time. The electoral decree of that year provided for one deputy for every 30,000 electors, the electorate being limited to men of 18 years of age and women of the same age holding at least an elementary school certificate.

The president nominates the prime minister; the prime minister chooses the members of his cabinet.

The party system is not highly developed; political groups tend to form round prominent individuals rather than to follow a set policy, and there are many independents.

Defense.—In 1955 the armed forces numbered about 25,000 and included only one motorized brigade. There was also a *gendarmerie* 4,000 strong.

Education.—There has been a steady increase in the provision of educational facilities. After 1919 the language of instruction in all government schools was Arabic, though during the mandatory period the French were criticized for giving too much encouragement to the study of French and to French schools, of which there were a large number. After 1945 the Syrian government exercised stricter control over the curriculums of foreign schools, and the number of them was reduced.

The Syrian University of Damascus, founded in 1923 (7,696 students and a teaching staff of 267 in 1957–58), had schools of medicine and law and later established schools of arts, science and engineering. There are two agricultural colleges and a college of engineering at Aleppo. The Arab academy at Damascus, financed by the government, is a recognized centre for the study of Arab history and literature.

ECONOMY

Agriculture.—Agriculture is the main industry and supports more than 50% of the population. But the proportion of cultivable land is low (about 40%) and less than half of this was under cultivation in 1947. In 1954 about 60% of the cultivable land was being worked, an increase due to extension of the acreage under cereals and, especially, under cotton. During the ten years following World War II the irrigated area was extended more rapidly than the rain-fed area by

TABLE II.—*Agricultural Production*

(In thousands of metric tons)

Crop	1934–38	1948–50	1952	1953	1954
Wheat	459	839	900	870	1,204
Barley	290	328	467	472	635
Rice, paddy	3	24*	5	16	25
Lentils	30	40	51	63	58
Broad beans	10	29	12	11	12
Chick peas	—	16	24	11	22
Olives	54	73†	33	49	36
Sugar beet	—	—	—	60	80
Grapes	170	170	167	244	257
Cotton, lint	6	36*	45	47	73

*1950. †1947–50.

the development of projects for flow irrigation and by the installation of pumps, largely by private enterprise. The production of principal crops is shown in Table II. Of the many varieties of fruits grown, the

apricot, introduced to Syria since the earliest times, forms one of the principal crops.

Sheep and goats are the most important in animal husbandry;

TABLE III.—*Livestock*
(In thousands of head)

	1947–50	1951	1952	1953	1954
Cattle	364	429	424	487	509
Sheep	2,954	—	3,085	3,560	3,746
Goats	1,185*	1,434	—	2,852	—
Horses	116	97	100	101	98
Mules	53	58	70	74	78
Asses	247	271	244	250	260

*1947.

other animals are employed for farm work or transport (horses, mules, donkeys, oxen, camels). Goats as well as cattle are raised for their milk. (*See* Table III.)

Industry.—As in agriculture so in industry most of the post-World War I expansion was due to private enterprise though the government gave much protection and some financial assistance. The textile industry is by far the most important; cotton spinning began in 1933 and by 1939 there were 10,000 spindles. In 1954 there were about 92,000 spindles with a capacity of 12,000 tons of cotton and rayon yarn. Small factories produce 40 tons of woolen yarn and 16 tons of silk yarn. Mechanical looms increased from 1,500 in 1939 to 5,500 in 1953; there were also 6,000 hand looms.

Food processing was the industry next in importance. It included sugar refining, production of vegetable oils, canning of fruit and vegetables, and the manufacture of confectionery (preserved fruits, chocolate, etc.). Other industries produce glass, cement, soap, matches and leather.

In 1951 the government took over the concessions of the electricity companies and increased the output from 47,000,000 kw.hr. in 1947 to 114,000,000 kw.hr. in 1953.

A concession to prospect for oil in all Syria north of Damascus is held by the Syrian Petroleum company, one of the Iraq Petroleum company groups. Pipelines of the Iraq Petroleum company and the Trans Arabian company (a subsidiary of Aramco) cross Syria to Tripoli and Saida respectively.

Foreign Trade.—In 1950 the customs union with the Lebanon was terminated by Syria mainly through the desire of the Syrian government to protect its industries and agriculture. Agreements regulating trade between the two countries were concluded in 1952 and 1953. (*See* Table IV.)

TABLE IV.—*Foreign Trade**
(In millions of Syrian pounds, valued at the free rate)

	1951	1952	1953	1954	1955
Imports (c.i.f.)	304.0	313.3	307.0	408.1	676.9
Exports (f.o.b.)	277.1	319.6	375.8	465.7	473.5

*Including gold.

There was, in fact, a trade deficit inasmuch as the import figures given in the table had to be adjusted upward by about 45%, to take account of the divergence between the official exchange rate and the free rate at which imports actually took place. The deficit seemed to have been covered mainly by receipts from the oil companies.

In 1954 the main suppliers of imports were, in order of importance, the United States, the United Kingdom, France, the German Federal Republic, Lebanon and Italy; Syria's main customers were Lebanon, France, the Netherlands, the United Kingdom and the German Federal Republic.

Communications.—The railways of Syria are part of the Syro-Lebanese system belonging to a French company, the Chemins de Fer Damas-Hama et Prolongements (D.H.P.), sections of the "Baghdad" railway and a portion of the Hejaz railway from Damascus to Deraa (76 mi.). Passenger traffic on the internal lines is extremely light except on that from Rayak (Lebanon) to Aleppo. The total length of roads in Syria in 1953 was about 6,260 mi., of which 1,515 mi. were asphalted. Main roads are: (1) from Alexandretta along the coast to Beirut and the frontier of Israel; (2) Aleppo to Homs and thence by alternative roads to Damascus; (3) Damascus to Kuneitra and thence to the Israel frontier; and (4) Damascus to Suwaida and Deraa. Desert roads connect Aleppo with Deir ez-Zor and northeastern districts and Damascus with Palmyra and Deir ez-Zor. The main desert road from Damascus to Baghdad (540 mi.) passes through the Iraqi post at Rutbah wells.

The airport at Mezze (El Mazza) 5 mi. W. of Damascus is used by the main air lines serving the middle east.

After 1956 the port facilities at Latakia were greatly expanded with the construction of a 4,000-ft. mole and a 2,000-ft. main jetty capable of dealing with six vessels at a time, and the reclamation from the sea of about 4 ac.

Between 1945 and 1954 the traffic of merchandise through the harbour of Latakia rose from 11,000 metric tons to 837,000 tons, because of the diversion of trade from Beirut.

Finance.—The monetary unit is the Syrian pound (L.S.), the value of which, by a decree of April 20, 1949, was fixed at 405.512 mg. of pure gold, corresponding to 7.967 U.S. cents. From that date the Syrian currency was no longer linked to the French franc. The free

TABLE V.—*Budget*
(In millions of Syrian pounds)

	1948	1949	1951*	1952†	1953‡	1954‡	1955‡
Revenue	107.00	103.53	209.70	270.1	194.9	208.0	261.0
Expenditure	76.37	77.62	234.80	267.7	208.0	208.0	261.0

*Jan. 1950 to June 1951. †July 1951 to Dec. 1952. ‡Estimates.

rate for the L.S. in Oct. 1955 was 4.634 U.S. cents. The currency in circulation in Dec. 1954 was about L.S. 360,000,000.

A budgetary surplus was recorded in every year of World War II and the reserve fund stood at L.S. 49,000,000 in 1945. A surplus of L.S. 8,000,000 was obtained in 1946. Table V shows revenue and expenditure from 1948 to 1955.

See also Index references under "Syria" in vol. 24.

BIBLIOGRAPHY.—G. Antonius, *The Arab Awakening* (1938, 1939); A. H. Hourani, *Syria and Lebanon, a Political Essay* (1946); E. E. Abouchid, *Thirty Years of Lebanon & Syria (1917–47)* (1948); George Kirk, *A Short History of the Middle East* (1948); Doreen Warriner, *Land and Poverty in the Middle East* (1948); United Nations, *Economic Development in the Middle East 1945–54* (1955). (K. Ç. B.)

SYRIAC LANGUAGE, belongs to the Eastern Aramaic group of the Semitic languages (*q.v.*). Since the name "Aramaic" came to mean "pagan," the Greek name "Syrian" or "Syriac" was adopted by the inhabitants of that region. It was the language of Edessa, present day Urfa. Edessa became a centre of Christianity at the end of the 2nd century and the intellectual metropolis of the Christian orient. As a result, the neighbouring Aramaic Christians adopted Syriac. Since the 5th century A.D., because of theological differences, Syriac-using Christians have been divided into Nestorians, or East Syrians under the Persian sphere of influence, and Jacobite (Monophysites) or West Syrians under the Byzantine sphere. The Nestorians were excluded from Edessa and settled in Nisibis in Persia.

These two groups became linguistically distinguished by certain differences of pronunciation, chiefly in the vowels. Thus, the Jacobites have *ō* instead of *ā*, *i* and *e* became confused into *i*, whereas *ō* and *ū* both became *ū*. The script likewise became differentiated.

Writing.—The earliest Syriac inscriptions (on stone) belong to the first half of the 1st century A.D. and to the second half of the 2nd century. The earliest extant documents not inscribed on stone date from A.D. 243 and come from Dura Europos. The early Syriac script was an offshoot of a cursive Aramaic script. The script reads horizontally from right to left. The Syriac alphabet consists of twenty-two letters, all of them having consonantal values. The order of the letters in the alphabet is the same as in Hebrew, but the names of some of them are slightly different; thus, *alaph* (Syriac), *aleph* (Hebrew); *gamal* (Syriac), *gimel* (Hebrew); *semkath* (Syriac), *samek* (Hebrew). The majority of the Syriac letters have different forms depending upon their position in a word, whether at the beginning, middle or end, and whether they stand alone or joined to others. As in other Semitic languages, the consonants ' (*alaph*), *w* and *y* were originally employed to express vowel sounds.

A highly developed vocalic system was introduced in the beginning of the 8th century. Three main vowel systems developed using a combination of small Greek letters and of dots placed above and below the letters. The most important variety of the Syriac scripts is "estrangelo" derived probably from *satar angelo*, the "evangelic" writing. After the schism in the church, the Syriac language and script split into two branches. The western *serta* or *serto* ("linear") developed later into two varieties, the "Jacobite" and the "Melchite." The eastern branch, called Nestorian, had greater importance in the history of writing. When Arabic became the speech of daily life, it was sometimes written in Syriac script; the term for it is *karshuni* or *garshuni*, but its meaning is uncertain.

Phonology.—The consonant phonemes of Syriac are: labials *p, b, m, w;* dentals and sibilants *t, d, ṭ, š, s, z, ṣ;* liquids *l, n, r;* palatals and velars *k, g, q;* laryngeals ', *h, ḥ, ʿ*. The stops *b, p, d,*

t, g, k become spirant *b, p, d, t, g, k* if they are preceded by a vowel. The vowels are *a, ā, e, ē, i, ī, o, ō, u, ū,* and a very short *ĕ*. The diphthongs are *aw, ay, āy* (Western Syrian *ōy*), *ĕw, īw*. No syllable can begin with a vowel. All the consonants except *h, ʿ* and *r* can be geminated (or long) in medial position. Gemination does not occur either initially or finally. There is no gemination in Western Syriac.

A few cases of partial and total assimilation as well as of metathesis occurring with the morpheme *t* of the reflexive-passive stem should be mentioned here. Metathesis occurs in the combination *tš, ts* becoming *št, st;* thus *'eštamlī* for *'etšamlī* "it was accomplished," *'estĕged* for *'etsĕged* "he was worshipped." Metathesis and partial assimilation occur in *tz, tṣ* becoming *zd, ṣt;* thus *'ezdĕqĕp* for *'etzeqep* "he was placed upright," *'eṣṭĕleb* for *'etṣĕleb* "he was crucified." Total assimilation occurs in the succession *t-'* becoming *tt* and *t-d* becoming *dd;* thus *'ettaqtal* for *'et 'aqtal* "he was killed," *'eṭṭašše* for *'ettašše* "hide me," *neddakrak* for *netdakrak* "he remembers you." Likewise *n* as first radical of the root is assimilated to any consonant with which it is in contact: *'appeq* for *'anpeq* "he brought out."

Morphology.—Syriac has a set of independent personal pronouns used for the expression of the subject, and a set of shortened forms suffixed to the nouns and to the verbs. With the verb they express the direct complement, occasionally the indirect complement; with the noun, they express the possession. The forms of the nouns and of the verbs with the suffixed pronouns very often change their syllabic structure.

The noun has three states, an absolute, a construct and an emphatic state. The absolute state is the pure stem and is used mainly for adjectival and participial predicates. The construct state has as a rule the same form as that of the absolute state and is used in the expression of possession (*mĕlĕk ʿālmā* "the king of the world"). The emphatic state has the suffixed element *-ā* of the article, but lost its determined value so that *malkā,* originally "the king" came to mean "king."

There are two genders in Syriac, masculine and feminine. The mark of the feminine is the suffixed *-t*. Its form is *-aṯ* with a noun in the construct state (*bīš-aṯ*), *-ēṯā* with a noun in the emphatic state (*bīš-ēṯā*), and *-ā* in the absolute state (*bīš-ā*). Several noun classes (animals, plants, parts of the body, utensils) are treated as feminine without having the feminine ending *-t*.

Syriac has two numbers, singular and plural. The forms of the plural are: for the masculine, absolute state *-īn* or *-ēn* (*bīš-īn*), construct state *-āy* (*bīš-āy*), emphatic state *-āy(y)ā,* mostly contracted into *-ē;* for the feminine the forms are, absolute *-ān,* construct *-āṯ,* emphatic *-āṯā*.

The Verb.—There are biradical, triradical and quadriradical verbs, the majority being triradicals. The basic stem of the regular triradical verb has the forms *qĕtal* (mostly transitive), and *qĕtel* (mostly intransitive). The verbs that have a first radical *w, y, n* or laryngeals as 1st, 2nd or 3rd radical have different syllabic structures in the various tenses of the verb. The biradicals are of the types *qām, baz*(*z*) and *rĕmā,* whereas the quadriradicals have the form *qartel*. The derived stems are *qattel* (with geminated 2nd radical) serving mainly for the expression of the intensive (occasionally causative) and *'aqtel* (with the prefix *'a-*) for the causative. All these stems can form a passive-reflexive mood with the prefixed morpheme *'eṯ,* the forms being *'eṯqtel* (from *qtal, qtel*), *'eṯqattal* (from *qattal*) and *'ettaqtal* (from *'aqtel*).

There is a considerable number of Syriac verbs having the prefix *ša-* (*šaqtel*) for the expression of the causative, and *qatqat* (that is, a repeated biradical root) for the expression of the frequentative.

The past is expressed by the "perfect" which is formed by suffixed morphemes (thus, **kĕtab* "he wrote," **kĕtabū* "they wrote"). Habitual or continuous action in the past is expressed by the active participle followed by the perfect of *hĕwā* "he was" (thus *dāḥel hĕwā* "he was fearing"). The future is expressed by the "imperfect" formed by prefixes or suffixes or both (thus, **nektūb* "he will write," *nektĕbūn* "they will write"). The present is expressed by the active participle followed by the personal pronoun, mostly in contracted form (thus **kāteb 'at* or **kātbat* "you write").

The expression for the 1st person is common for the masculine and feminine, whereas for the 2nd and 3rd persons there are special forms for the masculine and feminine. Note that the prefix of the imperfect 3rd person is *n* against *y* of the other Aramaic dialects.

Like all the other Semitic languages Syriac has no verb "to have." The idea of possession is expressed by the verb *'īṯ* "there is, there are" followed by the indirect complement (thus *kĕmā laḥmīn 'iṯ lĕkūn* "how many loaves of bread do you have?" literally "how-many breads there-are to-you?").

The Syriac vocabulary is basically Semitic. Among the loanwords are some Hebrew words in the domain of religion and some Persian words from translations of Persian, but the great majority of loanwords are Greek.

Bibliography.—A. Ungnad, *Syrische Grammatik und Übungsbuch,* 2nd ed. (1932); T. H. Robinson, *Syriac Grammar* (1915); T. Nöldeke, *Kurzgefasste Syrische Grammatik,* trans. by James A. Crichton, *Compendious Syriac Grammar* (1904); L. Costaz, *Grammaire syriaque* (1955); C. Brockelmann, *Syrische Grammatik,* 7th ed. (1955).

(W. Lu.)

SYRIAC LITERATURE. The use of Syriac by Christians as a literary medium had its original centre in Edessa (Syr. Ūrhāi, modern Urfa), where, in all probability, the chief Syriac versions of the Bible were made. The use of the same dialect appears in the earliest Christian literature connected with such Mesopotamian cities as Nisībis, Amid, Mardīn, Taghrīth and Seleucia-Ctesiphon, as well as west of the Euphrates at such centres as Mabbōgh (Hierapolis) and Aleppo, northward at Malaṭiah and Maiperkat and in the districts of Lake Van and Lake Urmia, and to the east and southeast of the Tigris in many places which from the 5th century onward were centres of Nestorian Christianity within the Sassanian empire.

In Palestine and western Syria, the home of pre-Christian Aramaic dialects, the vernacular Semitic speech had under Roman dominion been replaced by Greek for official and literary purposes. Apparently this state of things lasted till after the Mohammedan conquest, for Barhebraeus tells us that it was the caliph Walid I. (A.D. 705–715) who, out of hatred to Christianity, replaced Greek by Arabic as the language of official documents at Damascus.

Syriac literature continued in life from the 3rd to the 14th century A.D., but after the Arab conquest it became an increasingly artificial product, for Arabic gradually killed its vernacular use.

Religious Literature.—In the literature as it survives many different branches of writing are represented—homilies in prose and verse, hymns, exposition and commentary, liturgy, apocryphal legends, historical romance, hagiography and martyrology, monastic history and biography, general history, dogmatics, philosophy and science, ecclesiastical law, etc. But the whole is dominated by the theological and ecclesiastical interest. When we put aside one or two exceptionally fine pieces, like the hymn of the soul in the apocryphal Acts of Thomas, the highest degree of excellence in style is perhaps attained in straightforward historical narrative—such as the account of the Perso-Roman War at the beginning of the 6th century by the author who passes under the name of Joshua the Stylite, or by romancers like him who wrote the romance of Julian; by biographers like some of those who have written lives of saints, martyrs and eminent divines; and by some early writers of homilies such as Philoxenus (in prose) and Isaac of Antioch (in verse).

For the general history of culture the work of Syriac writers as *translators* is, perhaps, as important as any of their original contributions to literature. Beginning with the earliest versions of the Bible, which seem to date from the 2nd century A.D., the series comprises a great mass of translations from Greek originals—theological, philosophical, legendary, historical and scientific. In a fair number of cases the Syriac version has preserved to us the substance of a lost original text. Often, moreover, the Syriac translation became in turn the parent of a later Arabic version. This was notably the case with some of the Aristotelian writings,

so that in this field, as in some others, the Syriac writers handed on the torch of Greek thought to the Arabs, by whom it was in turn transmitted to mediaeval Europe. The early Syriac translations are in many cases so literal as to do violence to the idiom of their own language; but this makes them all the more valuable when we have to depend on them for reconstructing the original texts. The later translators use greater freedom. It was not from Greek only that translations were made into Syriac. Of translations from Pahlavī we have such examples as the version of pseudo-Callisthenes' *History of Alexander,* made in the 7th century from a Pahlavī version of the Greek original—that of *Kalilah and Dimnah* executed in the 6th century by the perioudeutēs Bōdh—and that of *Sindbad,* which dates from the 8th century; and in the late period of Syriac literature, books were translated from Arabic into Syriac as well as vice versa.

The Cradle of Syriac Literature.—All our historical sources support the view taken above that Edessa, the capital of the kingdom which the Greeks and Romans called Osrhoene, was the earliest seat of Christianity in Mesopotamia and the cradle of Syriac literature. But as to the date and circumstances of its evangelization we have little reliable information. The well-known legend of the correspondence of Abgar Ukkāmā, king of Edessa, with Christ and the mission of Addai to Edessa immediately after the Ascension was accepted as true by the historian Eusebius (d. 340) on the faith of a Syriac document preserved in the official archives of the city. An amplified form of the same story is furnished by the *Doctrine of Addai,* an original Syriac work which survives complete in a St. Petersburg (Leningrad) ms. of the 6th century, and is also represented by fragments in other mss. of the 5th and 6th centuries. This work was probably written at Edessa about the end of the 4th century. But whether in its longer or its shorter form, the whole narrative must be pronounced unhistorical. In all probability the first king of Osrhoene to adopt Christianity was Abgar IX., son of Ma'nū, who reigned from A.D. 179 to 214 or 216, and the legend has confounded him with an earlier Abgar, also son of Ma'nū, who reigned first from B.C. 4 to A.D. 7 and again from A.D. 13 to 50. (*See* Lipsius, *Die edessenische Abgar-Saga,* 1880.) A contemporary of Abgar IX. at Edessa was the famous Bardaiṣān, himself a convert from heathenism, who was of noble birth and a *habitué* of the Edessene court. It was no doubt partly under his influence—also possibly in part through impressions received by Abgar during his visit to Rome about A.D. 202—that the king's conversion took place. But Christianity must have reached Edessa some 30 to 50 years earlier. Our oldest native historical document in Syriac—the account of a severe flood which visited Edessa in Nov. A.D. 201—mentions "the temple of the church of the Christians" as overthrown by the flood.[1] The form of this notice shows, as von Gutschmid and others have remarked, that Christianity was not yet the religion of the State; but it must for some time have had a home in Edessa. By a skilful piecing together of the data furnished by the oldest Syriac versions of the Bible—such as the derivation of the Old Testament version from the Jews, and the almost exclusive use of Tatian's Diatessaron as the gospel of the Syriac Church down to the beginning of the 5th century—F. C. Burkitt has shown it to be probable that the preaching of Christianity at Edessa reaches back to the middle of the 2nd century or even to about the year 135 (*Early Eastern Christianity,* Lecture II.).

The Syriac versions of the Bible are treated elsewhere (*see* BIBLE) and may here be dismissed with a brief summary of facts and opinions. The received Syriac Bible or Vulgate (called the Peshitta or "simple" version from the 9th century onwards) contains all the canonical books of the Old Testament. In the New Testament, 2 Peter, 2 and 3 John, Jude and the Apocalypse were originally left out, but Syriac versions were made at a later time. The Peshitta version of the Old Testament must have been originally made mainly by Jews, of whom we know there were colonies in Mesopotamia in the 2nd century. The translation was executed entirely from the Hebrew, but underwent later revision which brought it more into conformity with the LXX.—this to a greater degree in some books than in others. The Peshitta New Testament—according to the convincing theory which at present holds the field—is not the oldest form of the Syriac version, at least as regards the Gospels. From the beginning of the 3rd to the beginning of the 5th century Tatian's Harmony or Diatessaron —whether originally compiled in Syriac, or compiled in Greek and translated into Syriac—was the current form of gospel in the Syriac Church. The text of the Gospels underlying it "represents the Greek text as read in Rome about A.D. 170." Slightly later was made the Old Syriac version of the separate Gospels, which survives in two mss.—the Curetonian and the Sinaitic—in two differing forms: but this never obtained much currency. Its text "represents, where it differs from the Diatessaron, the Greek text as read in Antioch about A.D. 200." Then at the beginning of the 5th century, by the efforts of the masterful Rabbūlā, who was bishop of Edessa from 411–412 to 435, a new version or recension of the Gospels was made and incorporated in the Peshitta or Vulgate, the use of the Diatessaron being henceforth proscribed. Rabbūlā's text of the Gospels "represents the Greek text as read in Antioch about A.D. 400." The history of the Peshitta rendering of the Acts and Epistles is less clear.

Apocryphal Books.—Of the large number of Apocryphal books existing in Syriac the majority have been translated from Greek, one or two (such as *Bar Sira* or *Ecclesiasticus*) from Hebrew, while some (like the *Doctrine of Addai* above referred to) are original Syriac documents. Special mention may be made here of the tale of Aḥīḳār—the wise and virtuous secretary of Sennacherib, king of Assyria—and of his wicked nephew Nādhān. This is the Syriac version of a narrative which has had an extraordinary vogue in the world's literature. It is now known to have existed in Aramaic as far back as the 5th century B.C., appearing in Jewish papyri which were lately discovered by the German mission to Elephantine. It appears to be traceable in its Greek dress in writings of the philosopher Democritus and the dramatist Menander; it was certainly known to the author of Tobit and perhaps to the author of Daniel; some would trace its influence in the New Testament, in the parable of the wicked servant and elsewhere; it was known to Mohammed and is referred to in the Koran; it has been included among the tales in the *Arabian Nights;* and it survives in a good many versions ancient and modern. The old Syriac version, which is to be found in a number of mss., was probably made from an early Aramaic version, if not from the original itself (which must surely have been Semitic). The Syriac has in turn become the parent of the Arabic, Armenian and Ethiopic—possibly also of the Greek and Slavonic-versions. (*See* F. Nau, *Histoire et sagesse d'Ahikar l'Assyrien,* 1909.)

Another deeply interesting Syriac Apocryphon is the *Acts of Judas Thomas* (*i.e.,* Judas the Twin), which is included in the collection of *Apocryphal Acts of the Apostles.* The *Acts of Thomas* is now generally recognized to be an original Syriac work (or "novel," as Burkitt calls it), although a Greek version also exists. It seems to have arisen in Gnostic circles, and its tendency is wholly in favour of asceticism and celibacy. Among its peculiarities is the fact that Judas Thomas is regarded as the twin brother of Christ. The author has incorporated in it the finest poem to be found in all Syriac literature, the famous Hymn of the Soul.

Lives of the Saints.—Lives of saints and martyrs form a large group among Syriac books. Among such documents connected with the early history of Edessa we have, besides the *Doctrine of Addai,* certain martyrdoms, those of Sharbēl and Barsamyā assigned to the reign of Trajan, and those of Guryā and Shāmōnā and of the Deacon Habbībh under Diocletian and Lincinius. All these documents, like *Addai,* belong probably to the 2nd half of the 4th century, and are unreliable in detail for the historian though they may throw some light on the conditions of life at Edessa under Roman government. There are also accounts of martyrdoms at Samosāta (Assemani, *Acta Mart.* ii. 123–147), including that of St. Azazail published by Macler (Paris, 1902). But the great bulk of the Syriac martyrdoms have their scene farther east, within the Persian dominions.

The life and writings of Bardaiṣān, "the last of the gnostics," and in some sense the father of Syriac literature and especially of Syriac poetry, have been treated in a separate article. *The Book of the Laws of the Countries,* which embodies his teaching

was re-edited in 1907 by F. Nau (this also in the 2nd volume of Graffin's *Patrologia*). Mention may here be made of the valuable edition and translation by the late C. W. Mitchell of *S. Ephraim's prose refutations of Mani, Marcion and Bardaisan. Transcribed from the palimpsest B. M. Add. 14628* (1912–21).

An early Syriac document, probably of the 2nd or 3rd century, is the *Letter of Mārā son of Serapion*, which was edited by Cureton in his *Spicilegium Syriacum*. It is almost the only exception to the rule that all surviving Syriac literature is Christian.

By the beginning of the 4th century much progress had been made with the organization of the Christian church not only within the Roman district of Mesopotamia, but also to the east and south-east within the Sassanian empire, round such centres as Seleucia-Ctesiphon on the Tigris (near Baghdad), Karkā dē-Bēth Sēlōkh (modern Kerkuk) and Bēth Lāpāt or Gundēshābhōr (in the modern province of Luristan). The adoption of Christianity by Constantine as the official religion of the Roman empire had an unfortunate effect on the position of the Christians in Persia. They were naturally suspected of sympathizing with the Roman enemies rather than with their own Persian rulers. Accordingly when Sapor II. (310–379) declared war on Rome about 337, there ensued almost immediately a somewhat violent persecution of the Persian Christians, which continued in varying degrees for about 40 years. One result of this and later persecutions of the same kind has been to enrich Syriac literature with a long series of *Acts of Persian Martyrs*, which, although in their existing form intermixed with much legendary matter, nevertheless throw valuable light on the history and geography of western Persia under Sassanian rule.

4th and 5th Centuries.—The two most important 4th-century writers—Aphraates and Ephraim—are dealt with in separate articles. The importance of the former lies in the simple cast of his religious thought, his independence of theological formulas, his constant adherence to the letter of Scripture, his quaint exegesis, and the light he throws on the circumstances of his time, especially (1) the feeling between Jews and Christians, and (2) the position and sympathies of the Christian subjects of Sapor II. The position and character of Ephraim are very different. He is the typical exponent in Syriac of unbending Catholic orthodoxy.

Before leaving the 4th century we may mention two other writers who probably both lived on into the 5th—Balai and Cyrillōnā. The former was the author of a good many poems; the longest—which is however by some attributed to Ephraim—is the work in 12 books on the history of Joseph, of which a complete edition was published by Bedjan in 1901. Other poems of his were edited by Overbeck in *S. Ephraemi Syri*, etc., *opera selecta*, pp. 251–336; and these have since been supplemented by Zetterstéen's edition of a large number of his religious poems or metrical prayers (*Beiträge zur Kenntniss der religiösen Dichtung Balais*, Leipzig, 1902). His favourite metre was the pentasyllabic. Cyrillōnā composed a poem on the invasion of the Huns in 395.

The 5th century was a time of storm and conflict in the churches of Mesopotamia and Persia, as in other parts of the Christian world. The teaching of Apollinarius that in Christ the Divine Word took the place of the human rational soul, thus seeming to do away with his possession of a true humanity, had led to a reaction by Paul of Samosāta, Diodore of Tarsus, Theodore of Mopsuestia, and Nestorius of Constantinople. Though with some points of difference, they agreed in emphasizing the permanence of the two separate natures in Christ, united but not mingled or confused, and laid stress on the reality of our Lord's human experience. One question on which great contention arose was as to the propriety of applying to the Divine nature attributes which belonged to the human nature—*e.g.*, birth from a human mother—and vice versa. Hence the great dispute about the application to the Virgin Mary of the epithet *θεοτόκος*. It seems to have been the objection of Nestorius to the use of this expression which mainly led to his condemnation and deposition at the Council of Ephesus (431) under the influence of Cyril, although as patriarch of Constantinople (428–431) he had distinguished himself by his zeal for Nicene orthodoxy. New light on the theological position of Nestorius is to be obtained from the long-lost *Book of Heraclides*, a work of his own which has turned up in a Syriac version and was published by Bedjan in 1910.

At Edessa the result of the conflict between the Nestorians and their opponents was long doubtful. When Rabbūlā, the fierce anti-Nestorian and friend of Cyril, died in 435, he was succeeded in the bishopric by Ibas, who as head of the famous "Persian school" in the city had done much to inculcate on his pupils the doctrines of Theodore of Mopsuestia. But the feeling against the Nestorian party grew in strength, till on the death of Ibas in 457 the leading Nestorian teachers were driven out of Edessa. The Persian school continued to exist for another 32 years, but was finally closed and destroyed by order of the emperor Zeno in 489. The Nestorian teachers then started a great school at Nisībis (which had been under Persian rule since Jovian's humiliating treaty of 363). By the energetic efforts of Barṣaumā, bishop of that city, practically the whole church of Persia was won over to the Nestorian creed.

Great Churchmen.—At the beginning of the 5th century one of the most able and influential men in the Syriac-speaking Church was Mārūthā, bishop of Maiperḳaṭ or Martyropolis. Without entering on the details of his ecclesiastical activity, we may note that he was twice associated with embassies from the Roman emperor to Yazdegerd I. (399–420); that along with Isaac, patriarch of Seleucia (390–410), he obtained from the Persian monarch a concordat which secured a period of religious toleration; and that he arranged for and presided at the Council of Seleucia in 410, which adopted the full Nicene creed and organized the hierarchy of the Persian Church. As a writer he is chiefly known as the reputed author of a collection of martyrologies which cover the reigns of Sapor II., Yazdegerd I. and Bahram V. By his history of the Council of Nicaea he made a great contribution to the education of the Persian Church in the development of Christian doctrine.

The next bishop of Edessa, Ibas, who succeeded in 435 at the death of Rabbūlā, proved himself a follower of the Nestorian doctrine. (*See* above.) As a teacher in the Persian school of Edessa he had translated, probably with the help of his pupils, certain works of "the Interpreter," *i.e.*, Theodore of Mopsuestia. Among these may have been the commentary on St. John of which the complete Syriac version was published by Chabot in 1897. He may possibly have translated a work of Aristotle. To the Nestorian movement in Persia he rendered useful service by his letter to Mārī of Bēth Hardashēr, in which he maintained the tenets of Diodore and Theodore, while allowing that Nestorius had erred. On the ground of his writings he was condemned and deposed by the "robber synod" of Ephesus (449), but was restored by the Council of Chalcedon (451), after he had anathematized Nestorius. His death in 457 was followed by a strong anti-Nestorian reaction at Edessa, which led to the expulsion of many of the leading teachers.

On Isaac of Antioch, "one of the stars of Syriac literature," *see* the special article. In spite of his over-diffuseness, he is one of the most readable of Syriac authors.

A Nestorian contemporary of Isaac, Dādhīshō', who was catholicus of Seleucia from 421 to 456, composed commentaries on Daniel, Kings and Ecclesiasticus. His chief importance in the history of the Persian Church lies in his having induced a synod of bishops to declare that church independent of the see of Antioch and of the "Western Fathers." (*See* J. Labourt, *Le Christianisme dans l'empire perse*, 1904, p. 122 *seq.*)

The most powerful missionary of Nestorianism during the 2nd half of the 5th century was Barṣaumā of Nisībis, whom his opponents called "the swimmer among the reeds," *i.e.*, the wild boar. Born probably between 415 and 420 he imbibed Nestorian doctrine from Ibas at the Persian school of Edessa, but was driven out in 457 on the death of his master, and went to be bishop of Nisībis. In a succession of missionary journeys he succeeded, partly by persuasion and partly (if his enemies are to be believed) by violence, in attaching to Nestorianism nearly all the Christian communities of Persia, with the exception of Taghrīth, which was always strongly Monophysite. He had many quarrels with

his ecclesiastical superior the catholicus of Seleucia, but finally made peace with Acacius soon after the accession of the latter in 484. Among other severities towards the Monophysites, he persuaded the Persian king Pērōz (457–484) to banish many of them into the Roman dominions. One of his great aims was to secure for the Nestorian clergy freedom to marry, and this was finally sanctioned by a council at Seleucia in 486 (Labourt, *op. cit.*, chap. vi.). Barṣāumā must have been bishop of Nisībis for nearly 40 years, but was dead by 496. His writings seem to have been chiefly liturgical; he gave the first set of statutes to the school of Nisībis, which was founded during his bishopric.

His fellow-worker Narsai, whom the Jacobites called "the leper," but the Nestorians "the harp of the Holy Spirit," apparently accompanied Barṣāumā from Edessa to Nisībis, where according to Barhebraeus he lived for 50 years. Barṣāumā appointed him head of the new school, where he taught rigidly Nestorian doctrine. He was a copious writer, especially in verse. Many of his poems have now been published. (*See* Mingana, *Narsai, homiliae et carmina*, 2 vols. Mosul, 1905.) His theological position is clearly defined in a homily on the three doctors —Diodore, Theodore and Nestorius—published by the Abbé Martin in the *Journal asiatique* for July 1900.

To about the same period belongs *The book of Hierotheus* by the "pantheist" Stephen bar Sūdh-aite, which has recently been edited and translated by F. C. Marsh (1927).

Early Monophysites.—Among the early Monophysites were two of the best of Syriac writers—Jacob of Sērūgh and Philoxenus of Mabbōgh, who have been treated in special articles. The one wrote mainly in verse, the other in prose. *See* also JOSHUA THE STYLITE.

Another early Monophysite was Simeon of Bēth Arshām, who by a series of journeys and disputations within the Persian empire did all he could to prevent the triumph of Nestorianism among the Persian Christians. He had considerable success at the time, but the ground he had won was soon reconquered by his opponents, except at Taghrīth and the surrounding district. It was after a successful disputation in presence of the Nestorian catholicus Bābhai (497–502/3) that Simeon was made bishop of Bēth Arshām, a town near Seleucia. He made several journeys to Constantinople, where he enjoyed the favour of the empress Theodora. It was there he died, probably about 532–533. His biography was written by John of Asia in the collection of lives of eastern saints which has been edited by Land (*Anecd. syr.* vol. ii.). His literary productions consist only of a liturgy and two exceedingly interesting letters. The one has for its subject Barṣāumā and the other Nestorian leaders in Persia, and gives a highly malicious account of their proceedings. The other, which has been often edited, is an account of a severe persecution which the Himyarite Christians of Najrān in south-west Arabia underwent in 523, at the hands of the king of Yemen. (See *The book of the Himyarites*, a Syriac work edited and translated by A. Moberg, Lund, 1924.) As Simeon had repeatedly visited al-Ḥīrah and was in touch with the Arab kingdom which centred there, his letter is a document of first-rate historical importance.

Mention should be made of two other early Monophysite leaders who suffered persecution at the hands of the emperor Justin I. (518–527). The one, John of Tellā, author of 538 canons, answers to questions by the priest Sergius, a creed and an exposition of the Trisagion. His life was written by his disciple Elias, and also by John of Asia. The other, John bar Aphtōnyā, was the founder of the famous monastery of Kenneshrē, opposite Jerābīs on the Euphrates, and wrote a commentary of the Song of Songs, a number of hymns and a biography of Severus, the Monophysite patriarch of Antioch (512–519).

The life of the great missionary bishop Jacob Burdĕ'ānā or Baradaeus, from whom the Monophysite Church took its name of Jacobite, belongs rather to ecclesiastical than to literary history. (*See* H. G. Kleyn, *Jacobus Baradaeus, de Stichter der Syrische monophysietische Kerk*, Leyden 1882.)

Translators from the Greek.—In Sergius of Rās'ain we find one of the best Greek scholars and ablest translators whom Syria has produced. Of his life little is known, and that little not wholly creditable. He wavered curiously in his ecclesiastical views, and ended by helping the persecutors of the Monophysite Church, to which he himself had belonged. He seems to have lived as a priest and physician at Rās'ain in Mesopotamia most of his life. About 535 he travelled on various ecclesiastical missions, and finally made a journey to Rome and thence to Constantinople (in this latter accompanied by the pope Agapetus). The result was to bring about the deposition and banishment of the Monophysites from the latter city. Sergius died almost immediately afterwards, in 536. Among the works which he translated into Syriac and of which his versions survive are treatises of Aristotle, Porphyry and Galen, the *Ars grammatica* of Dionysius Thrax, the works of Dionysius the Areopagite, and possibly two or three treatises of Plutarch. His own original works are less important, but include a "treatise on logic, addressed to Theodore (of Merv), which is unfortunately imperfect, a tract on negation and affirmation; a treatise, likewise addressed to Theodore, *On the Causes of the Universe, according to the Views of Aristotle, showing how it is a Circle;* a tract *On Genus, Species and Individuality;* and a third tract addressed to Theodore, *On the Action and Influence of the Moon, explanatory and illustrative of Galen's* Περὶ κρισίμων ἡμερῶν, bk. iii., with a short appendix 'On the Motion of the Sun' " (Wright). According to the historical compilation which passes under the name of Zacharias Rhetor, he also wrote a treatise on the faith. Some of his translations were revised at a later time by Honain ibn Iṣḥāḳ (d. 873).

Another translator from Greek was Paul, Monophysite bishop of Callinīcus or ar-Raḳḳah, who, being expelled from his diocese in 519, retired to Edessa and there occupied himself in translating into Syriac the works of Severus, the Monophysite champion who was patriarch of Antioch from 512 to 519. This version appears to be quite distinct from that used by the compiler of the chronicle of Zacharias, and also from the version of "the 6th book of the select letters of Severus" which was made by Athanasius "presbyter of Nisībis" in 669, edited by E. W. Brooks (1902–04).

That important legal work, *The Laws of the Emperors Constantine, Theodosius and Leo*, which was composed in Greek about 475, and "which lies at the root of all subsequent Christian oriental legislation in ecclesiastical, judicial and private matters" (Wright), must have been repeatedly translated into Syriac. The oldest form is contained in a British Museum ms. which dates from the earlier part of the 6th century, and this was edited by Land (*Anecd. syr.* i. 30–64). A later (probably Nestorian) recension is contained in a Paris ms., which was used along with the other by Bruns and Sachau in their exhaustive edition (*Syrisch-römisches Rechtsbuch*, Leipzig, 1880). In *Notulae syriacae* (privately printed 1887) Wright edited the surviving fragment of a 3rd recension which is preserved in a 13th-century ms. at Cambridge. Finally Sachau has published three new redactions of the treatise from a ms. found at Rome in 1894 (*Syrische Rechtsbücher*, vol. i., Leipzig, 1907).

Anonymous Works.—We may here take note of three important anonymous works, of which the first probably and the other two certainly belong to the 6th century.

The Mĕ'arrath gazzē or *Cave of Treasures*, translated and edited by C. Bezold (Leipzig, 1883–88), is akin (as Duval remarks) to the *Book of Jubilees*. It is an imaginary history of the patriarchs and their descendants.

The tripartite narrative which is known as the *Romance of Julian* (the Apostate) has no claim to be regarded as an historical document. Its hero is Jovian, one of the feeblest of Roman emperors, and Julian is everywhere exhibited in flaming colours as the villain of the story.

A valuable historical source, though of small dimensions, is the *Chronicle of Edessa*, which gives a record of events from 132–131 B.C. to A.D. 540—at first exceedingly brief, but becoming somewhat fuller for the later years. It appears to be thoroughly reliable wherever it can be tested. It has been three times edited—first by Assemani in the *Bibliotheca orientalis* (i. 388–417), secondly by L. Hallier (Leipzig, 1892) with a translation, introduction and abundant notes, and thirdly by Guidi with a Latin version

(in *Chronica minora*, Paris, 1903).

On John of Asia or Ephesus, the eminent Monophysite bishop and earliest Syriac church historian, *see* the separate article.

An historical work of somewhat similar character to John's is the compilation in 12 books which is generally known by the name of Zacharias Rhetor, because the anonymous Syriac compiler has incorporated the Syriac version or epitome of a lost Greek history written by that author. The Syriac work exists (not quite complete) in a British Museum ms. of about the beginning of the 7th century: this can be in part supplemented by an 8th-century ms. at the Vatican. From the latter Guidi published the interesting chapter (X. 16) which contains the description of Rome. The entire text of the London ms. was published by Land in the third volume of his *Anecdota syriaca;* and there is now an English translation by Hamilton and Brooks (1899), and a German one by Ahrens and Krüger (Leipzig, 1899).

Of the other 6th-century Jacobite writers we need mention only Moses of Aggēl (*fl. c.* 550–570) who translated into Syriac some of the writings of Cyril, and Peter of Callinīcus, Jacobite patriarch of Antioch 578–591, who wrote a huge controversial treatise in 4 books, each of 25 chapters, against Damian, patriarch of Alexandria, as well as other less important works.

The Nestorian writers of the 6th century were numerous, but as yet we know little of their works, beyond what 'Abhdīsho' tells us in his *Catalogue*. It will be sufficient to mention one or two. Joseph Hūzāyā (*i.e.*, of al-Ahwāz or Khūzistan), who came third in succession to Narsai as head of the school of Nīsībis, was the first Syriac grammarian and invented various signs of interpunction. Mārūthā, who was Nestorian catholicus of Seleucia from about 540 to 552 (*see* Labourt, *op. cit.*, pp. 163–191), and a man of exceptional energy, made the only known attempt, which was, however, unsuccessful, to provide the Nestorians with a Bible version of their own. He was the author of many commentaries, homilies, epistles, canons and hymns. Paul the Persian, a courtier of Khosrau Anōsharwān, dedicated to the king a treatise on logic which has been published from a London ms. by Land in the 4th volume of his *Anecdota*. Bōdh the periodeutes is credited with a philosophical work which has perished, but is best known as the author of the old Syriac version of the collection of Indian tales called *Kalīlah and Dimnah*. He made it doubtless from a Pahlavī version. His translation, which was edited by Bickell with an introduction by Benfey, must be distinguished from the much later Syriac translation made from the secondary Arabic version and edited by Wright in 1884, of this there is an English translation by Keith Falconer (1884). Ḥannānā of Ḥědhaiyabh, who nearly produced a disruption of the Nestorian Church by his attempt to bridge over the interval which separated the Nestorians from Catholic orthodoxy, was the author of commentaries.

Decay in the 7th Century.—"With the 7th century," as Wright remarks, "begins the slow decay of the native literature of the Syrians, to which the frightful sufferings of the people during the great war with the Persians in its first quarter largely contributed." The same process of decay was greatly promoted by the Arab conquest of Persia, achieved through the victory of Ḳādisīya in 636–637. The gradual replacement of Syriac by Arabic as the vernacular language of Mesopotamia by degrees transformed the Syriac from a living to a dead language. Apart from a few leading writers—such as Jacob of Edessa, the anonymous historian whose work has passed under the name of Dionysius of Tell-Maḥrē, Thomas of Margā, Dionysius bar Ṣalībī, and Barhebraeus (*qq.v.*)—there are few names of interest.

1. *Theology.*—Here we may first mention George, bishop of the Arabs (d. 724), who wrote commentaries on Scripture, and tracts and homilies on church sacraments, and finished the *Hexaēmeron* of Jacob of Edessa. Bābhai the Elder, a leading Nestorian in the beginning of the 7th century and a prolific author, wrote many commentaries and theological discourses. Īshō'yabh III., Nestorian catholicus from 647 to 657/8, wrote controversial tracts, religious discourses and liturgical works. Elias of Merv, who belongs to the 2nd half of the 7th century, compiled a *Catena patrum* on the Gospels and wrote many commentaries. Timothy I., catholicus 779–823, wrote synodical epistles and other works bearing on church law. Moses bar Kēphā (d. 903), one of the most fertile of 9th-century authors, wrote commentaries, theological treatises and many liturgical works. Other important contributors to this sphere of literature were Īshō' bar Nōn (d. 827/8), John bar Zō'bī (beginning of the 13th century), Jacob bar Shakkō (d. 1241), and the great Nestorian scholar 'Abhdīshō' (d. 1318).

2. *History.*—Besides the important writers treated in separate articles, we need mention only four. Elias bar Shīnāyā, who in 1008 became Nestorian bishop of Nisībis, was the author of a valuable *Chronicle*, to which are prefixed numerous chronological tables, lists of popes, patriarchs, etc., and which covers by its narrative the period from A.D. 25 to 1018. Of this work, which exists in only one imperfect copy, the later portion was edited by Baethgen in 1884, and the earlier by Lamy in 1888. Another important *Chronicle* is that of Michael I., who was Jacobite patriarch from 1166 to 1199. Its range extends from the Creation to the author's own day, and it was largely used by Barhebraeus in compiling his own *Chronicle*. Till recently it was known only in an abridged Armenian version which was translated into French by V. Langlois (Venice, 1868): but the Syriac text was found in a ms. belonging to the library of the church at Edessa, and has been published by J. B. Chabot. A work rather legendary than historical is the *Book of the Bee*, by Solomon of al-Baṣrah, who lived early in the 13th century. Lastly, acknowledgment must be made of the great value of the Catalogue of Nestorian writers, by 'Abhdīshō' of Nisībis, the latest important writer in Syriac. It was edited by Assemani in the 3rd part of his *Bibliotheca orientalis*, and has been translated into English by Badger.

3. *Biography, Monastic ,History, etc.*—Besides the important work by Thomas of Margā (*q.v.*) the following deserve special mention. Sāhdōnā, who was a monk in the Nestorian monastery of Bēth 'Ābhē (the same to which Thomas of Margā belonged two centuries later) and afterwards a bishop early in the 7th century, wrote a biography of and a funeral sermon on his superior Mār Jacob who founded the monastery, and also a long treatise in two parts on the monastic life of which all that survives has been edited by P. Bedjan (Paris, 1902). Whilst accompanying the catholicus Īshō'yabh II. (628–644) on a mission to Heraclius, Sāhdōnā was converted, apparently to Catholicism (*see* H. Goussen, *Martyrius-Sahdonas Leben und Werke*, Leipzig, 1897), and thereby caused much scandal in the East. The chief events in his life are narrated by Ishō'dĕnaḥ. Another Nestorian who, a few years later, wrote ecclesiastical biographies and other theological works was Sabhrīshō' Rusṭam, who lived at Mount Īzlā and other monasteries. In the beginning of the 8th century David of Bēth Rabhan, also a Nestorian monk, wrote, besides a geographical work, "a monastic history, called *The Little Paradise*, which is frequently cited by Thomas of Margā." A more important work is *The Book of Chastity*, by Ishō'dĕnaḥ, who according to 'Abhdīshō' was bishop of Ḳaṣrā—but read Baṣrā—about the end of the 8th century. This work is a collection of lives of holy men who founded monasteries in the East, and is a valuable historical source. The work itself, or an abridgment of it, was discovered and published for the first time by J. B. Chabot (Rome, 1896). As the last under this head we may mention a late anonymous biography, that of the catholicus Yabhalāhā III. (1281–1317), which throws much light on the relations of the early Mongol kings with the heads of the church in their dominions. Among other interesting features it contains information about the Nestorian Church of China in the 13th century.

4. *Philosophy and Science.*—Special mention may be made of 'Ānānīshō' of Ḥědhaiyabh (middle of 7th century) well known as the author of a new recension of the *Paradise* of Palladius, and also the author of a volume on philosophical divisions and definitions; Romanus the physician (d. 896), who wrote a medical compilation, a commentary on the Book of Hierotheus, a collection of Pythagorean maxims and other works; Moses bar Kēphā, the voluminous writer above referred to; the famous physician Ḥonain ibn Iṣḥāḳ (d. 873), who wrote chiefly in Arabic, but deserves mention here by his services to Syriac grammar and lexicography, and still more by his translations of Greek philosoph-

ical and scientific works into Syriac and from Syriac into Arabic, becoming in a sense the founder of a school of translators; and Jacob bar Shakkō, whose work called the *Dialogues* treats of grammar, rhetoric, poetry, logic, philosophy and science.

5. *Grammar and Lexicography.*—Several of the authors in this department have already been mentioned. The more important, besides Jacob of Edessa and Barhebraeus, are 'Anānīshō' of Hēdhaiyabh, Ḥonain ibn Iṣḥāḳ, his pupil Bar 'Alī, Bar Sarōshwai (early 10th century), Bar Bahlūl (middle of 10th century), Elias of Ṭirhān (d. 1049), Elias bar Shīnāyā (above), John bar Zō'bī (beginning of 13th century) and Jacob bar Shakkō.

See W. Wright, *History of Syriac Literature* (1895); R. Duval, *La Littérature syriaque* (3rd ed. 1907); A. Baumstark, *Geschichte der syrischen Literatur* (1922). (N. M'L.)

SYRINGA, the common name applied to several cultivated ornamental shrubs of the genus *Philadelphus*, of the family Saxifragaceae (*q.v.*). The name arose from the fact that the original generic name of these shrubs was *Syringa*, which, however, properly belongs to the lilac (*q.v.*). The garden syringas, better known as mock orange, are all species of *Philadelphus*. They possess conspicuous, usually strongly scented flowers and in a wild state are natives of the north temperate zone. There are about 40 species. The common garden syringa (*P. coronarius*) is Eurasian, but is found in many horticultural varieties. Noteworthy North American species are the scentless syringa (*P. inodorus*), the large-leaved syringa (*P. grandiflorus*) of the southeastern U.S., and the western syringa (*P. lewisi*), state flower of Idaho, found from Montana to British Columbia and Oregon.

SYRINGE, a hydraulic instrument, based on the principle of the pump, for the drawing up and ejecting of liquids. The ordinary form is that of a glass or metal tube ending in a pointed nozzle and fitted with an airtight piston rod and handle. The nozzle is inserted in the liquid, which enters the cylinder by atmospheric pressure when the piston rod is drawn up. On pushing back the piston the fluid is ejected in a jet through the nozzle. In sizes varying from the needle-pointed hypodermic syringe to the abdominal syringe, it is a common surgical implement used for the injection of fluids into the body or for the washing of wounds and cavities. A large syringe of metal with a flat perforated nozzle is used as a garden implement for watering plants.

SYRINX, the Greek name for the musical instrument more commonly known as the panpipes (*see* PANPIPE). Syrinx is also a term used in the fields of medicine and anatomy, meaning: (1) a tubelike structure; (2) a fistula; and (3) a kind of secondary larynx found in birds.

SYRYENIANS (also Sirianian, Syrjenian, Zyrenian, Zirianian, Zyrian and Zirian), a tribe of the Permian division of the eastern Finns. Their headquarters are at Ust-Ishma, at the junction of the Ishma and Pechora. Formerly they spread farther to the west. They are of moderate stature, blond and grey-eyed, energetic and inclined to trade. They were converted to Christianity about 1350 and their language was reduced to writing. They call themselves Komi and are not sharply distinguished from the tribes known as Permian, the languages being mutually intelligible. The archaeological remains in the governments of Perm and Vyatka called Chudish by Russians are probably Syryenian.

SYZRAN, a town of the Russian Soviet Federated Socialist Republic, U.S.S.R., in the Kuibyshev *oblast*, in latitude 53° 10′ N., longitude 48° 33′ E., on the right bank of the Volga. Pop. (1959) 148,866. It is a railway junction on the Moscow-Orenburg and Kazan-Volgograd lines.

The town, founded in 1683, was originally a fort to protect settlers from Tatar raids, but developed into a manufacturing centre, with river and rail trade in exported grain and imported timber and manufactured goods. Industries include the manufacture of building materials and locomotives and petroleum refinery.

SYZYGY, in astronomy, either of the points at which the moon is most nearly in a line with the earth and sun. The moon passes its syzygies, or is in a syzygy, at new and full moon.

SZCZECIN (German STETTIN), the chief town of western Pomorze (Pomerania) Szczecin *wojewodztwo* (province), Pol., and the main Polish seaport, lies on the left bank of the Oder (Odra) 17 mi. above its entrance into the Gulf of Szczecin and about 40 mi. from the Baltic sea. Pop. (1939 census) 268,915; (1950 census) 178,207; (1960 census) 268,900. Area 110 sq.mi.

Outside the old walled city, which was built high above the river level, stands the church of SS. Peter and Paul. Founded in 1124, it is the oldest Christian church of western Pomerania; it was restored in the 15th century and again in 1816–17. Within the city walls, dismantled in 1874, the Pomeranian dukes of the 13th century built a castle, which was often rebuilt and enlarged. The old city contained also the huge Gothic churches of St. John (founded 1240) and St. James (14th century), both destroyed in World War II. The gates of the fortifications built by Frederick the Great still stand.

The modern city of Szczecin was built to the north and west of the old city in the second half of the 19th and the beginning of the 20th century. Well-planned, with wide avenues and many parks and open spaces, its central point is the Grunwald square from which radiate eight streets. Since 1945 a medical school, a college of engineering and a school of agriculture have been established. The town has three theatres and issues two local daily newspapers and a literary weekly paper. The local broadcasting station covers the whole coastal area of Poland and the Baltic countries.

About 80% of the old city and 50% of the new were destroyed by two massive British air raids in Aug. 1944 and by the Russian heavy artillery fire during the last stages of World War II. By 1957 reconstruction of the city had made considerable progress.

Szczecin was first mentioned in a document known as *Dagome index*, dating from the first half of the 10th century, by which Prince Mieczyslaw or Mieszko I, the founder of the Piast dynasty of Poland, put the town of Schinesghe under the Holy See's protection. The name of the town comes from *szczecina* (bristle) and it was known to the Danes as Børstaborg (brush castle). In 1124 Christianity was introduced by Otto, bishop of Bamberg, whose mission was sponsored by Boleslaw III Wrymouth, king of Poland, and the Slavonic population, consisting largely of fishermen and craftsmen, was baptized. In 1243 Szczecin received municipal autonomy from Duke Barnim I and it remained the capital of the dukedom of Pomerania or Pomorze (*q.v.*) until 1637. Already a leading centre of trade, it joined the Hanseatic league (*q.v.*) in 1360. From 1648 to 1720, when it was ceded to Prussia, it belonged to Sweden and it remained Prussian until the end of World War II. Catherine the Great, empress of Russia, was born in Szczecin. The Potsdam agreement of Aug. 2, 1945, restored the town to Poland and all Szczecin's German inhabitants who had not already fled were transferred to Germany. By 1957 the population figure was almost as high as in 1939 but the inhabitants were exclusively Polish.

Szczecin exports coal and imports iron ore from Scandinavia. It also handles transit traffic between the central European countries and the far east. After 1945 its docks were enlarged and installations were repaired. Goods loaded and unloaded during 1953–56 amounted to 6,323,000 metric tons annually. The port is linked with its hinterland by the navigable Oder and with overseas countries by nine regular shipping lines. Eight railway lines join it to the rest of Poland and from the airport at Zdroje, 2 mi. from the town, there are services to Warsaw and Poznan. Szczecin has one of the two biggest Polish shipyards, capable of building ships up to 10,000 gross tons. It is also an industrial centre; its largest enterprises are a paper and cellulose plant, a phosphatic fertilizer plant, a synthetic yarn mill, ironworks, a cement factory and a motor-vehicle repair shop. These were largely destroyed in World War II but were later entirely rebuilt.

SZÉCHENYI, ISTVAN, COUNT (1791–1860), Hungarian statesman, born at Vienna on Sept. 21, 1791, who was called "the greatest of the Magyars" by his political antagonist F. Kossuth, was the son of Ferencz Széchenyi and the countess Juliana Festetics. Entering the army at the age of 17, he fought with distinction at the battle of Raab (June 14, 1809), and on July 16 brought about the subsequent junction of the two Austrian armies by conveying a message across the Danube to Gen. J. G. Chasteler at the risk of his

life. Equally memorable was his famous ride through the enemy's lines on the night of Oct. 16–17, 1813, to convey to Blücher and Bernadotte orders for the impending battle of Leipzig. In May 1815 he was transferred to Italy, and at the battle of Tolentino scattered Murat's bodyguard by a dashing cavalry charge. From Sept. 1815 to 1821 he traveled widely, studying the institutions of the countries through which he passed. A second tour with his friend, Baron Miklos Wesselényi, taught him much about trade and industry.

In 1825, when he went to France to attend the coronation of Charles X, the canal du Midi attracted his attention and suggested to him the idea of regulating the Danube and Theiss rivers. At the diet of 1825, when the motion for founding a Hungarian academy was made by Pál Nagy, who bitterly reproached the Magyar nobles for so long neglecting their mother tongue, Széchenyi set the example by offering to contribute a whole year's income (60,000 florins) toward it. The formation (June 1833) of the Danube Navigation company, which eventually opened up the Danube from Buda to the Black sea, was the fruit of Széchenyi's initiative and personal study. Széchenyi was also the first to start steamboats on the Theiss, the Danube and the lake of Balaton.

All this time Széchenyi had been following, with some anxiety, the political course of Kossuth, whose extravagances he feared would plunge Hungary back into the chaos out of which he had helped to raise it. The majority, indeed, sided with Kossuth, but neither this fact nor the gradual loss of his popularity restrained Széchenyi, both in the diet and at county meetings, from fulminating conscientiously against the extreme demands of Kossuth. His views at this period are expounded in the pamphlet *Politikai programm töredékek* ("Fragments of a Political Program"). He held the portfolio of ways and communications in the first responsible Magyar administration (March 23, 1848) under L. Bátthyány, but he became mentally ill and on Sept. 5 he was taken to an asylum. He took his own life on April 8, 1860.

See Antál Zichy, *Biographical Sketch of Count Stephen Széchenyi*, 2 vol. (1896–97).

SZECHWAN (SSU-CH'UAN SHENG), a province of southwest China, almost 220,000 sq.mi. in area following the reaccession of Sikang (*q.v.*), 1955. Pop. (1953) 65,685,063 (including Sikang); (1957 est.) 72,160,000. It contains four autonomous areas formed since 1952 inhabited predominantly by non-Han peoples (under 2% of total provincial population) living in the mountainous west and southwest: Ahpa (A-pa), Kantse and Muli—all mainly Tibetan; and the Liangshan Yi area. Szechwan was once almost a separate country by size and population, as well as being self-sufficient in resources.

Geography.—Topographically it is distinctive with lofty mountains everywhere and tremendous rivers of unmeasured power potential. From cloud-ringed O-mei Shan, 9,900 ft. (famed for temples, monkeys and pines), southwest of Ch'eng-tu, to the infinite complexity of irrigated garden-patch farms, and the inspiring 1,000–2,000 ft. gorges of the Yangtze (*q.v.*), the marks of man's adjustment and transformation of the landscape are everywhere. The heart of the province is the Szechwan or Red basin that occupies a region that continued as a sea gulf well into the later Paleozoic period, and changed to a fresh-water lake during the Mesozoic with deposition of red sandstones and limestones that were later wrinkled into a northeast-southwest ridge and valley system. On the west are the young Great Snow mountains (Tahsüeh Shan) and the eastern plateau of Tibet, 13,000 ft. above sea level; to the north, the Tsinling Shan (*q.v.*) at 10,000 ft. runs into the Ta-pa Shan and Wu Shan on the east, while the south is marked by the limestone plateau of Kweichow. The largest flat area is the Ch'eng-tu plain in the northwest of the basin. Szechwan means "four streams"; they are usually regarded as the Min, T'o, Chia-ling and the Ch'ien or Wu; sometimes either the Yangtze or Fou Chiang is given instead of the Ch'ien. Most of the major streams flow south, cutting steep gorges through former Sikang, or widening their valley floors in the soft sediments of the Red basin to pour eventually into the Yangtze before it slices its precipitous gorge, 150 mi. due east, through the Wu Shan below Wan-hsien. Within the basin most of the rivers are navigable and provide the most common means of transportation. Inside the mountain ring, which provides a barrier against polar winds, a mild climate permits the growth of almost every crop and tree found in China. Precipitation averages 40 in., and the growing season is 11 months; snow and freezing are rare although the foggy winters are penetrating. Because of the high humidity and constant cloud cover, the Chinese say that in Szechwan the dogs bark when they see the sun.

Agriculture.—Given favourable conditions for agriculture the large population has converted lowlands and terraced hillsides into intensive gardens everywhere. Rice, sweet potatoes, sugar cane, silk, tea, citrus, tobacco, cotton and tung oil are the primary products, with wheat, corn, soy beans, white potatoes and barley as secondary. Bamboo and palms on the lowlands are matched by conifers and alpine meadows on the high mountains. Szechwan is China's largest rice producer with surpluses for transport to deficit areas, along with outshipments of sugar, citrus, tung oil, spices and medicinal products. Liang-shan writing paper made from bamboo is prized by Chinese artists. The world's best natural bristle is plucked from the backs of Szechwan hogs; goats are also reared. Within high, secluded bamboo glades in the west live the rare giant panda bears.

Industry.—Minerals of industrial importance are increasingly exploited. Salt has long been pumped from natural brine wells, from 300 to more than 1,000 ft. deep, at Tze Kung, while bituminous coal outcrops are numerous around Chungking. Iron ores are scattered; the largest deposit, estimated at 500,000,000 tons of about 40% iron content, is in the southeast of former Sikang province. Petroleum is reported extracted and refined on the middle Chia-ling river, less than 90 mi. N. of Chungking (*q.v.*), while in the salt well area natural gas is also being converted into petroleum. Szechwan's first major hydroelectric station (48,000 kw.) was built northeast of Chungking on the Lungch'i. From the mountains in the west, small amounts of gold, silver, copper, lead, antimony and asbestos are produced. The pace of industrialization continues to increase. Chungking, the major industrial centre of southwest China, has an automatic thermoelectric power plant, iron and steel plants (open hearth and electric), textile mills, chemical works and lighter industries. Electric trolley buses run in the widened downtown streets, while cable cars have replaced many of the long flights of stone steps down to the rivers. In 1952 Szechwan's first railway, 313 mi., between Chungking and Ch'eng-tu, the provincial capital, was opened, followed in 1958 by the Pao-chi to Ch'eng-tu line, 400 mi. across the Tsinling Shan to the north. Two more lines are growing south, one from Chungking to Kuei-yang in Kweichow (190 mi.), by a bridge across the Yangtze; the other from Nei-chiang, the sugar centre, to K'un-ming in Yunnan (420 mi.). Ch'eng-tu is the terminus of the motor roads to the west. That change will continue is obvious, for there remains the challenge of damming the lower end of the great Yangtze gorges, a project first drafted during World War II and taken up again in China in the late 1950s; this would be one of the world's largest multipurpose river control projects, with 10,000,000 kw. power available for a 300-mi. area.

History.—In the past the keynote of Szechwan was isolation. Non-Han peoples related to the Thai stocks had their centres of the Shu and Pa states at Ch'eng-tu and Chungking respectively at least as early as the Chinese settlement in the Yellow river valley. They were conquered by the Ch'in from the north in 316 B.C., augmented by a large number of transported Chinese settlers who enlarged Ch'eng-tu's great irrigation system and built the walled city. A possible remnant of the pre-Chinese exists southwest of the Min Chiang in the Liangshan Yi autonomous area formed in 1952. During the Three Kingdoms period (A.D. 221–265), Ch'eng-tu formed part of the base of Chu-ko Liang, one of China's shrewdest and most admired historical strategists. At the end of the Ming dynasty widespread massacres left vacant land that was occupied in the 18th century by whole outside communities—most notably a large contingent from Kwangtung whose descendants around Ch'eng-tu still speak Cantonese. Since the main routes into Szechwan were over the Tsinling Shan, with the shift of Chinese power east and then south the province was some-

what forgotten except as a rice granary and control base against the non-Hans to the west. During the warlord days of the republic, the area was practically impregnable. After the start of the Chinese–Japanese war in 1937, industrial enterprises and millions of refugees from coastal provinces were evacuated to Szechwan which resulted in rapid economic development of the area. Chungking, former treaty port and the wartime capital of China, made Szechwan the political and economic centre and became the site of important military bases and headquarters installations for Chinese and allied forces.

Although the Japanese bombing of Szechwan cities, especially the capital, was severe, the province was not entered by the invaders. However, in spite of other large cities like Ch'eng-tu, Luchow, Wan-hsien and I-pin (*qq.v.*), most of the population remains basically rural. (TE. H.)

SZEGED, a city in, but independent of, the county of Csongrád, Hung. It lies just below the confluence of the Maros and Tisza rivers. Pop. (1949) 83,207; (1960) 99,061 (mun.). It is modern, the old town having been almost destroyed by a Tisza flood in 1879. In early times it was important because it lay on the salt and timber route from Transylvania to the Adriatic. A large cotton mill was built there in 1947–49. The city also has a shoe factory and food-processing plants. It is the seat of a university.

SZÉKESFEHÉRVÁR (German STUHLWEISSENBURG), a town in Hungary, capital of the county of Fejér, lies in a marshy valley, 40 mi. S.W. of Budapest. The town was the coronation and burial place of the Hungarian kings from the 10th to the 16th century. The coronation church was built by St. Stephen, the first king of Hungary. From 1543 to 1688 it was under Turkish occupation. In 1601 it was taken by the Austrians and the church with its tombs was destroyed. In the 1950s it became a centre of the aluminum industry. Pop. (1960) 55,934 (mun.).

SZEKLERS (Magyar *Székely*), a people inhabiting the upper valleys of the Mures and Olt rivers in what was eastern Transylvania and is now Rumania. It has been estimated that they number about 450,000, but they are no longer distinguishable from the Magyars, with whom they are closely akin. Their origin has been much debated. According to their own tradition, repeated in Procopius's *De bello Gothico*, they were descended from Attila's Huns, while others have seen in them Black Ugrians. It is, however, now generally accepted that they are true Magyars, transplanted there probably by St. Ladislaus to guard the frontier, their name meaning simply "frontier guards."

SZÉLL, KOLOMAN (1845–1915), Hungarian statesman, was born on June 8, 1845. He studied at Budapest and Vienna, and in 1867 became deputy for the district of St. Gotthard. In 1875 he was finance minister in I. Tisza's cabinet, and began the task of restoring the credit of Hungary.

In 1878 he concluded with Austria the first economic *Ausgleich*. At that time the single Austrian bank was changed, in conformity with this arrangement, into the dualistic Austro-Hungarian bank, and Széll consolidated the Hungarian *rentes*, and nearly succeeded in balancing the state finances. As he feared that this balance would again be upset by the occupation of Bosnia and Herzegovina, he resigned from the cabinet, incurring thereby the displeasure of the crown. At the beginning of the 1880s Széll founded the Hungarian Mortgage Credit bank, of which he was governor until the end of his life. He opened entirely new sources of credit for Hungarian agriculture. He declined repeated offers of the portfolio of finance. Széll became prime minister after the fall of D. Bánffy in Feb. 1899. On the basis of the so-called Széll formula the new *Ausgleich* with Austria until the year 1907 was concluded after long negotiations. He was forced to resign on June 16, 1903. Széll was one of the Liberal seceders in 1904. Under the coalition cabinet of A. Wekerle, he was chosen president of the Constitutional party. He tried in vain to bridge the opposition between Tisza and J. Andrássy. He died on Aug. 16, 1915.

SZENT-GYÖRGYI, ALBERT (1893–), Hungarian-U.S. biochemist, who was awarded the 1937 Nobel prize for medicine and physiology "for his discoveries in connection with the biological combustion processes, with especial reference to vitamin C and the catalysis of fumaric acid." He was born in Budapest, Sept. 16, 1893. His medical education at Budapest was interrupted by World War I, but he used an interval of leave after being wounded to complete his studies and graduate M.D. (1917). After the war he held posts at Pozsony (later Bratislava), Prague, Berlin, Hamburg, Leyden and Groningen. In 1927 he transferred his laboratory to Cambridge, Eng., and subsequently spent a year in the United States. He returned to Hungary in 1932 and held the chairs of medical chemistry at Szeged and at Budapest. In 1947 he emigrated to the United States and became director of the Institute for Muscle Research at the Marine Biological laboratory at Woods Hole, Mass. Szent-Györgyi's greatest achievement was the isolation from the adrenal glands and from various plant juices of hexuronic acid (ascorbic acid) and the provision of definite proof that this substance is identical with vitamin C. He was also the first to demonstrate the presence of flavin (riboflavin or vitamin B_2) in animal tissues. With S. Ruznyak he discovered vitamin P, which regulates vascular permeability. (W. J. BP.)

SZOLD, HENRIETTA (1860–1945) founder of what became through her leadership the world's largest Zionist group, Hadassah, the Women's Zionist Organization of America, was born in Baltimore, Md., Dec. 21, 1860. She pioneered in Americanization work after 1882, originating (1891) the night school for immigrant workers in Baltimore. A founder of the Jewish Publication Society of America (1888), she served until 1916 as the society's editor general, translator of Jewish classics and writer. In 1912, with 12 members, Miss Szold founded Hadassah. As its president she organized in 1916 a medical unit for Palestine comprising staff and equipment for hospitals, nurses' school, clinics and epidemiological and teaching services. The first woman elected to the World Zionist executive and member of the Palestine Jewish National council, she held their portfolios of health, education and social welfare. She established Lemaan Hayeled (named after her death the Szold foundation) for child welfare and research in juvenile problems. In 1933 she became director of Youth Aliyah, a world-wide movement to rescue and rehabilitate in Israel young victims of Nazi and other persecution. She worked also for Arab-Jewish *rapprochement* through common social and political planning. Miss Szold died in Jerusalem Feb. 13, 1945. She was recognized as the outstanding Jewish woman of her time, a pioneer in international social service, education and Zionist thinking.

See Marvin Lowenthal, *Henrietta Szold, Life and Letters* (1942); Rose Zeitlin, *Henrietta Szold* (1952). (T. DE S. P.)

SZOMBATHELY, a town in western Hungary, that occupies the site of the Roman capital of Pannonia, Sabaria or Savaria. The town lies in the centre of a rich wine-producing downland and is an important agricultural centre with steam flour and saw mills and factories producing agricultural machinery and textiles in addition to foundries and a state railway workshop. It has been the see of a Roman Catholic bishop since 1777 and possesses a beautiful cathedral and several other ecclesiastical buildings. Pop. (1960) 54,465 (mun.).

SZYMANOWSKI, KAROL MACIEJ (1882–1937), the principal Polish composer of orchestral music and the leading figure in the Polish musical revival. Born on Oct. 6, 1882 (new style) at Timoshovka, near Elisavetgrad in the Ukraine, he studied with his father and Gustav Heuhas and in 1901 with Z. Noskowski in Warsaw. His early piano works include preludes, studies and the variations, opus 3. Though influenced by Scriabin and Richard Strauss, his style developed between 1909 and 1913 in his second symphony, second piano sonata and the opera *Hagith* (Warsaw, 1922). During World War I he wrote his third symphony, showing oriental affinities, *Métopes* for piano, the *Mythes* for violin, including *The Fountain of Arethusa*, and the first violin concerto. Szymanowski's later works were remarkable for their lyrical inspiration and ornate craftsmanship. The best of his songs were written from 1915 onward and include the cycle *Slopiewnie*. His style reached maturity in the opera *Król Roger*, containing the well-known "Song of Roxana" (Warsaw, 1926). Folk-song elements were used in his ballet *Harnasie*, and in his *Stabat Mater* and *Litania* he revived an archaic style. In 1926 he became director of the Warsaw conservatory. His last compositions include the *Symphonie Concertante* and the second violin concerto. He died at Lausanne, March 29, 1937.

See Stefania Lobaczewska, *Karol Szymanowski* (1950).

THIS letter corresponds to Semitic *tau* (meaning "mark"), the form of which on the Moabite stone was X. In all Greek alphabets, as well as in the Lydian and the Latin, the form was T. Etruscan had a form ⊤, from which probably were derived the Umbrian form ⊤ and the Faliscan ⊤ and ⅄.

Latin cursive gives a form that leans to the right ⊢. Carolingian ꞇ is based upon uncial ꞇ. Generally speaking, in medieval manuscripts the letter did not rise above the line and it is sometimes difficult to distinguish from a minuscule *c*. In modern handwriting and printing it is the custom to extend the vertical stroke above the horizontal.

NAME OF FORM	APPROXIMATE DATE	FORM OF LETTER
PHOENICIAN	B.C. 900	X
CRETAN	600	T
THERAEAN	700-600	T
ARCHAIC LATIN	700-500	T
ATTIC	600	T
CORINTHIAN	600	T
CHALCIDIAN	600	T
IONIC	403	T
ROMAN COLONIAL	PRE-CLASSICAL AND CLASSICAL TIMES	T T
URBAN ROMAN		T
FALISCAN		⅄
OSCAN		T T
UMBRIAN		⊤ ⊤ ⊤
CLASSICAL LATIN AND ONWARDS		T

THE DEVELOPMENT OF THE LETTER "T" FROM THE PHOENICIAN THROUGH THE CLASSICAL, DOWN TO THE PRESENT FORM

The sound represented by the letter throughout its history has been the unvoiced dental stop. In English this has become more alveolar than dental, that is to say, it is pronounced by the tongue pressing upon the gums rather than the teeth. English unvoiced stops are probably also to be distinguished from the corresponding sounds in ancient Greek by the addition of a slight puff of breath following the sound, thus bringing the sound near to being an aspirate. Greek would perhaps have transliterated the English unvoiced stops by the letters θ, ϕ and χ rather than τ, π and κ. This tendency toward aspiration is carried to much farther lengths in the Irish dialects of English.

For an account of the Germanic consonant shifts involving the letter T, according to Grimm's law, *see* GERMANIC LANGUAGES.

(B. F. C. A.; J. W. P.)

TAAFFE, EDUARD FRANZ JOSEPH VON, COUNT [11th Viscount Taaffe and baron of Ballymote, in the peerage of Ireland] (1833–1895), Austrian statesman, was born at Vienna, Feb. 24, 1833, second son of Count Ludwig Patrick Taaffe (1791–1855), who was minister of justice in 1848 and president of the court of appeal. As a child Taaffe was one of the chosen companions of the young archduke, afterward emperor, Francis Joseph. In 1852 he entered the public service; in 1867 he was *Statthalter* of upper Austria, and minister of the interior in Beust's administration. In June he became vice-president of the ministry, and at the end of the year he entered the first ministry of the newly organized Austrian portion of the monarchy. For the next three years he took a very important part in the confused political changes, representing essentially the wishes of the emperor. From Sept. 1868 to Jan. 1870 he was president of the cabinet. In 1870 Taaffe, with the other members of the cabinet who advocated concessions to federalism, resigned. He returned to office in April, resigned once more, and became *Statthalter* of Tirol.

On the breakdown of the Liberal government in 1879 he was again called to office, becoming minister-president in July, and held office until July 1893, when he was defeated on a proposal for the extension of the franchise and resigned, dying at Ellerschau, Bohemia, on Nov. 29, 1895. For the history of Taaffe's administration, *see* AUSTRIA, EMPIRE OF. Essentially an opportunist, he maintained office for many years by an unprecedented employment of the principle *divide et impera*. Usually resting on the Slavs and clericals, he was accused by the German Liberals of ruining Austria; but he united a deep devotion to the emperor with a singular gift for managing men and at least prolonged the life of an impossible system.

By the death of his elder brother Charles (1823–1873), a colonel in the Austrian army, Taaffe succeeded to the Austrian and Irish titles. He married in 1862 Countess Irma Tsaky, by whom he left four daughters and one son, Henry.

See Wurzbach, *Biographisches Lexicon Oesterreichs, Memoirs of the Family of Taaffe* (1856), privately printed; and Taaffe's *Political Correspondence* issued after his death.

TAAL, a municipality (with administrative centre and 48 *barrios* or districts), of the province of Batangas, Luzon, Philippines, on the Pansipit river. It is about 50 mi. S. of Manila and near to Lake Taal. Pop. (1959 est.) 31,905.

On an island in the lake of Taal is a volcano which has erupted many times, notably in Jan. 1911. The consequently fertile surrounding lands produce palay (rice), maize (corn), sugar and other crops. The manufacture of baskets, mats, harness, shoes, slippers, embroidery and bolos is extensive, while horse and cattle breeding and fishing are also carried on. Tagalog is the vernacular. (C. S. L.)

TABACO, a municipality and port of entry of the province of Albay, Luzon, Philippines, on Tabaco bay, a protected harbour about 20 mi. N. of Albay (Legaspi), the provincial capital. Pop. (1959 est.) 40,683.

The chief agricultural produce is abacá (Manila hemp). Cloth, baskets and mats are woven by the women. Tabaco is a port of entry in Albay province and has considerable trade. Bikol is the vernacular. (C. S. L.)

TABARI, AL- (ABU JA'FAR MOHAMMED IBN JARIR AL-TABARI) (*c.* 838–923), Arab historian and theologian, was the author of a history of the world generally known as the *Annals*. Born at Amol in Tabaristan he studied in Baghdad, Syria and Egypt. Most of the rest of his life he spent in Baghdad, first as tutor to the vizier's son, then as a teacher of law and Tradition.

Originally a follower of the Shafi'ite school of law he later established a school of his own; its influence was, however, short-lived. The *Annals* (*Tarikh al-Rusul wa-al Mulk*, "History of the Prophets and Kings") are a general history from the Creation to A.D. 915 and are famous for their detail and, in the later parts, their accuracy (*see* ARABIC LITERATURE: *History and Biography*).

His second great work, a commentary on the Koran (*Jami' al-Bayan* or *Tafsir*) is marked by the same fullness of detail as the *Annals*.

Bibliography.—The *Annals* were ed. by M. J. de Goeje, 13 vol. (1879–1901). The commentary on the Koran was published in 10 vol. (1905–12), ed. by M. M. Shakir and A. M. Shakir, 3 vol. (1954–). *See* also D. S. Margoliouth, *Lectures on Arabic Historians* (1930).

TABASCO, a gulf coast state of southeast Mexico. Area 9,783 sq.mi. Pop. (1950) 362,716. The surface is generally low and flat, largely covered with lagoons, watercourses and swamps. In the south and southeast there is a higher area. Tabasco, a region of very heavy rainfall, is covered by dense forests which provide valuable fine woods and dyewoods. It is traversed by two large rivers, the Grijalva and Usumacinta. The Grijalva, also called Tabasco, is navigable for 62 mi. The Usumacinta forms the boundary between Guatemala and Chiapas until the frontier of Tabasco is reached, then it flows into the Grijalva—the two rivers having a common outlet. The Usumacinta, including its headstreams, is about 500 mi. long; it is navigable for about 310 mi. but only about 180 mi. by large steamers. There are few good roads, and the rivers are the principal means of transportation in the state. There are also air and rail lines. The state capital is Villahermosa (pop. [1950] 33,587), on the Grijalva river. The next most important town is Frontera (pop. [1950] 8,449), a port 3 mi. within the mouth of the Grijalva.

Juan de Grijalva explored the area in 1518, and in 1519 Cortés first clashed with the Indians, from whom he acquired his faithful Doña Marina. Francisco de Montejo partially subdued them after constant uprisings in the 1530s and 1540s. Tabasco, an Indian name meaning "damp earth," became a state in 1824. The potential oil wealth of the area began to be exploited in the 1950s. Cattle, copra, bananas, cocoa, coffee, rice, sugar cane, rubber, vanilla, tropical fruits and corn are the main agricultural products.

(J. A. Cw.)

TABERNACLE, the name given in the English Bible to the portable sanctuary, erected by Moses in the wilderness as the place of worship of the Hebrew tribes (Exod. xxv. *sqq.*). It stood within a rectangular court, measuring 100 cubits by 50, say 150 feet by 75, which formed the centre of the camp in the wilderness. Of the two equal squares into which the court may be divided, the more easterly was that in which the worshippers assembled. In the centre of this square stood the altar of burnt-offering, a hollow chest of acacia wood overlaid with bronze. In the western square stood the tabernacle itself. The essential part of the structure was that termed in the original the *mishkān, i.e.,* dwelling. It was formed of ten curtains, in two sets of five, of the finest linen with inwoven coloured figures of cherubim spread over a series of open frames of acacia wood overlaid with gold, each 10 cubits in height by 1½ in breadth. These frames, 48 in all, were so arranged as to form the southern, western and northern sides of a rectangular structure, 30 cubits in length and 10 in breadth and height, whose eastern end, forming the entrance, was closed by a special portière suspended from five pillars. The dwelling was divided into two parts by a second hanging, the "veil," 10 cubits from the western end. These two parts were termed respectively the holy place, and the most holy place or "holy of holies." Within the latter stood the ark of God, in which were deposited the two stone tables of the decalogue or "testimony." On the ark lay a solid slab of the finest gold, the propitiatory or mercy-seat, from which rose the figures of two golden cherubim, forming the innermost shrine of the wilderness sanctuary, the earthly throne of the God of heaven.

The furniture of the holy place consisted of the table of shewbread, the altar of incense—both, like the ark, of acacia wood overlaid with gold—and the golden "candlestick," or seven-branched lamp-stand. As a protection the delicate and artistic curtains of the dwelling were covered by two similar sets of goats'-hair curtains; these, in their turn, were protected by a double covering, the one of rams' skins dyed red, the other made of the skins of a Red Sea mammal, probably the dugong (Exod. xxvi. 14).

The aim of the "priestly" writers, to whom we owe this conception of the tabernacle, was to provide a sanctuary and a ritual worthy of the higher conceptions of the Deity, which had grown up as the fruit of the discipline of the exile. The thought of the almost unapproachable holiness of the Deity underlies not only the gradation of the parts of the tabernacle—court, holy place and holy of holies, each marked by an ascending degree of sanctity—but also the careful gradation of the materials employed in its construction. The whole is to be regarded as the expression of a religious ideal. Building on the traditions of the simple Mosaic "tent of meeting," the priestly idealists followed the example of Ezekiel, and elaborated a sanctuary to serve as the model for the worship of the theocratic community of the future. "Let them make me a sanctuary, that I may dwell among them" (Exod. xxv. 8).

Bibliography.—*See* "Tabernacle" by Kennedy in Hastings' *D.B.,* with which may be compared the corresponding articles in the *Ency. Bib.* by Benzinger, and in the *Jewish Encycl.* by König; also the commentaries on Exodus. (A. R. S. K.)

In architecture, the term loosely expresses a niche for a statue, with a canopy (*q.v.*) over it, especially in the mediaeval styles. When niches are arranged in vertical rows, with the canopy over one statue acting as the pedestal for the one above, the whole composition is often termed tabernacle work. The same term is also used for any richly decorated tabernacle or canopy-like forms, whether true tabernacles or not.

TABERNACLES, FEAST OF. A famous Jewish festival which commences on 15 Tishri (five days after the Day of Atonement), and lasts seven days. Of these the first two (Tishri 15 and 16) are full festival days, the last five (Tishri 17–21) half-holidays. The "eighth day" Festival (Shemini Aisereth), which like other festival days is doubled (Tishri 22 and 23; the second day in this case is called *Simchath Torah*), closes the celebration. The whole nine days bear one descriptive designation, "season of our rejoicing." The festival is one of venerable antiquity. Its observance is commanded in the Mosaic Law (Lev. xxiii. 34), and its purpose is there explained as to commemorate the way in which the Israelites dwelt in booths (*sukkoth*) in the wilderness.

Every Jew who owns a court or garden is required to erect a booth or something more or less equivalent, and to dwell in it—or at least have meals in it—while the feast lasts. In order that the character of the original booth may as far as possible be retained, the modern counterpart is very lightly constructed. It "must not be covered with fixed boards and beams or with canvas, but with detached branches of trees, plants, flowers and leaves, in such a manner that the covering is not quite impenetrable to wind and rain or starlight" (Friedlander). The booths are required to be made during the days that intervene between the Day of Atonement and the Feast itself; they are adorned with garlands, flowers, and the like; often the text "Ye shall dwell in booths seven days" (Lev. xxiii. 42), is displayed in a prominent place within the booth. On the eve of the Feast the members of the household attend Synagogue, and on returning assemble in their booth and partake of a meal. On this evening (the first of the Feast) the meal (as in the case of the weekly Sabbath on Friday evening) is preceded by Qiddush or the solemn sanctification, first over a cup of wine which is then handed round, all drinking of it, and then over two wheaten loaves, specially baked and placed on the table covered with a cloth; these the head of the family then proceeds to cut into pieces, which are distributed to and eaten by all present. A special blessing—which is repeated before every meal during the seven days—follows ("Blessed art Thou, O Lord . . . Who hast commanded us to dwell in a booth"). At the conclusion of the meal, and on leaving the booth the following prayer is said by the head of the household:

> May it please Thee, O Lord my God, and God of my fathers, that in like manner as I have this time obeyed Thy command and have been sitting in the booth, so in the coming year I may be counted worthy to sit in the booth of Leviathan.

It should be observed that the Leviathan plays a prominent part in haggadic legend, especially in connection with the Messianic time. The monster is to be killed and the flesh is to furnish food for the righteous (at the Messianic banquet). From the

hide tents will be made by God for the pious of the first rank. This last feature will explain the allusion in the text of the prayer.

On the occasion of the ceremony, the Synagogue is decorated with plants and fruits and certain ceremonies are carried out, viz., palm-branch processions culminating on the seventh day at the *Musaf* or additional service. (G. H. B.)

TABES DORSALIS (Locomotor Ataxia) is a form of neurosyphilis characterized by chronic degeneration in the posterior columns of the spinal cord. The disease results in loss of muscular co-ordination.

The first symptom, which occurs between about 15 to 25 years after infection, usually is a lessening or the loss of the ankle and knee reflex. As the disease progresses, the patient gradually loses his ability to walk and stand normally. Treatment is the same as for other forms of late syphilis. *See* also Venereal Diseases: *Syphilis: Late Symptomatic Syphilis.*

TABLE (Lat. *tabula*), in furniture, a flat, oblong slab supported upon legs or pillars. Of the many kinds of tables according to their uses, the following are outstanding as pieces of furniture and in artistic possibilities: (1) artist's table, (2) billiard table (*q.v.*), (3) breakfast table, (4) card table, (5) china table, (6) console table, (7) dining table, (8) dressing table, (9) library table, (10) side table, (11) sofa table, (12) tea table, (13) wine table, (14) work table, (15) writing table. The table was known in a small and rudimentary form to the Egyptians, who used wood for its construction; the Assyrians certainly employed metal and possibly other materials in its manufacture. Grecian tables were also often of metal, with three or four legs and of considerable variety of form; they were small and low. By Roman times the table had apparently become somewhat more common. The favourite form was the tripod, but one and four legs were also used. Already the shape varied considerably, and in addition to wood, there were tables of marble, ivory, bronze and the precious metals. The more costly examples were carved, inlaid or otherwise ornamented; cedar and the finely marked or grained woods generally were much sought after. As in Greece the tables were low; they were intended for reclining, rather than sitting; their legs were those of wild beasts, or were formed of sphinxes, termini and other figures. Some of those which remain are of extreme grace and most delicate workmanship; to them the Empire style is enormously indebted. In antiquity, tables of any kind can only have been the appanage of the rich. In the early middle ages, although there was variety of form—the circular, semicircular, oval and oblong were all in use—tables appear, save in rare instances, to have been portable and supported upon trestles fixed or folding, which were cleared out of the way at the end of a meal. The custom of serving dinner at several small tables, which is often supposed to be a very modern refinement, was certainly followed in the French châteaux, and probably also in the English castles, as early as the 13th century. For persons of high degree, fixed tables were reserved. Even at a period when domestic furniture was of a very primitive character and few modern conveniences had been evolved, costly tables were by no means unknown—some dim traditions of Rome's refinements must necessarily have filtered through the centuries. Thus Charlemagne possessed three tables of silver and one of gold—no doubt they were of wood covered with plates of the precious metals. Before the 16th century the number of tables properly so called was small; hence very few of earlier date than the middle of that century have come down to us. In the chapter house of Salisbury cathedral is a restored 13th-century example which stands practically alone. In point of age it is most nearly approached by the famous pair of trestle tables in the great hall at Penshurst.

When the table became a fixed and permanent piece of furniture the word board, which had long connoted it, fell into disuse save in an allusive sense, and its place was taken by such phrases as "joyned table" and "framed table"—that is, jointed or framed together by a joiner; sometimes people spoke of a standing or dormant table. They were most frequently oblong, some 2 ft. or 2 ft. 6 in. wide, and the guests sat with their backs to the wall, the other side of the table being left free for service. Sometimes they were used as side tables, or furnished with a cupboard beneath the board; they were supported on quadrangular legs or massive ends and feet, full of Gothic feeling, and were several inches higher than the dining table of the 20th century. Heavy stretchers or footrails were fixed close to the floor—for the avoidance, no doubt, of drafts. Oak was the usual material, but elm, cherry and other woods were sometimes used. Soon the legs became bulbous, and were gadrooned or otherwise ornamented, and the frame began to be carved. The introduction, before the 16th century closed, of the drawing table marked the rapidity with which this piece of furniture was developed. This was the forerunner of the extending dining table. Of the three leaves of which these tables were composed two were below the third; they drew out and were supported by brackets, while the slab proper dropped to the same level. Somewhat later legs became excessively bulbous; this ugly form gave place soon after the middle of the 17th century to baluster-shaped legs. Hitherto tables had, generally speaking, been large and massive—little in the nature of what is now called the occasional table seems to have been provided until some years after the Restoration. About that time small tables of varying sizes and shapes, but still of substantial weight, began to be made; many of them were flap tables, which took up little room when they were not in use. These, however, had been known at an earlier date. Charles II had not long been on the throne when the idea of the flap table was amplified in a peculiarly graceful fashion. Two flaps were provided instead of one, the result being the rather large oval table of the gate-leg variety that has remained in use ever since, in which the open gate supports the flap. Toward the end of the reign, tables began to have the graceful twisted legs joined to the flat serpentine stretchers, which produced, almost for the first time in English furniture, a sense of lightness and gaiety. The walnut tables of the end of the Stuart period were often inlaid with marquetry of great excellence. The number and variety of the tables in well-to-do households were now increasing rapidly, and the console table was imported from the continent contemporaneously with the common use of the mahogany side table.

As mahogany came into general use, about the beginning of the second quarter of the 18th century, an enormous number of card-tables were made with plain or cabriole legs and spade or claw and ball feet, often with lions' heads carved upon the knees; the top folded up to half its size when open. The Chippendale school introduced small tables with carved openwork galleries around the edges (to protect china and other small objects), and clustered legs; Gothic forms and Chinese frets were for a time fashionable. Later in this century, so prolific in new forms of furniture, tables were frequently made of rosewood and satinwood; side tables, often highly elaborate, adorned with swags and festoons and other classical motives, supported by termini or richly carved legs, were gilded and topped with marble slabs or inlaid wood. The Pembroke table, of oblong form, with two semicircular or oblong leaves, with edgings of marquetry, was a characteristic feature of late 18th-century English furniture, and still retains its popularity. During the Empire period the taper was replaced by the round leg, rosewood grew commoner, and brass mountings became the rule. *See* also Furniture Design and Interior Decoration, History of.

TABLE MOUNTAIN (Afrikaans, *Tafelberg*), a name frequently given in South Africa to flat-topped hills and mountains, which are a characteristic feature of the scenery. Occasionally such hills are called *plat; i.e.*, flat, *bergen*. Specifically Table mountain is the mountain which rises from the shores of Table bay, Cape Town lying at its seaward base and on its lower slopes. The mountain forms the northern end of a range of hills which terminates southward in the Cape of Good Hope. It is formed of gently inclined sandstones and quartzites of the Table Mountain Sandstone series, resting unconformably on granites, which are intruded into shales of the Malmesbury series. The sandstones descend toward sea level on the shores of False bay, but are too high to be accessible as building stone for Cape Town. The northern face of the mountain, overlooking Table bay, extends like a great wall some two miles in length, and rises precipitously to 3,567 ft. The face is scored with ravines, a particularly deep

cleft, known as Platteklip gorge, affording the shortest means of access to the summit. East and west of the mountain and a little in advance of it are lesser hills, the Devil's peak (3,567 ft.) being to the east and Lion's Head (2,178 ft.) to the west. Lion's Head ends seaward in Signal hill (1,100 ft.). The western side of Table mountain faces the Atlantic, and forms an escarpment, with a broken sky line, known as the Twelve Apostles; to the south Hout's Bay Nek connects it with the remainder of the range; on the east the mountain overlooks the Cape flats. On this side its slopes are less steep, and at its foot are Rondebosch, Newlands, Wynberg and other residential suburbs of Cape Town. The ascent of the mountain from Wynberg by Hout's Bay Nek is practicable for horses.

The southeast winds which sweep over Table mountain frequently cause the phenomenon known as "the table-cloth." The summit of the mountain is then covered by a whitish-gray cloud, which is being constantly forced down the northern face toward Cape Town, but never reaches the lower slopes. The clouds (not always caused by the southeaster) form suddenly, and the weather on the mountain is changeable. Rainfall on the summit is heavy, 72.14 in. a year being a 12-year average. This compares with an average of 54.63 in. at Bishop's Court, Newlands, at the foot of the mountain on the east, and with 25.43 in. at Cape Town (*q.v.*) at the northern foot of the mountain.

TABLES, MATHEMATICAL: *see* MATHEMATICAL TABLES.

TABLE TENNIS, an extremely fast indoor game, similar in principle to lawn tennis, was apparently derived from real or royal (indoor) tennis about 1880. It was known also under various trade-marks ("Gossima," "Ping-Pong," etc.) until 1926, when the present name was adopted by the International Table Tennis federation.

The federation, founded in Germany in Jan. 1926, called a meeting that was attended by representatives of Austria, England, Germany, Hungary, and later that season of Czechoslovakia and Sweden. Thirteen congresses were held between the years 1926–27 and 1938–39 in European cities and in Cairo, Egypt. Suspended during World War II, it held its first postwar congress in Paris in March 1947. Congresses were later held (1948–57) in Europe, India and Japan. The World Congress of 1957 was held in Stockholm, and thereafter it was held biennially beginning with the congress in Dortmund, Ger., in 1959. In that tournament Japan continued to dominate the world championships, though the men's singles championship was won by a player from Canton, China. Table tennis is a popular tournament sport in the orient and Europe but there is limited interest in tournament competition in the United States. On the other hand, it is a popular participant sport in the U.S., where tables and equipment are widely available at schools and public recreational facilities and in many private clubs and recreation rooms.

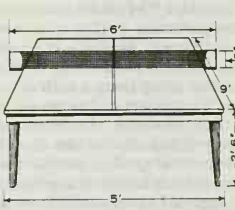


TABLE TENNIS TABLE

The Game.—Table tennis is played by striking a pale, hollow celluloid ball (4½–4¾ in. circumference, 37–39 g. weight, or 37–41 g. weight in the U.S.) with a racket over a net (6 in. high, 6 ft. long) fixed across a dark, nonreflecting table top (9 by 5 ft., 30 in. high). The blade of a modern racket is made of wood and is usually faced on both sides with stippled rubber to control the ball and to impart spin. Japan won the world championship in 1954 with a sponge-faced racket, an innovation in equipment that was banned under rules adopted by the I.T.T.F. in 1959.

The server's racket hits the ball behind his end of the table so that it bounces first on his side of the table, then over the net on his opponent's side. The ball is then struck to and fro directly over the net, after bouncing (volleying is illegal) until a point is scored by failure to make a good return. Service, after every 5 points, alternates and winner is the player who first scores 21 points; but at score 20-all service alternates after each point, and the winner must score two consecutive points. In doubles, the ball must first touch the server's right half court and after passing over the net the receiver's right court; once in play, it may be placed anywhere on the table, as in singles, except that partners must alternate in returning the ball. One U.S. close law prevents strictly defensive play; another prohibits server's imparting spin to the ball except by means of his racket.

The orthodox grip is assumed by "shaking hands" with the racket, as in lawn tennis, except that the table tennis racket or bat is not held at the end of the handle. The forefinger rests along the blade on the rear (backhand) side of the table tennis racket and the thumb on the forehand side. Slight variations from this grip are common, though "penholder" and other unorthodox grips are seldom used in tournament competition.

The basic shots are the push or half-volley, usually executed by meeting the ball with the racket in a near vertical position as the ball rises from the table; the forehand and backhand drives (characterized by topspin); and the forehand and backhand chops (which impart underspin). There are many variations.

In the United States, the U.S. Table Tennis association interprets the rules of International Table Tennis federation and publishes U.S.T.T.A. closed laws. In the early 1960s there were about 80 comparable national organizations throughout the world.

(P. W. R.)

TABLINUM, in architecture, one of the most important rooms of a Roman house, usually of large size, opening off the atrium (*q.v.*), in the centre of the side opposite the entrance. At its rear there was frequently a large window looking out on the peristyle or colonnaded inner court. It was frequently decorated with busts or pictures of the family ancestors, and took the place of the modern, formal drawing room.

TABOR, a small town of Ceske Budejovice, Czech., on the Lužnice, 52 mi. S. of Prague, is mainly of historical importance since it was founded in 1420 by the more advanced Hussites, who came to be known as Taborites. On the summit of an isolated hill, separated from the surrounding country by the Lužnice and protected by marshes, the market place in the centre of the town is approached by narrow streets to render the approach difficult in wartime. In the market place stands a statue of Žižka, one of the most famous Hussite leaders. There also are the early 16th-century Gothic diaconal church and Gothic town hall with its museum of Hussite memorials. The surrounding district produces the hardier cereals and potatoes. Tabor is a railway junction and an industrial centre. Pop. (1950) 18,914.

TABRIZ, the chief city of East Azerbaijan and second city of Iran, is built at a height of 4,400 ft. in the valley of the river Aji Chai ("bitter water"), which flows westward to Lake Urmia. It is about 100 mi. from the Turkish and 60 mi. from the Soviet frontiers. Tabriz is situated in a pleasant fertile valley surrounded with gardens, but it lies in an earthquake zone and suffers frequent shocks of considerable intensity. The climate is continental with hot dry summers and cold winters. Pop. (1956) 289,996.

The early history of Tabriz (ancient Tauris) is unknown. It became the capital of the Armenian king Tiridates III in the 3rd century A.D. In A.D. 791 it was rebuilt by Zobeidah, wife of Harun al-Rashid, after being destroyed by an earthquake, but was again destroyed in A.D. 1042 with the loss of 40,000 lives. In the 13th century Tabriz resisted the Mongol invasions but was taken finally and became the Mongol capital under Khan Ghazan (*c.* 1295–1304). Its bazaars, mosques, walls and baths were extensively rebuilt at this time. In 1392 it was sacked by Tamerlane, but remained a provincial centre and soon regained its prosperity. About 30 years later the Turkomans made Tabriz their capital and Jahanshah built (1436–1467) the Masjiid i-Kubud ("Blue mosque") which is now partly ruined. It remained the capital under the first Safawid Shah Ismail I, but his successor moved to Kazvin which was less exposed to Turkish attacks. In the centuries that followed, Tabriz was the scene of constant warfare between the Turks, Russians and Persians, and changed hands several times. Russian troops occupied the town in 1900–05,

1909, 1914 and 1915–17, and again in 1941–45.

In 1721 and 1780 the town was wrecked by earthquakes and has been the scene of recurring epidemics of plague and cholera. In 1908 it was the starting place of the revolution against the Shah Mohammed II.

The only ancient buildings of note are the Blue mosque and the Ali Shah citadel (Arg) situated in the southwestern quarter of the town. The mosque, partly destroyed by earthquakes, has a blue tile decoration, one of the finest in Persia. The citadel walls are 25 ft. thick and stand some 120 ft. high. The ruins of Shanab-Ghazan, the royal resting place of Ghazan, are situated to the northwest of the town.

In the middle of the modern city there is a large public park called Gulistan garden, with pools, ornamental fountains and flower gardens, while 3 mi. outside the city, there is a large artificial pool called the King's lake (Shah Kouli). An island has been constructed there which, with its large restaurants, forms a summer resort for the inhabitants of Tabriz. Within the city many new buildings had been erected or were in the course of construction by the late 1950s; a number of these, including local government and other official buildings, are in the European style. Among those being built were the railway station, expected to be the biggest in Persia, and new buildings for the university which has schools of medicine, literature and agriculture.

The principal industry in Tabriz is carpet making; carpets are woven in the city, which is also the centre of trade for carpets woven in the surrounding villages of Azerbaijan. They form the bulk of exports from Tabriz. After weaving, the most important industries are the preparation of leather and the manufacture of shoes. Tabriz is also noted for its dried fruits and the manufacture of soap, paint and matches.

Tabriz is a road centre, and, after the completion of the railway from Tehran, will become a junction on the Asiatic-European route to Istanbul. An obstacle to the expansion of the city is the lack of good water; local water is brackish and supplies have to be piped a considerable distance.

See Sir Percy Sykes, *History of Persia* (1939).

(A. Kn.; M. V. S.-W.)

TABU. The basic meaning of the word tabu (taboo) in English is "forbidden," "prohibited," but it usually refers to actions barred by rules of manners or morals rather than of law; *i.e.*, by social convention in which there are some irrational elements. It has also been extended to apply to any ritual setting-apart of persons or things sacred or accursed.

Origin and Scope.—The term is of Polynesian origin. The form *tabu*, which Capt. James Cook noted when he first discovered its use at Tonga (Friendly Islands) in 1771, was later used by the Tongans when they adopted a written alphabet, but they later came to write the word as *tapu*. This is the most common Polynesian form with *kapu* as a variant, and various speculative etymologies have been given for the literal meaning; *e.g.*, *ta*, "to mark," "strike," *pu*, "exceedingly." In Melanesia, corresponding terms in form and meaning are *dabung* (Busama) and *tambu* (generally). The term and the concept are not always equivalent even in Polynesia: in Samoa, the notion of "forbidden" is expressed by *sa* and some similar Oceanic words have quite different meanings. *Tabu* in the Trobriands is a kinship term applying among others to a female cross-cousin or other relative appropriate for marriage, while in Dobu it means a magic charm to cause disfigurement or disease. On the other hand, notions very similar to tabu are found widespread ethnographically and expressed by a great range of linguistic expressions, as for example in Rhodesia by *tonda* among the Ila and *bwanga* among the Bemba, or *haram* in the Arabic-speaking world.

The concept of tabu was soon adopted into European languages and, by its suggestion of mystery and of being related to primitive religious ideas, provoked an extensive literature. A major contribution was made by Sir James Frazer in his article on "Taboo" in the 9th edition of the *Encyclopædia Britannica* (1875–89) and in *The Golden Bough* (1890). He regarded tabu as a system of religious prohibitions, attaining its fullest development in Polynesia but traceable under different names in most other parts of the world, and, as something sacred, he contrasted it with the concept of *noa* (a Maori word meaning "profane"). He also regarded tabu as the negative aspect of magic, the rules of prohibition as contrasted with the positive aspect of magic, the charms and formulas for securing success. Frazer gave an elaborate classification of things and acts tabooed and his work was valuable in bringing together a great mass of ethnographic information, in showing the range of the phenomenon and in stimulating enquiry into primitive religion, but his treatment was not sociological and many of his general ideas have not found modern support.

It should be emphasized that the meanings of tabu current in English and other western languages may be very different from those current in any one of the Polynesian or other communities in which the notion has been identified. Tabus have often been spoken of as if they were current for Polynesia as a whole. Frazer comments that "the rule which forbids persons who have been in contact with the dead to touch food with their hands would seem to have been universal in Polynesia." But, while this rule was valid for many Polynesian societies, it did not apply, *e.g.*, to Tikopia, where it was quite proper for someone who had handled a corpse to prepare and partake of food, even without washing his hands. Again, tabu has been treated as if it represented an integral, closely knit system of prohibitions, but this is not so even in any single Polynesian society. The range of items and activities to which *tapu* may be applied need have no consistent general pattern.

Nature and Functions.—One simple usage of *tapu* (in Tonga) is as a conventionalized expression of respect: at the beginning of a public address the speaker may say *Tapu mo hou 'eiki* ("My respects to the chiefs"). In conversation when something has been said which by Tongan ideas of propriety may perhaps be regarded as disrespectful or improper—such as a reference to a part of the human body or to an animal—the expression *Tapu mo koe* (equivalent to "If you will excuse my mentioning such a matter") is interpolated.

To modern Polynesians, *tapu* may mean no more than a thing which must not be interfered with. So a Tikopia man may tell his child not to touch a visitor's camera or fountain pen, saying "It is *tapu*," by which is meant not that the article in question has any particular quality of its own, but that interference with it is not allowed. Similarly in modern Melanesia a board set up on a tree with the word *tambu* on it may be simply a warning to people against going along a path or into an orchard, broadly equivalent to "Trespassers will be prosecuted."

But the notion of tabu is usually much more complex. Generically it includes the idea that the object or action interdicted has some ritual quality, or that the sanction against a breach of the rule is extra-human or supernatural. Very frequently this sanction is regarded as automatic in its action: a breach of tabu is followed logically and necessarily by some kind of trouble to the offender; *e.g.*, lack of success in fishing or hunting, sickness, death of a relative. In ordinary terms this would be regarded as accident or bad luck, but to the people concerned it is regarded as a punishment for breaking the tabu. A person meets with an accident or has no success in his pursuit and, in seeking for its cause, he or others infer that he has in some manner committed a breach of tabu.

With some tabus, however, the sanction is not automatic, but is invoked by the person controlling the tabu, and is believed to consist in the power of a god or spirit to punish the offender. From this point of view, Frazer's equation of tabu with negative magic is inadequate, according to his distinction between magic and religion, for this kind of tabu should come in the religious category. Another theoretical equation of tabu with negative *mana* (*q.v.*) by R. R. Marett, also fails. These two concepts in their Polynesian setting belong to different categories: standing for efficacy and power of a supra-normal kind, *mana* is not necessarily connected with tabu.

In general, tabu has a dual significance. Much attention has been paid to the apparent inconsistency between the notions of holiness or sacredness apparent in many kinds of tabu (the head of a Polynesian chief is tabu as part of his general character

as a sacred leader) and the notions of uncleanness in others (contact with a menstruating woman may be tabu in a defiling sense). But common to both these ideas is respect, associated with the idea of not touching. It is not true, as Frazer thought, that primitive peoples can make no moral distinction between the conceptions of holiness and of pollution, nor is it paradoxical that sacredness and prohibition should attach to the same object. Things which are sacred have some elements of prohibition attaching to them, but all things which are prohibited are not necessarily of sacred quality.

Tabus are often thought to be entirely irrational, even absurd, yet in some societies regular tabus have important social functions; *e.g.*, those against incest. Others have economic effects; *e.g.*, food tabus which restrict the diet of pregnant women or of totemic clans. Some kinds of tabu are not always in operation, but are set up for specific purposes; *e.g.*, those prohibiting fishing or picking fruit at certain seasons may be valuable in food conservation.

Many tabu rules have a social utility and, as a part of the general system of social control, may be a substitute for laws. They may also have importance as a system of prohibitions which forms part of the moral code of the society.

Theories.—There have been many attempts at explanations of the origin of such beliefs and behaviour. One of the most ingenious was that of Sigmund Freud, who associated the concept of tabu with that of neurosis. He believed that tabus were generated in ambivalent social attitudes in which the social components largely replaced the sexual elements of the neurosis. The basis of tabu then is a forbidden action for which there exists a strong inclination in the unconscious. But this ignores the fact that an essential feature of tabu is the obedience of the individual to restrictions imposed by social tradition. Moreover, much of the desire to break through the prohibitions of the tabu may be quite conscious *e.g.*, as, when tabus are placed upon food supplies. Understanding of unconscious factors may be very important—as in an explanation of incest tabu—but for the concept of tabu in its wider range this alone is inadequate.

Social anthropologists have not put forward a completely acceptable theory, but there is broad agreement that the tabus current in any society tend to relate to objects and actions which are significant for the social order. They may do this directly, as in emphasizing the status of chiefs and religious objects, or indirectly, as in the prohibitions on the killing or eating of totemic species which in themselves are of no particular significance but which, as indices of social units, form part of a symbolic framework for the society. It is not always possible to state why a certain action or article has been singled out for such respect behaviour, but a large number of tabus can be understood in such terms.

See MANA; MORALITY, PRIMITIVE; PASSAGE RITES; PURIFICATION.

See also Index references under "Tabu" in the Index volume.

BIBLIOGRAPHY.—J. G. Frazer, *The Golden Bough*, 3rd ed., 12 vol. (1907–15), especially vol. 2, *Taboo and the Perils of the Soul*, and abridged edition, pp. 19–22 and ch. xviii–xxii (1922); W. Robertson-Smith, *Lectures on the Religion of the Semites*, pp. 446–456, 3rd ed. (1927); E. Durkheim, *Les Formes élémentaires de la vie religieuse* (Eng. trans. 1915); S. Freud, *Totem and Taboo* (1918); E. S. C. Handy, *Polynesian Religion* (1927); Raymond Firth, *(Primitive) Economics of the New Zealand Maori*, pp. 246–262, 2nd ed. (1959); F. R. Lehmann, *Die polynesischen Tabusitten* (1930); R. R. Marett, "Tabu," *Encyclopaedia of Religion and Ethics*, ed. by J. Hastings, vol. 12, pp. 181–185 (1921); M. Mead, "Tabu," *Encyclopaedia of the Social Sciences*, vol. vii, pp. 502–505 (1930–35); A. R. Radcliffe-Brown, *Taboo* (1939); H. Webster, *Taboo, A Sociological Study* (1942); W. Howells, *The Heathens*, ch. iii (1949); F. Steiner, *Taboo* (1956); W. A. Lessa and E. Z. Vogt (eds.), *Reader in Comparative Religion: an Anthropological Approach* (1958). (R. F.)

TABULARIUM, the technical name of the record office or national archives building of ancient Rome, which dominates the west end of the Forum. It is built along the side of the Capitoline hill and was bordered, on the side toward the Forum, by two stories of long arcades, the lower decorated with engaged Doric columns, and the upper with engaged Ionic columns. Each bay of the arcade is covered by a square, cloistered or four-sided vault. The tabularium is ascribed to Quintus Lutatius Catulus (*c.* 78 B.C.). It thus offers one of the earliest Roman examples of the use of engaged columns in connection with an arcade.

TABULATING MACHINES are mechanisms controlled by perforated cards which are used as mechanical aids to business methods. The first modern device of this class was designed for the automatic tallying or tabulation of statistics; but with extensive development modern tabulating machines are capable of automatically figuring and printing a report based on the information given to it in the form of perforations in cards or tape. Soon after the introduction of the Jacquard loom, in which the hooks lifting the warp threads were controlled by cards perforated to the desired pattern, Charles Babbage applied the principle to a calculating mechanism. Between 1834–54 he designed and partly built an "analytic engine" in which both the amount to be operated upon and the nature of the operation were entered by perforated cards. His machine was never completed. Development of the modern tabulating machine was accomplished by H. Hollerith.

Hollerith's first contribution was a means of tallying census returns through electrical reading of the data punched on cards and accumulating the totals in separate registers. The system was successfully used in the U.S. census of 1890. During the next decade notable improvements were made including automatic feeding of the cards under brushes for reading and a machine for sorting the cards rapidly into desired groupings. Consequently the 12th U.S. census in 1900 furnished particularly outstanding evidence of the possibilities of the system. The population volume was ready 1 year and 7 months from the start of enumeration; whereas it was estimated that hand tabulation of the 3 factors alone of sex, nativity and occupation would have required the services of 100 clerks for 7 years and 11 months. Since 1901 its commercial utility has been greatly increased. Whenever a transaction involves a number of amounts which are subsequently to be used as units two or more times in compiling various totals, it has been found to be generally economical to "translate" it on to a perforated card. If, for instance, the data connected with the cost of production of a certain part are recorded on a card this can then be used in computing totals of individual wages, department costs, machine costs, productive hours, etc. The tabulating method is now to be found in general use throughout the world.

The card upon which the data are recorded in the form of perforations has a column for each number or symbol to be entered. The digits 0 to 9 are denoted by holes punched proportionate distances from the bottom edge of the card starting with the 9 position. Above the 0 position is space for two additional perforations which are used principally for actuating certain control operations of the machine. These additional positions may also be used for special classification numbers, or, in combination with perforations in the regular digit position, to denote alphabetical characters. The card is divided vertically into "fields" each denoting a particular fact of the total information recorded. Certain information is descriptive while other is quantitative. The former class controls sorting and indicating, while the latter represents amounts to be totalled. Sorting and totalling are accomplished by different machines. Three separate mechanical devices are necessary to the system: a perforating machine or punch, a sorter and a tabulator. The ordinary form of perforating machine has a set of 12 punches under which the card is advanced a column at a time. By means of a skip bar or stops, certain fields can be skipped where no punching is to be done. There are also more complete types of perforating machines arranged to permit automatic duplication of information already on one card on to others, to "gang" punch any predetermined number of cards with identical data, to number cards consecutively as they are punched. The speed with which trained operators can perforate the cards depends on the nature of the information being recorded as well as its extent. An output of 350 punched cards per hour might be considered an average performance.

Sorting the cards into the desired groups for tabulation is done at the rate of about 400 cards per minute for each digit of the class number. One well-known form of sorting machine has 13 pockets, one for each possible columnar perforation and one for "rejects." The tabulator, or accounting machine, which totals the amounts perforated in given fields is capable of several different applications. The simplest form of this device consists of a set of counters actuated through a reading mechanism. Two methods of reading the cards are in use: electrical and mechanical. In the former, brushes make electrical contact through the perforations thereby actuating magnets to trip the counters. The mechanical method employs a full set of pins which are brought down upon the card momentarily at rest and which penetrate the perforations to actuate the counters mechanically. Printing tabulators are arranged to either total the cards at high speed—about 150 per minute—and then print the total, or list each card individually at slower speed followed by a total for the group. By means of special control features the machines automatically take totals of several fields after each sorted group and also store the amounts for grand totals. A great deal of flexibility is attainable in the way the

items and totals may be made to appear on the printed sheet. Mechanism has lately been incorporated accomplishing direct subtraction without the necessity of perforating the complements of the numbers as formerly required. This permits of reports showing total credit and debit amounts and net balance. By using combinations of two or more holes in a column it is possible to list alphabetical characters as well as numbers. With alphabetical printing, addition or subtraction automatically sensed by the machine and devices added to the printing section for handling report and bill forms, tabulating machines have become automatic accounting and billing machines. (B. To.)

TACANAN, a small group of tribes of South American Indians, constituting an independent linguistic stock. The Tacanas, from whom the stock takes its name, and their related tribes, live in northwestern Bolivia, in the region just east of Cuzco, on the upper Madre de Dios river and its tributaries, and extend eastward, between 13° and 15° S. lat., as far as the Beni. They are an agricultural as well as hunting and fishing folk, of the sedentary type, showing, apparently, considerable differences in cultural development between the different tribes. In spite of their close proximity to Cuzco, no significant study of this group has ever been made.

TACHEOMETRY, meaning "rapid measurement," is a method of surveying by rapid and indirect determination of horizontal and vertical distances to a point from an instrument station. The term is generally applied to operations in which distance is obtained by the measurement of a small angle lying opposite a short known base. Tacheometric instruments usually make use of a detached base; but, as in the case of the range finder (*q.v.*), the base line is sometimes an integral part of the instrument. Because of its size, the military range finder is not useful to the surveyor.

The instruments that make use of an external base may be divided into two groups: (1) those that make use of a fixed angle to measure an intercept on a distant staff, and (2) those that measure an angle subtended by a fixed base. In the latter case, the horizontal angle, subtended by a standard "subtense" bar of two metres and oriented at right angles to the line of sight, is measured on the circle of a theodolite. In the former case, the theodolite is equipped with stadia wires that intercept a portion of the length of a graduated rod held vertically at the distant point by two rays that include the fixed angle. A double-image type of tacheometer causes a portion of the image to be displaced with respect to another and the linear displacement is read on a horizontal graduated staff. The tacheometric instruments are generally theodolites or transits modified by the addition of the optical distance measuring devices and implemented by various self-reducing devices for finding the horizontal and vertical components of the slope distance.

See Range Finders; Surveying. (K. S. Cs.)

TACHINID FLY, the name given to any member of the family Tachinidae, of the order Diptera (*q.v.*). The tachinids play an extremely important part in keeping other insects under control and certain kinds have been transplanted from one country to another to help control imported and native pests. Among those imported into the United States is *Compsilura concinnata* for control of the gypsy moth; it attacks a large variety of native caterpillars; *Centeter cinerea* has been introduced from Japan for control of the Japanese beetle; *Microphthalma michiganensis* has been introduced into New Zealand from Canada for control of white grubs; *Archytas cirphis,* from Mexico to Hawaii for control of army worms; *Ptychomyia remota,* from Malaya to Fiji for control of *Levuana* caterpillars in Fiji. Many other successful importations have been made, but the great value of tachinids is found in their attacks upon native insects. They attack chiefly, caterpillars, beetle larvae and adults, sawflies, true bugs and earwigs. Their method of reaching the host is most varied; some scatter over grass thousands of eggs, which are eaten by the host; some lay eggs on the host, often inside it; some deposit living maggots, which attach themselves to passing victims. Those that lay eggs on adult beetles usually have a piercing ovipositor. Some tachinids are associated with army ants, apparently attacking insects fleeing the ants. Many species fly only at dusk, a few are nocturnal; most of them are sun lovers and are found on foliage; many visit flowers and are important in their pollination. The size ranges from three-eighths to three-quarters of an inch in length. While most are dull coloured, or shiny black, many are marked with red and yellow, and some are metallic green or blue. Some authors recognize several families and many subfamilies but there is no good way to separate them. The Dexiidae, with plumose arista, are the most frequently separated group.

The classification is highly technical and many of the species difficult to separate. The literature is voluminous and contained in many publications. The tachinids are one of the largest families of flies, more than 5,000 species being known from all parts of the world, and many new ones are added each year.

Bibliography.—For British species *see* C. J. Wainwright, "The British Tachinidae," in *Trans. Entomological Soc. of London* (1928); for European species *see* P. Stein, "Die Verbreitsten Tachiniden Mitteleuropas nach ihren Gattungen und Arten," in *Archiv fur Naturgesichta,* vol. 90 (1924). *See* also C. H. T. Townsend, *Manual of Myiology,* pt. i–xii (1934–44), containing keys and descriptions of families and genera of the world. (C. H. Cn.)

TÁCHIRA, a state in the Andean region of Venezuela bordering on Colombia. Area 4,286 sq.mi. Pop. (1950) 304,181; (1958 est.) 360,128. Agriculture dominates the economy and coffee and sugar are leading commercial crops. Rural industry, closely related to agriculture, is to a considerable degree carried on in homes. San Cristóbal, the state capital and leading commercial centre is on a trans-Andean highway about 34 mi. W. of the border town of San Antonio. Relatively large deposits of coal are found at Labatera. La Alguitrana claims the distinction of having the first oil well (1878–79) in Venezuela.

Certain inhabitants of Táchira belong to privileged groups who take great pride in the purity of their European ancestry. (J. J. J.)

TACHOMETER, an instrument that indicates the speed of rotation of a shaft or machine element. The major types of tachometers are classed as (1) centrifugal, (2) electrical, (3) vibrating reed and (4) electronic.

Centrifugal tachometers employ a spindle driven by the shaft whose speed is to be measured. As the spindle turns, centrifugal force causes fly-weights that are attached to the spindle to be displaced away from the axis of rotation. These weights are restrained by springs. The position of the fly-weights and the rotational speed of the shaft are interrelated. Through a suitable linkage the position of the weights is calibrated in terms of rotational speed and is indicated by a pointer which is on the instrument scale.

One type of electrical tachometer uses a small permanent-magnet-field direct current generator driven by the shaft whose speed is to be measured. As the voltage output of the generator is proportional to the speed of the shaft, a voltmeter, which can be remotely located, is graduated to indicate speed. Another electrical type uses a small ammeter actuated by the current flowing in a circuit due to the charging and discharging of a condenser. The instrument spindle, which is driven by the shaft, in turn drives a switch to open and close the circuit. The current flowing is proportional to the number of "makes" and "breaks" of the switch and the ammeter is graduated in revolutions per minute.

The vibrating reed type of tachometer uses a series of thin steel reeds with natural frequencies ranging from low to high and corresponding to certain speeds. When an instrument of this type comes into contact with the machine one or more of the reeds will vibrate in sympathy with the machine and indicate the rate of speed. Such a tachometer is not driven by the shaft but responds to vibrations of the machine.

Electronic tachometers are of two types. The strobotac employs the persistency of human vision to detect the change in speed viewed under a flashing light. When the flashes are synchronized with the shafts's rotational speed, the shaft appears stationary. There is no need for mechanical contact with the shaft. This feature is important for devices of very small power output, for contact between spindle and shaft actually slows down such machines. The other electronic-type tachometer is basically a revolution counter that receives its impulse from a pickup (often magnetic), counts the revolutions in each second and displays the count directly as revolutions per minute. (R. A. O.)

TACHYLYTE (TACHYLITE), in petrology the glassy variety of basic igneous rock such as basalt (*q.v.*) or diabase. Tachylytes are black in colour, with a pitchlike or resinous lustre, and in the thinnest sections are typically brown and translucent, the glass being crowded with granules of magnetite. They are found only under conditions which imply rapid cooling, and are much less common than the corresponding acid volcanic glasses, principally on account of the greater fluidity and greater power of crystallization which basic lavas possess. The principal mode of occurrence of tachylyte is as a chilled edge or selvage to thin dikes or sills of basalt or diabase. This edge may be only a fraction of an inch in thickness and merges inward into crystalline basalt. Tachylytes of this nature are common among the Tertiary igneous rocks of the Inner Hebrides of Scotland (Skye and Mull).

Exceptionally, tachylyte constitutes almost entire lava flows, as in the Hawaiian Islands. The rapid cooling of the highly fluid lavas of this region has inhibited crystallization, giving rise to vast floods of basaltic glass containing only minor amounts of crystalline material. Lastly, tachylytes occur as scoria, or bombs, thrown out by basaltic volcanoes. These are well known at Stromboli, Etna and in Iceland. Tachylytes readily undergo weathering and alteration, and are converted by oxidation and hydration into palagonite, a red, brown or yellow cryptocrystalline material. (C. E. T.)

TACITUS, CORNELIUS (*c.* 55–120), Roman historian, lived through the reigns of the emperors Nero, Galba, Otho, Vitellius, Vespasian, Titus, Domitian, Nerva, and Trajan. All we know of his personal history is from allusions to himself in his own works, and from 11 letters addressed to him by his very intimate friend, the younger Pliny. The exact year of his birth is a matter of inference, but it may be approximately fixed near the close of the reign of Claudius. Pliny indeed, though himself born in 61 or 62, speaks of Tacitus and himself as being "much of an age" (Pliny, *Epp.* vii.20), but he must have been some years junior to his friend, who began, he tells us, his official life under Vespasian (*Hist.* i.1), no doubt as quaestor, and presumably tribune or aedile under Titus (80 or 81), at which time he must have been 25 years of age at least. Of his family and birthplace we know nothing certain; we can infer nothing from his name Cornelius, which was then very widely extended; but the fact of his early promotion seems to point to respectable antecedents, and it may be that his father was one Cornelius Tacitus, who had been a procurator in one of the divisions of Gaul, to whom allusion is made by the elder Pliny in his *Natural History* (vii.76). But it is all matter of pure conjecture, as it also is whether his "praenomen" was Publius or Gaius. The most interesting facts about him to us are that he was an eminent pleader at the Roman bar, that he was an eye-witness of the "reign of terror" during the last three years of Domitian, and that he was the son-in-law of Julius Agricola. This honourable connection, which testifies to his high moral character, may very possibly have accelerated his promotion, which he says was begun by Vespasian, augmented by Titus, and still further advanced by Domitian, under whom we find him presiding as praetor at the celebration of the secular games in 88, and a member of one of the old priestly colleges, to which good family was an almost indispensable passport. Next year, it seems, he left Rome and was absent till 93 on some provincial business, and it is possible that in these four years he may have made the acquaintance of Germany and its peoples. His father-in-law died in the year of his return to Rome. In the concluding passage of his *Life of Agricola* he tells us plainly that he witnessed the judicial murders of many of Rome's best citizens from 93 to 96, and that being himself a senator he felt almost a guilty complicity in them. With the emperor Nerva's accession his life became bright and prosperous, and so it continued through the reign of Nerva's successor, Trajan, he himself, in the opening passage of his *Agricola*, describing this as a "singularly blessed time"; but the hideous reign of terror had stamped itself ineffaceably on his soul, and when he sat down to write his *History* he could see little but the darkest side of imperialism. He was an academic republican, like many of those who perished in that reign of terror. They were innocent in many cases of plotting against the emperor's life; and, no doubt, many of the plots were imaginary. But some had been real, and Domitian was determined to take no risks. Hence academic sympathizers with republican ideals fell under suspicion and were involved in the indiscriminate slaughter of those three years.

He was apparently convinced that Tiberius had been a ruler of the same type as Domitian; and that prejudiced him against the great, but grim, successor of Augustus. Moreover, some of the memoirs, contemporary with Tiberius' reign, were not likely to depict him in a favourable light, and on those Tacitus had largely drawn. (For the value of this evidence *see* Furneaux, Tac., Ann. vol. i. p. 20 *ad fin.* For Tacitus' judgment on Tiberius see *ibid.* Introduction ch. viii.) But, apart from personal prejudice, the Romans of the upper classes who lived in Rome or resorted thither tended to dislike imperial rule both socially and politically, and did, as a fact, see so much of the worst of it that they failed to appreciate the blessings it had conferred on the world outside. To his friend, the younger Pliny, we are indebted for the little we know about his later life. He was advanced to the consulship in 97, in succession to a highly distinguished man, Verginius Rufus, on whom he delivered in the senate a funeral eulogy. In 99 he was associated with Pliny in the prosecution of a great political offender, Marius Priscus, under whom the provincials of Africa had suffered grievous wrongs. The prosecution was successful, and both Tacitus and Pliny received a special vote of thanks from the senate for their conduct of the case. It would seem that Tacitus lived to the close of Trajan's reign, as he seems (*Ann.* ii.61; iv.4) to hint at that emperor's extension of the empire by his successful Eastern campaigns from 115 to 117. Whether he outlived Trajan is a matter of conjecture. It is worth noticing that the emperor Tacitus in the 3rd century claimed descent from him and directed that ten copies of his works should be made every year and deposited in the public libraries. He also had a tomb built to his memory, which was destroyed by order of Pope Pius V. in the latter part of the 16th century.

Pliny, as we see clearly from several passages in his letters, had the highest opinion of his friend's ability and worth. He consults him about a school which he thinks of establishing at Comum (Como), his birthplace, and asks him to look out for suitable teachers and professors. And he pays (*Epp.* vii.33.) him the high compliment, "I know that your *Histories* will be immortal, and this makes me the more anxious that my name should appear in them."

The following is a list of Tacitus' remaining works, arranged in their probable chronological order, which may be approximately inferred from internal evidence:—(1) the *Dialogue on Orators*, about 76 or 77; (2) the *Life of Agricola*, 97 or 98; (3) the *Germany*, 98, published probably in 99; (4) the *Histories* (*Historiae*), completed probably by 115 or 116, the last years of Trajan's reign (he must have been at work on them for many years); (5) the *Annals*, his latest work probably, written in part perhaps along with the *Histories*, and completed subsequently to Trajan's reign, which he may very well have outlived.

The *Dialogue on Orators* discusses, in the form of a conversation which Tacitus professes to have heard (as a young man) between some eminent men at the Roman bar, the causes of the decay of eloquence under the empire. There are some interesting remarks in it on the change for the worse that had taken place in the education of Roman lads. The style of the *Dialogue* is far more Ciceronian than that of Tacitus' later work, and critics have attributed it to Quintilian; but its genuineness is now generally accepted. It is noticeable that the mannerisms of Tacitus appear to develop through his lifetime, and are most strongly marked in his latest book, the *Annals*.

The *Life of Agricola*, short as it is, has always been considered an admirable specimen of biography. The great man, with all his grace and dignity, is brought vividly before us, and the sketch we have of the history of Britain under the Romans gives a special interest to this little work.

The *Germany*, the full title of which is "Concerning the geog-

raphy, the manners and customs, and the tribes of Germany," describes with many suggestive hints the general character of the German peoples and dwells particularly on their fierce and independent spirit, which the author evidently felt to be a standing menace to the empire. The geography is its weak point; much of this was no doubt gathered from vague hearsay. Tacitus dwells on the contrast between barbarian freedom and simplicity on the one hand, and the servility and degeneracy of Roman life on the other.

The *Histories,* as originally composed in 12 books, brought the history of the empire from Galba in 69 down to the close of Domitian's reign in 97. The first four books and a small fragment of the fifth, giving us a very minute account of the eventful year of revolution, 69, and the brief reigns of Galba, Otho, and Vitellius, are all that remain to us. In the fragment of the fifth book we have a curious but entirely inaccurate account of the Jewish nation, of their character, customs and religion, from a cultivated Roman's point of view, which we see at once was a strongly prejudiced one.

The *Annals*—a title for which there is no ancient authority, and which there is no reason for supposing Tacitus gave distinctively to the work—record the history of the emperors of the Julian line from Tiberius to Nero, comprising thus a period from A.D. 14 to 68. Of these, nine books have come down to us entire; of books v., xi. and xvi. we have but fragments, and the whole of the reign of Gaius (Caligula), the first six years of Claudius, and the last three years of Nero are wanting. Out of a period of 54 years we thus have the history of 40 years.

The principal mss. of Tacitus are known as the "first" and "second" Medicean—both of the 10th or 11th centuries. The first six books of the *Annals* exist nowhere but in the "first Medicean" ms., and an attempt was made in 1878 to prove that the *Annals* are a forgery, by Poggio Bracciolini, an Italian scholar of the 15th century, but their genuineness is confirmed by their agreement (*see* Introduction to vol. i., of Furneaux's edition of the *Annals* of Tacitus, Clarendon Press series, 1884) in various minute details with coins and inscriptions discovered since that period. Moreover, Ruodolphus, a monk, writing in the 9th century, shows that he is acquainted with a ms. of Tacitus containing at least the first two books. Add to this the testimony of Jerome that Tacitus wrote in 30 books the lives of the Caesars and the evidence of style, and there can be no doubt that in the *Annals* we have a genuine work of Tacitus.

Much of the history of the period described by him, especially of the earlier Caesars, must have been obscure and locked up with the emperor's private papers and memoranda. As we should expect, there was a vast amount of floating gossip, which an historian would have to sift and utilize as best he might. Tacitus, as a man of good social position, no doubt had access to the best information and must have talked matters over with the most eminent men of the day. There were several writers and chroniclers, whom he occasionally cites but not very often; there were memoirs of distinguished persons—those, for example, of the younger Agrippina, of Thrasea, and Helvidius. There were several collections of letters, like those of the younger Pliny; a number, too, of funeral orations; and the "acta senatus" and the "acta populi" or "acta diurna," the first a record of proceedings in the senate, the latter a kind of gazette or journal. Thus there were the materials for history in considerable abundance, and Tacitus was certainly a man who knew how to turn them to good account. He has given us a striking, and on the whole doubtless a true, picture of the empire in the 1st century. The rhetorical tendency which characterizes the "silver age" of Roman literature gives perhaps exaggerated expression to his undoubtedly strong sense of the badness of individual emperors, but he assuredly wrote with a high aim, and we may accept his own account of it (*Ann.* iii., 65): "I regard it as history's highest function to rescue merit from oblivion, and to hold up as a terror to base words and actions the reprobation of posterity." He is convinced of the degeneracy of the age, though it be relieved by the existence of truly noble virtues; and he connects this degeneracy more or less directly with the imperial régime. But it is difficult to dogmatize as to Tacitus' political ideals. He is primarily concerned rather with ethics than with politics; though he may feel that the world is out of joint—with whatever sentimental sympathy he may regard the age of "liberty" and admire the heroic epoch of the republic—yet he appears to realize that the empire is a practical necessity, and to the provinces even a benefit. Like the Stoics, with whom otherwise he has little in common, he censures rather individual rulers than the imperial system. But "the key to the interpretation of Tacitus," it has been well said, (Dill, *Roman Society from Nero to Marcus Aurelius,* Bk.i.ch.1.) "is to regard him as a moralist rather than a politician." Perhaps the strongest work in the *Annals* and *Histories* is the delineation of character.

Tacitus gives us no certain clue to his religious belief. His expressions of opinion about the government of the universe are difficult to reconcile with each other. There seems to have been a strange tinge of superstition about him, and he could not divest himself of some belief (*Ann.* vi. 21,22.) in astrology and revelations of the future through omens and portents, though he held these were often misunderstood and misinterpreted by charlatans and impostors. On the whole he appears to have inclined to the philosophical theory of "necessitarianism," that every man's future is fixed from his birth; but we must not fasten on him any particular theory of the world or of the universe. Sometimes he speaks as a believer in a divine overruling Providence, and we may say confidently that with the Epicurean doctrine he had no sort of sympathy.

Tacitus' style is discussed in the article LATIN LANGUAGE. Whatever judgment may be passed on it, it is certainly that of a man of genius, and cannot fail to make a deep impression on the studious reader. Tacitean brevity has become proverbial, and with this are closely allied an occasional obscurity and a rhetorical affectation which his warmest admirers must admit. He has been compared to Carlyle; and both certainly affect singularity of expression. But they are alike only in the brevity of sentences; and the brevity of Carlyle is not that of an artist in epigram. Tacitus was probably never a popular author; to be understood and appreciated he must be read again and again, or the point of some of his acutest remarks will be quite missed.

Tacitus has been many times translated, in spite of the very great difficulty of the task; the number of the versions of the whole or part is stated as 393.

Murphy's translation (we should call it a paraphrase) was for long one of the best known; it was published early in the 19th century. On this was based the so-called Oxford translation, published by Bohn in a revised edition. Messrs. Church and Brodribb's translation and Professor Ramsay's (1904) (the latter of *Annals* i.–iv.) are much better. Of the many commentaries *see* Ritter's (1864); Nipperdey's (1879); Heräus's (*Histories,* 1885); Furneaux's (*Annals,* i. vi., 1884; xi.–xvi. 1891; *Germania,* 1894); Spooner's (*Histories,* 1891). The last two editors' introductions are particularly useful. By far the best edition of the *Agricola* in English is that of Furneaux as revised by Anderson (Clarendon Press, Oxford). Of works relating to Tacitean Latinity, Draeger's *Syntax und Stil des Tacitus* is the best.
(W. J. Bk.; X.)

TACITUS, MARCUS CLAUDIUS, Roman emperor from Sept. 25, A.D. 275 to April 276, was a native of Interamna (Terni) in Umbria. In the course of his long life he held various civil offices, including that of consul in 273. Six months after the assassination of Aurelian he was chosen by the senate to succeed him, and accepted by the army. During his brief reign he set on foot some domestic reforms, and sought to revive the authority of the senate, but, after a victory over the Goths in Cilicia, he succumbed to hardship and fatigue (or was slain by his own soldiers) at Tyana in Cappadocia. Tacitus, besides being a man of immense wealth (which he bequeathed to the state), had considerable literary culture, and claimed descent from the historian.

See Life by Vopiscus in *Historiae Augustae Scriptores;* also Eutropius, ix. 10; Aurelius Victor, *Caesares,* 36; Zonaras xii. 28; H. Schiller, *Geschichte der römischen Kaiserzeit,* i. 1883.

TACK: *see* NAIL MANUFACTURE.

TACNA, department of Peru, to which by the terms of the 1929 settlement of the Tacna-Arica question it was transferred by Chile. Bounded north by Moquegua and Puno, east by Bolivia,

south by Tarapacá, Chile, west by the Pacific. Area 5,701 sq.mi. Pop. (1958 est.) 51,920. It belongs to the desert region of the Pacific coast. There are a few fertile spots near the mountains, where mountain streams afford irrigation and drinking water. Few of its streams cross the entire width of the department; most are lost in desert sands. The department is part of the former area, Tacna and Arica, which once formed part of the Peruvian department of Moquegua. Its capital is Tacna (pop. 1958 est. 17,230), a small inland town 48 mi. by rail from Arica, in a fertile valley among the foothills of the Andes.

TACNA-ARICA DISPUTE, a controversy involving title to territory that belonged to Peru before the War of the Pacific (*q.v.*). The treaty ending the conflict between Chile and Peru in 1883 provided that Tacna-Arica should be governed by Chile for a period of 10 years and at the end of that time a plebiscite should be taken to decide title to the territory, the party suffering loss of the territory to receive $10,000,000 from the other. Failure to agree on terms of the plebiscite ended in the breaking off of negotiations many times—in 1893, 1901, 1905, 1910 and 1913. In July 1922 the two parties accepted U.S. Pres. Warren G. Harding as arbitrator. After Harding's death, Pres. Calvin Coolidge took the responsibility. In 1925 he recommended that the plebiscite be held and for this purpose named a commission headed by Gen. John J. Pershing. The commission went to the area but failed to work out a solution. Following this failure, the proposal by the U.S. secretary of state, Frank B. Kellogg, that Tacna-Arica be ceded to Bolivia was accepted by Chile but rejected by Peru. In July 1928 Chile and Peru resumed direct negotiations and on Feb. 21, 1929, reached a settlement providing that Tacna should go to Peru and Arica to Chile. Chile agreed to pay Peru $10,000,000 and to carry out certain construction works in the city of Arica. *See* also ARICA; TACNA. (J. J. J.)

TACOMA, a city in western Washington, U.S., a port of entry and the seat of Pierce county, is located on Commencement bay, head of navigation on Puget sound, at the mouth of the Puyallup river, 30 mi. S. by water of Seattle. From Tacoma can be seen Mt. Rainier, 50 mi. E. (*see* MOUNT RAINIER NATIONAL PARK); via the Narrows bridge the city has easy access to the Olympic National park (*q.v.*).

From Commencement bay in 1841 Lt. Charles Wilkes, commander of a U.S. exploring expedition, launched the first survey of the Pacific Northwest. Wilkes commented in his official journal upon the surrounding waterways: "Nothing can exceed the beauty of these waters, and their safety . . . I venture nothing in saying, there is no country in the world that possesses waters equal to these."

Except for the activities of the Hudson's Bay company at Fort Nisqually, there was little in the way of white settlement of the area before 1868. In that year, its original name—Commencement City, after the bay which Wilkes described—was changed to Tacoma, when Gen. Morton Matthew McCarver platted a new city in anticipation of its selection as the western terminus for the Northern Pacific railway. The new name was a corruption of "Tahoma," the Indian name for Mt. Rainier.

Tacoma was incorporated as a city in 1875 under the laws of Washington territory. It operates under a council-manager form of government in effect since 1953. Its population of 147,979 in 1960 was about 95% white; many of its citizens are of Scandinavian heritage, typical of the coastal cities of Washington. Tacoma is the central city of a standard metropolitan statistical area comprising Pierce county, with a population in 1960 of 321,590, a gain of 16.6% in the decade. (For comparative population figures *see* table in WASHINGTON: *Population.*)

Tacoma's principal manufactures include lumber and lumber products, aluminum, copper and copper by-products, ferroalloys, castings and mineral wool. Two large electrochemical plants produce caustic and chlorine. Several candy manufacturers ship nationally and overseas, and the city has a large men's wearing apparel plant. Heavy cranes and logging equipment are manufactured to supply the needs of forest industry. Boatbuilding yards produce fishing boats and pleasure craft as well as watercraft to supply the armed services, marking Tacoma as one of the largest boatbuilding centres on the Pacific coast. Several major armed services installations, including Fort Lewis, McChord air force base and Madigan army hospital, were established nearby.

The University of Puget Sound (Methodist, founded 1888) and Pacific Lutheran university (established 1890) are located in the Tacoma area. The city also serves as headquarters for the Washington State Historical society, whose five-floor museum overlooks Commencement bay. Tacoma's public library system embraces a main library, branch libraries, a library service station and bookmobile service. The city maintains over 35 parks, comprising a total of more than 1,700 ac.; the largest is Point Defiance park, whose grounds contain not only a public yacht basin, a zoo and landscaped gardens but also the restoration of old Fort Nisqually. (B. M. L.)

TACONITE. The name taconite was applied originally to an iron-bearing formation of supposedly Taconic age (Cambrian to Ordovician). Though this was an error the name was associated thereafter with iron-bearing rocks on the Mesabi range in northern Minnesota consisting of chert (fine-grained quartz), magnetite, haematite, several silicates as minnesotaite, greenalite, stilpnomelane and amphibole and siderite. The rocks were originally chemical precipitates which now form beds of a maximum thickness of 750 ft. and a length of about 100 mi. along the strike. Their extent down the dip is unknown. In this taconite are found the high-grade iron-ore bodies which have supplied most of the iron of the U.S. As all taconite contains on an average 27% of chemically combined iron, it is considered the largest potential reserve of ore in the U.S. About 5,000,000,000 to 6,000,000,000 tons of it are near enough to the surface and contain sufficient amounts of magnetite so that its concentration to a high-grade product was in an advanced stage of development by the second half of the 20th century, and shipments of taconite concentrate and agglomerate were being made.

See also MINNESOTA: *Mining and Lumbering.*

See J. W. Gruner, *The Mineralogy and Geology of the Taconites and Iron Ores of the Mesabi Range, Minnesota* (1946). (JN. W. G.)

TACTICS. Tactics is the art and science of fighting battles. It is concerned with the approach to combat, the disposition of troops, the use made of various arms and the execution of movements for attack or defense. It is distinguished from strategy, in a military sense, which may be shaped by political and economic as well as purely military factors. Strategy consists of planning for the war as a whole, the disposition of forces among the theatres of operation and the direction of the campaigns leading up to battle (*see* STRATEGY). Karl von Clausewitz, the 19th-century Prussian military theorist, defined tactics as "the formation and conduct of single combats in themselves" and strategy as "the combination of single combats with each other. . . . In other words, strategy forms the plan of the war, maps out the proposed course of the various campaigns which compose the war and regulates the battles to be fought in each." (*On War*, trans. by J. J. Graham, vol. i, p. 86, Kegan Paul, Trench, Trubner, London, 1911.)

No matter how precisely the two terms may be defined, however, there is seldom any fixed line of demarcation. Military analysts themselves are at odds as to where the one leaves off and the other begins. Strategy is frequently limited by the tactical potentialities of the forces involved, and tactics in turn depend on such strategic considerations as the planning and conduct of campaigns. Actually, the two overlap more often than not. In recognition of this fact, the term "grand tactics" gained currency during the Napoleonic era to denote the transition from strategy to tactics, and 20th-century military commentators have used the phrase "operational strategy" to describe concentrations of such major units as corps and divisions in preparation for battle.

Designations of this sort, unfortunately for clarity, are the exception. As a rule, the student of war must appraise each operation individually and supply his own distinctions.

Following a discussion of the elements of tactics, the principles, tactical formations, attack and offensive and defensive tactics, this article is divided into the following sections:

I. Tactics of the Ancient World
1. The Greek Phalanx

2. The Tactics of Epaminondas
3. The Graeco-Macedonian Tactical System
4. World Conquest
5. The Rise of Rome
6. The Battle of Cannae
7. Tactics of the Cohorts
8. Tactics of the Empire
II. The Cavalry Cycle
1. The Byzantine Horse-Archer
2. The Byzantine Tactical System
3. Tactics of the Franks
4. The Decline of Infantry
5. Medieval Tactics of Europe
6. Longbow and Halberd
7. Early Victories of Gunpowder
III. The Military Renaissance
1. The Fall of Constantinople
2. Pike and Arquebus
3. Maurice of Nassau
4. Gustavus Adolphus and the Thirty Years' War
IV. The Age of Linear Tactics
1. The New Age of Moderation
2. The Wars of Louis XIV
3. The Tactical Theories of Marshal Saxe
4. Frederick the Great
V. Napoleon and 19th-Century Tactics
1. The American Revolution
2. French Revolutionary Tactics
3. Bonaparte's Early Victories
4. Tactics of the Empire
5. Wellington's Tactics
6. Half a Century of Tradition
7. The American Civil War
VI. World War I
1. The Battle of the Marne
2. Deadlock on the Western Front
3. The Offensives of 1917
4. The Mobile Warfare of 1918
VII. World War II
1. The Post-World War I Reaction
2. Early German Victories
3. The War in the Pacific
4. U.S. Amphibious Victories
5. Wide Variety of World War II Tactics
VIII. The Korean War
1. Intervention in Korea
2. Tactical Innovations of Korea
3. Tactics of the Atomic Age

Opposite Poles of Tactics.—The elements of tactics are dual in character. Attack and defense, cohesion and dispersion, protection and mobility, missile and shock effect, moral and material values—these elements have been changeless throughout the ages, regardless of variations in arms and communications. The barbarian's bow and arrow was as legitimate a missile weapon as the machine gun, and elephants were esteemed for shock effect many centuries before the invention of the tank.

Seldom do armies decide arbitrarily on the tactics they will adopt. The swing from one pole to another is nearly always influenced by geographical, climatic and economic factors. Nomads of the plains become cavalry warriors from necessity rather than choice, just as mountaineers fight on foot. Emerging tribes and nations, unsure of their military skill, were drawn toward the cohesion of the human wall, or phalanx. As such peoples grew in strength and confidence they tended to adopt formations of dispersion, which are better suited to aggression. Thus the Romans of the early city-state fought in an eight-rank phalanx, but the Romans of the conquering republic developed that marvelously flexible tactical instrument known as the legion (*q.v.*).

One of the most cherished delusions of the ancient world was the belief that a formation had a shock effect in proportion to its depth. But adding ranks to the phalanx did not add to the material might of the attack; it merely contributed to the moral impetus of the attackers. Shock effect is rarely the product of an actual physical impact. The shock is communicated to the soldier's mind rather than his body, and during the cavalry age he was morally defeated by his dread of being trampled, though horses seldom ran down foot soldiers who stood their ground.

Principles of Tactics.—Basic tactical principles are few and simple. The complexity and difficulty appear when such principles as surprise, secrecy, deception, security and the gathering of intelligence are put into actual practice.

So obvious are the tactical advantages to be gained by deception and surprise that they are sought instinctively by savages. Both principles are combined in that most primitive of tactics, the ambush. Not only savages but even wild animals know enough to hide while lying in wait to attack. Yet the ambush is one of the most effective tactics of 20th-century warfare; it was used by both British and German armoured columns in the complicated North African desert campaigns of World War II.

It is hardly necessary to add that surprise depends on the degree of secrecy with which a general masks his intentions and capabilities. The moral turning point of the American Revolution came in 1776 when George Washington crossed the ice-choked Delaware river with a supposedly beaten army and killed or captured nearly all troops of three German mercenary regiments sleeping off a Christmas celebration at Trenton. This surprise would not have been possible without the American general's accurate intelligence of the enemy's vulnerability. Nor could it have occurred if the German officers had enforced security precautions in keeping with the defense of an outpost. A famous example of tactical deception was provided a week later when Washington cheated the British of an almost certain victory. Trenton was the locale again as Lord Cornwallis brought his adversary to bay with a materially superior army. A British attack had been ordered for the following morning, but the Americans left their campfires burning and slipped away in the darkness. Cornwallis' first intimation of the escape came at daybreak when he heard cannonading ten miles in his rear as Washington defeated a detachment of British reinforcements in the small battle of Princeton.

So essential is intelligence to the tactician that a good spy is sometimes worth more to an army than a good general. Reconnaissance patrols, prisoner interrogations and observation or listening posts also play an important part in arriving at an estimate of an enemy's intentions and capabilities.

Security, like secrecy, is maintained in all military ages by means of eternal and unremitting vigilance. Although the firepower of missile weapons has often been given the sole credit for the prolonged stalemate in World War I, violation of age-old principles of tactics cannot be overlooked. Several promising Allied offensives met failure because surprise and deception were sacrificed to make possible the tremendous artillery "preparations" which had no decisive effect on an enemy burrowed deep into the earth. Such large quantities of shells were required for these bombardments that the area and extent of the impending offensive were revealed to the enemy by aerial photographs of ammunition dumps.

The plain lesson of history is that improvements in weapons and tactics cannot compensate for flaunting basic tactical principles.

Tactical Formations.—Far back in the racial past the most intelligent of our primitive ancestors perceived that superior numbers did not always make for superior force. A great deal depended on the tactical direction and speed with which this force was applied, and the solution was to group fighting men so that they could give one another the most support while doing the most harm to the enemy. This was the beginning of tactical wisdom.

The two basic formations are the line and the column, which differ only by 90° of the compass. In other words, a column of men in four files, facing north, has only to execute a "right face" to become a line of four ranks, facing east.

Neither of these formations is useful for all tactical purposes. Although the column is best adapted to marching and maneuver, its narrow front and vulnerable flanks are liabilities in combat. The line lacks depth and mobility, but it is more suitable for deployment on the battlefield. Mention should also be made of two other formations which appear less frequently in the chronicles of war—the hollow square and its counterpart, the 360° perimeter. Some stout stands have been made by men facing outward in all directions, but the defect of such formations is that ability to maneuver is sacrificed.

In the prehistoric past it was evident that forming warriors into a line or column was not enough. Some attention also had to be paid to problems of deploying them so as to make the best

of both shock and missile weapons. When men learned to fight on horseback, a cavalry arm came into being. The "eyes" of the primitive army, when used for scouting and reconnaissance, horsemen were also effective for pursuing a broken foe. A crude artillery arm appeared as early as the 8th century B.C., when the Old Testament dates engines of war (*q.v.*). Not even the tactical ancestor of the tank was lacking in the ancient world. Armoured chariots manned by archers and spearmen led the infantry attack in the battles fought on the plains of Mesopotamia, as depicted by bas-relief friezes dating back many centuries before the Christian era.

The Tactics of Attack.—Descriptions of battles are apt to give the impression that victories are won entirely by physical pressure, as measured by the casualties inflicted. It has been the experience of history, on the contrary, that moral forces play a much larger part in the outcome than material means. An army's combative spirit is not often broken because the ranks have been so thinned that the survivors are crushed by sheer human tonnage. They are beaten because of their conviction that further resistance is hopeless, and it is the business of the tactician to convince rather than crush his opponents.

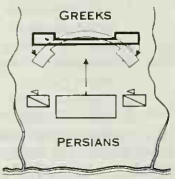


FIG. 1.—BATTLE OF MARATHON (490 B.C.), SHOWING TACTICS OF CONCENTRIC ATTACK

The test may be found in the fact that casualties of 30% are usually the most an army can endure without losing combative spirit. Yet seven men out of ten are still able to fight, and it may be that they outnumber the entire force of adversaries who accepted battle with a numerical disadvantage. History records instances of losers being routed while still possessing a two-to-one numerical superiority. They were not crushed; they were simply convinced.

The methods of creating a conviction of defeat may be reduced to a few plans of attack which repeat themselves throughout history. Most simple of all is the plain frontal assault all along the line. Justifiable and even desirable under some circumstances, it is as a rule the resort of unimaginative generalship. As for the direct rear attack, it is so rare as to be associated only with ambush or pursuit.

At an early age it was discovered that armies, like the individuals composing them, dread an attack from the flank or rear. An assault from one side or another is not enough, since an adversary can face in the new direction, whereupon the effort becomes a mere frontal attack. In order to deliver a true flank attack, part of the offensive force must fix the enemy's front, perhaps even defensively, while the assault force swings around to strike simultaneously from the side. If this force can make a wide enough swing to threaten an opponent's rear, the attack becomes a single envelopment. And in instances when the rear is assailed from both sides with concentric thrusts, a double envelopment is the result. Neither a flank attack nor envelopments may be possible when a general secures his flanks with natural obstacles. In that case it is sometimes possible to create two new flanks by rupturing an adversary's front and driving through the gap to strike at his rear. That is the battle of penetration and it has some distinguished victories to its credit.

In any event, the effectiveness of the attack is in direct proportion to the conviction of defeat it can instill. As evidence, history is filled with occasions when the casualties of a beaten army were the result rather than the cause of defeat.

Offensive and Defensive Tactics.—The relative merits of the offensive and defensive have been subjects of argument in all military ages. From one century to another each has dominated in turn, the new cycle being brought about by the introduction of new weapons or tactics.

The attacker is usually able to choose the time of the assault and sometimes the place as well. If there are several points where the blow might fall, he has an opportunity to utilize the principles of deception and surprise against an opponent who must prepare as best he can to meet every alternative.

The defender often has the advantage of choosing and fortifying his own ground, but at least temporarily he must concede the initiative to the enemy. In most instances, of course, he expects to regain it by means of a decisive counterattack after exhausting his adversary or placing him in an unfavourable position. Passive defenses are seldom the resort of a general who retains freedom of action, though there are exceptional occasions when the object is to inflict crippling casualties on a rash enemy.

Cordon defenses are put into effect along a frontier where troops can be thinly dispersed and quickly reinforced by reserves in the event of trouble. The mobile, or elastic, defense is based on the theory that if the enemy makes a penetration his flanks will be open to counterattack. As for the so-called defensive-offensive action, it merely implies that the defender has planned from the beginning to launch a counterstroke when he has an opportunity.

Breaking off contact with the enemy is a delicate operation calling for that specialized form of the defense known as a delaying action. The rear guard force holds just long enough to compel the pursuer to deploy, then pulls back to the next prepared position before becoming seriously engaged. In a two-line delaying action the first and second lines retire through each other in turn until the main body of troops is out of danger.

The ideal instrument would be an army equally formidable in offensive and defensive tactics, but there have been few such in military chronicles. Probably the most effective combination throughout history has been the strategic offensive and tactical defensive. Invading armies are often able to choose their ground and await an attack, since the enemy is obliged for the sake of national morale to defend the homeland. No better example of this compulsion could be found than the situation faced by Roman generals after Hannibal's crossing of the Alps into Italy (*see* HANNIBAL). One after another, they met him in headlong attack, only to go down to crushing defeat. Finally the proud republic was reduced to avoiding battle in order to recuperate.

Whenever the balance between offensive and defensive tactics is tipped too far in either direction, the social progress of the age is affected. In the 13th century the Mongol horsemen who overran Asia left desolation and anarchy in the wake of an irresistible offensive. Meanwhile, with fortification holding the advantage over siegecraft in western Europe, the consequences of an all-powerful defensive were political and economic stagnation.

Until the 19th century, offensive and defensive cycles were of long duration. Improvements in weapons have come so rapidly since the Napoleonic era that the swing from one to another is a matter of only a few years. Generals of all armies were confident at the outbreak of World War I that the offensive power of the arms of 1914 would prevail in a few months. But missile weapons favour the defensive, as a rule, and the machine gun had the leading part in turning the conflict into a stalemate. Not until the last few months of the war did the tank provide the victors with an offensive weapon capable of breaking the deadlock. Even so, a school of military thought held after the war that the defensive remained powerful enough to curb an aggressor in the next conflict. Thus, it was a sad disillusionment when the German armies of 1940 drove on to a succession of early victories.

I. TACTICS OF THE ANCIENT WORLD

1. The Greek Phalanx.—Two great tactical systems, the Graeco-Macedonian and the Roman, dominated the ancient world (*see* ARMY). Citizen-soldiers of the Greek city-states fought in an eight-rank phalanx of heavy infantry supported by contingents of light foot. Although the *Iliad* mentions horse and chariots a millennium before the Christian era, cavalry was not an arm suited to a mountainous peninsula, and the incessant strife of the city-states was usually decided by the clash of opposing hedges of infantry spears.

Greek warfare was distinguished more for brawn and fortitude than the intellectual brilliance which made for greatness in other directions. Even the double envelopment which won the battle of Marathon (*see* fig. 1) for the Greeks in 490 B.C. seems to have been the product of a fortuitous accident rather than generalship.

Attacking at a fast pace, before the Persian superiority in cavalry and missile weapons could become effective, the Greek *hoplites* (heavily armed infantrymen) of both wings hurled back the lightly armed invaders in their path. Meanwhile, their own hard-pressed centre gave ground before the enemy advance until it seemed that defeat was imminent. As it proved, this withdrawal created the opportunity for the heavy infantry of the Greek wings to converge on the flanks and rear of the Persian host. Thus the battle ended with the demoralization and rout of an army which probably outnumbered the Greeks at least two to one.

2. The Tactics of Epaminondas.—Defending their homeland against the Persians, the Greeks learned to place increased reliance on cavalry and missile weapons—mercenary horsemen from the plains of Thessaly and bowmen from Crete. The first Greek to distinguish himself as a tactician was Epaminondas (*q.v.*) in 371 B.C. Facing a Spartan army of 10,000 men with a force of 4,000 which included undependable allies, the Theban commander drew up a strange oblique order of battle by advancing his left wing and reinforcing it with his best troops to a depth of 48 ranks (*see* fig. 2). On the other wing he placed his doubtful allies. The battle of Leuctra opened with a cavalry clash won by the better-trained Theban horse. Then Epaminondas bore down with the infantry of his heavy left wing on the right of the Spartan phalanx, and the moral factor of surprise probably had more to do with the outcome than physical shock. At any rate, the Spartan right recoiled in disorder on its own centre, and a phalanx thrown into confusion was little better than a mob. The muscular tactics of Sparta were of no avail to demoralized men too crowded together to use hand weapons, and the Thebans gained a one-sided victory.

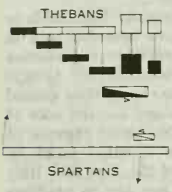


FIG. 2.—BATTLE OF LEUCTRA (371 B.C.), SHOWING TACTICS OF THE OBLIQUE ORDER

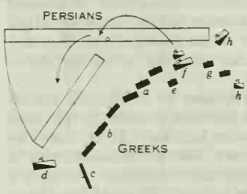
Epaminondas received a mortal wound while winning a second battle with his oblique order, and it remained for a barbarian prince to assemble the tactics of the disputatious Greeks into a system capable of conquering the world. Young Philip of Macedon was an attentive student of Epaminondas' methods while spending three years in Thebes as a hostage (*see* PHILIP II [of Macedonia]). He returned to his northern kingdom to organize the first army of ancient history to combine effectively such arms as infantry, cavalry and artillery.

3. The Graeco-Macedonian Tactical System.—The new tactical system was built about the phalanx. Philip strengthened the mobile human fortress by doubling the eight-rank formation of the Greeks and arming his warriors with a longer and heavier spear. He perceived that the two chief weaknesses of the phalanx were the vulnerability of its flanks and the inability to maneuver or even retain its solidarity on rough ground. As a tactical remedy for both defects, he organized and trained a light infantry such as the ancient world had never known before—a disciplined force of archers, slingers and javelin men. Their function was to act as tactical hinges on both sides of the phalanx, lending mobility as well as protection. Since the young nobles of Macedonia preferred to fight on horseback, Philip had no difficulty in organizing a heavy cavalry arm far superior to the establishments of the Greek city-states. Squadrons of light horse were also employed for scouting and skirmishing.

That the army had an efficient baggage train, medical corps and engineering section is evident from its accomplishments. In his development of "artillery," moreover, the Macedonian king went far beyond his predecessors who had found war engines useful only for siegecraft (*see* ENGINES OF WAR). He originated the idea of carrying only the essential parts on the march and depending on his engineers to cut timbers for the framework whenever needed.

On the battlefield the full phalanx of about 8,000 heavily armoured spearmen was doubtless a fearsome spectacle as it advanced with menacing steel points. Several such bodies of troops might be employed as the central bulwark of the battle line, the pivot either of offense or defense. On the wings were the heavy cavalry squadrons trained for shock effect. Between the phalangial centre and cavalry wings, the tactical hinges of light foot supplied the flexibility for maneuver.

4. World Conquest.—This was the army which Philip led to the conquest of Greece and his son Alexander (*q.v.*) made the instrument of world conquest. The Graeco-Macedonian tactical system met its greatest battlefield test on the field of Arbela, or Gaugamela, in 331 B.C. Darius III, the Persian monarch, had a numerical advantage of perhaps five to one over Alexander's 7,000 cavalry and 40,000 foot. The action took place near the ruins of ancient Nineveh on a plain suited to the Persian cavalry and chariots (*see* fig. 3). Despite the odds against him, Alexander rejected the plan of a night attack urged by his nobles. He opened the battle by seizing the initiative with a rapid advance of his right wing, followed in echelon by the rest of the army. The local superiority in numbers enabled him to drive so deeply between the enemy's left and centre as to reach the rear. A more perfect example of the battle of penetration can scarcely be found in the pages of history, for the Persian host was routed while still possessing a great numerical superiority.



a RIGHT OF PHALANX
b LEFT OF PHALANX
c LIGHT INFANTRY
d CAVALRY
e HYPASPISTS
f HEAVY CAVALRY
g LIGHT INFANTRY
h LIGHT CAVALRY

FIG. 3.—BATTLE OF ARBELA, OR GAUGAMELA, (331 B.C.), SHOWING TACTICS OF REAR ATTACK

Alexander's campaigns took him as far as modern India and Afghanistan during the six years after his conquest of Persia. Among his most troublesome foes were the Scythian horsemen who preyed upon his flanks and rear, seeking to cut off detachments. That the nomadic riders of the Asiatic plains could be formidable opponents was demonstrated in later centuries when no less than four Roman armies were lured to their destruction.

The Scythians and Parthians were made natural warriors by a harsh environment. Shunning action at close quarters, resorting to flight when pressed, then returning to the attack in swirling envelopments, they were able to discharge arrows while riding at top speed.

Alexander refused to be drawn into extended pursuits by feigned flights. Noting that the indiscipline of the Scythians made them reckless at times, he trained his light infantry to launch surprise counterattacks and beat the nomads at their own game by cutting off detachments. And once the lightly armed Scythians were attacked at close quarters, they were no match for Alexander's cavalry.

5. The Rise of Rome.—The conqueror's death was followed by the rapid crumbling of the great empire he had founded. But while his former generals disputed for mastery, a young and vigorous war power was rising in the west. Fighting in a simple phalanx of heavy infantry at first, the Romans learned by defeat that the human wall is not suited to battle on rough ground.

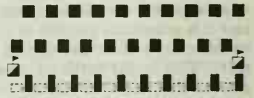


FIG. 4.—LEGION IN BATTLE ORDER: FIRST TWO LINES, HEAVY INFANTRY; THIRD LINE, CAVALRY AND LIGHT INFANTRY ALTERNATED WITH RESERVES

The new tactical system which gradually evolved was the legion (*see* fig. 4). Under the republic it consisted of small and supple infantry units known as maniples, each numbering 120 men in 12 files and 10 ranks. In a typical battle order the first two lines, drawn up in chessboard array, were composed of 10 maniples each. An interval equal to a maniple's front of 60 ft. separated the units, enabling the second line to merge with the first or the first to fall back into the second. The advantages of mobility could thus be exchanged for those of solidarity, since the two lines combined had a depth of ten ranks

and a front of 1,200 ft.

In the third line ten units of light foot were alternated with a like number of reserve units, the former numbering 120 men each and the latter 60. The three lines were 250 ft. apart, and from front to rear one maniple of each classification formed a cohort. This was the Roman equivalent of a battalion, numbering 420 men. Ten cohorts made up the infantry strength of a legion but 20 were customarily combined with a small cavalry force and other supporting units into a self-sustaining little army of about 10,000 men.

The two famous Roman infantry weapons were the *pilum* and *gladius*—the first a 7-ft. javelin used for both throwing and thrusting, and the latter a 20-in. cut-and-thrust sword with a broad, heavy blade. For protection the legionary had a metal helmet, cuirass and convex shield, but aggressive action was deemed a Roman's best defense. The first line attacked on the double, hurling javelins and diving in with swords before the enemy had time to recover. Then came the maniples of the second line, and only a resolute foe could rally from the two successive shocks.

FIG. 5.—BATTLE OF CANNAE (216 B.C.), SHOWING TACTICS OF REAR ATTACK

6. The Battle of Cannae.—When Hannibal descended from the Alps in 218 B.C. with an army of Iberian, Numidian and Gallic mercenaries, the republic paid the penalty of placing too much reliance on heavy infantry and failing to develop a cavalry arm and missile weapons. Two Roman armies, destroyed in the early encounters, were outmaneuvered and cut to pieces by the Carthaginian horse. Finally, Rome incurred a disaster which has made the name of Cannae a symbol ever since of the battle of annihilation and double envelopment (*see* fig. 5).

Hannibal provoked an action, though much outnumbered by a Roman army of 85,000 men. Adopting a convex formation of his foot in the centre, he gradually withdrew it in a feigned retreat until the Roman infantry rushed in recklessly, raising the victory shout. When the converging legionaries were too crowded together to use their weapons, the Carthaginian closed the trap by assailing the flanks and rear of the Roman infantry with his cavalry after driving the inferior enemy horse from the field. The result was a prolonged slaughter of helpless men which lasted until only about 15,000 scattered fugitives survived Rome's greatest disaster.

The victor managed to maintain himself for 16 years in a hostile country before the republic developed in Scipio Africanus a general who understood that a tactical system based on a single arm is as handicapped as a one-armed man would be in personal combat. He carried the war to Carthaginian territory in Spain and finally invaded Carthage. Hannibal returned in the autumn of 203 B.C. to defend his own country but was defeated on the field of Zama (202 B.C.), where the superior Roman cavalry delivered the *coup de grâce*.

7. Tactics of the Cohorts.—The patriotic citizen soldier of the republic's early years was replaced during the century of civil war by a professional soldiery serving for pay and donatives. The cohort of an average field strength of 360 men gradually replaced the maniple as the tactical unit of increasingly larger Roman armies. Ten cohorts made up a legion in the operations of Sulla and Caesar. The 3,600 heavy infantry were supported by enough horse and light infantry to bring the numbers up to 6,000 in all.

The maniple survived only as a unit of a cohort of eight or ten ranks. Four cohorts in the first line and three each in the second and third lines were drawn up in chessboard array, with intervals between units so that one line could merge with another (*see* fig. 6).

Seven legions in three lines, comprising about 25,000 heavy infantry, occupied a mile and a half of front. This was an average of about 11 men to a metre of front as compared to 28 in Alexander's army.

The practical good sense which distinguished Roman tactics was at its best in field fortification. No greater demonstration was ever seen in the ancient world than the 50 mi. of works with which Caesar gained the submission of Vercingetorix and 80,000 Gauls at Alesia. The Romans began by surrounding the strong natural defenses with a system of walls and trenches. When Gallic allies marched to the rescue of his opponent, Caesar surrounded his own army with works of circumvallation and beat off foes both in front and rear. Finally the relieving army disintegrated and Vercingetorix was starved into surrender.

8. Tactics of the Empire.—A gradual change in tactics was evident as Rome evolved from a conquering into a defending power. The spade and axe continued to be effective weapons of a cordon defense established along the frontiers, utilizing such natural barriers as the Rhine and Danube. Where they were lacking, the Romans built such works as the great walls in Britain and the wooden palisades (*Limes*) of Germany.

The cohort, increased to a field strength of 500 to 600 men, remained the unit of tactics which still trusted to the shock of *pilum* and *gladius*. But the 5,000 to 6,000 heavy infantry of the legion were combined with an equal number of supporting troops—cavalry, both heavy and light, and a light infantry made up of archers, slingers and javelin men.

During the first two centuries of the empire the proportion of mounted troops rose from one-seventh to one-fourth as the legion dealt with mounted barbarian raiders. Where war engines had been used chiefly in siegecraft, the empire relied upon them for the defense of fortified border outposts. By the 4th century A.D. according to Vegetius, as many as 10 catapults and 60 *ballistae* were assigned to each legion—a ratio of a "gun" or "howitzer" to 1,000 infantry. This is a higher proportion than the average of three pieces of artillery to 1,000 men in Napoleon's armies.

II. THE CAVALRY CYCLE

1. The Byzantine Horse-Archer.—After the fall of Rome the eastern or Byzantine empire managed to survive by means of bribing the most formidable barbarians and inducing them to move on to the occupation of Italy. The respite enabled Constantinople to develop during the reign of Justinian (A.D. 527–565) a cavalry tactical system which won victories on three continents.

This system was built about the horse-archer. The invention of the stirrup in the 6th century had given an impetus to cavalry warfare, and the historian Procopius has described the tactics of the mounted mercenaries whom Belisarius led to conquests in Persia, North Africa and Italy.

FIG. 6.—LEGION IN BATTLE ORDER: FOUR COHORTS IN THE FIRST LINE AND THREE EACH IN THE SECOND AND THIRD LINES WITH CAVALRY

"But the horsemen of the present time go into battle wearing corselets and fitted out with greaves which extend up to the knee. From the right side hang their arrows, from the other the sword. And there are some who have a spear also attached to them and, at the shoulders, a sort of small shield without a grip, such as to cover the region of the face and neck. They are expert horsemen, and are able without difficulty to direct their bows to either side while riding at full speed, and to shoot an opponent whether in pursuit or flight. They draw the bowstring along by the forehead about opposite the right ear, thereby charging an arrow with such impetus as to kill whoever stands in the way, shield and corselet alike having no power to check its force. Still, there are those who take into consideration none of these things, who reverence and worship the ancient times, and give no credit to modern improvements." (Procopius, *Works*, trans. by H. B. Dewing [Loeb Classical Library], vol. i, p. 7, Harvard University Press, Cambridge, Mass., 1914.)

Belisarius, landing in North Africa with 15,000 men, made easy work of the kingdom founded by the Vandals on the site of ancient Carthage. When he defeated the Ostrogoths in Italy and took Rome, another king was sent as a prisoner to join the captive Vandal monarch at Constantinople. It appeared for a time that Justinian's generals were about to restore the ancient grandeur of Rome, but the conquests were soon lost by an eastern empire ex-

posed to the onslaughts of Slav, Avar, Bulgarian, Persian, Saracen, Frank, Russian and Turk during five centuries of continual warfare.

2. The Byzantine Tactical System.—Byzantine tactics were eminently practical, the spirit of chivalry being unknown in an empire constantly fighting for existence. Greek in speech as well as tradition, Constantinople reduced war to a science with a series of treatises and manuals. The *Tactica* of Emperor Leo the Wise, written about 900, urged that operations should be varied to take advantage of each foe's peculiar weaknesses. Fraud and feigned flights were suggested without apology. If a foe could be bribed more cheaply than defeated, eastern Roman emperors deemed it no disgrace to pay tribute.

Since most of Constantinople's foes were mounted, the tactical system placed the emphasis on heavy cavalry, both archers and lancers. But even though the infantry ranked as a secondary arm, it was well-trained and combined with mounted troops in proportions suited to tactical demands. Squadrons of light horse were employed for scouting and reconnaissance.

Byzantine armies stood between the west and Asiatic invaders at a time when the impoverished Europe of the dark ages would probably have been overrun. Charles Martel and his Franks had difficulty in repulsing a single large-scale Moorish plunder raid on the field of Tours in 732. Yet Constantinople held its own in an 83-year war with the Saracens when Islam was at the height of its conquering momentum.

At the outset the fierce Moslems were cavalry warriors using much the same tactics as the Parthians and Scythians. Where they differed from other plains horsemen was in the ability to learn from their foes. After experience taught the need, the Saracens organized a heavy cavalry of their own and adopted many of the east-Roman tactics. Most astonishing of all, these former desert dwellers took to the sea and built ships large enough to launch two great land and water invasions which were repulsed outside the walls of Constantinople. Only that secret Byzantine weapon known as Greek fire (*q.v.*) saved the imperial city in 673.

3. Tactics of the Franks.—It was fortunate that Europe was spared a severe test, for as late as the 8th century the Franks could raise only masses of untrained and largely unarmoured men using the most primitive tactics. Even after Charlemagne (*q.v.*) introduced strenuous military reforms, he was not able until the final edict of his long reign (771–814) to forbid any man appearing for duty when armed only with a club.

The low esteem in which the Franks were held by skilled warriors is evident from Leo's *Tactica*. Admitting that the headlong charge of mounted Franks with lance and broadsword could be formidable, he advised that "it is best to decline a pitched battle with them until you have put all the chances on your side. You should take advantage of their indiscipline and disorder; whether fighting on foot or on horseback, they charge in dense, unwieldy masses which cannot manoeuvre because they have neither organization nor drill. . . . Hence they fall readily into confusion if suddenly attacked in front and rear—a thing easy to accomplish, since they are utterly careless and neglect the use of pickets and vedettes." (C. M. C. Oman, *A History of the Art of War in the Middle Ages*, vol. ii, p. 204. By permission of Methuen & Co., Ltd., London, 1924.)

Charles the Great made an unceasing effort to create a disciplined army of mailed cavalry out of the Frankish host of unarmoured foot. Ownership of land was the basis of military duty. Proprietors of large estates were commanded to send a mounted warrior equipped with a mail shirt, while masters of smaller tracts must at least combine to arm a man with spear and shield. Thus came into being the feudal system which was to shape the warfare of Europe during the next several centuries.

Despite the emperor's reforms, Europe was helpless for long against the plunder raids of the Vikings. Both the Greeks and Romans had depended on marines trained for shipboard duty and amphibious combat landings, and Constantinople had carried on the tradition. The Norsemen in their turn were unsurpassed as amphibious warriors. Their single-deck ships with carved prows and shield-hung bulwarks were fast and seaworthy, and experience had taught them effective tactics of seizing a beachhead after pushing up such rivers as the Loire and Meuse. The Vikings were formidable fighting men with axe and sword, and their forays grew so bold during the last half of the 9th century that they sacked London and laid siege to Paris.

There was not time to summon large armed forces from a distance against these fleet raiders. Each district had to fend for itself and the feudal system burgeoned as Europeans paid military duty to the local lord upon whom they relied for protection. Cavalry tactics were also given another impetus, for only mailed horsemen could cope with the Norsemen. Seldom has the social progress of an age been so largely shaped by tactical factors.

4. The Decline of Infantry.—In 1066, at the battle of Hastings, the last infantry host of western Europe went down to defeat in England before the descendants of Norsemen who had settled in France and learned to fight on horseback. The mounted and armoured men-at-arms had become virtually the tactical unit of artless battles which had as their first object the seizing of opponents to hold for ransom. Serfs who fought on foot were held in such low esteem that no count was taken of their numbers or losses.

Except for a few crossbowmen, Europe's armies of the crusades had no infantry worthy of the name. A similar situation prevailed in Asia, where the Mongols led by Jenghiz Khan built up a formidable cavalry system.

The bow, lance and scimitar were the weapons of fierce horsemen who used two varieties of bows and three "calibres" of arrows for various tactical purposes.

The Great Wall of China, with its 1,500 mi. of earth-filled brick or granite, proved to be no serious obstacle to invaders who swept like a pestilence from the Yangtze to the Vistula during the 13th century without incurring a single major defeat. Whether the conquerors could have continued to win victories in the mountains and forests of western Europe is open to speculation, for they returned to Asia after defeating the feudal horsemen of Hungary and Poland.

5. Medieval Tactics of Europe.—In contrast to the supremacy of the offensive in Asia, the defensive was all-powerful in medieval Europe. Concentric fortification added to the difficulties of besiegers by making it necessary to attack walls within walls as well as outer wards and foreworks, while the crossbow provided the garrison with a means of enfilade fire from rounded keeps.

Meanwhile, the gradual transition from chain to plate armour made the mailed knight and his barded charger more invulnerable to the weapons of the day (*see* ARMOUR, BODY).

The wretched communications of medieval Europe were a deterrent to maneuver even if the unruly armed forces of the 13th century had been disciplined enough to attempt it. It was seldom possible to force a battle on an unwilling opponent, and methods of supply were so primitive that an army often had to abandon a half-finished campaign after eating a province bare.

That the tactical superiority of the knight owed a great deal to his monopoly of the prerogative of bearing arms was demonstrated by despised Flemish burghers at the battle of Courtrai in 1302. French men-at-arms, putting down a revolt, launched their usual thundering charge only to bog down in disorder upon reaching marshy ground they had not troubled to reconnoitre. The Flemish foot clubbed the knights out of the saddle and dispatched them as they lay helpless in the mud, too burdened with armour to defend themselves. This was a disagreeable ending for a medieval battle as compared to being held for ransom.

6. Longbow and Halberd.—The use of gunpowder in the 13th century made it inevitable that the supremacy of the man of iron in his house of stone could not be long maintained. But the first handguns were not as effective as either the crossbow or English longbow, and it was actually the rise of disciplined forces of yeomen using infantry tactics which restored the balance between the offensive and defensive in medieval Europe.

Although a few weapons using gunpowder were present on the field of Crécy (*q.v.*) in 1346, the outcome owed more to the discipline and tactics of an English army in which every man re-

ceived a fixed and regular wage. For it was truly a national army that Edward III led to France, and the French knights would probably have been defeated even if the invaders had been armed with a less effective weapon than the 6-ft. longbow. Its superiority over the crossbow of the Genoese mercenaries was proved in a preliminary exchange of missiles. Edward had placed his dismounted men-at-arms in the centre as spearmen, and on both projecting wings the lightly armoured archers stood behind ditches and pointed stakes. They poured deadly flights of arrows into the front and flanks of the advancing French knights until 16 successive charges had been repulsed with heavy losses.

Another infantry tactical system was winning victories meanwhile in central Europe at the expense of the Austrian knights. Swiss burghers and peasants, fighting for their liberties, depended on a medieval phalanx composed of spearmen and warriors armed with the halberd—an 8-ft. shaft which had the blade of an axe, the point of a spear and a hook for pulling a rider out of the saddle.

The Swiss phalanx of the 15th century consisted of three-fifths halberds, one-fifth spears and one-fifth missile weapons, chiefly crossbows with handguns slowly gaining in favour. In the interests of mobility the men of the cantons discarded all armour except the helmet and cuirass. Marching 30 mi. a day to meet the enemy, they attacked with a celerity and discipline which were disconcerting to medieval adversaries. If hard-pressed, the Swiss "formed the hedgehog" by facing outward in all directions as men from the rear ranks filled any gap in front.

7. Early Victories of Gunpowder.—After beating mailed horsemen who had resisted tactical change, the English and Swiss in their turn were defeated because they lagged behind the military progress of the age. It had taken weapons of gunpowder a century and a half to have much influence on tactics, but at the siege of Orléans in 1428–1429 both sides used cannon firing stone balls. The French guns were better served than the mortars and bombards brought by the English, who met defeat at the battle of Patay after abandoning the siege.

Charles VII of France raised the first standing army of the middle ages—9,000 permanent troops, paid and armed by the king. This force gradually won back the French territory conquered by the English in the Hundred Years' War, and at the battle of Formigny in 1450 the invaders were decisively beaten after making their customary defensive stand with the longbowmen on both wings.

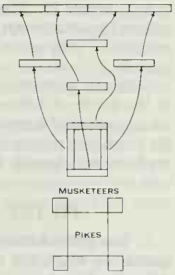


FIG. 7.—EVOLUTION OF THE SQUARE INTO THE LINE

Instead of launching their usual charge, the outnumbered French brought up culverins, or medieval fieldpieces, and cut the English archers to pieces with enfilade fire.

The Swiss phalanx was granted a longer period of grace, but eventually it too was to pay with defeat for clinging obstinately to outworn tactics.

Of all the early exponents of gunpowder weapons, perhaps the most remarkable was John Žižka's army of Bohemian peasants. Several crusades were proclaimed by Pope Martin V against these dissenting followers of John Huss, but the German and Hungarian knights encountered the most advanced tactical system of the new age of gunpowder. Combining the strategic offensive with the tactical defensive, the Hussite military genius drew up a line of "wagonforts"—armoured wagons affording cover to infantry armed with both handguns and crossbows. Between the vehicles were cannon mounted on four-wheeled carts and protected by infantry with pikes. Behind both wings the cavalry waited in readiness to deliver a counterstroke or lead a pursuit.

The combination of the three arms continued to prevail without a single reverse after the death of the blind leader. More than 50 battles and combats were won by the Hussites in 14 years, but civil war led to the destruction of armies which had been invincible against the invading knights.

III. THE MILITARY RENAISSANCE

1. The Fall of Constantinople.—In 1453 it was evident that a new military age had dawned when the Turks breached the walls of Constantinople with a siege train of 70 cannon. After 40 days of bombardment enough towers had been leveled and holes opened in the masonry for a general land and sea assault to succeed.

If weapons of gunpowder could batter down the strongest walls of Europe, it was plain that a defensive based on iron men in stone castles was no longer dominant. By the middle of the 15th century the plate armour panoply had become so heavy that special breeds of horses, the ancestors of Percherons and Belgians, were reared as chargers. Only armour of prohibitive weight could stop a ball from a handgun at close range, and men were beginning to suspect that mobility offered more protection than iron.

Fortification (*q.v.*) and siegecraft had changed but little in 2,000 years. Down to the 14th century an Assyrian of 700 B.C. would not have found much to astonish him in the attack and defense of walled strongholds in Europe. But after the fall of Constantinople it was realized that the wall was obsolescent. Works of the future must be low and screened, burrowing into the earth for security. Besiegers must dig their way forward in a maze of parallels and approaches affording shelter from the fire of the besieged.

About this time the ponderous handgun was being superseded by the arquebus—a comparatively portable matchlock which could be fired from the shoulder. Improvements in the manufacture of wheeled cannon had resulted in slender guns of cast iron or bronze firing iron balls. The principle of the limber had been discovered thus increasing mobility, and the invention of the trunnion permitted the muzzle to be elevated or depressed for more accurate aim. (*See* ARTILLERY; SMALL ARMS, MILITARY.)

The new weapons were utilized by brilliant captains in the struggle between France and Spain for control of Italy early in the 16th century. Tactical values changed from one battle to another with bewildering rapidity. On the field of Cerignola in 1503 the French infantry and Swiss mercenary pikemen were mowed down by the fire of Spanish arquebusiers who counterattacked and captured the enemy's entire artillery train. Nine years later, at the battle of Ravenna, the Spaniards again put their trust in entrenched arquebusiers. But this time the French moved 24 cannon to the enemy's flank and surprised the Spanish infantry with a deadly enfilade fire which decided one of the bloodiest battles of the age.

Again in 1515 the French artillery upheld its reputation by putting an end to the Swiss legend of invincibility. After winning their independence the men of the cantons had become the competitors of the German *landsknechts* as Europe's leading mercenaries. And though the Swiss had increased the proportion of firearms in their phalanx, they still depended mainly on the "push of pike" for victory. In the battle of Marignano, as Spanish hirelings, they paid the penalty of obsolescent tactics when the French repulsed 30 charges by blasting holes in the Swiss ranks with cannon and launching cavalry counterattacks.

Ten years later, on the field of Pavia, the French artillery seemed on the way to another triumph. At the critical moment, however, the victory was won by a surprise advance of 1,500 Spanish arquebusiers. Taking cover behind walls and hedges to fire at will, these elusive marksmen cut down *landsknechts* who had no opportunity to use their pikes.

2. Pike and Arquebus.—Not only had the Spaniards won Italy; they had also founded a new school of infantry tactics which was to rule the battlefields of Europe for a century and make incredibly daring conquests in the new world. Although the battle of Pavia had been won as much by mobility as firepower, the agile combat groups of that victory were gradually replaced by the first phalanx of the age of gunpowder—solid columns of infantry known as battles. Various combinations of "shot and pike" were used, but usually the arquebusiers were drawn up on the corners of battles 25 ranks deep. After firing at the word of command, each rank withdrew to the rear to reload and gradually moved forward by successive volleys until its turn came again. It was the function of the pikes to protect the arquebusiers while reloading.

When the enemy's ranks were broken by firepower, the pikemen evolved from square into line and advanced, shoulder to shoulder, in a massive charge calculated to sweep the field (*see* fig. 7).

The *tercio* of about 3,000 foot was the basic Spanish tactical unit. Consisting of 12 companies of about 250 men each, the fighting strength in 1534 was divided equally between pikemen and arquebusiers. Cavalry was represented in one-twelfth of infantry strength, and artillery by an average of one gun to each 1,000 foot.

The value of tactics and discipline was illustrated during the reign of the emperor Charles V (*q.v.*) when small bands of Spaniards won amazing victories over barbarian hordes in Mexico, Peru and the islands of the Caribbean. The natives were not wanting in sacrificial courage, and their advantage in numbers compensated for inferiority in weapons. One Aztec might have been a match for a Spaniard in single combat, but a thousand Aztecs were routed by a hundred Spaniards because they lacked the tenacity instilled by discipline as well as tactical system and precision.

3. Maurice of Nassau.—Tactical history repeated itself when the Dutch fighting for independence organized the first cohorts of the age of gunpowder to oppose the first phalanx. The Spanish forces had won the early battles of the Netherlands revolt with ease over larger armies of Dutch students, rustics and burghers. Defeat taught the harsh lesson that patriotism was not enough, and Maurice of Nassau (*q.v.*) trained a dependable Dutch standing army. Since the Spanish battles were unwieldy in maneuver, he revived Roman mobility by forming infantry regiments of 800 to 1,000 men, consisting of ten companies of pikemen and musketeers in equal numbers.

During the Wars of Religion the arquebus had been replaced by the musket, which was simply a heavier and more effective matchlock. Accuracy was increased by the introduction of a portable iron fork serving as a rest when the soldier took aim, though it was still necessary to protect him with pikes while reloading.

Another 16th-century invention, the wheel-lock pistol, dated a revival of cavalry tactics. Fired by a steel-toothed wheel which struck sparks from pyrites after the release of a tightly wound spring, the new pistol led to the cavalry tactic known as the caracole, designed to harass squares of pikemen. Advancing in successive waves, the horsemen fired their wheel locks in the faces of the enemy, then swerved off to the rear to reload.

The war in the Netherlands was decided chiefly by siegecraft and fortification. But at Nieuport in 1600 the mobile Dutch cohorts defeated the Spanish phalanx in a battle fought among seaside dunes. Maurice's companies, drawn up in chessboard array, had a great advantage in maneuver over the three Spanish battles of about 3,000 men each. The Dutch prince used heavy artillery to open the action and brought up fieldpieces to accompany his infantry and give close support. When the Spanish squares broke, Maurice's small but well-trained cavalry led the pursuit.

4. Gustavus Adolphus and the Thirty Years' War.—It remained for Gustavus Adolphus (*q.v.*) to improve upon Dutch reforms and merit the title, "father of modern tactics." The Swedish king organized his infantry into brigades, each consisting of two regiments which in turn included two battalions. Each battalion of 1,000 men was divided into eight companies of an average strength of 75 muskets and 60 pikes.

Swedish armourers had developed a lighter musket and Gustavus is credited with being first to introduce paper cartridges to speed reloading. On the firing line his six ranks advanced in open order and closed up into three, the first rank kneeling and the other two standing while firing a volley.

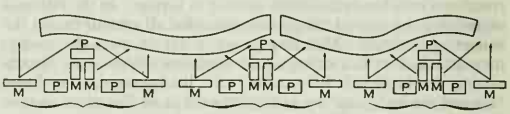
The Swedish artillery also benefited from innovations. Advances in metallurgy having made possible a lighter fieldpiece, Gustavus recalled most of his old ordnance. It was recast into four-pounders of 3 ft. 10 in. in length, 400 lb. in weight and 2:6 in. calibre. Four men or a single horse could handle the new gun on the field, though the four-pounders of other armies weighed half a ton or more.

The first artillery cartridge was developed as ammunition—a thin wooden case holding a prepared powder charge wired to the ball. Improvements were also made in the "bomb" or shell—a hollow iron sphere filled with gunpowder and fired with a lighted fuse. Reloading was speeded to such an extent that the Swedish four-pounder could be discharged eight times while the best musketeers fired six volleys. This was the first regimental piece of history, and each Swedish regiment of 1,000 men brought two four-pounders into the field.

It was characteristic of the forthright Swedish king to replace caracole tactics by training his cavalry entirely for shock effect with pistol and sabre. As light cavalry he organized Europe's first regular dragoon corps—a tactical hybrid made up of men armed with carbine and sabre who could fight either in the saddle or on foot.

The decline in body armour was accelerated when Gustavus himself discarded all protection save a leather buff coat. Field fortification was also reduced to its simplest elements by an army priding itself on the boast, "The Swedes do not defend their men with walls; the Swedes defend their walls with men."

In the battle of Breitenfeld (1631), the first large-scale test, Gustavus opened the action by committing 100 cannon to a long-distance artillery "preparation." Then his self-sustained battalions, each a mobile combat group of pikemen and musketeers



FROM CAPT. HART, "GREAT CAPTAINS UNVEILED" (BLACKWOOD)

FIG. 8.—COMPANIES OF MUSKETEERS AND PIKEMEN DRAWN UP IN COMBAT FORMATION BY GUSTAVUS ADOLPHUS

with their own four-pounder, were able by virtue of rapid shifts to concentrate fire where it was most effective (*see* fig. 8). Of the 33,000 imperial troops, 7,000 were captured and 12,000 killed or wounded. Only about 2,000 survivors could be rallied after the pursuit.

Although Gustavus inflicted another defeat on the imperialists in the battle of Lützen the following year, he paid with his life. In two years of campaigning he had overrun Germany, but after his death the army soon degenerated into another of the plundering hosts which made the war one of the most destructive of history.

IV. THE AGE OF LINEAR TACTICS

1. The New Age of Moderation.—It has aptly been said that the tactics which end one war are those which will begin the next. After the bloody battles of the Thirty Years' War resulted in a prolonged stalemate, the vicomte de Turenne (*q.v.*), commanding mercenaries paid by France, cleared the enemy from Bavaria in a campaign of maneuver, compelling the king to sue for an armistice. Avoiding battle unless he had an overwhelming advantage, the French tactician relied upon caution, precision and surprise to drive the enemy into ravaged areas while retaining his own sources of supply.

Thus the savage excesses of the war were made to appear futile by tactics which foreshadowed the most moderate era of strife in modern history. The founding father of the new military age was the marquis de Louvois, minister of war at the court of Louis XIV of France. Recognizing that tactics are dependent on supply, he organized a quartermaster general's department and a trained commissariat. As fixed points of maneuver, he established magazines at strategic centres for storage of reserves of food, ammunition and equipment.

The French minister also foresaw that two recent improvements in weapons would lead to tactical change. The first was the socket bayonet, which superseded the pike by providing every musketeer with his own means of protection. Equally important was the invention of the flintlock musket, making it possible to dispense with the clumsy lengths of smoldering "match."

Even Gustavus had not been able to reduce the depth of his formations to fewer than six ranks. But now that it was no longer

necessary to combine pikes with muskets, Louvois perceived that the day of linear warfare had dawned. His famous drillmaster, Gen. Jean Martinet, whose name has lived as a byword, trained the troops of the new French armies to deploy from column into line and advance in three ranks, pausing to fire platoon volleys at the word of command. The effect depended on precision of approach, and the men were trained to move forward in unison, keeping step at a stately cadence of 80 paces a minute.

Although the volunteer system prevailed in theory, the soldiers of the new military era were actually recruited for the most part by force, fraud or various forms of legalized blackmail. Since it took from three to five years to train a man for linear tactics, Louvois made a study of morale with a view to reducing the high ratio of desertions prevailing in the Thirty Years' War. The French army became the first to be completely uniformed, and the war minister was a pioneer in the awarding of medals for valour. He set another tradition with his annual war games in which large bodies of troops were trained to maneuver under simulated combat conditions.

2. The Wars of Louis XIV.—The French army had no equal in numbers as well as quality when Louis XIV began the aggressions which were to involve him in wars throughout his long reign. But the battle of Seneffe in 1674 made it plain that grievous losses were to be expected by the best troops when opposing lines traded volleys at 50 paces as a prelude to charging with the bayonet. Only iron discipline could fit men to stand firm and fill the gaps in the ranks, for the death casualties alone in this clash amounted to 17% for both armies.

Large-scale actions depended virtually on mutual consent, since a general could refuse by withdrawing to country too hilly or wooded for the precision demanded by linear tactics. As all armies adopted the magazine system, fortification and siegecraft became more decisive than battle, and some thrifty victories were won by Sébastien le Prestre de Vauban (*q.v.*) in attacks on enemy strongholds. He prevailed on Louvois to increase the number of mortars and organize battalions of troops armed with grenades. It became the military fashion to select only the tallest men for this dangerous duty, so that the grenadiers were recognized as the elite of the infantry.

3. The Tactical Theories of Marshal Saxe.—As compensation for the heavy battle losses, pillage and devastation were reduced to a minimum by disciplined armies of long-term troops. Payment was made for supplies even in enemy territory, and generals sought to gain their ends by maneuver rather than hard fighting.

"I am not in favor of giving battle," declared Marshal Saxe (*q.v.*), who won a memorable victory over a French and Dutch army at Fontenoy in 1745. "I am even convinced that a general can wage war all his life without being compelled to do so."

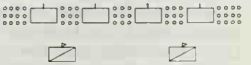


FIG. 9.—MARSHAL SAXE'S COMBINATION OF INFANTRY, LIGHT INFANTRY AND CAVALRY

Saxe foresaw the need for a trained light infantry and managed to raise companies for his Netherlands campaigns in the War of the Austrian Succession (*q.v.*). His victory at Fontenoy was made possible by redoubts from which his troops poured a deadly enfilade fire into their assailants. He also experimented with the unit which became the tactical ancestor of the division (*see* fig. 9). Two brigades of infantry and one of cavalry were combined with artillery into a unit intended to be tactically self-sufficient in local operations, yet capable of concentrating with like units for battle.

4. Frederick the Great.—Saxe's theories might have found widespread adoption sooner if they had not been eclipsed by the dramatic victories of Frederick the Great (*q.v.*). On the field of Leuthen in 1757, he won one of history's most famous victories by defeating an Austrian army of 72,000 troops with 43,000 Prussians. As a tactical innovation, Frederick brought up a battery of ten fortress guns to support his infantry battalions, each of which dragged a six-pounder to supplement the heaviest concentration of artillery ever seen on a European battlefield up to that time. The blue-clad Prussian infantry advanced in staggered alignment, so that the left was refused while the heavily reinforced right bore down on Austrians outnumbered at the point of contact. This was the "oblique order" which won fame for Frederick, though it was nothing more mysterious than a well-timed and executed flank attack which collected a victory already half-won by the fortress guns and six-pounders (*see* fig. 10).

Frederick's greatest contributions to tactics, in fact, were in artillery. Regimental pieces had previously been drawn into battle by soldiers tugging at ropes, but Frederick blamed his crushing defeat in the battle of Kunersdorf on lack of artillery mobility. As a remedy, he ordered that four horses in file be hitched to each piece, making it possible to bring larger quantities of ammunition into the field. He also placed more emphasis on howitzer fire in battle than predecessors who had regarded this type of cannon as being merely a weapon of siegecraft.

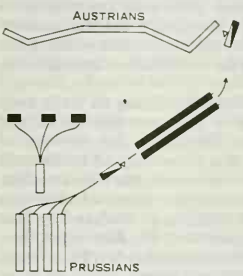


FIG. 10.—BATTLE OF LEUTHEN (1757), SHOWING TACTICS OF FREDERICK'S OBLIQUE ORDER

V. NAPOLEON AND 19TH-CENTURY TACTICS

1. The American Revolution.—Many of the tactics of the Napoleonic era had their roots in the American Revolution (*q.v.*). La Fayette and two of Napoleon's future marshals, Jean Baptiste Jourdan and Louis Alexandre Berthier, returned from service in that war with glowing accounts of new light infantry tactics. On the British side, Sir John Moore (*q.v.*), who was to shape the tactics of Wellington's armies, participated as a youthful subaltern. It was evident to these observers that the American frontier rifle ranked as history's first infantry firearm of precision. Although it took twice as long as a smoothbore musket to reload, it had from two to three times the effective range in addition to an accuracy which was beyond comparison.

Although a majority of the Americans carried smoothbore muskets, the farmers and villagers of the new world were accustomed to hunting small game at a time when laws against poaching and possessing firearms prevailed in Europe. The greenest American recruits took aim instinctively while European regulars were trained to point the piece in the general direction of the foe and fire at the word of command. Thus at the battle of Bunker Hill (*q.v.*) most of the British shots droned harmlessly over the heads of Americans who inflicted casualties of 45% with their limited ammunition by taking "aim at the handsome waistcoats."

Because of the broken country, light infantry tactics flourished in the American Revolution. And since all the armies were small, a two-rank line gained approval as a means of adding to infantry firepower.

2. French Revolutionary Tactics.—The "horde tactics" of the French Revolutionary armies evolved from military poverty. Lacking magazines, the sans-culottes gained in mobility by becoming less dependent on bases and lines of communication. Lacking tents, they learned to bivouac in the field, where all units could be more swiftly concentrated. Lacking supply wagons, they learned to march at a more rapid cadence than their opponents. Lacking training for the precision of linear tactics, they gradually developed a more effective attack of their own.

La Fayette had urged that a rifle company be included in each battalion of the revolutionary forces, and Jourdan tried to instruct his troops in light infantry tactics. Not much came of these efforts, but at least two lessons of the American conflict were put into effect—fire at will, as distinguished from volley fire, and taking natural cover on the skirmish line. Upon coming in contact with the enemy, the bolder spirits of a revolutionary army swarmed out to fire from ditches or behind hedges. The platoon volleys of linear tactics were not effective against these elusive French skirmishers, who took to their heels when threatened with a bayonet charge. Fighting of this sort was so interesting and com-

paratively safe that more and more men participated. Finally, at the battle of Hondschoote in 1793, General Jourdan's forces concentrated swiftly and almost literally mobbed an allied rear guard detachment with a wild bayonet attack.

Fortunately for the sans-culottes, the royal artillery and engineers, as technical branches, had never attracted the nobility. Many of the officers adopted the cause of the Revolution, and by 1793 the French were able to open a battle with a respectable cannonading while the skirmishers were peppering the enemy with harassing fire. Skill and confidence came with experience until the Revolutionary masses of attack bore more resemblance to columns and less to rioting mobs. The allies found it disconcerting, moreover, when a French army doubled in size within a few hours as distant units hastened to the battlefield at the quick-step from all sides.

3. Bonaparte's Early Victories.—It is taking no credit away from Bonaparte's genius for war to point out that he adapted rather than originated his tactics. He was the heir not only of French Revolutionary tactics but also of an army of battle-hardened veterans who had won some notable victories over the allies. From the old royal army he inherited the system of dispersion and concentration of divisions and brigades, as developed by Saxe and Marshal de Broglie. And on the eve of the Revolution, Jean Baptiste de Gribeauval's artillery reforms had reduced the weight and increased the mobility of regimental guns.

Bonaparte was not a great battlefield tactician. He was at his best in that hybrid between tactics and strategy known as "grand tactics"—the assembling of units prior to battle for a last-minute concentration by forced marches on the battlefield itself. No longer could an adversary refuse combat, as in the days of linear tactics. Bonaparte forced such an opponent to fight by sending a column around to threaten his line of communications. When the enemy was brought to bay, Bonaparte could summon his own outlying columns as mobile reserves in less time than could the armies trained for linear tactics.

The preliminary swarms of skirmishers, the cannonading by regimental guns firing grapeshot at close ranges, the battlefield concentration and the attack by bayonet columns of infantry—these were the four stages of Bonaparte's early victories. Some of them were gained by envelopment, some by penetration and others by plain frontal assault, according to the dictates of a marvellous intuition.

4. Tactics of the Empire.—Every conqueror runs the risk of teaching the art of war to his opponents, and Napoleon, as emperor, found his own tactics being used against him with increasing skill. In the battle of Eylau (1807) 7,000 Russian reinforcements appeared at the critical moment after the delay of the French concentration by a snowstorm. Napoleon barely escaped defeat at an appalling cost in killed and wounded.

Two years later, in the battles of Aspern-Essling and Wagram, the archduke Charles (*q.v.*) and his Austrian army gave Napoleon some anxious moments. France was weary of war, and as the quality of infantry conscripts declined, the emperor placed more reliance on bulk. He used artillery in great masses, and on the field of Wagram he sent an enormous hollow square of 30,000 men to advance on the Austrian centre behind 104 guns. This was the decisive blow, but the victory had been won at a price of 23,000 French killed and wounded.

5. Wellington's Tactics.—At a time when other generals were imitating Napoleon's methods, the duke of Wellington defeated French columns with linear tactics and depended on seaborne supplies while outmaneuvering an enemy living on a country eaten bare. He too was an adapter rather than an originator, for the British tactical system was largely the creation of Sir John Moore. In 1803, training troops to meet an expected Napoleonic invasion, he reduced the standard three ranks to the two-rank line of the American Revolution. He applied another American lesson by organizing three regiments of riflemen for his light division, soon to win fame as the most celebrated unit of British battles in Spain.

Moore was killed in his first Peninsular operation, but Wellington applied his tactics so successfully as to become one of history's few unbeaten generals. The French, as Europe's masters, felt obliged to attack. Their opponent profited by being able in every major battle to choose his own ground and draw up his "thin red line" behind a low ridge. This gave him the double advantage of protecting his men from ricocheting cannon balls and compelling the French bayonet columns to advance uphill. A screen of skirmishers from the light division was deployed along the forward slope. Since their rifles outranged French muskets, they were able to harass the enemy columns with aimed fire while the British artillery thinned them with shrapnel. The two-rank line gave the redcoats a wider front for musketry, and Wellington met the main French shock with steady platoon volleys. If the enemy wavered, he ordered a British bayonet counterattack.

This was the tactical formula, repeated with few variations, which won victories over Napoleon's marshals for five years in Spain. Wellington proved to be a match for the French in swift concentrations, and on one occasion his light division completed a march of 43 mi. in 22 hours. But it was with infantry firepower that he beat opponents who trusted too much in the bayonet, for Napoleon had refused to adopt the rifle and neglected musketry training in spite of his assertion, "The power of infantry lies in its fire." (*See* also PENINSULAR WAR.)

6. Half a Century of Transition.—At Waterloo, Wellington needed only the tactics that had won for him in Spain. But an era came to an end on that June day in 1815, when infantry, cavalry and artillery were in close contact and could be controlled by the general as easily as a regiment a century later. All this was changed by the widespread adoption of the percussion cap and long bullet, which multiplied the range as well as accuracy of the rifle and made it a universal military weapon. No longer was it possible to place the guns in front of the infantry, as was done sometimes during the Napoleonic era, with the cavalry close behind. No longer was it possible to mass cannon only 500 yd. from the enemy, as at Borodino, and blast a bloody path to victory. Thus the old battle order, which had changed but little in essentials since the day of Gustavus Adolphus, was relegated to the past. To command an army as Wellington did at Waterloo was out of the question; to render its separate parts co-operative demanded staff work and not merely a powerful voice.

During the half century preceding the American Civil War, inventions came so rapidly that soldiers could not keep pace. The revolver and repeating carbine opened up new vistas for the cavalry and rendered obsolescent the charge of lancers. Rifled and breech-loading fieldpieces worked a revolution in artillery tactics, just as the telegraph, the railroad and the steamship made it necessary to revise all previous ideas of military transport and communication.

Few precedents were available when green armies of citizen-soldiers met in the opening engagements of the American conflict. But if they had more to learn, they had less to forget than the tradition-ridden soldiers of Europe. And before the four-year struggle ended, the Americans had put into effect many of the tactics that would shape the outcome of World War I.

7. The American Civil War.—Four million men bore arms in the American conflict. More than 2,000 combats took place, of which 149 were of enough importance to be called battles. Yet these far-flung operations were dominated throughout by the defensive power of the rifle, supported by the axe and the spade.

The railway gun, the land mine, the wire entanglement, the observation balloon, the hand grenade, the field telegraph—these were some of the innovations or revivals of the Civil War. But they played second fiddle to the demonstrated ability of the man in the rifle pits to hold his own against three attackers. Since most of the operations occurred in wooded country, soldiers of both sides learned to cut "head logs" with firing slits to give added protection along the parapets of trenches. Generals were taught by experience that marksmen in log-faced earthworks, supported by rifled artillery, could seldom be dislodged except at a ruinous cost in casualties. Lee at Malvern Hill and Gettysburg, Grant at Vicksburg and Cold Harbor—the best tacticians of the war met with bloody repulses when they underestimated the firepower of opposing riflemen.

On the other hand, paradoxical as it may seem, the defensive power of the entrenched rifleman enabled Lee to launch successful

attacks on opponents with a material advantage. Relying on the proved ability of one rifleman to stand off three adversaries, the Confederate general dared repeatedly to divide his army in the face of a more numerous enemy. At Second Manassas, for instance, Jackson "contained" more than double his own strength in a holding battle until Lee could arrive the next day and deliver the decisive counterstroke. It was Lee's turn at Chancellorsville to fight a defensive battle with half his army while Jackson led the other half around the Union right in a surprise flank attack, thus winning one of history's few authentic victories over two-to-one numerical odds.

Grant became the first general to command the army of a million men. But European soldiers professed to see in the Civil War only the incoherent clashes of armed mobs, and in 1870 the French and Prussians made it evident that they had learned nothing from the American precepts. They sent cavalry in suicidal charges against infantry armed with breech-loading rifles with an effective range of 1,000 yd.; they took needless losses from rifled artillery as the penalty of close formations; and they persisted in infantry frontal attacks despite the fact that neither side succeeded in taking a single position by such tactics.

VI. WORLD WAR I

1. The Battle of the Marne.—It was the tragedy of World War I that the century of comparative European peace since Waterloo seemed to have made romanticists out of the professional soldiers who held in their hands the lives of millions of conscripted citizens. Despite the lessons of the American conflict, confirmed by the machine guns of the Russo-Japanese War, France's generals of 1914 had a mystical faith in the *offensive à outrance*—the headlong offensive, compounded of mass, velocity and morale. They believed that the prewar scale of two machine guns to a battalion was adequate; and as late as 1915 the British commander, Sir Douglas Haig, asserted that it was "a much overrated weapon" and agreed that the French scale was "more than sufficient." (B. H. Liddell Hart, *A History of the World War, 1914–1918* [enlarged ed. of *The Real War, 1914–1918*], p. 143, Little, Brown & Co., Boston, 1935.) By that time, nearly a million men had been killed by machine guns on the western front.

The German generals were more realistic, but they too had complete faith in the offensive and regarded any reference to defensive advantages as bordering on defeatism. Both the French and German tacticians considered a defensive action excusable only when "each man will die in place rather than give ground."

The Germans, who were better prepared at the outset, seemed to be proving their case as they swept through Belgium and northern France. The French soldiers, uniformed in red breeches for the sake of morale, endured frightful losses. And since the German generals held that "the abandonment of close-rank formations is an evil which should be avoided whenever possible," their troops in turn provided the French artillery with lucrative targets.

In the battle of the Marne the French stopped the invaders by taking advantage of blundering last-minute revisions in the German envelopment plan. Then began the "race to the sea" which was to result in a continuous front of trenches from the Atlantic to the Alps. The German losses of the first four months were 747,465 killed and wounded; the French had 854,000 casualties during the same period.

2. Deadlock on the Western Front.—In the spring of 1915 the entente nations believed that artillery was the key to the deadlock. But even after a prolonged bombardment by 1,140 pieces of heavy artillery, 53 French infantry divisions were able to advance only 3,000 yd. on a 12-mi. front. The enemy simply weathered the storm of steel in deep dugouts until the barrage lifted, then manned the machine guns which cut down attackers plodding across no-man's land under heavy burdens of equipment.

The Germans risked the censure of neutral nations in an attempt to win a decision with poison gas in the Neuve Chapelle sector. But the Canadians held out until they could be reinforced, and millions of soldiers were soon wearing gas masks as all armies resorted to chemical warfare. Steel helmets also became generally adopted as protection against shell fragments.

The entente nations made a bid for victory in 1915 by launching a land-and-sea assault on the Dardanelles for the purpose of knocking Turkey out of the war and battering down "the unlocked back door" of the central powers. A costly failure, after an eight-month campaign, was blamed on the supposed impossibility of conducting a major amphibious operation against modern weapons. Actually, as time was to prove, British and French repulses on the Gallipoli peninsula resulted from violations of basic tactical principles. (*See* DARDANELLES CAMPAIGN.)

By 1916 both sides recognized the dependence of ground operations on air warfare (*q.v.*). The two great campaigns of the year were the German offensive at Verdun and the Franco-British offensive on the Somme. New infantry tactics were introduced by the Germans as specially trained combat groups, armed with flame throwers, advanced after a terrific bombardment. Verdun became known as "the sausage grinder" as the struggle continued for ten months with losses to the French of 442,000 men and to the Germans of 278,000. Although it cannot be said that either side won the battle of attrition, the French heavy artillery and chain of underground forts stopped the enemy. At Verdun, moreover, sweeping changes in tactics were indicated when the French demonstrated that an army could be supplied by motor transport.

The Franco-British offensive on the Somme turned into another heroic slaughter. German defenses in depth, extending as far back as 10 mi., frustrated every attempt at a worthwhile penetration, and "gains" of a few hundred yards were recorded in London headlines. The British made a premature and unsuccessful use of tanks, but lack of planning and preparation actually served to retard the progress of armoured warfare.

3. The Offensives of 1917.—The western front was not the only theatre of war, but it was recognized that France would be the area of ultimate decision. Entente generals still trusted in the spring of 1917 that by increasing the density of artillery the infantry could make headway behind a rolling barrage timed to the advance. But even a ratio of a gun to 6 yd. of front was not enough, and the German machine guns continued to reap their deadly harvest.

It had been necessary for a civilian, David Lloyd George, to provide the British army with a realistic scale of machine guns and trench mortars in spite of the reluctance of General Haig and Lord Kitchener. Another civilian spoke for the dead of France when Premier Georges Clemenceau remarked that "modern war is too serious a business to entrust to soldiers."

Although Belgian engineers had warned against destroying the drainage system of reclaimed lands, 103,000 tons of shells were fired in the preliminary British bombardments of the third battle of Ypres, known to the soldiers as Passchendaele. Mustard gas was added to the horrors of the 100-day battle as both armies wallowed in the swamp created by 4,300,000 British shells. Haig's gains were negligible despite 300,000 casualties, and the Germans considered Passchendaele "the supreme martyrdom."

From the entente viewpoint, one of the few consolations of a gloomy year was the British breakthrough at Cambrai on a 6-mi. front with 381 tanks supported by infantry. Attacking by complete surprise, after a last-minute hurricane barrage, the clanking cavalcade ripped through defenses in depth and took 7,500 prisoners and 120 guns at a low cost in casualties. Adequate infantry forces had not been provided by high-level British generals who envisioned a cavalry exploitation, and the territorial gains of Cambrai were lost to a German counterattack. Even so, the operation had demonstrated that armour was the key to a decision on the western front. (*See* also CAMBRAI, BATTLE OF.)

The Germans had meanwhile developed a key of their own in a new concept of offensive tactics known as "infiltration." It called for the rapid advance of dispersed "storm groups," armed with automatic rifles and light mortars, preceding the main body of infantry. These specialists were to probe for weakness and break through wherever they could find an opening. Their mission was to surprise enemy machine guns and artillery from the flanks and rear while leaving the strong points to be enveloped frontally by following waves of infantry, closely supported by artillery and tactical aircraft. These new tactics of infiltration, carried

out by secretly trained troops, promised to create enough demoralization for a penetration as far back as corps command posts and communication centres.

Leaders of both sides realized early in 1918 that it was destined to be the year of decision. All the nations which had participated from the beginning were war-rotted except Great Britain and Germany, and it was a question whether the newly trained U.S. divisions could cross the Atlantic in time to compensate for the defection of Russia, convulsed by revolution.

4. The Mobile Warfare of 1918.—Infiltration won the first round. The German spring offensive in Picardy resulted in gains for a few days such as the western front had not known since 1914. Within a week a salient had been extended to a maximum depth of 40 mi. in the British sector by German columns which threatened to drive to the sea. Then, suddenly, the attack ground to a halt and the front was stabilized once more.

This loss of momentum was no coincidence. What had happened was that the Germans had outrun their supplies, for the rail transportation and horse-drawn artillery of 1918 were not adequate to nourish a rapid advance. General Ludendorff made four bids for a decision and each time the early gains merely resulted in creating another salient and two more German flanks.

The last effort, on July 15, 1918, was a complete failure. Attacking on the Champagne front, the Germans were trapped in the French elastic defenses planned by General Pétain. Here, for the first time in the war, land mines were used on a large scale, and the attackers never reached the main French positions.

The riposte came three days later when Americans and French Moroccans counterattacked on July 18 along the right face of the salient created by the German drive to the Marne. Complete surprise was achieved by rushing the assault troops up in trucks under cover of darkness. Without any preliminary bombardment, the infantry advanced at dawn with 352 tanks and penetrated so deeply as to threaten the whole salient when other French and American troops attacked the opposite face. Before the Germans could pull out, they left 30,000 prisoners behind and suffered about the same number of killed and wounded.

The precept of motor transport set at Verdun in 1916 bore further fruit as entente generals learned that they could gain surprise unfailingly by these tactics. Dispensing with bombardment and bringing the assault forces up in trucks meant that there would be neither ammunition dumps nor troop concentrations to warn the enemy by means of his aerial observation.

The Amiens salient was next to be reduced as 13 British divisions attacked on Aug. 8 with 450 tanks while armoured cars manned by automatic riflemen spread confusion in German rear areas. On Sept. 2 another British attack struck the enemy in Flanders, and on the 12th an army composed of new American divisions wiped out the St. Mihiel (*q.v.*) salient.

The end was in sight during October as the British on the left and French in the centre rained blow after blow, striking by surprise when and where they were least expected. These victories stemmed in large measure from the pressure exerted on the enemy as the Americans on the right slugged their way forward in the 46-day battle of the Meuse-Argonne against an enemy using a fourth of his entire strength on a 20-mi. front in difficult terrain. (*See* also under names of various battles.)

VII. WORLD WAR II

1. The Post-World War I Reaction.—During the postwar years the reaction of the former entente nations was in the direction of defensive tactics. The casualties inflicted by the machine gun were remembered when the victories of the tanks were forgotten. This tendency went so far as to resemble in some instances the wishful tactical thinking of 1914 in reverse.

The French, with traumatic memories of two German invasions in half a century, translated their theories of the all-powerful defensive into the concrete and steel of the Maginot line—a bristling belt of underground forts and tank traps extending from the Vosges westward to the Ardennes, where it was believed that the hills and woods would turn back armoured columns. The British, as their modern counterpart of the Great Wall of China, had a defensive military philosophy, the main tenet of which was that postwar improvements in weapons added to the risks and perils awaiting a future aggressor.

The tactical euphoria of the former entente nations was shaken in the 1930s by successful Japanese, Italian and German aggressions. When the Spanish Civil War served as a rehearsal for Russian as well as German and Italian tactics, it was belatedly realized that improvements in weapons favoured the offensive to such an extent as to endanger laggards in preparations. Tanks were many times as fast and powerful as those of 1918; motorized artillery, on rubber tires, showed an enormous gain in mobility over horse-drawn, iron-wheeled vehicles; and motor trucks poured from the assembly lines by the thousands to solve the transport and supply problems of a new lightning warfare.

2. Early German Victories.—Any doubts as to the effectiveness of German preparations were dispelled during the first two years of World War II as Hitler's armoured columns swept on to victory after victory. The Maginot line was outflanked by German tanks which broke through in the Ardennes and drove from the Meuse to the Channel in a week, cutting off 335,000 men of the British expeditionary force and attached French units. It seemed an authentic miracle when these beleaguered troops were evacuated by sea from Dunkirk, leaving their heavy arms behind, as the German columns raced on to the occupation of Paris.

From the late summer of 1939 to the autumn of 1941 the Germans never met with a serious reverse in land warfare. Their triumphs were gained with such swiftness, such precision and economy of means that the blitzkrieg, or lightning war, was popularly accepted as a brilliant new tactical system. Actually it was the legitimate offspring of the infiltration tactics of 1918 mated with the tank tactics. There was nothing new that could not be explained by the gains of two decades in the offensive powers of such weapons as the tank (*q.v.*), motorized artillery and tactical aircraft. With a few exceptions such as antiaircraft guns, defensive weapons had lagged in the armament race while armour and artillery were leaping ahead from the 3 m.p.h. of 1918 to the 30 m.p.h. of 1938.

The French army of 1940 was essentially a World War I instrument, patched here and there with tactical changes, just as the Maginot line was in effect a modernized western front. Despite warnings that the enemy might strike through the gap in the Ardennes between the French and Belgian fortified lines, the French in the Sedan area were unprepared for German "tactics of the space and gap." Only three penetrations on a 12-mi. front, each no wider than 2,000 yd., were enough as the starting point of a major breakthrough. Like a flood enlarging holes in a dam, tanks and combat groups in armoured cars poured into the gaps, depending on surprise and speed as well as Stuka dive-bombers for flank protection. Their direction of attack was lateral as well as forward, the object being to widen the gaps for following German forces. Within four hours the invaders were overrunning command posts and communications centres in French rear areas too demoralized for effective resistance. The Maginot line had simply been by-passed and left behind by the torrent of armoured invasion which inundated northern France in a week.

Predictions that Hitler would meet his downfall in Russia were not upheld by military results of the first few months. The Russian army, like the prehistoric mastodon, seemed doomed to extinction by its very bulk as German armoured envelopments swallowed up prisoners by the division and corps. Political factors influenced tactics when Hitler's doctrines of racial superiority led to cruelties which drove Russian peasants into guerrilla warfare. Effective resistance of this sort, combined with Soviet scorched earth tactics and a seemingly endless reserve of man power, enabled the Soviet army to survive until the German transport bogged down in autumnal mud.

Even so, the tactics of Hitler's armies had been so successful during the first two years that a serious defect might easily have been overlooked. That was the apparent lack of preparation for amphibious tactics in a global war. Although the British Isles were desperately short of arms after the Dunkirk evacuation, the Germans had neither the training nor the equipment to land troops

across a few miles of water. In Crete they proved their mastery of airborne attack techniques, but a clumsy attempt at an amphibious combat landing ended in disaster.

3. The War in the Pacific.—After the United States and Japan entered the lists as enemies, the war in the Pacific necessarily consisted in large part of amphibious operations. And despite the tradition that the English-speaking peoples are never prepared for war, events showed that the U.S. forces had developed an effective system of amphibious combat tactics. Three continents were opened up to invasion by this system, which won a score of major victories without suffering a single major defeat. In the 1920s the "Orange" plans, adopted by the United States for offensive operations in the event of war with Japan, called for attacks on that nation's shield of fortified islands in the Pacific. By a directive (1927) of the joint board of the U.S. army and navy, the marine corps was given the mission of preparing for the conduct of landing operations.

There were few 20th-century precedents, though the Gallipoli failure of 1915 supplied no lack of object lessons. Weapons and techniques had to be improvized where none had existed before. Strange landing craft and amphibian vehicles evolved as inventors were encouraged to submit designs. Large-scale simulated assault landings were made annually in fleet exercises of the 1930s held at Caribbean and Pacific bases. New bases were established, new amphibious units formed, and new types of ammunition developed. As a result, the United States possessed on the eve of World War II what Maj. Gen. J. F. C. Fuller has described in his history of that conflict as "in all probability . . . the most far-reaching tactical innovation of the war." (J. F. C. Fuller, *The Second World War*, p. 207, Eyre & Spottiswoode, London, 1948.)

4. U.S. Amphibious Victories.—Within a year the outer barrier of Japan's island defenses was breached by U.S. ship-to-shore assaults. Months of detailed planning and rehearsal preceded each operation. Supply ships were "combat-loaded" so that tanks, trucks, artillery, fuel, rations and water would be made available to the landing force in the order needed. Intricate firing plans were drawn up for artillery and rocket bombardments to scorch the target area while armoured landing craft took the assault troops to the beaches under cover of air strikes. The transition from one element to another was the critical moment. Maneuver was seldom possible during the controlled chaos of the landings and repulse meant a disaster which could hardly be retrieved. Once the beachhead was won, special units unloaded cargo ships, established supply dumps and regulated traffic while the assault troops drove inland with the support of tanks and artillery as soon as they could be landed.

U.S. ship-to-shore assaults were conducted by troops using navy and marine techniques. The army trained special units for shore-to-shore assaults, utilizing the same landing craft. New Guinea was thriftily reduced by these troops in coastal landings which by-passed strong Japanese garrisons and left them behind to "wither on the vine."

The amphibious offensive had become by 1943 the most decisive tactical system of World War II, not only in the Pacific theatre but also in North Africa and Europe. Neither the Japanese nor the Germans ever found the clue to a successful defense. Landings in Morocco, Algeria, Sicily and Italy were followed by the climactic operation of June 6, 1944, when 4,100 ships put ashore five divisions complete with tanks and artillery on the beaches of Normandy. Nine days later in the Pacific an equally successful amphibious combat landing on Saipan brought American bombing planes within striking distance of Tokyo and led up to the assaults on Iwo Jima and Okinawa which made Japanese defeat inevitable.

5. Wide Variety of World War II Tactics.—So varied were the geographical and climatic conditions that it is difficult to generalize about the tactics of a war in which a score of major campaigns were being waged simultaneously in different parts of the world.

Tank battles on the plains of Russia began with tremendous artillery "preparations" and ended with armoured envelopments in which as many as 20 divisions were surrounded. In such instances the beleaguered forces were sometimes able to form a "hedgehog," or circular formation, and fight their way out of the encirclement with the aid of airborne supplies. In the long run it was a grinding war of attrition in which the Russians gained the upper hand largely as the result of arms and equipment delivered by Britain and the United States in spite of heavy losses suffered by sea convoys.

Armoured columns in the desert warfare of North Africa achieved a mobility which makes their operations instructive to the student of tank tactics. But General Rommel's defeat was foreordained by the inability of the Germans to mount an amphibious assault on Malta. The campaign in the desert was actually a battle of supply, and while the British held Malta they could provide air escorts for convoys of cargo ships in the Mediterranean.

On all fronts, special units equipped for special tasks were employed to an extent never known before. Infantry was the arm of decision in most operations, but it became standard practice to organize task forces ranging in size from platoons to divisions, reinforced by tanks, artillery, rocket launchers or such other support weapons as might be indicated for the specific mission.

So potent were the missile weapons of World War II that survival for the individual soldier depended on cover and concealment. Small infantry units took on increasing importance as experience taught that latitude of decision must be left to troops in direct contact with the enemy, since victory often depended on the initiative and intelligence displayed by platoon commanders.

Some of the infantry combat teams became standardized. The British commando units, made up of soldiers, sailors and marines organized on a permanent basis, conducted raids throughout the war. In Burma the British developed special penetration groups of infantry to operate behind the enemy's lines, attacking headquarters and supply dumps while being supplied by air.

All armies made an unceasing effort to gain a decision by the introduction of new weapons and tactics. But the German V-1 jet-propelled bomb came too late to change the outcome, as did the more formidable V-2 rocket after the Allies had broken through in Normandy and swept across France to the German frontier. (*See* ROCKETS.)

On the Allied side, the 2.36-in. rocket launcher and the 57-mm. and 75-mm. recoilless rifles provided the infantry with weapons which were tank-killers under favourable circumstances. The VT (variable time) artillery fuse enabled a shell to be exploded automatically in electronic reaction to waves reflected from the target itself. These tactical innovations seemed trivial, however, after President Truman's radio announcement of Aug. 6, 1945:

"Sixteen hours ago an American airplane dropped one bomb on Hiroshima, an important Japanese army base. That bomb had more power than 20,000 tons of T.N.T. . . . It is an atomic bomb. It is a harnessing of the basic power of the universe." (*See* ATOMIC ENERGY; *Bomb Explosions.*)

Thus, at the conclusion of World War II the participants were confronted by the fact that "the bomb" might render obsolescent all the tactics of victory. Even the discovery of gunpowder does not represent a comparable turning point in tactical history, for the first handguns were not as deadly as the crossbow. Nor did gunpowder have a rapid evolution, considering that the T.N.T. of the 20th century was only about twice as powerful as the explosives of the middle ages. Yet the world of 1945 had to adjust itself overnight to the necessity of revising all military standards of the past. (*See* ARMY; ARTILLERY; INFANTRY.)

VIII. THE KOREAN WAR

1. Intervention in Korea.—It is one of history's ironies that the first war of the strident new atomic age should have been waged with tactics reminiscent of the western front in 1917. Where the public had envisioned a "push-button" war of guided missiles, combat in Korea (1950–53) consisted for the most part of infantry fights with small arms on the platoon and company levels. (*See* KOREAN WAR.)

Called a "police action" at first, the intervention of 1950 proved during its first year to be the fourth most costly military effort of U.S. history. After a preliminary period of mobility, the United Nations and Chinese Communist forces faced each other in lines

of trenches and defenses in depth extending across the peninsula. Neither side attempted to hold its front in force. Bunkers and outposts on the main line of resistance sheltered combat groups whose duty it was to maintain contact with the enemy and warn of a large-scale attack. Patrols ranging in size from an infantry squad to a platoon crept through no man's land in the darkness as both sides found the ambush an effective tactic.

The Korean conflict taught that even in the atomic age an Asiatic peasant army using semiguerrilla tactics can be formidable in mountain terrain suited to its operations. (*See* MOUNTAIN WARFARE.) Atomic weapons would not have been effective in such a war, even if they had been considered, and the Korean experience made it appear unlikely that sound infantry tactics would soon become outdated.

2. Tactical Innovations of Korea.—The two great tactical innovations of the war were the combat helicopter and the revival of infantry body armour. Rotary wing aircraft had been developed too late for an extensive test in World War II, but in Korea the helicopter proved its value immediately in tactical as well as logistical respects. A U.S. brigade commander found it possible with the use of rotary-wing transportation to maintain combat control over units 25 mi. apart. When a U.S. squadron of transport helicopters reached the front, it was demonstrated that a battalion of infantry could be lifted in a few hours to the main line of resistance in mountainous country where an all-day march would have been necessary. On an occasion when an infantry company was surrounded, six helicopters flew in ten tons of ammunition and evacuated casualties, thus enabling the encircled troops to fight their way out of the trap.

The revival of body armour, after a lapse of several centuries, was made possible by 20th-century plastics. Nylon pads and thin slabs of Fiberglas were combined in a sleeveless vest weighing about eight and a half pounds and allowing freedom of movement. The armour was capable of stopping .45-calibre automatic pistol bullets, submachine gun slugs and most shell, mortar or grenade fragments of a velocity less than 1,100 ft. per second.

All U.S. combat troops in Korea were equipped with the garment before the conflict ended. On the basis of U.S. army and marine corps surveys, it was established that body armour prevented from 60% to 70% of chest and abdominal wounds while reducing the severity of 25% to 30% of the wounds resulting from penetrations. Aside from humanitarian aspects, this meant that a large proportion of the enemy's most effective antipersonnel weapons were rendered useless. The tactical value of body armour was increased by its contribution to morale. Troops on the defensive had their foxholes and trenches, but the vest-protected men advanced in the open during an attack, when protection was most needed. The result, according to U.S. infantry officers in Korea, was a marked increase in combative spirit.

3. Tactics of the Atomic Age.—At the end of the Korean war, the military problems of the atomic age were no nearer to solution because of lessons taught by that war. There was some reassurance, however, in the fact that neither side in an ideological conflict had even threatened to use atomic weapons, though it was known that Soviet Russia as well as the United States had conducted successful atomic tests. As further evidence that warring nations could practise restraint, it might have been cited that poison gas, one of the most deadly weapons of World War I, was not used in the next two wars, though all armies were in possession of far more destructive gases than those of 1918. Nor was any attempt made to gain a military advantage by the possibilities for wiping out entire populations offered by germ warfare.

During the first decade of the atomic age, the U.S. tests conducted at Bikini atoll in the Pacific and Yucca Flat in Nevada encouraged soldiers to believe that tactics might be adjusted to the new weapon by means of troop dispersion on a scale never envisioned before. Then came the announcement of the hydrogen bomb and predictions that weapons with fantastic destructive power measured in terms of 45 megatons (45,000,000 tons of TNT) were being developed. (In 1961 the U.S.S.R. detonated a device of even higher yield.) Contemplating such a prospect military men could only admit that they knew of no tactical antidote to a force capable of wiping out a province or a metropolitan city with a single explosion. It was no longer war when such weapons were used; it was sheer insane destruction. And there was not much consolation in the thought that an enemy could be paid back in kind.

Fears of intercontinental atomic bombing raids led in the late 1940s to popular clamour for great fleets of strategic aircraft. But during the years following the Korean conflict, future trends were indicated by the development of an entire new family of low-yield atomic weapons suited to battlefield use against tactical targets. Although the nature of most of these weapons was kept a military secret, it became public knowledge that nuclear artillery and rockets were included.

A result of these trends was the pentomic infantry division announced by the U.S. army in 1957. Firepower, mobility and dispersion suited to the atomic age were the goals sought in the most sweeping changes made by any army up to that time. The three infantry regiments of the triangular division, with their nine battalions, were replaced by five battle groups, each comprising approximately 1,400 officers and men. Although total numbers were reduced, the "foxhole strength" of the pentomic division was increased.

In 1961 the pentomic division, designed for employment in nuclear warfare, was abandoned because it proved too cumbersome and inflexible for use in "brushfire" wars involving limited forces and conventional weapons. Its place was taken by a new divisional structure consisting of an "all-purpose" divisional base with three brigade headquarters to which could be attached varying numbers and types of combat maneuver battalions (*see* DIVISION, MILITARY). This reorganization provided a division whose size (10,500 to 19,950 men) and equipment could be tailored to each specific mission and to the type of terrain encountered in its accomplishment. Other results of the change were improved balance between nuclear and non-nuclear firepower, greater internal flexibility and an improved command and control structure.

At about the same time a unified command was established to include the strategic army corps (STRAC), designed for fast deployment by air, and the tactical air command (TAC) of the U.S. air force. Thus was created a mobile striking force of ground troops adequately supported by aviation units. This new command helped alleviate an army complaint that ground forces were without adequate tactical air support since the diversification of command caused by the separation of the army and air force in 1947.

Still another change in tactical thinking occurred during the early 1960s with the establishment of special forces for carrying on small-unit guerrilla warfare behind enemy lines or co-operating with indigenous populations in antiguerrilla operations.

See ARMY; LOGISTICS; STRATEGY; *see* also references under "Tactics" in the Index.

BIBLIOGRAPHY.—J. F. C. Fuller, *A Military History of the Western World,* 3 vol. (1954–56; pub. in Great Britain and Canada as *Decisive Battles of the Western World and Their Influence Upon History*); Lynn Montross, *War Through the Ages,* rev. ed. (1946). *Weapons:* T. Wintringham, *The Story of Weapons and Tactics* (1943); J. R. Newman, *The Tools of War* (1942). *Tactical Theory:* K. von Clausewitz, *On War,* Eng. trans. by J. J. Graham, rev. ed. with introduction and notes by F. N. Maude, 3 vol. (1940); A. du Picq, *Battle Studies* (1921); J. F. C. Fuller, *The Reformation of War* (1923). *Ancient Warfare:* T. A. Dodge, *Alexander,* 2 vol. (1890), *Hannibal,* 2 vol. (1891), *Caesar,* 2 vol. (1892). *Medieval:* C. W. C. Oman, *A History of the Art of War in the Middle Ages,* enl. ed. (1924). *Early Modern:* C. V. Wedgwood, *The Thirty Years' War* (1938); W. S. Churchill, *Marlborough: His Life and Times,* 5 vol. (1933–37). *Napoleonic Era:* T. A. Dodge, *Napoleon,* 2 vol. (1904); B. H. Liddell Hart, *The Ghost of Napoleon* (1933); C. W. C. Oman, *History of the Peninsular War,* 6 vol. (1902–20). *Nineteenth Century:* D. S. Freeman, *R. E. Lee,* 4 vol. (1934–35); J. F. C. Fuller, *The Generalship of Ulysses S. Grant* (1929); C. von der Goltz, Eng. trans. by Joseph T. Dickman, *The Nation in Arms* (1914). *World War I:* W. S. Churchill, *The World Crisis,* 5 vol. (1923–29); B. H. Liddell Hart, *The Real War, 1914–1918* (1930). *World War II:* C. Wilmot, *The Struggle for Europe* (1952); J. F. C. Fuller, *The Second World War, 1939–1945* (1948); J. A. Isely and P. A. Crowl, *The U.S. Marines and Amphibious War* (1951). *Korean War:* R. Gugeler, *Combat Actions in Korea* (1954). For detailed descriptions of the pentomic infantry division, see *Infantry: Official Quarterly Publication of U.S. Army Infantry School,* vol. 47, no. 2 and 3 (April and July, 1957). *See* also U.S. army and marine corps official histories of operations in World War II and Korean war, and official British, Canadian and Australian histories.

(LN. MS.)

TACTICS, NAVAL: *see* NAVAL STRATEGY AND TACTICS.

TACUAREMBÓ, the largest department of Uruguay, located north of the Río Negro in the north central part of the nation, has an area of about 21,000 sq.km. (8,114 sq.mi.) and a population of

136,421 (1962 est.), of whom approximately 21,000 live in the departmental capital, also called Tacuarembó. Tacuarembó is a typical ranching department of the Uruguayan interior, with the 1951 livestock census counting 817,126 cattle and 2,338,538 sheep. Corn, tobacco, peanuts, rice and forage crops are the leading crops. Construction of the hydroelectric dam at Rincón del Bonete on the Río Negro, which supplies power for Montevideo, created a large lake in the department. (M. I. V.)

TACUTU (TAKUTÚ), a tributary in the Amazon river system. It rises in the Serra Uaçari on the Brazil-British Guiana border and flows northward and then westward and southward to its junction with the Uraricoera river about 20 mi. above Boa Vista, where the two rivers form the Rio Branco. On its northward course the Tacutu flows along the western edge of the Rupununi savannas and forms a section of the Brazil-British Guiana boundary. Upon turning westward it receives the waters of the Ireng (or Maú) river, which also forms part of the Brazil-British Guiana boundary. (D. R. D.)

TAD: *see* DORGAN, THOMAS ALOYSIUS.

TADZHIK (TAJIK), the original Iranian population of Afghanistan and Turkistan. Tadzhik constitute a majority of the population of Soviet Tajikistan (*see* TADZHIK SOVIET SOCIALIST REPUBLIC); in the 1960s they numbered about 1,400,000 in the U.S.S.R., with more than 300,000 in Uzbekistan. They comprised about one-third of the population of Afghanistan, where their communities have been reported near Kabul in the Herat region, and in a number of localities north of the Hindu Kush.

Tadzhik were the heirs and transmitters of the central Asian sedentary culture that diffused in prehistoric times from the Iranian plateau into an area extending roughly from the Caspian sea to the borders of China. Here were built villages of flat-roofed mud or stone houses along streams and rivers that irrigated fields of wheat, barley and millet, and gardens famous for melons and a variety of fruits. Crafts were highly developed, and towns along the caravan routes linking Persia, China and India were centres of trade networks in the area.

Turks, and later Mongols, flowed westward into the area, and Pushtu speakers (Afghans) gradually pushed northward. As these nomadic peoples settled, they adopted Iranian culture, but in Turkistan many of the Iranians adopted Turkic speech.

The name Tadzhik refers to the traditionally settled Caucasoid people who have retained Iranian speech. Most speak dialects of Persian, but the Mountain Tadzhik or Galcha in the Pamir mountains speak dialects of two eastern Iranian languages, Pamir and Yagnobi.

See also AFGHANISTAN: *The People.*

See D. N. Wilber *et al., Afghanistan: Its People, Its Society, Its Culture* (1962); T. Fitzsimmons *et al., U.S.S.R.: Its People, Its Society, Its Culture* (1960). (EL. B.)

TADZHIK SOVIET SOCIALIST REPUBLIC (TADZHIKSKAYA SOVETSKAYA SOTSIALISTICHESKAYA RESPUBLIKA; TADZHIKISTAN, or TAJIKISTAN), a republic of the U.S.S.R. which achieved the status of a federal union republic on Dec. 5, 1929. Situated in the southeast of Soviet central Asia, the republic is bordered on the south by Afghanistan and on the east by the People's Republic of China. Area 55,212 sq.mi. The capital is Dushanbe, which, from Dec. 1, 1929, to Nov. 11, 1961, was called Stalinabad.

Physical Geography.—In general the Tadzhik S.S.R. is high mountain terrain. Its whole eastern part covers the high-level valleys or *pamirs*. The country is traversed by many mountain ranges generally extending east-west, and including the Zeravshanski and Gissarski ranges in the west, and the Peter I (Petra Pervogo), Trans-Alai (Zaalaiski) and North and South Alichurski ranges in the Pamirs (*q.v.*). The north-south Sarykolski range forms the eastern frontier of the republic. Communism peak (formerly Pik Stalina; 24,590 ft.) and Lenin peak (Pik Lenina; 23,405 ft.) in the north of the Pamirs, and Karl Marx peak (22,067 ft.) in the south are the highest mountains in the Soviet Union.

South of Mt. Communism lies the 50-mi. Lednik Fedchenko, the longest continental ice stream in the world. With the exception of the Tadzhik section of the Syr-Darya (Jaxartes) in the Fergana valley, of the Zeravshan river which loses itself in the Uzbek desert near Bukhara, and of the Markansu river which flows into Sinkiang, all the glacier-fed streams are right-bank tributaries of the Amu-Darya (*q.v.*, or Oxus). The Amu-Darya rises in the Afghan Hindu Kush as the Vakhan-Darya and, as the Pyandzh (from Persian *panj*, meaning five; *i.e.*, the five streams which it collects on its way), it forms most of the frontier between Afghanistan and the Tadzhik S.S.R. It then receives the Vakhsh river, which rises in the Kirgiz S.S.R. and is called first Kyzyl-Su ("red water" in Uzbek), and lower down Surkh Ab ("red water" in Persian).

There are many lakes in the Pamirs formed either by glaciers or as a result of avalanches (produced by earthquakes) which dammed existing streams. The largest glacier-formed lake is Kara-Kul, 141 sq.mi. and 780 ft. deep. Among lakes of such tectonic origin the largest and most recent is Sarezkoye Ozero, 38 mi. long and 1,657 ft. deep, formed in 1911.

The country has three types of climate. In the lowlands it is hot and dry, with a yearly rainfall averaging 10 in. and temperatures varying between 31° C. (88° F.) in July and 3° C. (38° F.) in January. In the foothills there is more rainfall and it is cooler. In the high mountains above 10,000 ft. there is only 3 in. of yearly precipitation and the yearly average temperature is below freezing point.

The northwestern lower slopes are well covered with forest. Many varieties of vegetation are represented, from saxaul scrub and jungle grass, through deciduous and coniferous forest to alpine pasture. Animal life is also varied. In the southern lowlands tigers, deer and many kinds of water birds are found. In the higher plateaus there are antelopes and huge lizards up to 4½ ft. long, and in the Pamirs there are bears and mouflons.

History.—The Tadzhik (*q.v.*) peoples are of Persian-speaking Iranian stock. They were included in the empires of Persia and of Alexander the Great of Macedonia. In the 8th century the Arabs crossed the Amu-Darya, conquered Sogdiana (*q.v.*; Transoxiana) which was inhabited by Tadzhiks, and called it Mavera un nahr ("land beyond the river"). Two centuries later the first Turkic invaders (from the northeast) seized Mavera un nahr, and in time, as both conquered and conquerors were Muslim, many Tadzhiks, especially those in the valleys of Syr-Darya and Amu-Darya, became turkicized.

Until mid-18th century the Tadzhiks were part of the amirate of Bukhara (*q.v.*), but then the Afghans conquered lands south and southwest of the Amu-Darya with their Tadzhik population, including the city of Balkh (*q.v.*), an ancient cultural Tadzhik centre. Another partition of the Tadzhik territory took place in 1895 when an Anglo-Russian commission fixed the northern frontier of Afghanistan on the Pyandzh river and separated the Russian and Indian empires by an Afghan corridor on the Vakhan-Darya extending to the Chinese frontier.

After the Russian Revolution of 1917 a considerable part of the Tadzhik people was included in the Turkistan Autonomous Soviet Socialist Republic established in April 1918. On Aug. 23, 1920, the revolution was extended to the khanate of Bukhara, which embraced most of the territory occupied by the modern Tadzhik S.S.R. In Oct. 1920 the Bukhara People's Soviet Republic was declared in being, and early in 1921 the Soviet army captured Dushanbe and Kulyab. From 1922 to the summer of 1923 Tadzhikistan was the scene of the Basmachi revolt, and rebel bands under Ibrahim Bek operated in Eastern Bukhara until 1926. In 1924 the Tadzhik A.S.S.R. was created as part of the Uzbek S.S.R.; in Jan. 1925 a Special Pamirs *oblast* was created out of the Kara-Kirgiz and Tadzhik parts of the Pamirs; and in Dec. 1925 this *oblast* was renamed the Gorno-Badakhshan Autonomous *oblast*. Tadzhikistan was thus composed of the following former administrative units: Eastern Bukhara, 7 volosts (rural districts) of the Samarkand *uyezd* (county), the 5 southern volosts of the Khodzhent (later Leninabad) *uyezd*, and 6 volosts of the Pamirs region. On Dec. 5, 1929, the status of the republic was raised to that of Soviet Socialist Republic.

Collectivization of agriculture was carried out in 1928 and 1929

in the face of considerable opposition. There were later periodical purges of the council of ministers and the party apparatus, including one in 1961.

Population and Administration.—According to the 1959 census the population was 1,979,897, including 1,051,164 Tadzhiks (53.1%), 454,433 Uzbeks (23%), 262,610 Russians (13.3%), 56,893 Tatars (2.9%), 26,921 Ukrainians (1.4%), 25,635 Kirgiz (1.3%), and 12,551 Kazakhs (0.6%). The density of population was 13.9 per sq.km. (about 36 per sq.mi.). The most thickly populated areas are the Leninabad district in the north of the republic and the Gissar and Vakhsh valleys. One-third of the population are town dwellers living in 14 towns and 31 "settlements of town type." The main towns are Dushanbe, Leninabad (*qq.v.*), Ura-Tyube, Kulyab and Kurgan-Tyube.

In the absence of any official statistics it can be assumed that all the inhabitants of the republic other than Russians and Ukrainians are actual or nominal adherents to the Muslim religion.

The government of the Tadzhik S.S.R. is the standard one for union republics. Control over every form of government, economic, social and cultural activity is exercised by the communist party. After the administrative reorganization carried out in 1962 the republic consisted of the Gorno-Badakhshan Autonomous *oblast* (*q.v.*) and a number of *raions* under direct republican administration. Leninabad *oblast* was abolished.

Education and Culture.—The culture of the Tadzhiks, who constitute a little over half of the population, is Iranian, their language closely resembling the Persian spoken in Iran and Afghanistan, although it has been written exclusively in Cyrillic (Russian) characters since 1940. About 27% of the population is made up of various Turkic elements, mainly Uzbek, and these, with the exception of a few small tribes in the south, stubbornly resist Iranization.

Several newspapers are published in the Uzbek language which is taught to Uzbeks attending school. During the Soviet period the Tadzhiks have laid claim to some of the principal Persian poets such as Rudaki and Omar Khayyam. Sa'di, Hafiz and Jami are described as "organically entering" the history of Tadzhik literature.

In the early 1960s there were more than 2,600 schools with more than 330,000 pupils, and about 200 kindergartens with 16,000 children. There were also 80 schools for young industrial workers with 8,500 pupils, more than 100 schools for young agricultural workers with 2,700 pupils, 25 "middle" special training establishments with 4,200 pupils, and 7 higher educational establishments (including the Tadzhik State university, established 1948) with 18,200 students. A branch of the U.S.S.R. Academy of Sciences was established in Dushanbe in 1941 (this became an independent academy in 1951). In Dushanbe there are radio and television transmitters broadcasting in Tadzhik, Russian and Uzbek languages.

The Economy.—The economy of Tadzhikistan is mainly concerned with the growing and processing of cotton. Industry is principally light and apart from cotton ginning and textile weaving it is connected chiefly with food processing. In agriculture the main crop is cotton, but fruit is grown extensively as well as food grains and technical crops. There is a considerable cattle-breeding industry.

In the early 1960s there were about 350 kolkhozes (collective farms) and 40 sovkhozes (state farms) in Tadzhikistan. Half of the total sown area was under irrigation. The main canals are the Vakhsh, the Great Gissar, the Great Fergana, and the Northern Fergana. A large area is under cotton and the yields per acre are among the world's highest.

On the unirrigated lands, principally on the northern slopes of the Turkistan range and in the southeast of the republic, food grains (wheat, barley and millet) are grown and oil grains (mainly flax) are cultivated in the southeast. Stock breeding includes sheep, horned cattle, pigs, karakul sheep, angora goats, Lokai horses and yaks (in the eastern Pamir).

There are cotton-ginning factories in all the cotton-producing valleys, the principal ones being in Dushanbe, Regar and Kurgan-Tyube. Most of the textile weaving is carried on in the Dushanbe cotton *kombinat* (combine). Leninabad is an important centre of the silk-weaving industry. In the early 1960s large factories for the manufacture of carpets and woolen goods were built at Ura-Tyube. Cotton-seed oil processing is allied to cotton ginning, the principal centres being Kanibadam, Leninabad, Kurgan-Tyube and Shaartuz.

There are fruit canneries at Kanibadam, Leninabad, Chkalovski and Regar and wineries at Dushanbe, Leninabad and Ura-Tyube. Minerals that are being exploited include lead and zinc (Kansai), fluorspar (Takob) and brown coal (Shurab); oil is produced at Kim and Nefteabad. In its hydroelectrical resources the Tadzhik S.S.R. occupies the second place in the U.S.S.R. (after the Russian S.F.S.R.), the main power stations being at Kairak-Kum on the Syr-Darya and at Perepadnaya on the Vakhsh canal. During the 1959–65 seven-year plan a gas pipeline from the Uzbek S.S.R. was planned and under construction.

Communications.—There are 258 km. (160 mi.) of standard- and 300 km. (186 mi.) of narrow-gauge railway lines. The main line, connecting the national network, is from Termez to Ordzhonikidzeabad. From Dushanbe to the Vakhsh valley and thence to Kulyab there is a narrow-gauge line. There are main highways between Dushanbe and Khorog, Khorog and Osh, Dushanbe and Ura-Tyube; Kurgan-Tyube and Kulyab also have road links with Dushanbe.

The Amu-Darya is navigable below Faizabadkala. Air transport is important and airlines join Dushanbe with Moscow, Tashkent, Baku and other towns. (G. E. Wr.; K. Sm.)

TAEGU, a city in the Republic of (South) Korea, 60 mi. N.W. of Pusan on a tributary of the Naktong river. It is an important commercial centre for southeastern Korea. Textiles are a major industry, with more than 200 factories engaged in silk spinning and about 100 in cotton spinning. The city is an important market centre for apple growers, and a large cattle market is also held there periodically. In 1950 during the Korean War, Taegu was captured by the Communist forces, but was later recaptured by United Nations troops. In 1960 the population was 678,277.
(S. McC.)

TAEJON, a city in the Republic of (South) Korea, 125 mi. N.W. of Pusan and 90 mi. S. of Seoul. Taejon is a relatively new town constructed as the junction point of a large rail network developed during the Japanese control of Korea (1910–45). This large and important commercial and industrial centre suffered extensive battle damage during the Korean War, but has been largely rebuilt. Textile factories and consumer goods industries are Taejon's major sources of employment. Its central location in southern Korea makes it a rail and air centre. The population in 1960 was 229,393. (S. McC.)

TAEL. The tael (Chinese, *liang*), a Chinese unit of weight, when applied to silver was long used as a currency unit. There have been many taels of varying weight, generally in the neighbourhood of 1⅓ oz.

The currency tael was seldom minted in the form of a coin, but served as a standard money of account, actual transactions being completed with ingots of silver, with bank notes or checks expressed in taels, or with silver coins, especially the Spanish and Mexican dollars which flowed into China in great volume in the 18th and 19th centuries.

Bar silver imported into China was remelted and cast in ingots weighing about 50 taels, of a special shape called "shoes" (from the Dutch word *schuyt* meaning boat) of *sycee* (literally, fine silk, hence, fine silver). Each shoe was marked by the public assaying office with its weight and fineness, from which data its value in currency taels was computed. Sycee formed a large part of bank reserves down to 1933.

The most important currency tael was the Shanghai tael, adopted in 1857 to replace the Spanish dollar which was becoming scarce. The fine-silver equivalent was 518.512 grains. Since China was on a silver standard while the rest of the world was on a gold standard, the exchange value of the Shanghai tael fluctuated with the price of silver in London and New York, ranging from around 6*s.* or $1.50 before 1873 to as low as 1*s.* 2*d.* or 28 cents in 1931. It was the basis for wholesale trade and for-

eign exchange transactions in China's most important commercial city.

Of the numerous other taels used as moneys of account at various times and places, two deserve mention. The *k'uping* or treasury tael was used under the Manchu dynasty for the payment of taxes; 100 *k'uping* taels equalled 109.60 Shanghai taels. The *Haikwan* or customs tael was adopted in 1858 for the valuation of imports and exports and for the fixing of duties; 100 *Haikwan* taels equalled 111.40 Shanghai taels.

From ancient times the money of the masses for small transactions was the cash (*q.v.*) but from the latter part of the 19th century retail trade gradually began to be conducted with Mexican and, later, Chinese silver dollars, fractional silver coins, and 10-cash copper pieces.

Finally, in April 1933, the tael was officially abolished and replaced by the new Chinese standard dollar or *yuan* at a fixed rate of Shanghai taels 71.50 equal 100 yuan. The new coin weighed 412 grains, 0.880 fine, and its exchange value, of course, fluctuated with the world price of silver.

See CHINA: *The Economy: Trade and Finance.*

(D. H. L.; X.)

TAENIODONTA, an extinct order of North American mammals of the Paleocene and Eocene epochs (a period extending from 40,000,000 to 75,000,000 years ago). The single known family Stylinodontidae includes two subfamilies, Conoryctinae and Stylinodontinae. Conoryctinae, rather generalized forms with no outstanding peculiarities, did not survive the Paleocene. Stylinodontinae became progressively and more curiously specialized, culminating in forms with extremely short and broad skulls; deep, massively built lower jaws; rootless teeth; and limbs of the digging type, tipped by very large, long, laterally compressed claws. The canines were the outstanding features of cranial and dental evolution, the rest of the dentition becoming accessory and to a degree subordinate to them in a manner unique among mammals. Claws and teeth together suggest adaptation to a specialized diet, the nature of which is unknown. Some superficial resemblance exists between taeniodonts and ground sloths, but this is certainly due to similar habits and not to common ancestry. Taeniodonts may represent an independent and isolated offshoot of the earliest Insectivora.

See also MAMMAL; PALEONTOLOGY.

See A. S. Romer, *Vertebrate Paleontology,* 2nd ed. (1945); B. Patterson in G. L. Jepsen *et al.* (eds.), *Genetics, Paleontology and Evolution,* pp. 243–278 (1949). (B. PA.)

TAFFETA, a term strictly denoting a light, thin and plain texture of silk possessing a high sheen or gloss and finished in a manner that imparts to it the "rustle" or "scroop" peculiar to certain types of closely woven, plain silk textures produced from spun silk yarn. The warp and weft may be of corresponding denier or counts, approximately, while the number of warp threads and picks per inch may also correspond.

Taffeta is sometimes embellished with a very small brocade or "float" figuring of warp or weft, and may either be woven from skein-dyed yarn, or else piece-dyed after weaving. Some qualities of taffeta fabrics are artificially "loaded," *i.e.,* weighted, with metallic salts, and are therefore more liable to split or crack when creased or folded. It is used chiefly for dress, millinery and similar purposes.

The word "taffeta," however, is now applied as a description of many other varieties of fabrics produced from silk, wool, cotton and combinations of these materials.

Chiffon taffeta is of a softer texture and finish than that of ordinary taffeta. One variety of taffeta is produced from a silk warp and Botany wool weft; another variety, described as "wool taffeta," is woven entirely of fine wool yarn both for warp and weft, which makes a fine and close texture of light weight, with a smooth surface.

Some varieties of taffeta are produced from a cotton warp and fine Botany weft, while others are woven entirely with twofold cotton yarn of fine counts, and finished with a stiff "finish" to simulate the "scroop" peculiar to silk fabrics.

TAFILÁLT or TAFILET (*i.e.,* "The Country of the Filáli," as its inhabitants are called, because descended from the Arabian tribe of Hilál, settled there in the 11th century), important oasis of the Moroccan Sahara, ten days' journey south of Fez, across the Atlas. It is celebrated for its dates, to the successful cultivation of which, soon after the arrival of an ancestor of the reigning dynasty of Morocco (hence called the Filáli Sharifs, *i.e.,* descendants of Mohammed) *c.* A.D. 1250, the dynasty owed its rise.

Tafilált is a succession of ksurs (fortified villages) and palm groves which correspond to the zone irrigated by the Wad Ziz, and stretch over a length of about 50 km. The largest centre is Bu-Am, an important market.

The sedentary natives, who were exploited by the Aït-Atta nomads, live in poverty on their palm groves. The industry is the tanning of goatskins to obtain *filali* leather, which has a great reputation. The *takaut* (*Tamarix articulata*) yields gall for tanning. After 1648 it was the custom of Moorish sultans to dispatch superfluous sons and daughters to Tafilált. In Ifli, the central portion, formerly existed the town of Sigilmasa, founded by Miknása Berbers in 757 B.C. It was on the direct caravan route from the Niger to Tangier, and attained a considerable degree of prosperity.

The first European to visit Tafilált was Rene Caillie (1828), the next Gerhard Rohlfs (1864). A later visit to the oasis by W. B. Harris is described in his book *Tafilet* (London, 1895).

TAFT, LORADO (1860–1936), U.S. sculptor, lecturer and teacher, was born at Elmwood, Ill., on April 29, 1860. He graduated from the University of Illinois, Urbana, where he first worked in sculpture, in 1879, and then studied at the École des Beaux-Arts in Paris until 1883, remaining in that city with but brief interruption until 1885. Establishing himself in Chicago in 1886, he taught sculpture at the Art institute and began his long career of public lecturer. He was elected to the National Academy in 1911. He participated actively in the formation of plans for public education in art, and served (1914–17) as director of the American Federation of Arts.

Aside from many portraits, included among his works are: "Sleep of the Flowers" and "Awakening of the Flowers" for the Columbian exposition (1893), "Despair" (1898), "Solitude of the Soul" (1900), the colossal "Black Hawk" at Oregon, Ill. (1912), Thatcher Memorial fountain, Denver, Colo. (1918), and the "Fountain of Time" (1920), intended as part of a monumental complex of sculpture in Chicago.

In 1903 he published *The History of American Sculpture,* the first comprehensive work on the subject. His *Some Modern Tendencies in Sculpture* was published in 1920.

He died at Chicago on Oct. 30, 1936. (J. C. T.; A. T. G.)

TAFT, ROBERT ALPHONSO (1889–1953), United States senator from Ohio and son of William Howard Taft, 27th president of the United States, was born in Cincinnati, O., on Sept. 8, 1889. He attended the public schools of Cincinnati and the Taft school at Watertown, Conn.; was graduated from Yale university with a B.A. degree in 1910 where he stood first in his class; graduated from Harvard university with an LL.B. degree in 1913 where he was first in his class and editor in chief of the *Harvard Law Review.* In 1913 he passed the Ohio bar examination with highest honours in the state. In 1914 he married Martha W. Bowers.

After practising law in Cincinnati, Taft served during World War I, 1917–18, as assistant counsel for the United States Food administration. In 1919 he was counsel for the American Relief administration in Europe. From 1920 through 1938 he practised law in Cincinnati, establishing a reputation as one of the ablest lawyers in Ohio. During this period, 1921–26, he was a member of the Ohio house of representatives where he became Republican floor leader and speaker of the house. He served in the Ohio senate in 1931 and 1932.

Shortly after his election to the United States senate in 1938, Taft became a national political figure. He was re-elected to the senate in 1944 and re-elected again in 1950 by a majority of 431,000 votes, one of the largest ever given a candidate for U.S. senator from Ohio. With each U.S. senator being responsible only to one of 48 constituencies, leadership in the senate is often more titular than real. Taft, however, excelled in the essential political

art of compromise. His exceptional intellect and almost legendary capacity for work gave him a mastery of legislative facts far beyond the narrow range of committee specialization. His honesty, integrity and patriotism were never responsibly questioned. His Republican colleagues in the senate selected him in 1947 as chairman of the senate Republican policy committee, a position he held until his election as senate majority leader in Jan. 1953. It is difficult for anyone outside the congress to evaluate precisely the worth of one of its members. Much of the actual work is done in closed committee meetings and in informal conferences. The congress itself is the only point at which all the myriad facts and pressures coalesce. Perhaps the most meaningful tribute to Senator Taft's stature is that he was regarded by his colleagues, Democrats as well as Republicans, with unprecedented respect and esteem.

Taft lacked many of the traits commonly associated with a good politician in the United States. He was thought by many to be cold and austere. His genial personality and gentle humour were reserved for friends and for direct personal contacts. He had no trace of demagoguery. His appeal was to logic, rarely to emotion. He championed unpopular causes. He disdained the use of wit, eloquent speech and brilliant metaphor. He spoke bluntly, armed only with facts and simple statements of his principles. His willingness to make an unequivocal statement on any controversial issue frequently embarrassed his political managers. Typical was his forthright condemnation of the Nuremberg trials as vengeance by the victors decked out in the trappings of justice. Whether or not enemy leaders were tried under *ex-post facto* laws was not an issue that would ever come before the senate. His bold expression of his convictions contrary to prevailing public opinion flouted all conventional norms of good politics.

Taft was generally regarded as a conservative in American politics. In recognition of his leadership of the more conservative political party, he was nicknamed "Mr. Republican" by the American press. But even in the field of social-welfare legislation, he was not inflexibly wedded either to the status quo or to his former position. He was an original sponsor of legislation providing for public housing for low-income groups. He sponsored legislation authorizing federal aid to state governments for the improvement of medical care and education and of the living standards of the aged.

On the other hand, Taft was an uncompromising opponent of legislative measures which, in his judgment, undermined personal incentives for economic betterment or demanded excessive government intervention in the lives of its citizens. He consistently opposed compulsory health insurance on the ground that government control of the medical profession made the promised benefits illusory. He opposed peacetime price and wage controls. His political philosophy was perhaps best stated in 1948 when he described the "one great issue in the world today" in these words: "It is that of government operated by a free people on the basis of freedom and justice for every individual, on the one hand, against the totalitarian state on the other—the kind of state which concentrates in a central government power to direct the lives of all its people, its agriculture, its commerce and its industry." Taft often defined the basic domestic issue in American life as one of "liberty versus socialism."

It was in the field of labour law that Taft won his greatest legislative fame. In the nation-wide transportation strike of 1946, Pres. Harry S. Truman made a personal appeal to the congress for legislation authorizing him to draft striking workers into the army "on such terms as he might impose." The house of representatives passed the bill immediately, and the senate seemed likely to follow suit. Notwithstanding the intense pressure for drastic action, Taft blocked it in the senate by utilizing its rule of unlimited debate.

When the Republican party won control of congress in the elections of 1946, Taft had the choice of becoming chairman of either the senate finance committee or the senate labour committee. Typically, he chose the more politically explosive assignment. As chairman of the labour committee he guided to passage over a presidential veto the Labor-Management Relations act of 1947, popularly known as the Taft-Hartley law. He defended it as imposing on labour unions obligations comparable to those imposed on management. Labour union leaders, almost without exception, denounced it as a "slave labour" law. They mobilized their forces to defeat him in 1950. No senatorial campaign in U.S. history attracted so much nation-wide interest. Taft's overwhelming victory was interpreted by his followers as proof that no economic group in America is cohesive enough to decide the outcome of an election through bloc voting on isolated issues.

Prior to United States participation in World War II, Taft opposed measures calculated to increase the chances of involvement. For that stand, he was frequently criticized as an "isolationist." After the war he recognized the United States' new position of world leadership. He urged ratification of the United Nations charter. Because of Soviet Russia's intransigence, however, Taft had little faith in the U.N. as a shield against aggression. In opposition to the Truman administration and many Republicans as well, he advocated reduced foreign aid expenditures, greater emphasis on retaliatory air power and less on ground defense forces in western Europe, increased attention to far eastern problems, and congressional approval of executive decisions to send American troops abroad.

Taft was three times an unsuccessful candidate for the Republican nomination for president. In the Republican national convention of 1940, he lost the nomination to Wendell Willkie; in 1948 to Thomas E. Dewey; and in 1952 to Dwight D. Eisenhower. Despite the bitterness of the prenomination struggle in 1952, Taft and his followers united wholeheartedly behind Eisenhower. It was widely predicted that the differences between Eisenhower and Taft, particularly on foreign policy issues, would make close cooperation between the White House and the Capitol impossible. Taft, however, was determined to make the first Republican administration in 20 years an unqualified success. Most observers agreed that he was eminently successful in bridging the constitutional gap dividing executive and legislative authority which is so essential in the American ideal of freedom and yet so often productive of discord and inaction.

Taft died of cancer in New York city on July 31, 1953. His body lay in state in the rotunda of the Capitol—an honour accorded to only 12 others. He was buried near his home in Cincinnati in the cemetery of the Indian Hill church.

BIBLIOGRAPHY.—*See* indexes for the *Congressional Record*, vol. 84–99, under heading "Taft, Robert A. (a Senator from Ohio)"; Robert A. Taft, *A Foreign Policy for Americans* (New York, 1951). (J. W. BR.)

TAFT, WILLIAM HOWARD (1857–1930), the 27th president and tenth chief justice of the United States, was born in Cincinnati, O., on Sept. 15, 1857. His father, Alphonso Taft, was attorney general in Pres. U. S. Grant's cabinet (1876–1877) and minister to Austria (1882–84) and to Russia (1884–85). William Howard Taft graduated second (salutatorian) in his class at Yale college in 1878, and at the law school of Cincinnati college, in 1880, dividing the first prize for scholarship. In the same year he was admitted to the Ohio bar. In 1881 he was appointed assistant prosecuting attorney of Hamilton county (in which Cincinnati is situated), but resigned in 1882 on being appointed U.S. collector of internal revenue for the first district of Ohio.

In 1883 he returned to the law. From 1885 to 1887 he served as assistant solicitor of Hamilton county, and in the latter year was appointed judge of the superior court of Ohio, to fill a vacancy, being elected in the next year. In 1890, he was appointed solicitor-general of the United States by Pres. Benjamin Harrison. In 1892 he was appointed U.S. circuit judge for the sixth circuit.

From 1896–1900 Taft also served as professor and dean of the law department of the University of Cincinnati.

Governor of Philippines.—In 1900 he was appointed by Pres. William McKinley to the presidency of the Philippine commission. He served as such from March 13, 1900, to Feb. 1, 1904, becoming governor ex officio on the establishment of civil government, July 4, 1901 (*see* PHILIPPINES, REPUBLIC OF THE). The delicate matter of confiscated church lands was arranged by Taft in an interview with Pope Leo XIII, in the summer of 1902. The pope sent a special delegate to appraise the

lands, and the sum of $7,239,000 was paid for them in Dec. 1903.

Secretary of War.—In Feb. 1904, Taft returned to the United States to become secretary of war. In Sept. 1906, on the downfall of the Cuban Government and the intervention of the United States he took temporary charge of affairs in that island (Sept.–Oct.). In the next year (March–April) he inspected the Panama canal and also visited Cuba and Porto Rico. He again visited the Philippines to open the first legislative assembly (Oct. 16, 1907). On this tour he visited Japan, and on Oct. 2, at Tokyo, made a speech which had an important effect in quieting the apprehensions of the Japanese on the score of the treatment of their people on the Pacific coast. While secretary of war, Taft entirely reorganized the work of carrying on the construction of the Panama canal, and selected Col. George W. Goethals, whom President Roosevelt put in entire charge of the work. From that moment the enterprise assumed a new vigour, the work was completed and was opened to traffic in 1913.

President.—In 1908 Taft was favoured by President Roosevelt as his successor, and in the November election, was victorious over William Jennings Bryan, the Democratic nominee, by a popular majority of 1,269,900 votes (Taft, 7,679,006; Bryan, 6,409,106) and by a majority of 159 votes in the electoral college. In accordance with his pre-election pledge, he called the Congress to meet in extra session on March 15, 1909, to revise the tariff. The final bill known as the Payne-Aldrich Act was approved by the president on Aug. 5, 1909, although in many respects it was not the measure he desired. The income of the Government under the Dingley tariff had proved inadequate to meet its current expenses, the estimated deficit for the fiscal year ending July 1, 1909, being $100,000,000. He therefore recommended to the Congress the adoption of resolutions to amend the Constitution by expressly granting Congress power to levy an income tax without apportionment among the several States. For the purpose of meeting the immediately anticipated deficiency in the revenues, he recommended the adoption of a provision in the Tariff Bill imposing an annual special excise tax upon corporations organized for profit, measured by a percentage on their net incomes. Pursuant to these recommendations a proposed amendment to the Constitution was passed by both Houses of Congress, afterwards ratified by the requisite number of the States, and became the 16th amendment, and there was included in the Payne-Aldrich Act a provision imposing the proposed annual special excise tax upon corporations, measured by 1% of their net incomes.

Payne-Aldrich Tariff Act.—While the Payne-Aldrich Act was much criticized, because it failed to reduce duties on imports to the extent expected, yet it marked a distinct step towards a lower scale of import duties, and opened the door to the free exchange of commerce with the Philippines. By the enactment of the maximum and minimum tariff provisions, it placed a powerful weapon in the hands of the American Government for the protection of American commerce against unduly discriminatory practices by foreign Governments. It also, for the first time in American history, provided for the establishment of a tariff board to study the questions which lie at the foundation of the protective system.

While Congress was considering the Tariff Bill, the president called upon the heads of the departments of Government to prepare, in advance of the usual time, their estimates for the fiscal year ending June 30, 1911, admonishing them to reduce expenses wherever it could be done. These estimates were then considered by the president and cabinet, and for the first time in American history, a comprehensive budget was prepared. As a result of these efforts, a cut was made of about $55,000,000 in the estimates for the fiscal years 1910 and 1911. In June 1910, the Congress granted the president's request for an appropriation to enable him to employ competent persons to study the methods of transacting the public business of the executive departments and other Government establishments, and to recommend such legislation as might be necessary to carry into effect changes found to be desirable that could not be accomplished by executive action alone. President Taft caused a commission to be organized, and as a result of its studies, sent to Congress four special messages transmitting the reports of the commission, dealing with particular subjects, pointing out defects in the organization and conduct of the business and changes that might be made advantageously.

Federal Budget Urged.—In his message of Jan. 17, 1912, the president called attention to the fact that the United States was the only great nation whose Government was operated without a budget, and recommended the adoption of an annual budget. Nothing was done by Congress in these matters, but many, if not most, of the improvements in the organization and conduct of the administrative branch of the Government which have been adopted in more recent years were first formulated and recommended by President Taft. He also recommended to Congress the enactment of amendments to the interstate commerce law, the establishment of a postal savings bank system, the enactment of laws giving the interstate commerce commission the power to determine upon the uniform construction of appliances used by railway trainmen in the operation of trains and to intervene and investigate the reasonableness of a proposed increase in railway rates before they became effective; of amendments making the Employer's Liability Act more easy to enforce; of laws carrying out with intelligent discrimination the principle of conserving the national resources; the establishment of a national health bureau; requiring publicity of campaign contributions; regulating the procedure of Federal courts concerning the issue of injunctions; and readjusting the regulations concerning postage of second class mail matter. Most of these recommendations were embodied in laws passed by the Congress. By executive order, the President placed the positions of secretaries in the diplomatic service in much the same position as those in the consular service, and introduced into both the principle of appointment for reasons of ascertained fitness and promotion by merit.

Arbitration Treaties.—In 1910 and 1911, President Taft negotiated with the Governments of France and Great Britain identical treaties, signed on Aug. 3, 1913, intended to be models for a series of treaties providing firstly, for the arbitration of all differences which might arise between the parties which it was not possible to adjust by diplomacy, and which were justiciable in their nature, and secondly, for the creation of a joint high commission of inquiry, to which should be referred for impartial investigation, any justiciable controversy between the parties before submission to arbitration, and also any other controversy, even if the parties were not agreed that it fell within the scope of the first article; the report of such commission not to be regarded as a decision on the questions submitted. The first of these provisions embodied in effect the agreement reached in the treaty between the United States and France of Feb. 6, 1928. The second furnished a model for the treaties negotiated by Secretary Bryan in 1914, known as the Bryan Conciliation Treaties. Together, they represented a greater advance in provisions for the peaceful settlement of international controversy than any the world had known. The U.S. Senate qualified its approval by such extensive reservations, that the president withdrew the treaties from further consideration. In 1910, President Taft initiated negotiations with the British Government for the establishment of a reciprocal trade agreement between the United States and the Dominion of Canada. Instead of embodying the desired tariff changes in the formal shape of a treaty, it was agreed that the Governments of the two countries should use their utmost efforts to bring about such changes by concurrent legislation at Washington and Ottawa. It was further agreed that a considerable list of articles produced in both countries should be reciprocally free. As to another group, common rates of duties upon importation from either country into the other were fixed. The proposed legislation was, however, defeated in the dominion parliament, and the effort failed.

As the construction of the Panama canal approached completion, a bill was introduced in Congress to provide for its operation and for the government of the Canal Zone. Taft was of the opinion that as the United States had built the canal at its own

expense, it was entitled to permit the use of the canal by American merchant ships without payment of tolls. Great Britain claimed that this would be a violation of the Hay-Pauncefote Treaty. While not admitting the contention, the president expressed his willingness to arbitrate the question. The bill, as passed, embodied the president's views, but the next Congress modified the act by removing the discrimination.

Other Events in Administration.—President Taft's administration was characterized by a systematic and vigorous enforcement of the antitrust laws. The decisions by the U.S. Supreme Court in 1911, of the prosecutions which had been initiated by Pres. Theodore Roosevelt against the Standard Oil trust and the American Tobacco combination, and which were argued in the Supreme Court by Atty. Gen. George Wickersham, gave an authoritative interpretation to the act, which was at once put into effect by the Department of Justice, and for the first time after the enactment of the law in 1890, a careful, systematic effort was made by the Government to enforce its provisions. Both the oil trust and the tobacco combination were dissolved in such manner as to restore healthy competition in those industries without destruction of the value of their securities held by innocent parties. A large number of combinations in restraint of trade were either put an end to by judicial process, or voluntarily desisted from a continuance of the practices which had been adjudged to be illegal.

During Taft's administration, many vacancies occurred in the Federal judiciary. He was called upon to appoint six justices of the Supreme Court, including a chief justice. On the death of Chief Justice Fuller, in 1910, President Taft nominated as his successor, Associate Justice Edward D. White, a Democrat, a former Confederate soldier, and a Roman Catholic—a nomination which was confirmed by the Senate at once.

Conservation of the natural resources of the nation was one of the subjects which had greatly occupied the attention of the president and his administration. While Taft was in full sympathy with the objects of this policy, he did not approve of some of the means adopted by that administration. In making up his cabinet, he did not include James R. Garfield, secretary of the interior under President Roosevelt, and on Jan. 7, 1910, he dispensed with the services of Gifford Pinchot, the forester. These men, with a number of other adherents of Roosevelt, initiated a campaign against President Taft, based upon the representation that he was opposed to the conservation policy of Roosevelt. Their first attack upon his Administration was directed against Richard Ballinger, who had succeeded Garfield as secretary of the interior. A complaint by L. R. Glavis, an agent in the land office in the department of the interior, of Ballinger's action with regard to certain coal land entries in Alaska known as the Cunningham claims, which President Taft on investigation held to be unfounded, and which led to the dismissal of Glavis from the service, gave rise to a Congressional investigation into the conduct of the interior department by Ballinger. After prolonged hearings a report was made by the committee exonerating Ballinger, but failing to win the confidence of the public, he resigned in the spring of 1911.

In the regular Republican convention at Chicago, Ill., the total number of delegates summoned was 1,078, with 540 necessary to a choice. Taft had 561 votes on the first and only ballot and was declared the nominee. Roosevelt and his progressive supporters, claiming that part of his delegates had been stolen through the action of a hostile committee, refused to accept the results of the convention, organized one of their own, known as the "Bull Moose" convention, nominated Roosevelt for the presidency and carried on an aggresive campaign against Taft, which resulted in splitting the Republican vote in almost every State, and thus electing Woodrow Wilson, the Democratic nominee, as president. The electoral vote was: for Wilson, 435, Roosevelt, 88 and Taft, 8. The popular vote, however (Wilson, 6,285,214; Roosevelt, 4,126,020; Taft, 3,483,922), showed a much closer contest. This Party schism served in large measure to obscure the merits of President Taft's administration.

Subsequent Career.—On retiring from the presidency in 1913, Taft became Kent professor of law at Yale, but devoted much time to lecture engagements. In the same year he was elected president of the American Bar association, and in 1914 first president of the American Institute of Jurisprudence, organized to improve law and its administration. Taft was an active promoter of the League to Enforce Peace, but after U.S. entrance into World War I, he supported the Administration, taking the ground that victory was necessary to the attainment of lasting peace. In 1918, he was appointed by President Wilson a member of the national war labour board for arbitrating labour disputes. In 1919, he endorsed the peace treaty of Versailles, regarding its most important part to be the Covenant of the League of Nations. He spoke throughout the country on behalf of the League urging reservations if these would secure ratification. On the death of Chief Justice Edward White, on June 30, 1921, he was appointed chief justice by Pres. Warren G. Harding. He died in Washington on March 8, 1930.

In 1886 he had married Helen Herron (1861–1943).

Taft is the author of *Popular Government; Its Essence, Its Performance, and Its Perils* (1913); *The Anti-Trust Act and the Supreme Court* (1914); *The United States and Peace* (1914); *Ethics in Service,* Yale lectures (1915); *Our Chief Magistrate and His Powers,* Columbia lectures (1916) and *The Presidency: Its Duties, Its Powers, Its Opportunities and Its Limitations,* lectures at the University of Virginia (1916). (G. W. Wi.)

TAGALOG, the second largest ethno-linguistic group in the Philippines, inhabiting central and southern Luzon and numbering, in 1948, 3,730,028; they are considered the most cultured and westernized Filipino, a pre-eminence resulting largely from their proximity to Manila, the great city of the Philippines which was the focal point of more than 300 years of Spanish contact and a half-century of American influence. Literacy among the Tagalogs in either English or the vernacular is the highest in the Philippines, exceeding 70%. A majority of the national leaders in all fields have come from this group. In 1940 Tagalog became the basis for the Philippine national language and by 1948, persons able to speak Tagalog numbered 7,126,913.

The Tagalogs are primarily rural agriculturists, rice and coconuts being the major crops, and they generally live in barrios (small villages) of from 100 to 700 persons. Nevertheless, many Tagalogs are found in the urban centres, providing a professional and business class.

At least 83% of the people in the Tagalog area are Roman Catholics; the remainder are affiliated with the Philippine Independent Church (about 7%), with one of a number of Protestant churches or evangelical groups (total about 3%) or with some other religious body. The elementary family and the sibling group are the basic units of their social organization and these have marked solidarity as reflected in the Philippine Civil Code. Life in the characteristic rural areas is highly co-operative, being based upon primary group relationships and a system of reciprocal obligations. (R. B. Fx.)

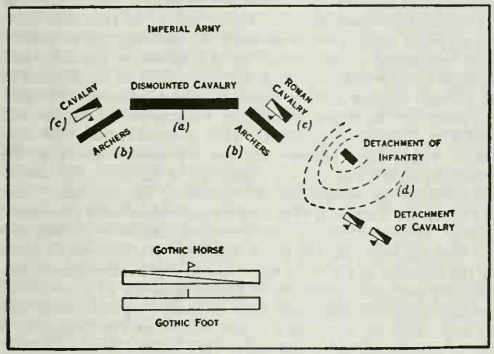
TAGANROG, a Russian seaport in the Rostov *oblast,* Russian Soviet Federated Socialist Republic, U.S.S.R., on the north shore of the northeastern arm of the Sea of Azov. Pop. (1956 est.) 189,000. It is connected by rail with the north, and also with Rostov-on-Don to the east. A colony was founded on the site by Pisan merchants in the 13th century, but was destroyed by the Mongols. Later it was occupied by the Turks and annexed to Russia in 1769.

TAGES, a minor Etruscan deity, the grandson of Jupiter and founder of the art of divination in Etruria. According to the story, during the plowing of a field near Tarquinii a being of boyish appearance sprang out of the furrow. The shouts of the plowman (Tarchon) brought a crowd to the spot and later all the people of Etruria, whom the boy proceeded to instruct in the art of divination. Having done this, he suddenly disappeared. His instructions were committed to writing and formed the 12 books of Tages, containing a complete system of Etruscan lore.

TAGINAE, BATTLE OF (A.D. 552). The battle of Taginae was fought in July 552. It terminated the Gothic Wars of the Emperor Justinian, and was won over the Gothic King

Baduila by the eunuch Narses, a general who in ability rivalled the great Belisarius, his contemporary. Taginae is situated below the Apennines near modern Gubbio. Narses had crossed the head of the valley of the Chiascio when Baduila arrived and seized its main outlet. The Goth ranged the whole of his horsemen in front, his infantry, mostly archers, he drew up in second line. His aim was to ride down the imperial army in one terrific charge.

Belisarius had noted the weakness in the Gothic tactics, namely, that there was no co-operation between lance and bow. Narses



BATTLE OF TAGINAE (JULY, A.D. 552), WON BY NARSES, GENERAL OF THE EMPEROR JUSTINIAN, OVER BADUILA, KING OF THE GOTHS

now aimed at this co-operation, but in a novel form which closely resembles the combined tactics made use of by Edward III. at Crécy. He dismounted his cavalry (*Foederati*) and formed a phalanx 8,000 strong, on the wings of which he drew up 8,000 archers with their outer flanks thrown forward. In rear of these he posted his Roman cavalry to combine fire and shock. His left flank he rested on a small hill, which he occupied by a force of infantry and at the foot of which he deployed a few squadrons of cavalry.

At noon Baduila, ignoring the archers, suddenly charged his enemy's centre, expecting to overwhelm it, for infantry he held in contempt. He was met by a tremendous converging fire of arrows which slew and dismounted hundreds of his knights. After several assaults he clinched with the imperial centre—its protective base of action—which held him firm whilst his flanks were plied with arrows. At length, exhausted, the Gothic horsemen broke back, when immediately Narses launched his mounted squadrons driving them in confusion on to their infantry which were ridden over. The victory was complete.

BIBLIOGRAPHY.—Procopius, *History of the Wars* (Loeb edition); Oman, *The Art of War in the Middle Ages* (1924); *Cambridge Mediaeval History*, vol. ii.; Gibbon, *Decline and Fall of the Roman Empire*; Finlay, *Greece under the Romans*. (J. F. C. F.)

TAGLIACOZZO, a town of Abruzzi e Molise, Italy, in the province of L'Aquila, 56 mi. by rail E.N.E. of Rome and 10 mi. W. of Avezzano. Pop. (1951) 3,903. It lies 2,428 ft. above sea level, at the mouth of the deep ravine of the Imele. It contains several old churches, notably S. Francesco, with a fine rose window in the façade, and medieval houses.

The palace, built at the end of the 14th century by the Orsini, is fine, with interesting paintings in the interior. At the end of 1268 a battle took place here between Conradin of Hohenstaufen and Charles of Anjou, which resulted in the defeat of Conradin.

Tagliacozzo is a resort and has felt and hosiery mills.

TAGORE, SIR RABINDRANATH (1861–1941), Indian poet and author, was born in Calcutta on May 6, 1861, the youngest son of Maharshi Devendranath and grandson of Prince Dwarkanath Tagore. After a private education in India he was sent to England in 1877 to study law, but soon returned to India, and while still quite young commenced writing for Bengali periodicals. In 1901 he established the famous Santiniketan, a school at Bolpur, 93 mi. from Calcutta, which developed into an important educational institution conducted on unconventional lines. In 1913 Tagore was awarded the Nobel prize for literature, and utilized the amount, £8,000, for the up-keep of his school. He visited Europe on several occasions and travelled also in Japan and the United States. He accepted a knighthood in 1915, but in 1919 resigned it as a protest against the methods adopted for the repression of disturbances in the Punjab. In later years, however, he offered no objection to the use of this title. Tagore was interested in politics only in so far as it concerned the deeper life of India, and he desired that the nationalist movement consider social reforms before political freedom. By his abundant writings, which are permeated by a sense of the beauty of the universe, by a love of children and of simplicity, and by a consciousness of God, Tagore did much to interpret for the west the more serious reflections of the people of Bengal. In 1929 he took up painting. He died Aug. 7, 1941.

His most important works which have been translated into English are *Gitanjali* (1913); *The Crescent Moon* (1913); *Chitra* (1914); *The Post-Office* (1914); *The Gardener* (1914); *One Hundred Poems of Kabir* (1915); *Fruit-Gathering* (1916); *Stray Birds* (1916); *Nationalism* (1917); *My Reminiscences* (1917); *The Parrot's Training* (1918); *The Home and the World* (1919); *Red Oleander* (1924); *Broken Ties* (1925); *Letters to a Friend* (1928); *The Religion of Man* (1931). *See* E. Thompson: *R. Tagore, Poet and Dramatist* (Oxford, 1926).

TAGUS, the second longest river of the Iberian peninsula. Its length is 565 mi., of which 171 are on or within the frontier of Portugal, and the area of its basin is 31,213 sq.mi. The basin is comparatively narrow, and the Tagus, like the other rivers of the Iberian tableland, generally flows in a rather confined valley, often at the bottom of a rocky gorge, as at Toledo and Alcántara, below the general level of the adjacent country. The river rises on the western slope of the Muela de San Juan (5,312 ft.), a mountain which forms part of the Sierra de Albarracin, 88 mi. E. of Madrid. Thence the Tagus flows at first north-westwards, but, after receiving the Gallo on the right, it flows west, and then south-west or west-south-west, which is its general direction for the rest of its course. Regular river navigation begins only at Abrantes, a few miles below which the Tagus is greatly widened by receiving on its right bank the impetuous Zezere from the Serra da Estrella. Passing Santarem, the highest point to which the tide ascends, and the limit of navigation for large sailing vessels and steamers, the river divides below Salvaterra into two arms which enclose a deltaic formation, a low tract of marshy alluvium known as the Lezirias, traversed by several minor channels. Both branches terminate in a broad tidal lake immediately above Lisbon (*q.v.*). The Tagus estuary, though partly blocked by a bar of sand, is one of the chief harbours of south-western Europe.

TAHITI, the largest and most important of the Society Islands (part of French Polynesia) in the central South Pacific ocean. The island lies in lat. 17° 38′ S. and long. 149° 25′ W. It is 33 mi. long, northwest to southeast, and consists of the deeply dissected cones of two ancient volcanoes joined by a narrow isthmus of low land, the whole fringed by a coral barrier reef lying from 1 to 2 mi. offshore. The northern cone is roughly circular in plan, about 20 mi. across and rises to a height of 7,339 ft. Numerous secondary cones rise from its ravined sides and there is a narrow plain encircling the base. To the southeast and joined to the larger cone by an isthmus about ¾ mi. wide, the second cone rises to a height of 4,341 ft. The oval-shaped peninsula of Taiarapu has a northwest-southeast length of 14 mi. and a breadth of 8 mi. For a further account of its physical features, *see* PACIFIC ISLANDS.

Tahiti has an area of 388 sq.mi. and a population (1956) of 36,326. The chief town and seat of the French administration is Papeete at the northwestern end of the island. Its commune population in 1956 was 17,247. All the inhabitants are French citizens (1,000 of metropolitan origin) except for the Chinese (5,242) and some other foreigners (300).

The island produces coconuts, bananas, sugar cane and other tropical plants. The chief export is copra of which 21,570 tons were exported in 1954.

History.—Tahiti seems to have been peopled in about the 14th century by Polynesians from the neighbouring island of Raiatéa. The first European to visit it was the Englishman Samuel Wallis (1767). Then came the Frenchman L. A. de Bougainville (1768) and the Englishman James Cook (*q.v.*; 1769, 1773 and 1777). The first English missionaries landed in 1796 but they were unsuccessful. Interisland wars were mainly responsible for the failures though the attempts were favoured by the support of the local rulers, Pomare I and Pomare II. The latter fled with the missionaries in 1808, but returned with them in 1812, having renounced heathenism, and regained his power in 1815. For a time missionaries made good progress—a printing press was established (1817), and coffee, cotton and sugar were planted (1819); but soon there came a serious relapse.

Pomare II died in 1824. His successor, Pomare III, died in 1827, and was succeeded by his half-sister Aimata, called Queen Pomare IV.

In 1836 French Catholics attempted to open a mission in Tahiti. Queen Pomare, advised by the English missionary and consul, George Pritchard, removed by force two priests who had landed surreptitiously. In 1842 a French frigate appeared under the command of Abel Dupetit-Thouars, and obtained from Pomare the right of settlement for Frenchmen. Pritchard opposed this, and caused Pomare to apply for British protection; but this was a failure, and the native chiefs compelled the queen to turn to France.

A convention was signed in 1843, placing the islands under French protection. Dupetit-Thouars now reappeared, and, alleging that the treaty had not been duly carried out, deposed the queen and took possession of the island. His highhanded action was not countenanced by the French government and the protectorate was re-established. Trouble broke out and it was only in 1846 that the queen, who had taken refuge in the Windward Islands, regained Tahiti.

Queen Pomare died in 1877, and her son Ariane (Pomare V) abdicated in 1880. In the same year Tahiti was proclaimed a French colony and became the centre of French Oceania.

Tahiti, with the other Society Islands, adhered to the "Free French" movement of Gen. Charles de Gaulle during the German occupation of France, 1940–44. In 1946 all the inhabitants became French citizens.

Bibliography.—E. Caillot, *Histoire de la Polynésie orientale* (Paris, 1910); Teuira Henry, *Ancient Tahiti* (Honolulu, 1928); F. Osborne (ed.), *The Pacific World* (New York, 1944; London, 1946); A. T'serstevens, *Tahiti et sa couronne* (Paris, 1950); J. Bourgeau, *La France du Pacifique*, 2nd ed. (Paris, 1954); H. Deschamps, *Tahiti* (Paris, 1955). Among narrative works may be mentioned Pierre Loti, *Le mariage de Loti* (Paris, 1881); Dora Hurt, *Tahiti: The Garden of the Pacific* (London, 1891). (Hu. De.)

TAHR (*Hemitragus jemlaicus*), a shaggy-haired, brown Himalayan wild goat, characterized by its short, triangular and sharply keeled horns. Besides the tahr the genus includes the wariatu or Nilgiri ibex (*H. hylocrius*), from south India, and the smaller *H. jayakeri*, from south Arabia. Tahr frequent the worst ground of all ruminants.

TAILLE, the equivalent of the English *tallage* (*q.v.*), was in France the typical direct tax of the middle ages, just as the word *tonlieu* was the generic term for an indirect tax. Other words used in certain districts in the same sense as *taille* were *queste* (*questa, quista*), *fouage* (*foragium*), *cote*. The essence of the tax denoted by these names was that the amount was fixed *en bloc* for a whole group of persons, and afterwards divided among them in various ways. In ancient French law we find three forms of *taille:* the *taille servile, taille seigneuriale* and *taille royale*.

The *taille servile* can scarcely be termed a tax; it was rather a tax which had degenerated into a source of profit for certain individuals. Every lord who possessed serfs could levy the *taille* on them, and originally this was done arbitrarily (*a volonté*) both as to frequency and amount. It always remained a characteristic feature of serfdom, but was limited and fixed, either by contracts or concessions from the lord (*taille abonnée*), or by the customs.

The *taille seigneuriale* was a true tax, levied by a lord on all his subjects who were neither nobles nor ecclesiastics; but the writer holds that, when feudalism was set up, the right to levy it did not belong to every lord, but only to the lord having the *haute justice*. But he levied it by right, without the necessity for any contract between him and those who paid it. He fixed the sum to be paid by each group of inhabitants, who then had to see that it was assessed, collected and paid to the lord, electing commissaries (*preud hommes*) from among themselves for this purpose. The seigniorial *taille*, like the servile, had the character of a personal tax (*taille personelle*), a rudimentary tax on income, every man being taxed according to his wages or other income. The king originally had only the right of levying the *taille* in places where he had retained the exercise of the *haute justice*. At that time there was no royal *taille*, strictly speaking; it was only the seigniorial *taille* transferred to the Crown, but it was one of the first taxes his right to levy which upon all the inhabitants of the domain of the Crown, whether serfs or *roturiers*, was recognized. The general *taille*, for the benefit of the king, became more frequent, and tended to become permanent. This transformation was confirmed, rather than effected, by the *ordonnance* of 1439. Its immediate object was, not the regulation of the *taille*, but the organization of the *compagnies d'ordonnance, i.e.*, the heavy cavalry which the king from that time on maintained on a permanent footing. Military expenses thus becoming permanent, it was natural that the *taille*, the tax which had long been devoted to meeting the expenses of the royal wars, should also become permanent. This was contained implicitly in the *ordonnance* of 1439, which also suppressed the seigniorial *taille*, as competing too closely with the royal *taille* by imposing a double burden on the taxpayer. A kind of seigniorial *taille* continued to exist besides the servile *taille*, but this kind presupposed a title, a contract between the taxable *roturier* and the lord, or else immemorial possession, which amounted to a title.

Throughout the *pays d'élections* the *taille* was almost universally personal (*taille personnelle*); *i.e.*, a tax on the whole income of the taxpayer, whatever its source. It was also a distributory tax (*impôt de repartition*); every year the king in his council fixed the total sum which the *taille* was to produce in the following year; he drew up and signed the *brevet de la taille* (warrant), and the contribution of the individual taxpayer was arrived at in the last analysis by a series of subdivisions.

In certain districts the *taille* was real (*taille reelle*); *i.e.*, a tax on real property. It was not an equal tax falling on all landowners, but the question as to whether a certain estate was to be taxed or not was decided according to the quality of the property, and not that of the owner. The *biens nobles* (*fiefs*) and the *biens ecclesiastiques* were exempt; *tenures roturières*, however, by whomsoever held, were taxed. A small part of the *pays d'élections* was also *pays de taille réele*. But it was the chief form of tax in the *pays d'états*, and even there an attempt had generally been made to check the exemption of nobles' property. It has been shown that in these districts the *taille* had originally been personal, having become real by a curious evolution. In these districts there were *cadastres*, or *compoix-terriers* (land registers), which allowed of a non-arbitrary assessment; and at the end of the *ancien régime* merely needed revision.

In certain provinces where the royal *taille* was levied there were neither *élections* nor *généralités*, and the whole administration of the tax was in the hands of the *intendants*. These were the provinces of the east and north, which were united to the Crown at a period when the power of the *intendants* was already fully developed; they were sometimes known as *pays d'imposition*.

See Auger, *Code des tailles* (1788); Henri Sée, *Les Classes rurales et le régime domanial en France au Moyen Âge* (1901).

(J. P. E.; X.)

TAILLEFER, the surname of a bard and warrior of the 11th century, whose exact name and place of birth are unknown. He accompanied the Norman army to England in 1066, and obtained permission from William to strike the first blow at the battle of Hastings, where he was killed. Mention of Taillefer is made by Guido, bishop of Amiens, in his *Carmen de bello Hastingensi*, v. 931–944 (in *Mon. Hist. Brit.*, 1848) and by Henry of Huntingdon in his *Historia Anglorum* (in *Rer. Brit. med. aevi script.*, p.

763, ed., Arnold, London, 1879); and his prowess is depicted on the Bayeux tapestry. The statement of Wace in the *Roman de Rou*, 3rd part, v. 8035–62, ed. Andresen (Heilbronn, 1879), that Taillefer went before the Norman army singing of Charlemagne and of Roland and the vassals who died at Roncevaux, has been considered important in demonstrating the existence of a comparatively early tradition and song of Roland.

See W. Spatz, *Die Schlacht von Hastings* (Berlin, 1896); Freeman, *History of the Norman Conquest*.

TAILOR, one who cuts out and makes clothes (Fr. *tailleur*, from *tailler*, to cut, Lat. *talea*, a thin rod, a cutting for planting). Formerly the tailor, or *cissor*, made apparel for both men and women, and not merely outer garments, but also articles of linen and the padding and lining of armour—whence the style "Taylors and Linen Armourers" applied to the Merchant Taylors Company of the City of London in their earliest charters. But the word is now generally limited to those who make the outer (cloth) garments for men, and less frequently for women, though a phrase such as "shirt-tailor" is occasionally met with. In modern usage, too, it commonly has the implication that the garments are made to the order, and to the measure, of the individual purchaser, as opposed to ready-made clothing, which means articles of apparel manufactured in large quantities in a series of stock or standard sizes, such that any purchaser may expect to find among them one that will fit him with more or less accuracy. The clothing trade was originally confined to goods of the poorest grades, but it has come, both in Great Britain and in America, to include articles of good, though not of the first, quality. It probably first came into existence at seaport towns, where, to meet the convenience of sailors returning from long voyages and requiring their wardrobes to be replenished at short notice, the "outfitters" kept stocks of ready-made garments on sale; but it made no considerable progress until after the middle of the 19th century, when the introduction of the sewing-machine brought about the possibility of manufacturing in large quantities. Its development was attended with gradually increasing subdivision of labour and, to a large extent, with the disappearance of the tailor as a skilled craftsman.

The first step was for a garment, such as a coat, to be completed by the joint efforts of a family. Then followed the "task system," which in America was the result of the influx of Russian Jews that began about 1875. Under it a team of three men, with a "presser" and a girl to sew on the buttons, divided the work between them. Payment was made by the "task," *i.e.*, a specified number of garments, the money being divided between the members of the team in certain proportions. Often several teams would be run by a contractor, who naturally selected the cheapest workshops he could find and packed them as full of workers as possible; and when through stress of competition he had to accept lower prices the plan he adopted was to increase the number of garments to a task, leaving the pay unaltered. The result was the introduction of many of the worst features of the "sweating system," the workers having to work excessively long hours in order to finish the task, which in some cases meant as many as 20 coats a day. In the "factory" or "Boston" system the subdivision is still more minute, and as many as 100 persons may be concerned in the production of one coat. The amount of tailoring skill required in a worker is even further reduced, but the premises come under the regulation of the factory laws. The factory system has also cheapened production in a legitimate way, because it has enabled mechanical power for driving sewing-machines, and also expensive labour-saving machinery, to be introduced to an extent not practicable in small shops.

TAIN, a royal and small burgh in the county of Ross and Cromarty, Scot., near the southern shore of Dornoch firth, 26½ mi. N.E. of Dingwall by road. Pop. (1961) 1,699. The name is from the Icelandic *thing* (assembly). It is the centre of a prosperous agricultural area specializing in raising and exporting seed potatoes (mainly to England) and in rearing the valuable Calrossie shorthorn cattle. There are well-known golf links. Anciently Tain was a place of sanctuary, the "Immunity of Tayne" being marked by four crosses, but the privilege was violated in 1306, when the earl of Ross abducted Robert Bruce's wife and daughter and handed them over to Edward I, and again in 1427 when McNeille of Creich burned the chapel of St. Duthac (Duthus) of which there are remains near the railway station. The church of St. Duthus was built by the bishop of Ross in 1370. In 1487 it became collegiate, the alteration being confirmed by a papal bull still preserved by the corporation. The building was used as the parish church until 1815 when it fell into disrepair, but it was restored in 1871–76. Three and a half miles southeast are the remains of the abbey of Fearn, founded at Edderton in 1230 by Farquhar, 1st earl of Ross, and transferred thither in 1338.

TAINE, HIPPOLYTE ADOLPHE (1828–1893), French critic and historian, son of Jean Baptiste Taine, an attorney, was born at Vouziers on April 21, 1828. J. B. Taine died on Sept. 8, 1840, leaving a moderate competence to his widow, his two daughters, and his son. Taine was educated at the Collège Bourbon, where he formed lifelong friendships with L. A. Prévost-Paradol, long his most intimate friend, Émile Planat, the future "Marcelin" of the *Vie parisienne*, and Cornélis de Witt, who introduced him to Guizot when the latter returned from England in 1846. Taine passed from the Collège Bourbon to the École Normale, but in 1851 he was refused the fellowship in philosophy which was universally regarded as his due. Nevertheless the minister of public instruction appointed him to a professorship at Toulon, which he exchanged for one at Nevers. But he refused to express explicit approval of the coup d'état of Dec. 2, 1851. Consequently he was removed to an inferior post; next year saw a further step downward, and he accordingly applied for, and received, indefinite leave of absence. In a few months his two dissertations, *De personis Platonicis* and the essay on La Fontaine's fables were finished, and May 30, 1853 he took his doctor's degree. This was the last act of his university career; his life as a man of letters was to begin.

No sooner had he deposited his dissertations at the Sorbonne than he began to write an essay on Livy for one of the competitions set by the academy. There again the moral tendency of his work excited lively opposition, and after much discussion the competition was postponed till 1855; Taine toned down some of the censured passages, and the work was crowned by the academy in 1855. In the beginning of 1854 Taine, after six years of uninterrupted effort, broke down and was obliged to rest. The year 1854 was an important one in the life of Taine. His enforced leisure, the necessity of mixing with his fellowmen, and of travelling, tore him from his cloistered existence and brought him into more direct contact with reality. He lived with his mother in the Ile Saint-Louis and once more associated with his old friends, Planat, Prévost-Paradol and Edmond About. He made the acquaintance of Renan, and through Renan that of Sainte-Beuve, and he renewed friendly relations with Ernest Havet, who for three months had been his teacher at the École Normale. These years (1855–56) were Taine's periods of greatest activity and happiness in production. In 1855 he published 17 articles in the *Revue de l'instruction publique*, and 20 in 1856 on the most diverse subjects, ranging from Menander to Macaulay. From 1857 onward he was a regular contributor to the *Journal des débats*.

But he was seeking a larger field. In Jan. 1856 his history of English literature was announced, and in 1855–56 he published in the *Revue de l'instruction publique* a series of articles on the French philosophers of the 19th century, which appeared in a volume at the beginning of 1857. In this volume he energetically attacked the principles which underlie the philosophy of Victor Cousin and his school. The book closes with the sketch of a system in which the methods of the exact sciences are applied to psychological and metaphysical research. The work itself met with instantaneous success, and Taine became famous. In 1858 appeared a volume of *Essais de critique et d'histoire;* in 1860 *La Fontaine et ses fables*, and a second edition of the *Philosophes français*.

In 1864 Taine succeeded Viollet-le-Duc as professor of the history of art and aesthetics at the École des Beaux Arts. Renan's appointment at the Collège de France and Taine's candidature for the Polytechnic school had alarmed Dupanloup, who in 1863 issued an *Avertissement à la jeunesse et aux pères de famille*, which consisted of a violent attack upon Taine, Renan and Littré.

Renan was suspended, and Taine's appointment to Saint Cyr would have been cancelled but for the intervention of the princess Mathilde. In Dec. 1863 his *Histoire de la littérature anglaise* was published, prefaced by an introduction in which Taine's determinist views were developed in the most uncompromising fashion. In 1864 Taine sent this work to the academy to compete for the Prix Bordin. M. de Falloux and Dupanloup attacked Taine with violence; he was warmly defended by Guizot; finally, after three days of discussion, it was decided that as the prize could not be awarded to Taine, it should not be awarded at all. After three attempts to secure admission to the academy Taine was elected in Nov. 1878.

The period from 1864 to 1870 was perhaps the happiest of Taine's life. He derived pleasure from his employment at the Beaux Arts and Saint Cyr, which left ample leisure for travel and research. In 1865 appeared *La Philosophie de l'art*, in 1867 *L'Idéal dans l'art*, followed by essays on the philosophy of art in the Netherlands (1868), in Greece (1869), all of which short works were republished later (1880) as a work on the philosophy of art. In 1865 he published his *Nouveaux essais de critique et d'histoire;* from 1863 to 1865 appeared in *La Vie parisienne* the notes he had taken for the past two years on Paris and on French society under the subtitle of "Vie et opinions de Thomas Frédéric Graindorge," published in a volume in 1867, the most personal of his books, and an epitome of his ideas. In 1867 appeared a supplementary volume to his history of English literature, and in January 1870 his *Théorie de l'intelligence*. In 1868 he married Mademoiselle Denuelle, the daughter of a distinguished architect. He had made a long stay in England in 1858, and had brought back copious notes, which, after a second journey in 1871, he published in 1872 under the title of *Notes sur l'Angleterre*.

Taine, deeply shaken by the events of 1870, felt that it was the duty of every Frenchman to work solely in the interests of France. On Oct. 9, 1870 he published an article on "L'Opinion en Allemagne et les conditions de la paix," and in 1871 a pamphlet on *Le Suffrage universel;* and his intention of writing on the French Revolution returned in a new and definite shape. He determined to trace in the Revolution of 1789 the reason of the political instability from which modern France was suffering. From the autumn of 1871 to the end of his life his great work, *Les Origines de la France contemporaine*, occupied all his time, and in 1884 he gave up his professorship in order to devote himself wholly to his task; but he succumbed before it was finished, dying in Paris on March 9, 1893.

Les Origines de la France contemporaine, Taine's monumental achievement, stands apart from the rest of his work. The problem which Taine set himself was to inquire why the centralization of modern France is so great that all individual initiative is practically nonexistent, and why the central power, whether it be in the hands of a man or of an assembly, is the sole and only power; also to expose the error underlying two prevalent ideas. Against the view that the Revolution destroyed absolutism and set up liberty, he points out that it merely caused absolutism to change hands. Opposing the view that the Revolution destroyed liberty instead of establishing it or that France was less centralized before 1789 than after 1800 he retorts that France was already a centralized country before 1789 and grew rapidly more and more so from the time of Louis XIV onward: the Revolution merely gave it a new form. The *Origines* differ from the rest of Taine's work in that, although he applies to a period of history the method which he had already applied to literature and to the arts, he is unable to approach his subject in the same spirit. He loses his philosophic calm, he cannot help writing as a man and a Frenchman, he lets his feeling have play, and what the work loses thus in impartiality it gains in life.

Taine was the philosopher of the epoch which succeeded the era of romanticism (1820–50) in France. The ideal of the newer generation was truth; their watchword liberty; to get as near as possible to scientific truth became their object. Taine must ever be regarded as one of the most authoritative spokesmen of this period.

Taine served science unfalteringly, without looking forward to any possible fruits or result. In his work we find neither enthusiasm nor bitterness, neither hope nor yet despair; merely a hopeless resignation. The study of mankind was Taine's incessant preoccupation, and he followed the method already described. He made a searching investigation into humanity and his verdict was one of unqualified condemnation. In "Thomas Graindorge" we see him aghast at the spectacle of man's brutality and woman's folly. In man he sees the primeval savage, the gorilla, the carnivorous and lascivious animal, or else the maniac with diseased body and disordered mind, to whom health, either of mind or body, is but an accident. Taine is appalled by the *bête humaine;* and in all his works we are conscious, as with Swift and Voltaire, of the terror with which the possibilities of human folly inspire him. It may be doubted whether Taine's system, to which he attached so much importance, is really the most lasting part of his work. For Taine was an artist as well as a logician, an artist who saw and depicted what he saw in vital and glowing language. From the artist we get his essay on La Fontaine, his articles on Balzac and Racine, and the passages on Voltaire and Rousseau in the *Ancien Régime*. Moreover, not only was Taine an artist who had not escaped from the influence of the romantic tradition, but he was by his very method and style a romanticist. His emotions were deep if not violent, his vision at times almost lurid. He sees everything in startling relief and sometimes in exaggerated outline, as did Balzac and Victor Hugo. Hence his predilection for exuberance, strength and splendour; his love of Shakespeare, Titian and Rubens; his delight in bold, highly-coloured themes.

Taine's influence was great, and twofold. On his own generation it was considerable; during the epoch in which he lived, while a wave of pessimism was sweeping over French literature, he was the high priest of the cult of misanthropy, in which even science was held to be but an idol, worthy of respect and devotional service, but not of faith. In its turn came the reaction against positivism and pessimism, and an attempt at spiritual renascence. Around a man so remarkable as Taine a school was certain to form itself; Taine's school, which was one of positivist doctrines, rigid systems and resigned hopelessness, was equally certain to produce at some time or another a school of determined opponents to its doctrines and system. If, therefore, the tone which pervades the works of Zola, Bourget and Maupassant can be immediately attributed to the influence that we call Taine's, it is also the influence of Taine which is one of the ultimate causes of the protest embodied in the subsequent reaction. (M. Bg.; X.)

BIBLIOGRAPHY.—The official life, *H. Taine, sa vie et sa correspondance*, was published in 3 vol. (1902–05), Eng. trans. by Mrs. R. L. Devonshire (1902–08). His friend, Émile Boutmy, published an appreciative study of Taine's philosophy in his *Taine, Scherer, Laboulaye* (Paris, 1901). *See* also Émile Faguet, *Politiques moralistes au XIXe siècle* (Paris, 1900); M. Barrès, *Taine et Renan* (1922); F. C. Roe, *Taine et l'Angleterre* (1923); A. Chevrillon, *Taine, formation de sa pensée* (1932).

T'AI-PEI (Japanese, TAIHOKU), one of the chief cities of Formosa (Taiwan). Founded in 1708, T'ai-pei after the middle of the 19th century became an important centre for overseas trade through its outport, Tan-shui. This trade was based upon tea and was dominated by British trading companies.

Situated in the centre of the largest agricultural basin in northern Formosa, T'ai-pei (population in 1956 was 668,093 exclusive of rural areas included in the T'ai-pei city administrative district) forms the nucleus of a major industrial area. The T'ai-pei industrial complex includes light and heavy industries within the urbanized area and also in several industrial suburbs, including Pan-ch'iao and Nan-chiang. Cotton and artificial fibre textiles, rubber goods, electrical appliances, alcoholic beverages, metal products, chemicals, coke and printed materials are produced. T'ai-pei also is the island's major communications centre, containing the central workshops of the government owned railway and Formosa's international airport. National Taiwan university and the Taiwan Provincial Normal university, as well as several professional and specialized training schools, are located in T'ai-pei. Because it contains more mainland Chinese than any other Formosan city, characteristic types of mainland entertainment, such as opera and fine regional restaurants, are found there.

Under Japanese rule (1895–1945) T'ai-pei was the seat of the Taiwan government-general; as a result it acquired the characteristics of an administrative centre, such as public buildings and extensive civil servant housing facilities. After the communists took China in 1949, it became the capital-in-exile of the Nationalist government of the Republic of China, headed by Chiang Kai-shek. (N. S. G.)

T'AI P'ING REBELLION. The T'ai P'ing rebellion of 1851–1864 in China was a great national religious-political upheaval. It inspired many other rebellions and revolutions in late 19th-century China, culminating in Sun Yat-sen's overthrow of the Manchu dynasty in 1911–12. Historical judgment of it has varied. The rebellion has been condemned as the work of "Cantonese bandits" or "long-haired bandits," and has been praised as a national agrarian and proletarian revolution. This great upheaval grew out of a complex background including political corruption of the Manchu regime, foreign aggression during and after the Opium War of 1839–42, agrarian distress, and severe famines that left many people homeless and ready to join the T'ai P'ings.

The titular leader of the rebellion was Hung Hsiu-ch'üan (1814–64), a Cantonese who had struggled in vain to gain office through government civil service examinations. After taking the metropolitan examination at Canton in 1836 he was given nine Christian tracts, which he read cursorily. His third attempt to pass the examination failed in 1837. He was so dismayed that he fell ill, suffered delirium, and had visions of an old man and a middle-aged man in heaven; the latter ordered him to kill demons on earth. After his vision, Hung continued to teach in a village school. Six years later, after his last examination failure in 1843, he decided to give up hope. His cousin then called his attention to the nine Christian tracts. Hung studied the booklets, which had the general title of "Good Words to Exhort the Age" (*Ch'üan-shih Liang-yen*), quotations and paraphrases mainly from the Old Testament, together with sermons. The compiler of the tracts, Liang A-fa (1789–1855), was the first Protestant convert in China. After reading the tracts Hung realized that the old man in his vision was God the Heavenly Father, the middle-aged man was Jesus Christ the Elder Brother, while he himself was the second son of God commissioned to destroy demons, and thus completing a new trinity. Hung and his cousin baptized each other and soon afterward organized the God Worshipers' society.

It was in 1850 that Hung ordered all God Worshipers in various places in mountainous Kwangsi province to resist government troops and to concentrate at Chin-t'ien. Their slogan—to share property in common—attracted many famine-stricken peasants, workers, miners and others. They took Yung-an in Kwangsi on Sept. 25, 1851, and Hung proclaimed himself the heavenly king of the T'ai-p'ing T'ien-kuo, or Heavenly Kingdom of Great Peace. The year 1851 was to be the first year of the new dynasty.

The T'ai P'ings thereupon sought to overthrow the Manchu dynasty, and advanced rapidly from Kwangsi through Hunan and Hupeh along the Yangtze valley. They formed themselves into separate battalions for men and women and behaved much better than the government troops. In March 1853 the T'ai P'ings established their "heavenly capital" at Nanking, and it lasted for 11 years. A northern expedition to take Peking failed but another expedition to reconquer the upper Yangtze valley scored many victories. Victory, however, was followed by a series of murders among the leaders of the rebellion in 1856. Two associate kings with hundreds of their followers were killed; an enormous army was taken away by the assistant king to southwest China.

Meanwhile, the inefficient government troops were replaced by a new force trained in Hunan by Tseng Kuo-fan (1811–1872). A Confucian scholar, Tseng employed his students as officers and created a Confucian army to combat the rebels. Tseng eventually managed to hem Hung's forces in Nanking. The T'ai P'ing attempts to take Shanghai in the 1860s then met with foreign intervention. Western volunteers organized the "Ever-Victorious army" under the command of F. T. Ward and then of C. G. Gordon (*q.v.*), who routed the T'ai P'ings in the greater Shanghai area. Nanking was taken by Tseng's force in July, 1864. By this time Hung Hsiu-ch'üan had committed suicide. Some T'ai P'ings fled to the southern coast and survived until 1866, and some, allied with the "Nien-fei" in North China, continued their anti-Manchu war until 1868. (*See* NIEN REBELLION.)

During their years of power the T'ai P'ings inaugurated many reforms. Their religion was a Chinese pseudo-Christianity. They knew only God the creator of the world, the Ten Commandments and baptism. They recognized the Sabbath and said grace before meals. They did not adopt the Lord's Supper and the Christian ideas of kindness, forgiveness and service to mankind. Their government system was a theocracy. The Heavenly King was both spiritual and temporal ruler, and his four associate kings and an assistant king were civil and military chiefs. All posts were hereditary. The legal code was severe. Organization of the army was elaborate, with strict rules governing soldiers in camp or on the march. The T'ai P'ings planned equal distribution of land according to a primitive form of communism, but did not put this system into practice. Families were graded in a co-ordinated system that took into account the exigencies both of peace and war. There was to be equality between men and women. Monogamy was the rule for the common people, but not for the kings. Prostitution, footbinding and slavery were prohibited, as well as opium smoking, adultery, witchcraft, gambling, and use of tobacco and wine. Chinese language simplification was practised. Westernization and industrialization were planned.

The T'ai P'ing rebellion failed because of poor leadership in military and civil administration, internal dissension, alienation of the Confucian literati, and the combined attack by Tseng's army and foreign-led forces. It ravaged 17 provinces and caused the loss of about 20,000,000 lives.

See also CHINA: *19th Century and Revolution.*

BIBLIOGRAPHY.—Ssu-yu Teng, *Historiography of the Taiping Rebellion* (1962); Eugene Boardman, *Christian Influence Upon the Ideology of the Taiping Rebellion, 1851–1864* (1952); Vincent Yu-chung Shih, "Interpretation of the Taiping T'ien-kuo by Non-Communist Chinese Writers," *Far Eastern Quarterly*, x, 2, pp. 248–257 (Feb. 1951); P. M. Yap, "The Mental Illness of Hung Hsiu-ch'üan," *Far Eastern Quarterly*, viii, 3, pp. 287–304 (May 1954); Kuo Ting-yee, "The Totalitarian Rule of the Taiping Heavenly Kingdom," *Annals of Academia Sinica*, No. 2, pt. 1, pp. 209–230 (May 1955); Franz Michael, "T'ai-p'ing T'ien-kuo," *The Journal of Asian Studies*, xvii, 1, pp. 67–76 (Nov. 1957); John S. Gregory, "British Intervention Against the Taiping Rebellion," *The Journal of Asian Studies*, xix, 1, pp. 11–24 (Nov. 1959); Yuan Chung Teng, "Reverend Issachar Jacox Roberts and the Taiping Rebellion," *The Journal of Asian Studies*, xxiii, 1, pp. 55–67 (Nov. 1963).
(S. Y. T.)

TAIT, ARCHIBALD CAMPBELL (1811–1882), English divine, archbishop of Canterbury, was born at Edinburgh on Dec. 21, 1811, of Presbyterian parents. He entered Balliol college in 1830 as a Snell exhibitioner from the University of Glasgow. He became fellow and tutor of Balliol and was also ordained deacon (1836) and priest (1838), and served the curacy of Baldon. He never sympathized with the principles of the Tractarian movement, and on the appearance of Tract 90 in 1841 he drafted the famous protest of the "Four Tutors" against it; but this was his only important contribution to the dispute. In 1842 he succeeded Arnold as headmaster of Rugby; and, after a serious illness in 1848, accepted a less strenuous appointment as dean of Carlisle in 1849. During his residence at Carlisle he served on the University commission, restored his cathedral and did much excellent pastoral work. There too he suffered the great sorrow of his life. He had married Catharine Spooner at Rugby in 1843; in the spring of 1856, within five weeks, five of their children died of scarlet fever. He was consecrated bishop of London on Nov. 22, 1856. He became archbishop of Canterbury in 1868. His last years were interrupted by illness and saddened by the death in 1878 of his only son Craufurd, and of his wife.

Tait was constant in his attendance in parliament, and pressed on many measures of practical utility. The modification of the terms of clerical subscription (1865), the new lectionary (1871), the Burials act (1880) were largely owing to him; for all of them, and especially the last, he incurred much obloquy at the time. He played a prominent part also in the royal commissions on ritual (1867) and on the ecclesiastical courts (1881).

Tait was less successful in dealing with matters which called

for the higher gifts of a ruler, and especially in his relations with the liberal trend in modern thought, and the Catholic revival. His object in dealing with questions of faith, as in dealing with the ritual question, was primarily a practical one: he wished to secure peace, and obedience to the law as he saw it. Consequently, after his sympathies had led him to express himself favourably toward some movement, he frequently found himself compelled to draw back. He expressed a qualified sympathy with some of the writers of *Essays and Reviews*, and then joined in the censure of it by the bishops (1861). The same kind of apparent vacillation was found in his action in other cases. Tait was concerned with the Catholic revival during the whole of his episcopate, and above all on the side of ritual. He had to deal with the St. George's-in-the-East riots in 1859, and the troubles at St. Alban's, Holborn, in their earlier stages (1867); he took part as assessor in the privy council judgment in the Ridsdale case (1877); he was more closely concerned than any other bishop with the agitation against confession in 1858, and again in 1877. His method throughout was the same: he endeavoured to obtain a compliance to the law as declared by the courts; failing this, he made the most earnest efforts to secure obedience to the ruling of the Ordinary for the sake of the peace of the Church; after this, he could do nothing. In fact Tait could hardly realize anything but the connection between the English Church and the state. From such a position there seemed to be no escape but in legislation for the deprivation of the recalcitrant clergy; and the Public Worship Regulation act (1874) was the result. A feeling of indignation spread not only among High Churchmen, but among many who cared little for the ritual practices involved. At length, when A. H. Mackonochie was on the point of being deprived of his benefice for contumacy, the archbishop, then on his deathbed at Addington, took steps which resulted in the carrying out of an exchange of benefices (which had already been projected), which removed him from the jurisdiction of the court. This proved to be the turning point; and the ritual difficulty was afterward dealt with from a different point of view, and the Public Worship Regulation act became practically obsolete. The archbishop died on Dec. 3, 1882.

See R. T. Davidson and D. Benham, *Life of Archbishop Tait,* 2 vol. (1891); A. C. Tait, *Catharine and Craufurd Tait* (1880).

TAIT, PETER GUTHRIE (1831–1901), Scottish mathematician and physicist known for his work on thermodynamics, was born at Dalkeith on April 28, 1831. After attending the academy at Edinburgh and spending a session at the university, in 1848 he entered Peterhouse, Cambridge, graduating in 1852. As a fellow and lecturer of his college he remained in Cambridge until 1854 and then left to take up the professorship of mathematics at Queen's college, Belfast. There he joined Thomas Andrews in researches on the density of ozone and the action of the electric discharge on oxygen and other gases. From 1860 to 1901 he was professor of natural philosophy at Edinburgh, where he died on July 4, 1901.

His earliest work dealt mainly with mathematical subjects, and especially with quaternions (*q.v.*), of which he may be regarded as the leading exponent after their originator, Sir William Rowan Hamilton. With the help of Hamilton and Philip Kelland he wrote two textbooks, *Elementary Treatise on Quaternions* (1867) and *Introduction to Quaternions* (1873). He also produced original work in mathematical and experimental physics. In 1864 he published a short paper on thermodynamics, and from that time his contributions to that and kindred departments of science became frequent and important. He worked on thermoelectricity and thermal conductivity, its variation with temperature and its relation to the electrical conductivity of the same material.

From 1879 to 1888 he was engaged on difficult experimental investigations, which began with an inquiry into the corrections required, because of the great pressures to which the instruments had been subjected, in the readings of the thermometers employed by the "Challenger" expedition for observing deep-sea temperatures; his investigations were extended to include the compressibility of water, glass and mercury. Between 1886 and 1892 he published a series of papers on the foundations of the kinetic theory of gases, and about the same time he carried out investigations into impact and its duration.

A selection from Tait's papers was published in two volumes (1898, 1900). With Lord Kelvin he collaborated in writing the well-known *Treatise on Natural Philosophy* (1867), only the first part of which was completed. He collaborated with Balfour Stewart in *The Unseen Universe* which was followed by *Paradoxical Philosophy* (1878).

TAIWAN: *see* FORMOSA.

T'AI-YÜAN (formerly YANGKÜ or YANG-CH'U), capital of Shansi province, China, stands at the northern end of a heavily populated plain on the upper course of the Fen Ho. Ancient Mongol invasion routes lead from the north off the Mongolian plateau and from the south via Ta-t'ung. To the east through the T'ai-hang Shan is the old imperial highway from Peking via Shihmen to T'ai-yüan and then south and west to Sian. Both routes have railways and motor roads. Although the whole area lies within the brown, eroded loess highlands, the fertile T'ai-yüan plain produces an abundance of wheat, cotton and northern fruits and vegetables; sheep on the surrounding hills are raised for wool. There, too, are China's largest known reserves of high-grade bituminous coal. The city was known as the home of nationwide private banking, as a textile and leather centre (gold lacquered leather boxes and trays) and for metal production. Late in the 19th century an arsenal was built using the abundant local coal and iron ore, followed by an ironworks, flour and textile mills. Under Japanese occupation, 1937–45, the city and its industries were vital for the control of north China despite the constant cutting of railway lines by Chinese guerrilla forces. Under the Chinese People's Republic (Communist) the population increased and T'ai-yüan expanded into a major centre for iron and steel, textiles, textile machinery, cement and a phosphorous fertilizer plant (100,000-ton capacity). It has an institute of technology. Pop. (1953) 720,700; (1957 est.) 1,020,000.

(TE. H.)

TAJ MAHAL, the mausoleum built on the south bank of the Jumna river, outside Agra in India, on the orders of the Mogul emperor Shah Jahan in memory of his beloved wife, Arjumand Banu Begum, called Mumtaz Mahal, "chosen one of the palace" (of which Taj Mahal is a corruption). She died in childbirth in the town of Burhanpur in 1631 after having been the emperor's inseparable companion since their marriage in 1612. The building was commenced in 1632, after plans had been prepared by a council of architects from India, Persia, central Asia and beyond; the credit for the final plan is given to one Ustad Isa, either Turkish or Persian, although the master builders, masons, inlayers and calligraphists, like the materials they worked with, came from all over India and central Asia. More than 20,000 workmen were employed daily to complete the mausoleum building itself by 1643, although the whole Taj complex took 22 years to complete, at a cost of 400 lakhs of rupees.

The complex consists of a rectangle measuring 634 yd. by 334 yd. aligned north and south. A central square garden area, 334 yd. on each side, leaves an oblong area at each end; that at the south consists of the sandstone entrance gateway with its attendant service buildings, while that at the north (river) end comprises the mausoleum itself, flanked on the west and east walls by two symmetrically identical buildings, the mosque and its jawab ("answer") respectively. All is enclosed within a high red sandstone boundary wall with octagonal pavilion turrets at the corners, while outside the enclosure at the south are ancillary buildings such as stables, outhouses and guard quarters. The whole complex is the begum's memorial; it was conceived and planned as an entity, since Mogul building practice allowed of no subsequent addition or amendment. Its northern end is the most significant architecturally, with mosque and jawab both facing the mausoleum itself. The mosque and jawab of red Sikri sandstone, with marble necked (not bulbous) domes and architraves and some restrained *pietra dura* surface decoration, contrast well with the mausoleum of pure white Makrana marble. This mausoleum, standing on a 312-ft.-square marble plinth 23 ft. high, is a square of 186 ft. with chamfered corners and with a massive arch in each face, rising to 108 ft.; over all is a bulbous double dome, supported on a tall drum,

the pinnacle of which stands 243 ft. above garden level. The sky-line rhythm is enhanced by parapets over each arch, corner pinnacles and domed kiosks over each corner. At each corner of the plinth stands a three-storied minaret, 138 ft. high to the crowning kiosk; the countersunk face joints of the marble bricks provide textural contrast with the highly finished marble of the mausoleum, but unfortunately recall factory chimney construction to the sophisticated western beholder. Inside the mausoleum is the octagonal chamber, embellished with low-relief patterns and fine *pietra dura*, containing the cenotaphs of the begum and Shah Jahan. These, of marble decorated with superb *pietra dura*, are enclosed by an exquisite perforated marble screen studded with precious stones. A vault below, at garden level, contains the true sarcophagi.

The minarets lead the observer's eye from the mausoleum to its setting; this is now somewhat unsatisfactory, as trees obscure the view of other structures, and from the garden there is some feeling of restlessness; but from the river or from the fort the balance restores itself, and justifies H. R. Nevill's remark that "the Taj . . . is within more measurable distance of perfection than any other work of man" (*Agra Gazetteer*, 1921). Its design, successor in the indigenous tradition to the mausoleums of Humayun and the Khan-i-khanan at Delhi, has been criticized as "too feminine" —thus complimenting the designers' intentions. The Moguls are said to have "built like Titans and finished like goldsmiths"; certainly the Taj Mahal is their finest jewel. *See* ISLAMIC ARCHITECTURE.

BIBLIOGRAPHY.—Percy Brown, "Monuments of the Mughal Period," *Cambridge History of India*, vol. iv (1937), *Indian Architecture: the Islamic Period* (1942); for the controversy over the attribution of the design to a European, *see* V. A. Smith, *History of Fine Art in India and Ceylon*, pp. 182–185 (1930); E. B. Havell, *A Handbook to Agra and the Taj*, appendix i, new ed. (1924). (J. B.-P.)

TAK, a mountainous province of Thailand on the Burmese border. Area 6,027 sq.mi. Pop. (1960) 165,104. The Ping river runs through its northeastern portion, and the provincial capital, Tak or Rahaeng (pop., 1957, 11,641), is the river's major commercial centre south of Chiangmai. The town is an important crossroads of the national highway system, and a motor road, following an ancient caravan route to Burma, now traverses the province from Tak to Maesot (pop., 1957, 9,615). (G. W. Sk.)

TAKAMATSU, capital city and commercial centre of Kagawa prefecture, Jap. Pop. (1960) 228,172. Serving as the transportation hub of Shikoku Island, Takamatsu has a small but flourishing port that is linked by railroad ferry to Uno on the Inland sea coast of Honshu. The water front is dominated by a breakwater, port facilities, a small industrial zone on reclaimed land and extensive salt fields. Federal agencies for Shikoku have their headquarters in Takamatsu. Ritsurin park, once the estate of the Matsudaira family, and Inland sea scenery viewed from Yashima, an elevated tableland noted for its relics of the 12th-century Taira clan, attract many tourists. (J. D. Ee.)

TAKAOKA, a town in Toyama prefecture, central Japan. Pop. (1960) 135,190. In 1609 Maeda Toshingaga, feudal lord of the Kaga clan, built a castle there and Takaoka developed. After his death the castle town flourished as a commercial and industrial centre. With the annexation of the old port, Fushiki, in 1922, Takaoka became one of the busiest port towns on the Japan sea. It is also an important railway junction. There are factories producing copper, iron and lacquer ware as well as cotton-spinning and textile-dyeing industries. (R. B. H.)

TAKASAKI, town in Gumma prefecture, east Japan, situated on a tributary of the Tone river. Pop. (1960) 142,152. Takasaki is a railway communication centre. A typical castle town, it became increasingly important as a commercial centre, especially after the Meiji era (1868–1912), with the growth of the railroads. Various manufacturing industries have developed, including flour milling and machinery. The Yorimasa shrine is located in Takasaki park. (R. B. H.)

TAKIN, a rare, little-known goat antelope (*Budorcas taxicolor*), which inhabits high, rough country from Bhutan to southern Shensi province, China. The takin is clumsily built, with yellowish-brown or gray hair and curiously curved horns, which recall those of the gnu. This shy animal is hunted for food by the people of its native regions, but very few specimens have been captured alive. Its nearest relative appears to be the musk ox (*q.v.*). *See* BOVIDAE.

TAKLA MAKAN, a desert in central Asia, the major central division of the Tarim basin. In general the Takla Makan is a region of drift sand and moving sand dunes, and is a waterless area except for the Khotan and Keriya rivers, which carry water northward into the desert. The sand dune desert is uninhabitable away from these rivers and from the Tarim river, which skirts the northern margins. The sand dunes occupy chiefly the south and southwest parts of the Takla Makan, where the full force of the northeast winds is felt. In this area there are two systems of dunes; one, of high chains, stretches from east to west, and the second, of transverse dunes, runs from north to south or northeast to southwest. The steeper sides of the dunes generally face toward the south, away from the direction of the prevailing winds, though there is local variation in alignment, facing and movement.

Between the lower Tarim river, in the northeast, and the Charchan desert, in the southeast, the surface generally is different from that of the rest of the Takla Makan. In this region the sand dunes are interrupted by tracts of level soil, or by areas of old channels, bare mesas and erosional remnants, swept clean by the wind. This is the zone in which the Tarim river and the lake, Lop Nor, have wandered repeatedly, so that in places there is fresh to brackish water available near the surface.

Vegetation and animal life are almost nonexistent in the dry heart of the Takla Makan, but along the permanent streams and locally in the eastern sector various steppe plants, reeds, tamarisks and poplars provide a thin scattering of plant life. On the lower Tarim river, and around Lop Nor, heavy reed growth does occur. The animal life of the region is concentrated at localities providing water and vegetation. Hares, rodents, foxes, wolves, wild pigs, a few antelope and, in a few places in the east, wild camels, are among the land animals. In general the animal population of the eastern sector is greater than that of the western zone. A variety of fish and birds are found along the lower Tarim river and throughout Lop Nor.

The climate of the Takla Makan is one of extremes. Winter temperatures as low as −23° or −24.5° C. (−10° or −12° F.) occur almost everywhere, and a temperature of −30° C. (−22° F.) has been observed in the southeast, where winter snowstorms sometimes occur. Temperatures as high as 30° C. (86° F.) have been recorded in late spring, but it is likely that higher temperatures occur in the interior dune areas in midsummer. Suffocating sandstorms occur, often enveloping the desert for days and producing an impenetrable dust haze.

See also SINKIANG; TARIM; LOP NOR. (J. E. Sr.)

TAKORADI: *see* SEKONDI-TAKORADI.

TALAVERA DE LA REINA, a town of central Spain, in the province of Toledo, on the right bank of the river Tagus, and on the Madrid-Cáceres railway. Pop. (1950) 21,728 (mun.). Talavera is of great antiquity, the Caesobriga of the Romans. Portions of the triple wall which surrounded it remain standing and the Arco de San Pedro is one of its Roman gates restored. Among the ancient buildings are the Torres Albarranas, built by the Moors in the 10th century, and the Gothic collegiate church. The bridge of 35 arches across the Tagus dates from the 15th century.

Talavera, Battle of, 1809.—For the operations which culminated in the famous battle of Talavera between the English and the French and those which followed that engagement, *see* PENINSULAR WAR. Sir Arthur Wellesley (afterward Duke of Wellington), the British commander, acting in co-operation with Lieutenant-General Cuesta's Spanish army, took position on July 27, 1809, on the Upper Tagus, protected by his advanced guard. His line, facing due east, ran north from the right bank of the river to a ridge running parallel to the Tagus, beyond which ridge, also parallel to the river, lay the Sierra de Montalban. Cuesta's men with their right flank resting on the river held Talavera itself and the close country to the northward of it; Wellesley's right

connected with Cuesta's left, and his line stretched away northward to the ridge. The Sierra was not occupied until the second day. On the first day the French attacked the allied positions but were unable to breach the defenses. On the 28th the French renewed their efforts and mounted several more attacks which were thrown back. Wellesley finally decided the day by a counter-attack with the 48th regiment, made with great intrepidity and steadiness. The guards, with splendid discipline, resumed their positions, and eventually the French fell back. Failure all along the line and heavy losses left King Joseph no alternative but to retire toward Madrid. The French lost 7,268 men out of 46,138 present, the British 5,363 out of 20,641; the Spanish losses were officially returned at 1,201 out of some 36,000 present.

TALBOT, the name of one of the oldest families in the English nobility, which traces its descent from RICHARD TALBOT (d. 1174–75), who held Eccleswall and Linton, in Herefordshire, in the time of Henry II. Records mention several persons called Talbot living earlier in England, and in Normandy before 1066, among them Geoffrey Talbot (d. 1140), an opponent of King Stephen; but since Talbot was at first a nickname and not a hereditary surname, their relationship is obscure.

The descendants of Richard Talbot, with whom the proved pedigree begins, ranked for several generations among the lesser families of the Welsh border counties. GILBERT TALBOT (1276–1346) took part with other Marcher lords against Edward II's favourites, the Despensers, who had wrested estates from his daughter-in-law Elizabeth Comyn. With his son, Richard, he was taken prisoner by the royal forces at Boroughbridge (1322) but was later pardoned and restored to his estates. He was afterward king's chamberlain and justice of south Wales, and was summoned to parliament as a baron (1332). His son, RICHARD (*c.* 1305–56), 2nd lord, claimed extensive lands in Scotland in right of his wife, Elizabeth, daughter of John Comyn of Badenoch, and in 1332 he joined Edward Balliol in his invasion of Scotland, being present at the victory of Dupplin (Aug. 12). Captured by the Scots in Sept. 1334, he was released for a ransom of £2,000. His wife, also a coheiress of the Valence earls of Pembroke, brought him the manor of Painswick in Gloucestershire and the castle of Goodrich on the river Wye; their grandson later acquired in her right the Irish lordship of Wexford (1389). Richard afterward fought with the English armies in France at Morlaix (1342) and Crécy (1346). His son, GILBERT (*c.* 1332–87) served in France and also Spain. The family estates in Herefordshire, Gloucestershire, Oxfordshire and Kent were further increased by the marriage of Gilbert's son, RICHARD (*c.* 1361–96) to Ankaret, daughter and heiress of John, 4th Lord Strange of Blackmere, who brought him Whitchurch in Shropshire and other lands. His son, GILBERT (*c.* 1383–1418), 5th lord, campaigned in Wales against Owen Glendower, and, after proving himself one of Henry V's most vigorous captains in France, died during the siege of Rouen. On the death of his infant daughter and heir (1421), he was succeeded as 7th lord by his brother, JOHN (1384–1453), whose marriage to Maud, heiress of the lords Furnival, had brought him wide estates in Yorkshire, including Sheffield castle, which became a favourite family residence until 1618. In reward for his distinguished services in France he was created earl of Shrewsbury (1442) and earl of Waterford (1446).

The 1st earl's son, JOHN (*c.* 1413–60) was chancellor of Ireland during his father's lifetime and treasurer of England (1456–58). A strong Lancastrian, he was rewarded with grants from the forfeited Yorkist estates (1459–60) and was killed fighting for Henry VI at Northampton (July 1460). JOHN (1448–73), 3rd earl, at first shared his father's loyalties and was present in Henry VI's army at the second battle of St. Albans and at Towton (1461), but afterward made his peace with the new king, Edward IV, and was appointed chief justice of north Wales in 1471. His son, GEORGE (1468–1538), 4th earl, began a family tradition of devoted service to the Tudor kings. He fought at the battle of Stoke (1487), was steward of the household (1506–38), commanded the army which invaded France in 1513, and, as king's lieutenant in the north, crushed the rebels of the Pilgrimage of Grace (1536). FRANCIS TALBOT (1500–60), who succeeded his father as 5th earl, joined in the invasions of Scotland (1542 and 1544) and spent much of his life on the Scottish borders, being president of the council of the north from 1549 until his death. His son, GEORGE (*c.* 1522–90), 6th earl, who was styled Lord Talbot during his father's lifetime, was selected by Elizabeth I for the unenviable position of custodian of Mary Stuart, who spent the years 1569–84 in his castles at Sheffield and Tutbury or at Chatsworth house. His task was complicated by the limited funds allowed him by Elizabeth, and by the malicious attempts to discredit him of his second wife, Bess of Hardwick (*see* SHREWSBURY, ELIZABETH TALBOT, Countess of).

The 6th earl was succeeded in turn by two of his sons, but in 1618 the direct male line descended from the first earl ceased, and the title and estates passed to the head of a cadet branch of the family founded by SIR GILBERT TALBOT (d. 1517) of Grafton, Worcestershire, a younger son of the 2nd earl. This was GEORGE TALBOT (1566–1630), 9th earl, a Roman Catholic like nearly all his descendants until 1856, who had been ordained priest before he succeeded to the title. His great great nephew, CHARLES (1660–1718), 12th earl, was raised to the dignity of a duke for his services to William III (*see* SHREWSBURY, CHARLES TALBOT, Duke of), but as he left no son this title died with him and the earldom passed to his cousin, GILBERT (1673–1743), a Roman Catholic priest who never assumed the title. From this time the direct line of Sir Gilbert Talbot of Grafton began to fail. A nephew three times succeeded to an uncle, and the title then passed to a cousin, BERTRAM ARTHUR TALBOT (1832–56), the 17th earl, who died unmarried. The lands were claimed under his will by Lord Edmund Howard, a younger son of the 14th duke of Norfolk and afterward Viscount Fitzalan, but the courts decided that the title and estates should remain together, and a successor to the earldom was found in HENRY JOHN CHETWYND-TALBOT (1803–68), Earl Talbot, head of another line of the descendants of Sir Gilbert Talbot of Grafton, who became the 18th earl of Shrewsbury. The head of this family a century before had been WILLIAM TALBOT (d. 1730), successively bishop of Oxford, Salisbury and Durham. His son, CHARLES (1685–1737), lord chancellor (1733–37), was created Baron Talbot of Hensol in 1733. Charles's son, WILLIAM (1710–82), was raised to the dignity of Earl Talbot in 1761, and William's nephew JOHN CHETWYND (1750–93) added the title of Viscount Ingestre (1784), now usually borne by the earl of Shrewsbury's eldest son. All these titles have since been united in the line descended from the 18th earl. (C. D. R.)

TALBOT, MARY ANNE (1778–1808), the "British Amazon," was born in London on Feb. 2, 1778. She believed herself to be the illegitimate child of the 1st Earl Talbot. Early in her career she eloped, in the disguise of a boy, with a captain. In 1792 she was a drummer in Flanders. In the capture of Valenciennes her lover was killed; and Mary Anne deserted and became cabin boy on a French lugger, which she asserted was captured by the British, who transferred her to the "Brunswick," where she served as a powder monkey, being wounded in Lord Howe's victory of June 1, 1794. For this she later received a small pension. When the wound healed she again went to sea, was captured by the French, and imprisoned for a year and a half. Her sex was not discovered until shortly after she was seized by a pressgang. She finally became a household servant to Robert Kirby, a London publisher, who included an account of her adventures in his *Wonderful Museum* (1804) and in *Life and Surprising Adventures of Mary Anne Talbot* (1809). She died on Feb. 4, 1808.

TALBOT, WILLIAM HENRY FOX (1800–1877), English pioneer in photography (*q.v.*), was born at Melbury house, near Evershot, Dorset, on Feb. 11, 1800, and died there on Sept. 17, 1877. He was educated at Harrow and at Trinity college, Cambridge. Before L. J. M. Daguerre in 1839 exhibited pictures taken by the sun, Talbot had obtained similar success, and as soon as Daguerre's discoveries became known he communicated the results of his experiments to the Royal society. In 1841 he made known his discovery of the calotype or Talbotype process and, after the discovery of the collodion process by F. S. Archer (1851), devised a method of instantaneous photography. With Rawlinson

and Hincks he was one of the earliest to decipher the cuneiform inscriptions from Nineveh.

His works include *Pencil of Nature* (1844); *Hermes or Classical and Antiquarian Researches* (1838–39); *Illustrations of the Antiquity of the Book of Genesis* (1839); *English Etymologies* (1846).

TALBOT OF HENSOL, CHARLES TALBOT, 1ST BARON (1685–1737), lord chancellor of England, the eldest son of William Talbot, bishop of Durham, a descendant of the 1st earl of Shrewsbury, was educated at Eton and Oriel college, Oxford, and became a fellow of All Souls college in 1704. He was called to the bar in 1711, and in 1717 was appointed solicitor general to the prince of Wales. Having been elected a member of the house of commons in 1720, he became solicitor general in 1726, and in 1733 he was made lord chancellor and raised to the peerage with the title of Baron Talbot of Hensol. He died on Feb. 14, 1737. Talbot enjoyed the reputation of a wit; he was a patron of the poet Thomson, and Butler dedicated his *Analogy* to the lord chancellor.

See Lord Campbell, *Lives of the Lord Chancellors and Keepers of the Great Seal*, 8 vol. (1845–69).

TALC is the mineral name for the familiar hydrous magnesium silicate, a common mineral occurring as veins, foliated masses and as a constituent in certain rocks in metamorphic terranes. Talc is distinguished from almost all other minerals by its extreme softness. The hardness of 1 (easily scratched with the fingernail) imparts a soapy or greasy feel which accounts for the name soapstone or steatite for the compact aggregates of talc. Soapstones have been employed since ancient times for carvings, ornaments and utensils. Assyrian cylinder seals, Egyptian scarabs and Chinese figure carvings are notable examples. The resistance to attack from most reagents and to moderate heat makes soapstones especially suitable for sinks and countertops. Talc is also used for lubricants, leather dressings, toilet and dusting powders and certain marking pencils. The major industrial uses of talc and soapstone are in ceramics, paint, rubber, insecticides, roofing and paper.

The colour of talc is white, gray, yellow or various shades of green with a pearly or silvery lustre. Talc is usually found in foliated masses of thin, small, sometimes hexagonal plates. The individual plates easily separate along the perfect basal cleavages into thin flexible but inelastic lamellae (cleavage fragments of muscovite (*q.v.*) which can be confused with talc are elastic). Pyrophyllite (*q.v.*) can be distinguished from talc only by optical or chemical methods.

Mode of Occurrence and Composition.—Talc is found as a metamorphic mineral in the lower facies of metamorphism and is often associated with other minerals such as serpentine, tremolite (amphibole) (*qq.v.*) and forsterite (olivine; *q.v.*) and almost always associated with carbonate: calcite, dolomite or magnesite (*qq.v.*). Talc, $Mg_3Si_4O_{10}(OH)_2$, can occur as an alteration of a pre-existing mineral such as tremolite or forsterite as evidenced by preserving the crystal outlines of the pre-existing minerals. Talc also occurs as veins and masses formed at the expense of and replacing pre-existing mineral aggregates. Soapstones have two general types of occurrences: (1) a large group of soapstones are found intimately associated with metamorphosed carbonate rocks probably originally rich in magnesium. By the addition of water and silica and the removal of carbon dioxide, talc can be formed:

$$3MgCa(CO_3)_2 + 2H_2O + 5SiO_2 \rightarrow 2Mg_3Si_4O_{10}(OH)_2 + 3CaCO_3 + 3CO_2$$

(2) a large group of soapstones are found intimately associated with olivine- and/or pyroxene-rich igneous rocks which have been extensively altered by the addition of water and carbon dioxide or water and silica to form talc, for example:

$$\underset{\text{Forsterite}}{3Mg_2SiO_4} + 2H_2O + 5SiO_2 \rightarrow \underset{\text{Talc}}{2Mg_3Si_4O_{10}(OH)_2}$$

or the reaction may be as follows:

$$\underset{\text{Forsterite}}{4Mg_2SiO_4} + H_2O + 5CO_2 \rightarrow \underset{\text{Talc}}{Mg_3Si_4O_{10}(OH)_2} + \underset{\text{Magnesite}}{5MgCO_3}$$

A sequence of mineral formation ending in talc is very commonly observed: olivine plus pyroxene is altered to tremolite, which is altered to serpentine, which in turn is altered to talc plus carbonate. The alteration of serpentine by the addition of silica to form talc is common, with the result that serpentine-rich rocks and soapstones are usually found together. The conditions for the formation of talc apparently are often met in metamorphism—as evidenced by the world-wide occurrence of soapstones in many metamorphic terranes.

One of the remarkable features about talc is its simple, almost constant composition. Unlike other silicates, even closely related ones, talc appears to be unable to accept iron or aluminum into its structure and form solid solution series (*see* MINERALOGY: *The Nature of Minerals: Solid Solution*). This is especially curious because an iron analogue of talc is known, and the structurally related chlorite (*q.v.*) forms at least a partial solid solution series between the iron and magnesium end members.

Crystal Structure.—The crystal structure of talc, a phyllosilicate, accounts for its physical properties. The phyllosilicates are characterized by a hexagonal sheet structure arrangement of the SiO_4 tetrahedral groups formed by the sharing of three oxygen atoms from each SiO_4 tetrahedron with neighbouring SiO_4 tetrahedra with the silicon atoms arranged at the corners of hexagons. The fourth (unshared) oxygen atom of the tetrahedron is capable of forming chemical bonds with other metal atoms. (For description of this structure *see* AMPHIBOLE; PYROXENE.) In most phyllosilicates, talc included, the unshared oxygens are on the same side of the sheet. Two such sheets (oriented such that their unshared oxygens face each other) are bonded together by the magnesium atoms with the remaining charges balanced by OH groups (located at the centres of the hexagonal groups) also bonded to the magnesium atoms. This arrangement results in a double-sheet structure which completely satisfies the valence demands of the atoms. Therefore, these double-sheet units cannot be bonded to adjacent units by other metal atoms satisfying remaining charges. Consequently, the talc crystal is made up of stacks of these double-sheet units held together in the stack by the weakest of chemical bonds—the van der Waals' forces. Thus, this stacking can be disturbed by slight forces causing the slipping of the double-sheet units along this perfect cleavage direction.

(G. W. DEV.)

TALCA, a province of Chile, bounded on the north by Curicó, east by Argentina, south by Linares and Maule and west by the Pacific. The area, since being separated from the province of Curicó, is 3,722 sq.mi.; pop. (1960) 205,448. In the east the Andean slopes cover a considerable part of its territory, and in the west another large area is covered by the coast range. Between these is the central valley of Chile in which the population and industries of the province are chiefly concentrated. The lower mountainous parts are well wooded. The intermediate plain is fertile and devoted to wheat, grapes and stock. The capital of the province is Talca (pop. [1960] 80,277), on the Río Claro, 156 mi. S. of Santiago by rail. The city, founded in 1692 and revived in 1742, has a variable climate.

TALCAHUANO, a seaport of the province of Concepción, Chile, on the bay of Talcahuano, 8 mi. N.W. of the city of Concepción. Pop. (1960) 99,231 (mun.). It has the best harbour on the Pacific coast of South America and is one of the most important ports of southern Chile. The Chilean government established its chief naval base there.

TALE, a general term, in the usual acceptance of the word, for fictitious narratives, long or short, ancient or modern. In this article "tale" is used as equivalent to the German *Volksmärchen* or the French *conte populaire*. Thus understood, popular tales mean the stories handed down by oral tradition from an unknown antiquity. So understood, popular tales are a subject in mythology. The Homeric epics, especially that of the *Odyssey*, contain adventures (for example, the husband who returns in disguise) which are manifestly parts of the general human stock of popular narrative. Other examples are found in the Rig-Veda and in the myths which were handled by the Greek dramatists. The *Thousand and One Nights* (*q.v.*) is full of popular tales, and

popular tales are the staple of the medieval *Gesta Romanorum* and of the collections of G. Straparola and other Italian storytellers. In all these the story, long circulated among the people, is handled with conscious art. *Histoires ou contes du temps passé* of C. Perrault (1697) is one of the earliest collections of tales which were taken down as they were told by nurses to children. The success of Perrault's tales brought the genre into fashion, and the Comtesse d'Aulnoy invented, or in some cases adapted, *contes*, which retain popularity. But the precise and scientific collection of tales from the lips of the people is not much earlier than the 19th century. The chief impulse to the study was given by the brothers Jacob and Wilhelm Grimm. The first edition of their *Kinder- und Haus-Märchen* was published in 1812. In addition to the *märchen* of Indo-European peoples, the Grimms became acquainted with some Malay stories, some narratives of Bechuanas, Negroes, American Indians and Finnish, Estonian and Magyar stories. *See* NOVEL. (A. L.; X.)

TA-LI, a small but historic city in southwest China, about 170 mi. N.W. of the Yünnan provincial capital of Kunming and located on the western shore of Erh lake. Erh lake occupies a deep structural basin; it is about 30 mi. long and from 2 to 5 mi. wide and its surface is 6,700 ft. above sea level. The west shore of the lake comprises a narrow plain, running the length of the lake and rising on the west to the Ts'ang Shan, a massive mountain range topping 14,000 ft.

Ta-li was founded about 746 A.D. as the capital of the Thai (Tai) people's state of Nan Chao. It remained the capital until the Mongols overran the region in the 13th century and the Thai moved southward. Both the ancient Burma-China caravan route and the World War II Burma road touched the southern corner of the lake plain at Hsia-kuan. Much of the city was destroyed by an earthquake in 1925 and was not rebuilt. The Ta-li plain is the home of the Min-chia people, related to the Thai, who have absorbed much of Chinese culture and who comprise about one-third the 20,000 residents of the city of Ta-li. Ta-li is essentially a regional trade and political centre notable for its fine marble products. In the late 1950s the region became a Min-chia autonomous area. (J. E. Sr.)

TALIESSIN: *see* WELSH LITERATURE.

TALLADEGA, a city of Alabama, U.S., the seat of Talladega county, is located 57 mi. E. of Birmingham in the foothills of the Blue Ridge mountains. The city is in an agricultural and dairying region, with marble and limestone quarries and iron mines. The chief industrial products are cotton textiles, cottonseed oil, machine parts, pipe fittings, soil pipe, lumber, woolen worsted fabric and yarn. Talladega college, a coeducational institution founded in 1867 and affiliated with the Congregational Christian Church (United Church of Christ), and the Alabama School for the Deaf and Blind are located there. Nearby are Cheaha State park and Talladega National forest. The name Talladega, also applying to mountains south and east of the town, is Indian for "border town" as its site was on the edge of the Creek Indians' land. Andrew Jackson defeated a large force of Indians there in 1813. Talladega was settled in the 1830s and incorporated in 1835; the county was established in 1832. For comparative population figures *see* table in ALABAMA: *Population.* (B. Cr.)

TALLAGE, a term employed, while it was in ordinary usage, in the general sense of "tax," and applied to various seignorial, royal and municipal taxes (most of them also known by other names such as "aid"), but particularly to a tax imposed by the lord of an estate upon his unfree tenants. Tallage of this kind was already an established seignorial perquisite in England, as elsewhere in western Europe, in the 12th century. Its arbitrariness was generally greater than that of similar aids taken by lords from free tenants. Its frequency or its amount or both might be at the lord's will, though on many estates in the 13th century it had already become a fixed charge. In England from the reign of Henry II a lord sometimes required specific royal authority to take a tallage—when, for instance, the estate to be tallaged was a former royal estate.

In England at the end of the 12th century, "tallage" established itself as the name of a royal tax (developed by Henry II but not in principle novel) levied on estates in the king's hand and on boroughs. It was assessed by the king's itinerant justices, who had administrative as well as judicial duties, or by special commissioners. The assessors usually negotiated with the community of a borough or estate for a lump sum, but sometimes took the less popular course of tallaging individuals separately. Royal tallages were most frequent during John's reign; objection to them found expression in the Magna Carta. Under Henry III an interval of two to four years was usual and the average yield seemed no greater than before. At its highest, the yield of a tallage exceeded £5,000, and payment was prompt. London's contribution was by far the largest single item, sometimes exceeding a third of the whole. In the reign of Edward I, as the summons of representatives of boroughs to parliaments became more frequent, parliamentary taxation of boroughs and the king's estates was preferred to tallage. Edward I took a tallage in 1304 after a lapse of 36 years. It was last taken in 1312. An attempt to revive it in 1332 was abandoned in favour of a parliamentary grant. (ER. S.)

TALLAHASSEE, a city of Florida, U.S., the capital of the state and the seat of Leon county, is in the northern part of the state 166 mi. W. of Jacksonville, 200 mi. E. of Pensacola and 20 mi. N. of the Gulf of Mexico. The city and the surrounding region are characterized by rolling hills, lakes, streams and giant magnolia and oak trees. Leon county and the 12 counties immediately to the west and east comprise the "big bend" of north central Florida, an agricultural area producing livestock, "shade-cured" (formerly called Spanish or Florida) tobacco, pecans, tung oil trees and diversified farm products. Among the industries of the city are lumber-planing mills, naval store refineries and pecan processing plants. It is a key wholesaling and distribution point for northwest Florida and southwest Georgia. It is the seat of two coeducational universities, Florida State university and the Florida Agricultural and Mechanical university (*see* FLORIDA: *Education*). Among its stately homes are the Grove (completed in 1830), the Columns (1835), Goodwood (1843) and the governor's mansion (1957).

In 1539–40, Hernando De Soto and a Spanish expedition spent the winter in the vicinity, this incident leading to a local assumption that the first observance of Christmas in the present United States was conducted in or nearby the site at that time. Other Spaniards, while fighting the Apalachee Indians of the region, are supposed to have fortified a hill near the site in 1638. By 1675 seven Franciscan missions were established in the region, San Luis de Talimali, in the western section of the present city, being the most important. This mission, sometimes called Fort San Luis, was destroyed and the Apalachee Indians were dispersed by Gov. James Moore of South Carolina during Queen Anne's War (the War of the Spanish Succession, 1701–13), when he invaded Florida to punish the Spanish for having incited Indian attacks against white settlers in South Carolina. Tallahassee then acquired its name from a Creek Indian word meaning "abandoned village." From this time until Spain ceded Florida to the United States in 1821, Tallahassee remained quite insignificant.

When Florida became a U.S. territory it had two capitals, one in St. Augustine and one in Pensacola. The first territorial legislature met in Pensacola in 1822, the second in St. Augustine the next year. Tallahassee was named the permanent capital in 1823 by the territory's American settlers; it was incorporated in 1825 and became the state capital when Florida entered the union in 1845.

Before the American Civil War Tallahassee became a trade centre for the rich cotton plantation country of middle Florida, partially because of its connection with the little port of St. Marks nearby. Negro slavery was more prevalent in Leon and adjoining counties than elsewhere in the state. Because it was so far removed from the significant battle areas, Tallahassee was the only state capital in the Confederacy east of the Mississippi river which was not occupied by Federal troops during the Civil War. In the late 1800s and early 1900s, when south Florida entered its boom era, Tallahassee declined as a commercial and business centre. After 1940 it regained importance because of its position as the governmental centre of the state. The city has a council-

manager form of government, in effect since 1920. For comparative population figures *see* table in FLORIDA: *Population.*
(W. T. Jo.)

TALLAPOOSA, a river of northwestern Georgia and east central Alabama, U.S., rises in the piedmont area of western Georgia about 30 mi. W.N.W. of Atlanta and flows southwest in an irregular, steplike course for about 268 mi. to where it joins the larger Coosa river about 10 mi. N. of Montgomery, Ala., to form the Alabama river. Its only large tributary is the Little Tallapoosa which joins it about midway in its course. All but the lower part of the drainage basin of about 4,500 sq.mi. lies in the Alabama-Georgia piedmont. Falls at Tallassee, Ala., chief city on the Tallapoosa, about 43 mi. upstream from its mouth, and a steep gradient above that prevented use of the river for navigation. Three private power dams (Martin, Yates and Thurlow) above Tallassee have created reservoirs such as Lake Martin (1,630,000 ac.ft.) for river control and recreation as well as power.

The name Tallapoosa is of Creek origin and many Indian villages were situated along the lower river before the 19th century. The Creek War battle of Horseshoe Bend (1814) was fought in a river bend near what is now the northern end of Lake Martin, about 9 mi. N. of Dadeville, Ala. (M. W. M.)

TALLCHIEF, MARIA (1925–), U.S. ballet dancer noted for the crisp perfection of her technique, was born in Fairfax, Okla., Jan. 24, 1925, of American Indian descent. After studying with Bronislava Nijinska she joined Ballet Russe de Monte Carlo in 1942 and soon became a leading dancer. In 1947 she was guest artist at the Paris Opéra, and in 1948, the year it was founded, she joined the New York city ballet as prima ballerina. Tallchief particularly excelled in the roles created for her by George Balanchine in *The Firebird, Orpheus* and *The Nutcracker.*
(LN. ME.)

TALLEMANT DES RÉAUX, GÉDÉON (1619–1692), French writer of informative and entertaining *Historiettes,* or short biographies. Born at La Rochelle on Oct. 2, 1619, of a rich Protestant family, Tallemant went early to Paris and began to frequent literary circles. In 1646 he married Élisabeth Rambouillet in the Protestant temple at Charenton. His later life was made miserable by family bankruptcy and by separation from his wife. She was received into the Catholic Church in 1665, he not until July 1685, about three months before the revocation of the Edict of Nantes. Tallemant died in Paris on Nov. 10, 1692.

Tallemant published nothing in his lifetime save a few poems in anthologies. His *Historiettes* were not published till 1834, from a folio manuscript that had appeared in a Paris auction. This contains a mass of particulars about leading men and women in Parisian society and French public life of the first half of the 17th century. Tallemant appears to have started writing in 1657 from data collected earlier. A good listener, he also took the trouble to seek out well-informed authorities for his lives (such as Boisrobert on Richelieu and Racan on Malherbe), with the aim of recording anything of interest that was not in the printed histories. His new, disturbing and apparently chatty information about dignified persons was at first decried, but research has done much to establish its reliability. These acute jottings, often so near to the actual event, may be considered a firsthand source and have been useful as a corrective to later interpretations. There are modern editions by G. Mongrédien, 8 vol. (1932) and in the "Pléiade Series" by A. Adam (1960–).

See E. Magne, *La Joyeuse Jeunesse . . .* (1921) and *La Fin troublée de Tallemant des Réaux* (1922); also E. Gosse, *Tallemant des Réaux or the Art of Miniature Biography* (1925). (W. G. ME.)

TALLEYRAND-PÉRIGORD, CHARLES MAURICE DE (1754–1838), French diplomatist and statesman, was born at Paris on Feb. 12 or 13, 1754, the son of lieutenant general Charles Daniel de Talleyrand-Périgord. His parents, descended from ancient and powerful families, were in constant attendance at the court of Louis XV. In his third or fourth year, while under the care of a nurse in Paris, he fell from a chest of drawers and injured his foot for life. This accident darkened his prospects; for though by the death of his elder brother he should have represented the family and entered the army, yet he forfeited the rights of primogeniture, and the profession of arms was thenceforth closed to him. At the age of 8 he was sent to the Collège d'Harcourt at Paris, and, at 13 to St. Sulpice, where he conceived a dislike of the doctrines and discipline thrust upon him. After a visit to his uncle, the archbishop of Reims, he returned to St. Sulpice to finish his preliminary training for the church, but in his spare time he read the works of Montesquieu, Voltaire and other writers who were beginning to undermine the authority of the *ancien régime,* both in church and state. As subdeacon he witnessed the coronation of Louis XVI at Reims, but he did not take priest's orders until four years later. While rejecting the authority of the church in the sphere of dogma and intellect, he observed the proprieties of life, and respected the outward observances of religion.

During his life at Paris he frequented the salon of Madame de Genlis, and there formed his ideas in favour of political and social reform. After taking his licentiate in theology in March 1778, he gave little more attention to theological studies. Nevertheless his ability and his social position gained for him in the year 1780 the position of agent-general of the clergy of France. The growing claims of the state on the exchequer of the clergy made his duties responsible. At the extraordinary assembly of the clergy in 1782 he made proposals, by one of which he sought, though in vain, to redress the most glaring grievances of the underpaid *curés.* Though the excellence of his work as agent-general in the years 1780–86 was fully acknowledged, yet he did not gain a bishopric until the beginning of the year 1789, probably because the king disliked him as a freethinker. He now became bishop of Autun, and was installed on March 13.

The Revolution.—The first important act of the new bishop was to draw up a program of the reforms which he desired to see carried out by the states-general of France. It comprised the following items: the formation of a constitution which would strengthen the monarchy by calling to it the support of the whole nation, the drafting of a scheme of local self-government on democratic lines, the reform of the administration of justice and of the criminal law and the abolition of the most burdensome of feudal and class privileges.

His influence in the estate of the clergy, however, was cast against the union of the three estates in a single assembly, and he voted in the minority of his order which in the middle of June opposed the merging of the clergy in the National Assembly. Thereafter he inclined more and more to the democratic side. In July he was chosen as one of the committee to prepare a draft of a constitution; and in the session of the assembly which the comte de Mirabeau termed the *orgie* of the abolition of privileges (Aug. 4) he intervened in favour of discrimination and justice. On Oct. 10 he proposed to the assembly the confiscation of church lands to the service of the nation, but on terms rather less rigorous than those in which Mirabeau (*q.v.*) carried the proposal into effect on Nov. 2. He identified himself in general with the left of the assembly. The course of events harmonized with the anticlerical views of Talleyrand, and he gradually loosened the ties that bound him to the church. He took little part in, though he probably sympathized with, the debates on the measure known as the Civil Constitution of the Clergy. When the Assembly sought to impose on its members an oath of obedience to the new decree, Talleyrand and three other bishops complied out of the thirty who had seats in the Assembly. The others, followed by the greater number of the clergy throughout France, refused, and thenceforth looked on Talleyrand as a schismatic. He did not long continue to officiate, as many of the so-called "constitutional" clergy did; for, on Jan. 21, 1791, he resigned the see of Autun, and in the month of March was placed under the ban of the church by the pope.

Just before his resignation he had been elected, with Mirabeau and Sieyès, a member of the department of Paris; and for some eighteen months he supported the cause of order in the turbulent capital. Though he was often on strained terms with Mirabeau, their views generally coincided. Talleyrand's reputation for immorality, however, was as marked as that of Mirabeau. While excelling Mirabeau in suppleness and dexterity, Talleyrand lacked the force of character possessed by the great "tribune of the

people"; and his influence was gradually eclipsed by the Girondins and the Jacobins.

Debarred from election to the second National Assembly (known as the Legislative) by the self-denying ordinance passed by the "constituents," Talleyrand, at the close of 1791, sought to enter the sphere of diplomacy. The condition of affairs on the continent seemed to French enthusiasts to presage an attack by the other powers on France. In reality those powers were far more occupied with the Polish and eastern questions than with the affairs of France; and the declaration of Pilnitz, drawn up by the sovereigns of Austria and Prussia, which appeared to threaten France with intervention, was recognized by all well-informed persons to be "a loud-sounding nothing." The French foreign minister, Delessart, hoping to confirm Pitt in his intention of keeping England neutral, sent Talleyrand to London. Talleyrand found public opinion there so far friendly that he wrote off to Paris, "Believe me, a *rapprochement* with England is no chimera." Pitt received him cordially; and to Grenville the envoy stated his hope that the two free nations would enter into close and friendly relations. Great Britain returned no definite answer, but Talleyrand was convinced that it would not intervene against France unless the latter attacked the Dutch Netherlands.

He returned to Paris on March 10, 1792, to persuade the foreign minister (Dumouriez now held that post) of the need of having a fully accredited ambassador at London. The ex-marquis Chauvelin was appointed, with Talleyrand as adviser. But the overthrow of the monarchy on Aug. 10 and the September massacres rendered hopeless all attempts at an *entente cordiale* between the two peoples; and the provocative actions of Chauvelin undid all the good accomplished by the tact and moderation of Talleyrand. The latter now sought to escape from France, where events were becoming intolerable; he landed in England on Sept. 23, avowedly on private business, but still animated by the hope of averting a rupture between the two governments. In this he failed. The provocative actions of the French Convention had brought the two nations to the brink of war, when the execution of Louis XVI (Jan. 21, 1793) made it inevitable. Talleyrand was expelled from British soil and made his way to the United States. There he spent thirty months in a state of growing uneasiness and discontent.

The Directory.—The course of events after the Thermidorian reaction of July 1794 favoured his return to France. Thanks to the efforts of Daunou and others his name was removed from the list of *émigrés*, and he set sail for Europe in November 1795. Landing at Hamburg in the January following, he spent some time there in the company of his friends Madame de Genlis and Reinhard and returned to Paris in Sept. 1796. He gained the post of foreign minister, entering on his duties in July 1797.

Talleyrand, despite the weakness of his own position (he was as yet little more than the chief clerk of his department), soon came to a good understanding with Bonaparte, and secretly expressed to him his satisfaction at the terms which the latter dictated at Campo Formio (Oct. 17, 1797). The *coup d'état* of Fructidor (September 1797) had perpetuated the Directory and led to the exclusion of the two "moderate" members, Carnot and Barthélémy; but Talleyrand saw that power belonged really to the general who had brought about the *coup d'état* in favour of the Jacobinical Directors headed by Barras.

With the renewal of war on the continent Talleyrand had little or no connection. His powers as minister were limited, and he regretted the extension of the area of war. Moreover, in the autumn of 1797 his reputation for political morality (never very bright) was overclouded by irregular dealings with the envoys of the United States sent to arrange a peaceful settlement of certain disputes with France. Talleyrand refused to clear himself of the charges made against him, and the incident probably told against his chances of admission into the Directory. A year later he resigned the portfolio for foreign affairs (July 20, 1799), probably because he foresaw the imminent collapse of the Directory. If so, his premonitions were correct. Their realization was assured by the return to France of the "Conqueror of the East" in October. The general and the diplomatist soon came to an understanding, and Talleyrand brought about the alliance between Bonaparte and Sieyès (*q.v.*) (then the most influential of the five Directors) which paved the way for the *coup d'état* of Brumaire.

Talleyrand's share in the actual events of the 18th, 19th Brumaire (9th, 10th of November) 1799 was limited to certain dealings with Barras on the former of those days. About midday he took to Barras a letter, penned by Roederer, requesting him to resign his post as Director; on that afternoon Barras left Paris. Talleyrand's reward for helping on the winning cause was the ministry for foreign affairs. In the great work of reconstruction of France now begun by the First Consul, Talleyrand played no unimportant part. His great aim was to bring about peace, both international and internal. After the battle of Marengo an Austrian envoy had to come to Paris in response to a proposal of Bonaparte, and Talleyrand persuaded him to sign terms of peace. These were indignantly repudiated at Vienna, but peace was made between the two Powers at Lunéville on Feb. 9, 1801.

As regards French affairs, Talleyrand used his influence to help on the repeal of the vexatious laws against *émigrés*, nonjuring priests, and the royalists of the west. He was also in full sympathy with the policy which led up to the signature of the Concordat of 1801–2 with the pope (*see* Concordat); but it is probable that he had a hand in the questionable intrigues which accompanied the closing parts of that complex and difficult negotiation. At the end of June 1802 the pope removed Talleyrand from the ban of excommunication and allowed him to revert to the secular state. On Sept. 10, 1803, owing to pressure put on him by Bonaparte, he married Madame Grand, a *divorcée* with whom he had long been living.

During the meeting of Italian notables at Lyons early in 1802 Talleyrand was serviceable in manipulating affairs in the way desired by Bonaparte, and it is known that the foreign minister suggested to them the desirability of appointing Bonaparte president of the Cisalpine Republic, which was thenceforth to be called the Italian Republic. In the negotiations for peace with England which went on at Amiens during the winter of 1801–2 Talleyrand had no direct share, these (like those at Lunéville) being transacted by Napoleon's eldest brother, Joseph Bonaparte (*q.v.*). On the other hand he helped the First Consul in assuring French supremacy in Switzerland, Italy and Germany. In Germany the indemnification of the princes who lost all their lands west of the Rhine was found by secularizing and absorbing the ecclesiastical states, and Talleyrand is known to have made large sums of money by trafficking with the claimants of church lands.

While helping to establish French supremacy in neighbouring states and assisting Bonaparte in securing the title of First Consul for life, Talleyrand sought all means of securing the permanent welfare of France. He worked hard to prevent the rupture of the peace of Amiens in May 1803, and he did what he could to prevent the sale of Louisiana to the United States earlier in the year. These events, as he saw, told against the best interests of France and endangered the gains which she had secured by war and diplomacy. Thereafter he strove to moderate Napoleon's ambition and to preserve the European system as far as possible. The charges of duplicity or treachery made against the foreign minister by Napoleon's apologists are nearly all unfounded. On assumption of the imperial title by Napoleon in 1804, Talleyrand became grand chamberlain of the empire.

The Empire.—Talleyrand had rarely succeeded in bending the will of the First Consul. He altogether failed to do so with the Emperor Napoleon. His efforts to induce his master to accord lenient terms to Austria in November 1805 were futile; and he looked on helplessly while that power was crushed, the Holy Roman empire swept away, and the Confederation of the Rhine set up in central Europe. In the negotiations with England which went on in the summer of 1806 Talleyrand had not a free hand; they came to nought, as did those with Russia which had led up to the signature of a Franco-Russian treaty at Paris by d'Oubril which was at once disavowed by the tsar. The war with Prussia and Russia was ended by the treaties of Tilsit (July 1807). Talleyrand had a hand only in the later developments of these negotiations.

Talleyrand had long been weary of serving a master whose policy he more and more disapproved, and after the return from Tilsit to Paris he resigned office. Nevertheless Napoleon retained him in the council and took him with him to the interview with the Emperor Alexander I. at Erfurt (September 1808). Talleyrand disapproved of the Spanish policy of Napoleon which culminated at Bayonne in May 1808; and the stories to the contrary may in all probability be dismissed as idle rumours. On Talleyrand now fell the disagreeable task of entertaining at his new mansion at Valençay, in Touraine, the Spanish princes virtually kidnapped at Bayonne by the emperor. They remained there until March 1814. At the close of 1808, while Napoleon was in Spain, Talleyrand entered into certain relations with his former rival Fouché (*q.v.*), which aroused the solicitude of the emperor and hastened his return to Paris. He subjected Talleyrand to violent reproaches, which the ex-minister bore with his usual ironical calm.

After the Danubian campaign of 1809 and the divorce of Josephine, Talleyrand used the influence which he still possessed in the imperial council on behalf of the choice of an Austrian consort for his master, for, like Metternich (who is said first to have mooted the proposal), he saw that this would safeguard the interests of the Habsburgs, whose influence he felt to be essential to the welfare of Europe. He continued quietly to observe the course of events during the disastrous years 1812–13; and even at the beginning of the Moscow campaign he summed up the situation in the words, "It is the beginning of the end." Early in 1814 he saw Napoleon for the last time; the emperor upbraided him with the words: "You are a coward, a traitor, a thief. You do not even believe in God. You have betrayed and deceived everybody. You would sell even your own father." Talleyrand listened unmoved, but afterwards sent in his resignation of his seat on the council. It was not accepted. He had no share in the negotiations of the congress of Châtillon in February–March 1814. On the surrender of Paris to the allies (March 30, 1814), the Emperor Alexander I. took up his abode at the hôtel Talleyrand, and there occurred the conference wherein the statesman persuaded the tsar that the return of the Bourbons was the only possible solution of the French problem, and that the principle of legitimacy alone would guarantee Europe against the aggrandizement of any one state or house. The tsar was convinced, and Talleyrand, on April 1, convened the French senate (only 64 members out of 140 attended), and that body pronounced that Napoleon had forfeited the crown. Ten days later the emperor recognized the inevitable and signed the Act of Abdication at Fontainebleau. The next effort of Talleyrand was to screen France under the principle of legitimacy and to prevent the schemes of partition on which some of the German statesmen were bent. Thanks mainly to the support of the tsar and of England these schemes were foiled; and France emerged from her disasters with frontiers which were practically those of 1792.

The Restoration.—At the congress of Vienna (1814–15) for the settlement of European affairs, Talleyrand, as the representative of the restored house of Bourbon in France, managed adroitly to break up the league of the Powers (framed at Chaumont in February 1814) and assisted in forming a secret alliance between England, Austria and France in order to prevent the complete absorption of Poland by Russia and of Saxony by Prussia. The new triple alliance had the effect of lessening the demands of those Powers and of leading to the well-known territorial compromise of 1815. Everything was brought into a state of uncertainty once more by the escape of Napoleon from Elba; but the events of the Hundred Days, in which Talleyrand had no share—he remained at Vienna until June 10—brought in the Bourbons once more; and Talleyrand's plea for a magnanimous treatment of France under Louis XVIII. once more prevailed in all important matters. On July 9, 1815, he became foreign minister and president of the council under Louis XVIII., but diplomatic and other difficulties led him to resign his appointment on Sept. 23, 1815, Louis, however, naming him high chamberlain and according him an annuity of 100,000 francs. The rest of his life calls for little notice except that at the time of the July Revolution of 1830, which unseated the elder branch of the Bourbons, he urged Louis Philippe, duke of Orleans (*q.v.*), to take the throne offered to him by popular acclaim. The new sovereign offered him the portfolio for foreign affairs; but Talleyrand signified his preference for the embassy in London. In that capacity he took an important part in the negotiations respecting the founding of the new kingdom of Belgium. In April, 1834 he crowned his diplomatic career by signing the treaty which brought together as allies France, Great Britain, Spain and Portugal; and in the autumn of that year he resigned his embassy. During his last days he signed a paper signifying his reconciliation with the Roman Catholic Church and his regret for many of his early actions. The king visited his death-bed. His death, on May 17, 1838, called forth widespread expressions of esteem for the statesman who had rendered such great and varied services to his country. He was buried at Valençay. He had been separated from the former Madame Grand in 1815 and left no heir.

The Talleyrand *Mémoires* were edited by the duc de Broglie in 5 vols. (Paris, 1891–92). They have been translated into English by A. Hall, 5 vols. (London, 1891–92). Of his letters and despatches the following are the chief collections:—G. Pallain, *La mission de Talleyrand à Londres en 1792* (Paris, 1889), and *Le ministère de Talleyrand sous le Directoire* (Paris, 1891); P. Bertrand, *Lettres inédites de Talleyrand à Napoléon*, 1800–09 (Paris, 1889); G. Pallain, *Talleyrand et Louis XVIII.* (Paris, 1881), and *Ambassade de Talleyrand à Londres* (1830–34), 2 vols. (Paris, 1891).

Among the biographies, or biographical notices, of Talleyrand the following are, on the whole, hostile to him: G. Touchard Lafosse, *Talleyrand, histoire politique et vie intime* (Paris, 1848); G. Michaud, *Hist. politique et privée de Talleyrand* (Paris, 1853); A. Pichot, *Souvenirs intimes sur Talleyrand* (Paris, 1870); Sainte-Beuve, "Talleyrand," in *Nouveaux lundis*, No. xii.; and Villemarest, *Talleyrand.* The estimate of him of Sir H. L. E. Bulwer Lytton in his *Historical Characters*, 2 vols. (London, 1867) and that of Lord Brougham in *Historical Sketches of Statesmen*, 3 vols. (London, 1845, new edition), are better balanced, but brief. Of recent biographies of Talleyrand the best are Lady Blennerhasset's *Talleyrand* (Berlin, 1894, Eng. translation by F. Clarke, 2 vols. London, 1894); *Talleyrand, a Biographical Study*, by Joseph McCabe (London, 1906); and Bernard de Lacombe, *La vie privée de Talleyrand* (1910); C. Dupuis, *Le Ministère de Talleyrand en 1814* (2 vols., 1919). (J. H. Ro.)

TALLIEN, JEAN LAMBERT (1767–1820), French revolutionary, was born in Paris in 1767. He entered a printer's office, and by 1791 he had reached the position of overseer of the printing department of the *Moniteur.* While thus employed he conceived the idea of the *journal-affiche*, and after the arrest of the king at Varennes in June 1791 he placarded a large printed sheet on all the walls of Paris twice a week, under the title of the *Ami des Citoyens, journal fraternel.*

This enterprise, financed by the Jacobin Club, made him well known to the revolutionary leaders; and he made himself still more conspicuous in organizing the great "Fête de la Liberté" on April 15, 1792, in honour of the released soldiers of Château-Vieux, with Collot d'Herbois. He was active in the events of Aug. 10, and was made clerk to the Revolutionary Commune of Paris. At the close of September he resigned his post on being elected a deputy to the Convention by the department of Seine-et-Oise. He took his seat upon the Mountain, and was one of the most vigorous Jacobins, particularly in his defence of Marat, on Feb. 26, 1793; he was elected a member of the Committee of General Security on Jan. 21, 1793. He took an active part in the *coups d'état* of May 31 and June 2, which resulted in the overthrow of the Girondists. On Sept. 23, 1793, he was sent with Claude Alexandre Ysabeau (1754–1831) to Bordeaux. This was the month in which the Terror was organized under the superintendence of the Committees of Public Safety and General Security.

Tallien showed himself one of the most vigorous of the proconsuls sent over France to establish the Terror in the provinces and soon awed the great city. Among his prisoners was Thérèse, the divorced wife of the comte de Fontenay, and daughter of the Spanish banker, François Cabarrus, one of the most fascinating women of her time, and Tallien not only spared her life but fell in love with her. Suspected of "Moderatism" on account of this incident, Tallien increased, in appearance, his revolutionary zeal, but from the lives Thérèse saved by her entreaties she received

the name of "Our Lady of Thermidor," after the 9th of Thermidor. Tallien was even elected president of the Convention on March 24, 1794. Robespierre began to see, however, that he must strike at his own colleagues in the committees if he was to carry out his theories; but they determined to strike first, and on the great day of Thermidor Tallien opened the attack upon Robespierre. Robespierre and his friends were guillotined; and Tallien, as the leading Thermidorian, was elected to the Committee of Public Safety. He suppressed the Revolutionary Tribunal and the Jacobin Club and fought bravely against the insurgents of Prairial. He was supported by Thérèse, whom he married on Dec. 26, 1794, and who became the leader of the social life of Paris. His last political achievement was in July 1795, when he was present with Hoche at the destruction of the army of the *émigrés* at Quiberon, and ordered the executions which followed.

Tallien's political importance came to an end with the Convention for, though he sat in the Council of Five Hundred, the moderates attacked him as terrorist, and the extreme party as a renegade. Madame Tallien also tired of him, and became the mistress of the rich banker Ouvrard. Bonaparte took him to Egypt in his great expedition of June 1798, and he edited the *Décade Egyptienne* in Cairo. But General J. F. Menou sent him away from Egypt, and he was captured by an English cruiser and taken to London, where he had a good reception among the Whigs and was well received by Fox. On returning to France in 1802 he obtained a divorce from his wife (who in 1805 married the comte de Caraman, later prince de Chimay), and was for some time without employment. At last he was appointed consul at Alicante, and remained there until he lost the sight of one eye from yellow fever. On returning to Paris he lived on his half-pay, and his latter days were spent in poverty. He died in Paris on Nov. 16, 1820.

Tallien left an interesting *Discours sur les causes qui ont produit la Révolution française* (1791) and a *Mémoire sur l'administration de l'Égypte à l'arrivée des Français. See* "Tallien et l'Expédition d'Égypte" in *La Révolution Française: Revue d'histoire moderne et contemporaine*, t. iii. p. 269. On Madame Tallien *see* Arsène Houssaye, *Notre Dame de Thermidor* (1866); J. Turquan, *Souveraines et grandes Dames: La citoyenne Tallien, témoignages des contemporains et documents inédits* (1898); and Louis Gastine, *La belle Tallien* (1909).

TALLINN, (Tallin, Reval), primary seaport and capital of Estonia, in 59° 26′ N., 24° 46′ E., on a bay in the south coast of the Gulf of Finland. Pop. (1959) 281,714. The port, frozen over about 45 days each winter, accommodates vessels drawing 30 ft. The chief exports are textiles, cereals, timber, paper and Portland cement, and the imports foodstuffs, cotton and coal. The town has textile, paper, cement and timber industries. The remains of the Toompea citadel (1539) and the ancient city wall with many of its towers make Tallinn picturesque. Many new buildings dominate the town, including a town hall. The Art Museum has a fine collection of modern Estonian painting and sculpture.

A Danish settlement on the high Silurian crag known as the Domberg existed in 1093, and the Danish king, Valdemar II, built a castle in 1219. It was captured by the Livonian knights in 1228, but returned to the Danes in 1237. Merchants from Lübeck and Bremen settled here in the 13th century, and it became a port of the Hanseatic league. It was fortified early in the 14th century and in 1343 sustained a siege by the revolted Estonians. Valdemar III sold Estonia to the Teutonic knights in 1346, but on the dissolution of the order, in 1561, Estonia and Reval surrendered to the Swedish king, Erik XIV. A great conflagration in 1433, the pestilence of 1532, the bombardment by the Danes in 1569, and the Russo-Livonian War destroyed its trade. The Russians besieged Reval in 1570 and 1577 and in 1710 it was surrendered to Peter the Great, who erected a port for his Baltic fleet. His successors continued to fortify the access to Reval from the sea, especially in the early years of the 19th century. It passed from Russian to Estonian rule in 1918, but fell to Russia in 1940, and to Germany in 1941. In World War II the shipping sections were heavily bombed.

TALLIS (Tallys, Talys or Tallisius), **THOMAS** (*c.* 1515–1585), justly styled "the father of English cathedral music," was born probably before 1515. It has been conjectured that, after singing as a chorister at old Saint Paul's under Thomas Mulliner, he obtained a place among the children of the chapel royal. He is known to have become organist at Waltham abbey, where, on the dissolution of the monastery in 1540, he received, in compensation for the loss of his preferment, 20s. for wages and 20s. for reward. In the library of the British Museum there is preserved a volume of ms. treatises on music, once belonging to the abbey, on the last page of which appears his autograph, "Thomas Tallys"—the only specimen known. Not long after his dismissal from Waltham, Tallis was appointed a gentleman of the chapel royal.

One of the earliest compositions by Tallis to which an approximate date can be assigned is the well-known *Service in the Dorian Mode*, consisting of the *Venite*, *Te Deum*, *Benedictus*, *Kyrie*, *Nicene Creed*, *Sanctus*, *Gloria in Excelsis*, *Magnificat* and *Nunc Dimittis*, for four voices, together with the *Preces*, *Responses*, *Paternoster* and *Litany*, for five, all published for the first time, in the Rev. John Barnard's *First Book of Selected Church Music*, in 1641, and reprinted, with the exception of the *Venite* and *Paternoster*, in Boyce's *Cathedral Music* in 1760. (Boyce's omission of the very beautiful *Venite* is hard to account for.) That this work was composed for the purpose of supplying a pressing need, after the publication of the second prayer-book of King Edward VI. in 1552, there can be no doubt. Written in the style known among Italian composers as *lo stile famigliare*, *i.e.*, in simple counterpoint of the first species, *nota contra notam*, with no attempt at learned complications of any kind, it adapts itself with equal dignity and clearness to the expression of the verbal text it is intended to illustrate. In self-restraint the *Litany* and *Responses* bear a close analogy to the *Improperia* and other similar works of Palestrina.

Tallis raised the English school to a height which it was to maintain until the death of its last representative, Orlando Gibbons, in 1625. Though this school is generally said to have been founded by Dr. Tye, there can be no doubt that Tallis was its greatest master, and that it was indebted to him alone for the infusion of new life and vigour which prevented it from degenerating, as some of the earlier Flemish schools had done, into a mere vehicle for the display of fruitless erudition. Like every other great musician of the period, he produced occasionally works confessedly intended for no more exalted purpose than the exhibition of his stupendous skill. In his canon *Miserere nostri* (given in Hawkins's *History of Music*) the intricacy of the contrapuntal devices seems little short of miraculous; yet the resulting harmony is smooth and normal, and only the irregular complexity of the rhythm betrays the artificiality of its structure. The famous forty-part motet, *Spem in alium*, written for eight five-part choirs, stands on a far higher plane, and the *tour de force* of handling freely and smoothly so many independent parts is the least remarkable of its qualities. It was edited by Dr. A. H. Mann in 1888 (London, Weekes and Company). (In the sixth volume of *Tudor Church Music* [Oxford University Press, 1928] all the forty parts are printed on one page.) The art with which the climaxes are built up shows that Tallis's object in writing for forty voices is indeed to produce an effect that could not be produced by thirty-nine. But this illustrates one phase only of Tallis's many-sided genius, which shines with equal brightness in the eight psalm tunes (one in each of the first eight modes) and unpretending little *Veni Creator*, printed in 1567 at the end of Archbishop Parker's *First Quinquagene of Metrical Psalms*.

In 1575 Tallis and his pupil William Byrd—as great a contrapuntist as himself—obtained from Queen Elizabeth royal letters patent granting them the exclusive right of printing music and ruling music-paper for twenty-one years; and, in virtue of this privilege, they issued, in the same year, a joint work, entitled *Cantiones quae ab argumento Sacrae vocantur, quinque et sex partium*, containing sixteen motets by Tallis and eighteen by Byrd, all of the highest degree of excellence. Some of these motets, adapted to English words, are now sung as anthems in the Anglican cathedral service. But no such translations appear to have been made during Tallis's lifetime; and there is strong

reason for believing that, though both he and Byrd outwardly conformed to the new religion, and composed music expressly for its use, they remained Catholics at heart.

Tallis's contributions to the *Cantiones Sacrae* were the last of his compositions published during his lifetime. He died on November 23, 1585, and was buried in the parish church at Greenwich, where a quaint rhymed epitaph, preserved by Strype, and reprinted by Burney and Hawkins, recorded the fact that he served in the chapel royal during the reigns of Henry VIII., Edward VI., Mary, and Elizabeth. This was destroyed with the old church about 1710; but a copy has since been substituted. Portraits, professedly authentic, of Tallis and Byrd, were engraved by Vandergucht in 1730 for Nicolas Haym's projected *History of Music*, but never published. One copy only is known to exist. A list of the printed and manuscript works of Tallis will be found in Grove's *Dictionary*.

TALLOW, the solid oil or fat of ruminant animals, but commercially obtained almost exclusively from oxen and sheep. *See* OILS, FATS AND WAXES.

TALLOW TREE, the common name of a small tree, *Sapium sebiferum*, family Euphorbiaceae, a native of China, but cultivated in other warm countries. The seeds are thickly coated with a white greasy substance—so-called vegetable tallow—from which candles are made, and which is also used in soapmaking and dressing cloth in parts of the orient. Elsewhere the tallow tree is used as a shade or ornamental tree. The butter tree or tallow tree of Sierra Leone is *Pentadesma butyracea*, family Guttiferae. The fruit, which is four to five inches long and about three inches in diameter, has a thick fleshy rind abounding in a yellow greasy juice which is used locally as butter.

TALLY. The wooden tally used in reckoning comes from two primitive notions; that of notching (scoring) a piece of wood for counting, as Robinson Crusoe did, and that of the broken stick shared between two parties to a bargain: the developed tally combines both. Use of this double tally, once prevalent all over Europe, is still frequent in the less advanced countries and not unknown in most. In England, though now nearly obsolete, it was exceedingly common mediaevally and its methods were highly developed; it endured long in certain connections; and its nomenclature survives in many words; to score at cricket (in Pickwick's day "to notch") derives from the single tally, and the verb "to tally" comes obviously from the double one; and other derivatives are many, if not quite so direct.

In England, however, the chief interest of the tally centres in its public use. This is earlier even than the very early "exchequer" organization (*see* EXCHEQUER); by a date not long after 1100 it was a settled system—a system, moreover, carefully differentiated from any private one and used for money only; and the tally continued to be the recognized form of receipt for payments into the Royal Treasury down to 1826. During this long period, though there were modifications of wording, it changed little outwardly save for a continually increasing length; the original 9 in. being extended in one extreme example which still survives at the Bank of England to 8 ft. 6 inches. This, however, is due solely to the increased number of thousands (notches of the thickness of a man's hand) which the tally might be required to show; and (apart from any difficulty due to the writing) a 13th century clerk could have interpreted it; the revolutionary changes lie, not in the tally's form, but in its employment.

The Assignment System.—At a very early date anticipation of the royal revenue became habitual and in this practice the tally was invaluable; nothing was easier than to "levy" a tally for a sum due for payment later and to "assign" this as payment to a royal creditor. Probably, at certain periods, creditors themselves had little objection to a practice which, in an age of difficult transport and clumsy currency, held many advantages. At the same time it might obviously lead to the lowering of royal credit and to endless confusion in accounts. Assignments were, in effect, payments; but tallies were essentially receipts; and as such, from the 13th century onwards, were all entered on "receipt rolls" and, later, figured in numerous supplementary series of records. Moreover an "assignment" tally frequently went wrong, and the mediaeval administrator could think of no better device than to substitute a fictitious "loan" on the roll, so as to square the account. The roll of "receipts" might thus include actual payments of cash, sums credited long before they were paid or due, and entries which indicate no payment at all, but rather a debt to be cleared later. The difficulties of the modern historian desiring to use these (exceedingly important) records, with none of the contemporary clerk's expert knowledge, needs no emphasis.

The Last Tally, 1826.—By Pepys' time tally-making at Westminster had become a ritual involving numerous officials, much delay and many fees, and from then onwards periodical attempts were made to destroy it. But vested interest successfully withstood reform till the 18th century; and it was not till 1826 that the last tally was levied and a system of "indented cheque receipts" substituted. A few years later (1834) reforming zeal celebrated its triumph by destroying (in the furnaces which heated the House of Lords) the whole accumulation of ancient tallies, with disastrous results; the old Houses of Parliament were burned and, incidentally (till the casual discovery in 1909, in the Chapel of the Pyx, of about 1,300 Exchequer tallies, practically all "stocks" of the 13th century), English historians were left without an essential clue to the understanding of some of the richest record series in the national collections.

See Hilary Jenkinson, "Tallies," *Archaeologia*, lxii (1911) and lxiv (1924). (H. JE.)

TALMA, FRANÇOIS JOSEPH (1763–1826), French actor, a rebel in politics and in art, was born in Paris on Jan. 16, 1763. He received part of his excellent education in England, where his father, a dentist, worked for several years. Although Talma practised dentistry after returning to Paris, he soon entered the School of Elocution for future *sociétaires* of the Comédie Française and made his debut at the famed theatre on Nov. 21, 1787, as Seide in Voltaire's *Mahomet*. Influenced by his friend the painter David, Talma became one of the early advocates of historical costuming when he appeared in a Roman toga and headdress in the small role of the tribune Proculus in Voltaire's *Brutus*. He found his great opportunity in Joseph Chénier's antimonarchical *Charles IX*, which not only provided him with an exciting role, acted with careful attention to psychological and historical detail, but also led him to challenge the monopoly of the aristocratic Comédie Française. The revolutionary sentiments of Chénier's play caused demonstrations in the theatre and aroused such dissension among the *sociétaires* that republican Talma established a rival troupe, known as the Théâtre de la République, on the site of the present Théâtre Français. There Talma developed realism in staging and costuming. When the two companies were reunited in 1799 at his theatre as the Comédie Française, Talma emerged the supreme tragedian of the era. Napoleon heaped honours on him; in 1808, for example, he had Talma perform at Erfurt before an audience of crowned heads. In both his scenic reforms and his acting, Talma was a leading precursor of 19th-century French romanticism and realism. He substituted intense emotionalism and realistic accuracy for the dignified repose of the French classical tradition. His artistic credo was "grandeur without pomp, nature without triviality." Talma died in Paris on Oct. 19, 1826.

Talma wrote *Reflections on the Actor's Art* as a preface to the *Memoires de Lekain* (1825). (H. K. CY.)

TALMUD (Heb. "learning"), a compilation that was revered by the Jews as a sacred book from the time it was compiled down to the 18th century, the laws expounded there being considered of divine origin on a par with those in the Torah (*see* PENTATEUCH). After the emergence of the *Haskalah* (Enlightenment) movement, and later upon the rise of the Reform movement, many Jews ceased to think of it as divinely inspired, regarding it nonetheless as of great value for the study of Jewish history, law, theology, folklore and ethics. A large segment of the Jewish population continues to regard the laws set forth in the Talmud as having been given by God, and the power of thought displayed in it by the Jewish sages of antiquity inspires profound respect in all students.

The Talmud comprises two main sections, the Mishnah and

the Gemara (the latter often called Talmud also), plus some additional laws called Baraitot and passages from another collection called Tosefta. The Mishnah ("repetition") is a lawbook, the production of the sages who lived in Palestine before the destruction of the Second Temple (A.D. 70) and a century and a half thereafter. The Gemara ("completion"), a commentary on the Mishnah, is the work of later sages, the *amoraim*. The Talmud deals, in the main, with legal matters, and this part is called *halaka* (pl. *halakot;* "law"); the nonlegal part is called *haggadah* ("saying," "narrative"). The *halakot,* the laws embodied in it, are based on Oral Law.

The Talmud is a storehouse of law, religion, history, ethics, metaphysical speculations, medical science, astronomy and folklore. It is an encyclopaedia covering every phase of human activity, a mine of information for the study of religion, history and civilization not only of the Jews but of the peoples of the entire middle east. It is important for a proper understanding of the origin of Christianity, since this literature came from men who taught at the time when Jesus lived. It is essential for a true comprehension of the controversies over the law between the Pharisees and Jesus as recorded in the Synoptic Gospels.

For the Jews the Talmud was not merely a lawbook, regulating life; its discussions, being divinely inspired, were the purpose and essence of life itself. When the Jews were persecuted and driven from public life during the middle ages, they found consolation in the study of the Talmud. It stimulated and cultivated their minds, saving them from stagnation. By its study they were assured of a reward in the future world, and this conviction was a religious incentive to every Jew. In the middle ages, when the great majority of the European population was illiterate, the Jews were saved from total ignorance by their intense devotion to the study of the Talmud.

ORAL LAW

Traditional Laws.—The Jewish people, like other peoples, in early times had unwritten as well as written laws. The written laws are those in the code called the Torah, the Pentateuch. But in the writings of the prophets as well as in the Hagiographa (*i.e.,* the Old Testament books exclusive of the Pentateuch and the Prophets) there are references to customs and laws that are not found in the Pentateuch. In the book of Jeremiah, for example, it is stated that when the prophet bought a field from Hanamel he wrote a deed in the presence of witnesses who affixed their signatures. There is no mention in the Pentateuch of the requirement of a deed and witnesses in the transfer of real property. In the book of Ruth it is said that Boaz told his kinsman that he would have to marry Ruth when he purchased the field from Naomi "in order to restore the name of the dead to his inheritance" (iv, 5). It is further stated that this was the custom "in former times in Israel concerning redeeming and exchanging: to confirm a transaction, the one drew off his sandal and gave it to the other, and this was the manner of attesting in Israel." This custom of transferring immovable property by the symbolic act of drawing off the shoe is not mentioned in the Pentateuch.

In ancient times, if a man did not pay his debts, his creditor had the right to take him into servitude; if he died and left the debt unpaid, the creditor could enslave his children. This is shown by a passage in II Kings (iv, 1) in which a woman cries to Elisha, "'Your servant my husband is dead; and you know that your servant feared the Lord, but the creditor has come to take my two children to be his slaves.'" This law is not mentioned in the Pentateuch.

It is evident from the book of Genesis that among the early Hebrews a man had to purchase his wife from her father. Jacob paid Laban in manual labour for the right to marry his daughters; David paid Saul 200 foreskins of the Philistines for the right to marry his daughter Michal. Yet the Pentateuch makes no reference to bride price. Nor is any writ in connection with marriage referred to in the Bible; nonetheless a custom of writing marriage documents prevailed among the early Hebrews. Such a document is recorded in the Elephantine papyri. Also, in the book of Tobit, it is said that when Raguel gave his daughter Sarah to Tobias he called her mother, "took a scroll and wrote out the contract; and they set their seals to it" (vii, 14).

The unwritten laws were as old as, and perhaps older than, some of those embodied in the Pentateuch. The sages of the Talmud recognized this when they said that thousands of laws or customs were forgotten during the time of mourning for Moses. Such nonpentateuchal laws came to be known as *halakot, torah shebe-al pe,* "Oral Law," while the laws of the Pentateuch were described as *torah shebikthab,* "written laws."

The unwritten law has been considered holy by all peoples. Sophocles made Antigone justify the burying of Polynices, against the order of King Creon, on the basis of "the immutable unwritten laws of god. They were not born today nor yesterday; they die not; and none knoweth whence they sprang." Aristotle spoke of written and unwritten laws, the latter of which were based upon nature and universally recognized. Philo, in his *Special Laws,* wrote, "For customs are unwritten laws, the decisions approved by men of old, not inscribed on monuments nor on leaves of paper which the moth destroys, but on the souls of those who are partners in the same citizenship." He differentiated between written and unwritten laws, stating, "Praise cannot be duly given to one who obeys the written laws, since he acts under admonition of restraint and the fear of punishment. But he who faithfully observes the unwritten deserves commendation, since the virtue which he displays is purely willed." Aristotle reflected a similar opinion in the *Rhetoric.*

Pharisees and Sadducees.—According to Josephus, who is upheld by the tannaitic literature, the Sadducees denied the authority of the Oral Law while the Pharisees adhered to it. But the Sadducees could not reject all the laws that were not mentioned in the Pentateuch, since no state can function without custom. The cleavage between the Pharisees and the Sadducees was due to the fact that the former maintained that the Oral Laws were as binding as the written laws and that transgressors of the unwritten laws should be punished as were those who transgressed the written laws. The Sadducees denied this, on the ground that the unwritten laws were less binding than the pentateuchal laws. The Pharisees also maintained that the pentateuchal laws must be amended and interpreted so as to conform to the changing culture and economic condition of the people; the Sadducees rejected such innovations. (*See* JEWISH SECTS DURING THE SECOND COMMONWEALTH: *Pharisees; Sadducees.*)

Revisions and New Laws.—In the course of time, however, all laws, written and unwritten, become outmoded. Some device or institution is needed to provide for amendment and interpretation in order to bring them into consonance with contemporary life. In the Jewish second commonwealth period such an institution was developed—the Sanhedrin (*q.v.*).

Before the Hasmonaean period the Jewish government was a theocracy. Religious authority was vested in the high priest, the representative of God. Theocracy was abolished after the establishment of the commonwealth, and nomocracy (the rule of the law) was instituted in its stead. A *bet din* ("court"), commonly known as Sanhedrin, was inaugurated in 141 B.C., with authority to interpret the pentateuchal laws and to enact the unwritten laws into statutory laws. Further, new laws were added by the *sopherim,* "scribes," of the Sanhedrin (*see* SCRIBES). Rabbi Joshua said that certain new laws were introduced by the *sopherim,* "and I cannot gainsay them" (*Tebul Yoṁ* 4, 6). The *sopherim* also modified the pentateuchal laws, holding that their modifications were in the tradition of the fathers. They amended, for example, the pentateuchal law concerning levitical purity. According to the Pentateuch, ritual cleansing could not be completed till after sunset. In later times this worked great hardship on the people, interfering with their daily lives. The sages therefore declared that the law applied only to priests in the matter of eating the sacred food. But the requisite of washing the whole body still worked hardship, so the sages reduced it to a mere washing of the hands before meals. Reference to this law is made in the Synoptic Gospels, where it is described as the "tradition of the elders."

It is stated in the Pentateuch that a Jew may not carry a burden on the sabbath, but Oral Law specified that things may be carried

in the house. According to the Pentateuch cattle must be slaughtered before their flesh is eaten, but the method of slaughtering is not described; the Oral Law defines ritualistic slaughter as cutting the throat.

In some cases Oral Law took precedence over the written law or even nullified it. The Pentateuch states that the blood of a fowl should be covered with sand, but according to the Oral Law it may be covered with anything that happens to be at hand. The sages took cognizance of the fact that the Oral Law expresses the spirit of the law, namely that the blood be covered.

Civil Law.—Civil as well as religious laws in the Pentateuch are so ambiguous and narrow that no civilized society could have been governed without the Oral Law, which supplemented and interpreted the pentateuchal laws, making them adaptable to changing conditions and expressive of the evolving views of society. There are laws in the Pentateuch regarding damages and injury, but there are no specifications of gradations in torts, injuries and liabilities. No law is given regarding damages that are not discernible. For example, a person might mix another person's wheat with "sacred wheat," making it impossible for the second man to use it and thus causing a financial loss; yet the damage was not recognizable in pentateuchal law. According to the Oral Law, however, the person committing this act was obligated to pay for the damages he caused.

The laws of partnership, of the means of acquiring property and of personal agency are not dealt with in the Pentateuch but are based on Oral Law. The pentateuchal laws dealing with inheritance, for example, assumed the inalienability of bonded property from the possession of the tribe. When a person dies his estate is inherited by his sons; where there are no sons it passes to his daughters, and where there are no daughters it goes to the nearest kin. It must always remain in the family of the tribe. The idea of testament is not found in the Pentateuch for the simple reason that the Pentateuch does not recognize that property may be transferred to a stranger. The Oral Law took cognizance of changed sociological conditions and gave a person the right to write a will devising part of his property to anyone, even to a stranger. According to the Oral Law, man is immortal but the title to property is limitable by time. A man may stipulate that his will take effect after his death and the court is obligated to carry out his bequest.

No reference is made in the Pentateuch to writs, whether in connection with property transactions or loans. It is clear from papyri that as early as the 5th century B.C. a debtor gave his creditor a note in which he not only obligated himself to pay the debt but also gave the creditor the right to seize his property if the debt was not paid; if the debtor died before the loan was paid, his children had to pay it. The right of seizure of a debtor's property was based on the unwritten law, and the promissory note nullified the biblical laws that gave the creditor rights over the person of the debtor and his children.

Use of promissory notes also modified the pentateuchal law in connection with the sabbatical year. The Pentateuch clearly specifies that a debt that outlasts the sabbatical year can no longer be collected. To encourage loans to the needy, Hillel introduced the *prosbul* (derived from the Greek *pros boule*, "before the council"), in which the creditor deposited with the court a note stating that a debt owed him by A might be collected at any time the creditor chose. The creditor thus secured his debt against the provision of the pentateuchal law concerning the sabbatical year. A loan for which no note was given was forfeited if not paid before the year of release.

The Pentateuch refers twice to the *lex talionis*, an eye for an eye, a tooth for a tooth (Ex. xxi. 24; Lev. xxiv. 20), but injury and bodily mutilation as well as theft were considered private wrongs. It was not for the state to punish the offender; the matter was to be settled between the person who inflicted the injury and the one injured. The man who suffered the loss of an eye or any other member of his body could absolve entirely the man who caused the injury or could demand any satisfaction, even to taking out the eye of the man who caused the loss of his eye. *Talio* was the extreme satisfaction the plaintiff might demand, but it was not mandatory; the injured person could obtain satisfaction with money if he so wished. On the principle that two different persons could not have exactly the same members of the body, the sages enacted a law by which the injured man could not demand an eye from the person who caused the loss of his eye but that he could demand the value of his eye. In this way, the *lex talionis* was abolished.

Shammaites and Hillelites.—The sages of the commonwealth endeavoured not only to interpret the pentateuchal laws so as to bring them into consonance with contemporary life but also to incorporate some of the old *halakot* into statutory laws—in other words, to make them written laws in order that they might be used as a basis for deducing new laws in accord with the demands of the times. Two schools of thought arose in regard to the soundness and the validity of interpreting the pentateuchal laws and the *halakot*. One, the Shammaites, represented the conservative, the other, the Hillelites, the liberal point of view. They were named after the great sages Shammai and Hillel (*qq.v.*), the fifth and last of the *zugot* ("pairs"), or joint heads of the Sanhedrin. Every sage enrolled himself as a Shammaite or as a Hillelite, and the two schools, which arose with the establishment of the commonwealth, continued until the time of the compilation of the Mishnah.

To establish a legal system by which the pentateuchal laws could be interpreted, and to legalize the modification, Hillel applied three hermeneutic devices. These are (1) *kal wa-homer*, inference *a minori ad maius*, from the less important (law) to the more important (law); (2) *gezerah shawah*, inference by analogy of words; and (3) *hikish*, inference by analogy of subjects. These three principles were not introduced by Hillel, but it was he who organized them into a juristic discipline by which the entire pentateuchal law could be interpreted. Later the three were developed into seven, and Rabbi Ishmael, who lived after the destruction of the Second Temple, increased them to 13. In accepting the hermeneutics as a legal system the sages found it necessary to limit them in certain ways. In regard to penal law, for example, the inference from *minori ad maius* could not be applied; no penalty could be inflicted that was based upon an inference. Again, the principle of inference could not be made from a *halaka;* it was applicable only to the laws of Scriptures and to those *halakot* that had become statutory.

The method of interpretation of the pentateuchal laws was called Midrash Mikra and that of interpretation of the old *halakot* was called Midrash Halakot. The sages who interpreted the pentateuchal laws were called *darshanim*, "interpreters," and the place where they assembled was known as the Bet ha-Midrash.

The idea of intention as a legal principle was the great innovation introduced by Hillel. An act without intention, although willed, was not considered a legal act. For example, if a man gives a ring or money to a woman, this does not make her his wife. The consummation of a marriage requires intention. The man has to give the ring to the woman with the intention of marrying her by open declaration. Hence, according to talmudic law, marriages of minors and lunatics are invalid, since such persons cannot exercise intention. The principle of intention had a far-reaching influence on all Jewish laws, ritual, civil and criminal. Shammai and his school strongly opposed this innovation, but it became an integral part of the legal system of the Talmud.

***Rabbon* and *Tanna*.**—With the liberal views of Hillel prevailing among the sages, there came a change in the leadership of the Sanhedrin. Previously two men had headed this institution, the *nasi* and the *ab bet din*. The office of the latter was now abolished. Gamaliel (*q.v.*), grandson of Hillel, became the sole leader of the Sanhedrin, with the title *rabbon*, "our *nasi*" (*rabba* is the Aramaic form of *nasi*). The sages who interpreted the law were called *sopherim*, "scribes," before the destruction of the Second Temple, and this term occurs in the Gospels; no title was appended to their names. After the destruction of the Temple the sages came to be called *tannaim*, and those who received authorization had the title rabbi.

The *tannaim* were jurists, as is exemplified in the *halakot* with regard to *jus in personam* and *jus in rem*. They held that the law

was made for man and not man for the law. They maintained that society cannot be governed without law, and that the purpose of law is to prevent men from doing what is wrong and to punish violators. They held however that equity should be applied to decisions and leniency should take precedence over strictness in the application of the law. The *tannaim*, who were the spiritual leaders of the people, aimed to promote the general welfare by regulating the relations between man and God and between man and man.

THE MISHNAH

After the destruction of Jerusalem by the Romans in A.D. 70, Jabneh (modern Jamnia, located on the Judaean plain of western Israel) became the centre of Jewish learning. With the collapse in A.D. 135 of the Bar-Cochba revolt, which led to the virtual decimation of the Judaean population, the centre was moved to Usha in Galilee (near modern Haifa). During this momentous period in the history of the Jews the sages decided that the *halakot* should be assembled and codified. There was disagreement as to how the collection should be organized, one school holding that the *halakot* should be collected in the order of the biblical verses from which they were interpreted, the other maintaining that they should be arranged according to the subject matter. The view of the latter school prevailed, and the assembling of the *halakot* on this plan, begun by Rabbi Akiba, was continued by his disciples (especially Meir) and completed in the early part of the 3rd century A.D. by Rabbi Judah ha-Nasi, the Prince, known as rabbi par excellence. This collection of the *halakot* was called Mishnah (Gr. Deuterosis), second to the Torah (the word is variously transliterated also as Mishne and Mishna). Soon after its completion it became the standard work for further interpretations of the *halakot* and was revered as second only to the Bible. The halakic decisions in the Mishnah became binding.

Orders of the Mishnah.—The Mishnah is divided into six "orders" (*sedarim*), each order into tractates (*masechot*) and each tractate into chapters (*perakim*). The *sedarim* are as follows:

1. *Zeraim* ("seeds"), 11 tractates. The first tractate, called *Berachot*, deals with worship and treats of benedictions and prayers and the time of the reading of the *shema*. In other tractates the biblical precepts are set forth relating to the rights of the poor, tithes given to the priests and the Levites, rules and regulations connected with agriculture and the laws connected with the sabbatical year.

2. *Moed* ("festivals"), 12 tractates. In this order the laws in connection with the sabbath and the festivals are set forth. One tractate is devoted to the laws connected with fasts. Another, *Shekalim*, treats of the half-shekel which every Jew had to contribute to the Temple.

3. *Nashim* ("women"), 7 tractates. This order deals with the laws of marriage and divorce and Levirate marriage and divorce. Also given are the regulations governing the relations between husband and wife and between the sexes in general. In this order the tractate *Nedarim* treats of vows and their annulment. Another tractate, *Nazir*, gives the laws concerning Nazirites.

4. *Nezikin* ("damages"), 10 tractates. The first tractate, called *Nezikin*, is divided into three parts: *Baba Kama* ("first gate"), in which the laws of damages and injuries and those regarding theft and robbery are set forth; *Baba Metzia* ("middle gate"), which treats of laws concerning found property, usury and hired labour; and *Baba Bathra* ("last gate"), which deals with *hazakah* (usucapion), the laws regarding the sale and purchase of immovable and movable property, and the laws of partnership and inheritance. *Sanhedrin* ("court") concerns the composition of the court and trial procedure, with the different punishments for various offenses, with the principle of corpus delicti and with the laws of arbitration. *Makkoth* ("stripes") deals with the lashings to which a person is condemned by a court. *Shebuoth* ("oaths") treats of the different kinds of oaths, those made in private life and those administered in court. *Aboda Zara* ("idolatry") treats of the laws concerning idols and the relation to idol worshipers. Two other tractates that bear no relation to the main subjects are included in this order: *Eduioth* ("testimonies"), a collection of traditional laws; and *Aboth* ("fathers"), a collection of ethical maxims of the sages.

5. *Kodashim* ("holy things"), 11 tractates. This order, devoted to the subject of sacrifices, contains the tractates *Tammid* ("daily sacrifice") and *Middoth* ("measurements"), which describe the daily sacrifices of the morning and of the evening, and of the area of the Temple, its courts, gates and the daily services of the priests and the Levites.

6. *Toharoth* ("purifications"), 12 tractates. It sets forth the laws of ritual cleanliness and uncleanliness, both in persons and in things, and the laws in connection with purifications. The tractate *Yadayim* ("hands") is devoted to the laws in connection with the ritual impurity of the hands, their defilement of the Holy Scriptures. In this tractate is related the controversy in regard to the canonicity of Ecclesiastes and the Songs of Solomon.

Character of the Mishnah.—The Mishnah is a code of *halakot* consisting of religious laws and laws governing society. These *halakot* in general are not based on the Pentateuch; biblical texts are not quoted to substantiate the *halakot*. In *Baba Bathra*, which consists of ten chapters dealing with civil law, there is not a single reference to the Bible. References are made to biblical verses only in some chapters of a few tractates, and this practice can be traced to the Midrash form applied by many sages in interpretation of *halakot*.

Two orders—*Kodashim*, which treats of sacrifices, and *Toharoth*, which sets forth the laws of ritual cleanliness and uncleanliness—are concerned with matters that had become obsolete. They were included in the Mishnah because, first, many of the tractates had already been compiled before Rabbi Judah's time, and, second, the Jews of the period had complete confidence that before long the Temple would be rebuilt and sacrifices restored.

The Mishnah, because of the fact that the tractates were compiled at various different times, shows a lack of consistency. The unevenness of the Hebrew language also can be attributed to the fact that it represents many periods of antiquity and of post-Hadrianic times. The sequence of many tractates within each order is of a later era and is mechanical. Rabbi Judah did not set the sequence.

The Mishnah, being a code, does not contain *haggadoth*, legends such as occur in the tannaitic literature and in the Talmud (*see* below, *The Talmuds*). Messianic and eschatological ideas that were prevalent among the Jews of the time of its compilation are largely ignored. It is stated that those who do not believe in the Revelation and those who deny resurrection (spiritual) will not share in the future world. No mention is made of those who deny the coming of the Messiah. No reaction may be seen in the Mishnah toward the Christians in general, or, with one exception, toward the Jewish Christians. In the tractate *Berachot* it is stated that, by law, when a Jew greets another Jew he should pronounce the Name, the Tetragrammaton, although pronouncing the name of God is prohibited in the Pentateuch (*e.g.*, Ex. xx, 7, "You shall not take the name of the Lord your God in vain"). Jewish Christians, on meeting their fellow Christians, greeted them, "Peace in the name of God and the Lord Jesus Christ." The sages, disregarding the injunction of the Pentateuch, introduced a law requiring pronunciation of the Tetragrammaton in order to signify that there is one Lord and that he is God, and that Jesus was not the Messiah; furthermore, this greeting would reveal believers in Jesus.

It may be asked, if the Mishnah is a corpus juris, why does it contain divergent opinions in connection with the laws? The answer is that Rabbi Judah believed that the opinions of the minority as well as of the majority should be recorded, even though the decision of the law had to follow the majority opinion. The minority opinion was included so that if at some future time new laws were enacted the rabbis would be able to find support for their decisions in these minority opinions. When Judah thought that the decision of the law should follow the opinion of one particular sage he recorded his opinion anonymously or in the name of the sages; *i.e.*, he stated it as a majority opinion. Other principles were laid down regarding decisions of the law; for example, when there was a controversy between the schools of Shammai and Hillel, the law was decided in accordance with the latter.

The Mishnah, though a code, cannot be compared to the Pentateuch, which was believed to be revealed by God to Moses and had to be administered by a high priest, a vicar of God. Nor can it be compared to the code of the Roman law—the sages did not derive their authority from secular rulers or magistrates. Their authority was based rather on their profound knowledge of the law and ability to interpret it, and on the authorization received from their spiritual mentors, which, according to tradition, came down unbroken from Moses.

Writing of the Mishnah.—A divergence of opinion exists with regard to the compilation of the Mishnah. Many rabbis of the middle ages, particularly of Spain, followed by many modern scholars, maintained (and maintain) that neither the Mishnah nor the *halakot* were committed to writing until the 6th century A.D. and that the Mishnah was organized only orally by Rabbi Judah. On the other hand, many medieval rabbis, particularly of France and Germany, supported by many other modern scholars, were (and are) of the opinion that the Mishnah was committed to writing by Judah. It can be deduced from internal evidence not only that the Mishnah was written by Judah but that many *halakot* had been put in writing long before his time. The injunction against the writing of *torah shebe-al pe* (the unwritten law) refers only to the customs and the oral laws that had not yet been made statutory.

Great *Tannaim*.—Among the *tannaim* responsible for the shaping of the *halakot*, two disciples of Rabbon Johanan ben Zakkai deserve particular mention: Rabbi Eliezer and Rabbi Joshua. Another sage was Rabbon Gamaliel, the grandfather of Rabbi Judah ha-Nasi, the successor of Johanan ben Zakkai as head of the academy of Jabneh and chiefly responsible for the organization of the *halakot*. Rabbi Akiba by his great intellect and acumen had great influence in the shaping of the *halakot*. He introduced new methods of interpretation of the Bible for the purpose of enacting new laws. He laid the groundwork for the compilation of the Mishnah that was accomplished later by Judah. His dialectic methods are recognized in the halakic discussions of his disciples, particularly those of Simon, Judah, Jose and Meir. The last had the keenest legal mind. To this group must be added Rabbi Simon ben Gamaliel, the father of Judah ha-Nasi, who headed the academy at Usha. These are only a few of the *tannaim* who flourished after the destruction of the Second Temple. The tannaitic period ended with the compilation of the Mishnah.

THE TANNAITIC LITERATURE

Midrash Mikra.—There are other collections of laws, the works of the sages who maintained that the *halakot* should be arranged according to the sequence of the pentateuchal passages. They are: (1) *Sifra* ("the book" par excellence), in which the *halakot* are arranged in the order of the passages in the book of Leviticus; and (2) *Sifre* ("the books"), in which the *halakot* are arranged in the order of the passages in the books of Exodus, Numbers and Deuteronomy. These collections are called Midrash Mikra. (*See* also MIDRASH.)

Tosefta and Baraitot.—Rabbi Judah collected *halakot* that he felt should be recorded in the Mishnah as a way of life for the Jews. He did not include those which at that time were out of date due to social and religious changes that had occurred in the life of the Jews. For example, the laws of impurity, held by the sages to be inapplicable to the pagans, had become obsolete after the destruction of the Second Temple, since pagans were now considered human beings like the Jews and the laws of impurity became applicable to them; likewise, the law that a woman can be among the seven called to the Torah was omitted, since at this time women were already segregated in the synagogues. Rabbi Judah's disciple Rabbi Hiyya made a collection, called Tosefta, of those *halakot* that were not recorded in the Mishnah. The sequence and the arrangement of the orders of the tractates in the Tosefta are similar to those in the Mishnah; three tractates—*Aboth*, *Middoth* and *Tammid*—were omitted since they did not contain *halakot*.

There are notable differences between the Mishnah and the Tosefta. In the Mishnah eschatological ideas are not recorded and there are no references to reward for the righteous and punishment for the wicked after death. The Tosefta, on the other hand, deals with these speculations, and it states that wicked Jews and wicked pagans will be judged in Gehenna for 12 months as a kind of purgatory. However the Jewish apostates and informers would never be released from punishment. Righteous pagans would share a portion in the future world, and thus were placed in the same category as Jews. None of this, not even the word Gehenna, is found in the Mishnah.

During the second commonwealth the Pharisees were inimical toward the apocalyptists, who were occupied with esoteric doctrines and metaphysical ideas and believed in a supernatural Messiah and a physical resurrection. The apocalyptists became influential after the conquest of Jerusalem by the Romans; some accepted Christianity while others infiltrated the ranks of the Jewish people. The *tannaim*, heirs of the Pharisees, continued their opposition, which is revealed in a story related in the Tosefta: Four men—Simon ben Azzai, Simon ben Zoma, Elisha ben Abuya and Rabbi Akiba—went into *pardes* ("paradise") and steeped themselves in mystic and esoteric philosophy. Only Akiba emerged without impairment to his faith or detriment to his intellect. Simon ben Azzai met with an untimely death, Simon ben Zoma became demented, and Elisha ben Abuya turned apostate.

The opposition of the *tannaim* toward apocalyptic doctrines was unsuccessful; indeed some of the *tannaim* themselves resorted to such speculations. Their strictures against superstition did not deter some *amoraim* from adopting these beliefs. Under the stress of personal misfortune and national calamity, people tend to seek solace in superstition and mysticism, and such has been the inclination of many Jews throughout the ages. Medieval Cabala (*q.v.*) was an outgrowth of apocalyptism, and many rabbis of the middle ages regarded Cabala as a menace to true Judaism, as their predecessors, the *tannaim*, had regarded apocalyptism. (*See* also APOCALYPTIC LITERATURE.)

A story in the Tosefta illustrates the endeavours of the followers of Jesus to convert the Jews to belief in him as the true Messiah, and also the disposition of the *tannaim* toward the Jewish Christians. It concerns a nephew of Rabbi Ishmael, Eliezer, who was bitten by a serpent. A certain Jacob sought to heal him in the name of Jesus, son of Panthera. Rabbi Ishmael forbade Eliezer to accept the ministrations of Jacob, and though Eliezer died, yet Rabbi Ishmael exulted, because his nephew had not succumbed to the Christian faith. The story led the 3rd-century Christian writer Origen, in his treatise against the pagan Celsus, to accuse the Jews of spreading false tales regarding the genealogy of Jesus by maintaining that he was the son of Panthera.

A number of additional *halakot* not included in the Tosefta but found in the Talmud are laws of the tannaitic period called Baraitot (sing. Baraita; from a word meaning "outside," thus *halakot* outside the Mishnah). The Tosefta and the Baraitot, since they contain *halakot* that were already obsolete at the time of compilation of the Mishnah, as well as other *halakot* that arose afterward, are more important than the Mishnah itself for the history of the religious and social life of the Jews. The use of Roman legal terms in the Tosefta and Baraitot shows the influence of Roman law on the tannaitic *halakot*.

Both the Tosefta and the Baraitot in some respects differ from the Mishnah. The latter, since it is a code, presents the laws in a legal form. The others are more informal, and in the Tosefta some laws are illustrated by parables. For example, a law which states that the owner of a field must not throw stones from the field onto public property is illustrated in the Tosefta by a story that stresses the moral. Once a proprietor in clearing a field threw the stones onto the public road. A pious man, observing this, told the owner that he was really throwing stones onto his own property. Subsequently the owner, for economic reasons, was compelled to sell his field. Walking on the road where he had thrown the stones he stumbled on them and was injured.

THE TALMUDS

The Mishnah attained canonical authority soon after its completion. The sages and teachers who expounded it were called

amoraim, and the work that resulted from their endeavours was the Gemara, known as the Talmud. The Gemara constitutes both a commentary on and an interpretation of the Mishnah. There are two Gemaras—the Palestinian (*Talmud Yerushalmi*), written in Western Aramaic and replete with many Greek and Latin terms and expressions; and the Babylonian (*Talmud Babli*), written in the Eastern Aramaic dialect and sprinkled with Persian words. The Babylonian Talmud or Gemara is often called the Talmud.

Contents.—The Palestinian Gemara consists of 39 tractates: the entire orders of *Zeraim*, *Moed*, *Nashim* and *Nezikin* (with the exception of *Eduioth* and *Aboth*); and a section of the tractate *Niddah* of the order *Toharoth*. It contains none of the tractates of the order *Kodashim*.

The Babylonian Gemara consists of 37 tractates: one tractate, *Berachot*, of the order *Zeraim;* all of the order *Moed* except the tractate *Shekalim;* all of the order *Nashim;* all of *Nezikin* except *Eduioth* and *Aboth;* all of *Kodashim* except the tractates *Middoth* and *Kinnim;* none of the order *Toharoth* except *Niddah*.

The reason for the omissions is not known; perhaps they were lost. Although the Palestinian Talmud contains more tractates than the Babylonian, the latter is much fuller and larger.

Talmudic Academies.—The disciples of Judah ha-Nasi took up the study of the Mishnah after his death. Tiberias became the centre of Jewish learning, but there were other *yeshibot*, or academies, in Caesarea and Acco. The talmudic academies in Palestine soon fell into decay because of political and economic conditions. The Palestinian Jews at first were under the yoke of the Roman despotism. Later, when Christianity became the dominant religion of Rome, their condition worsened. They were not even allowed to communicate with their coreligionists in Babylonia and were compelled to employ a code to notify them of the intercalations of the year. By the end of the 4th century all of the academies in Palestine had ceased to exist and the compilation of the Palestinian Talmud came to an abrupt end.

Abba Arika, known as Rab, and Mar Samuel, disciples of Judah ha-Nasi, returned to Babylonia, their native land, after the death of their master, and Mar Samuel became the head of the *yeshiba* in Nehardea on the Euphrates. After that city was sacked by the Palmyrians a *yeshiba* was organized at Sura, headed by Rab (A.D. 219). Rab Judah ben Ezekiel, a disciple of Samuel, in 247 organized the great *yeshiba* in Pumbeditha which lasted more than seven centuries. These two *yeshibot* became the leading centres of Jewish learning. The *yeshiba* of Pumbeditha was noted for the keen dialectics employed in the halakic discussions, particularly prominent in which were Judah ben Ezekiel, Rabbah bar Nahmani and his nephew Abaye and his colleague Rabba. Their discourse at times soared into the realm of the incomprehensible.

In the *yeshibot* the studies were designed for those persons, called *haberim*, who devoted their lives to research in the *halakot*. Sessions were also organized, however, to disseminate knowledge among laymen. There was a maxim among the *tannaim*, adhered to by the *amoraim*, that knowledge is superior to observance. These sessions were held in the spring in the month of Adar (March) and in the autumn in Elul (September), and the laity came to the *yeshibot* and participated in talmudic discussions.

Authors.—All the *amoraim* contributed to the Talmuds, but the chief architect of the Palestinian Talmud (compilation of which ceased late in the 4th century) was Johanan bar Napacha, while the Babylonian Talmud was chiefly compiled by Rabbi Ashi (*q.v.*), the head of Sura academy in the early part of the 5th century A.D. It was completed by Rabina *c.* 499.

The *saboraim* ("those who reflect"), who flourished in the 6th and the first part of the 7th century, explained the terminology and interpreted many passages. They also established rules to help in the understanding of the Talmud. Their interpretations and explanations became a part of the work. While some of their interpretations help in the understanding of intricate passages, others lead to confusion. It is difficult to know what passages came from the *amoraim* and what passages were interpolated by the *saboraim*, since no early manuscripts exist. The work of the *saboraim* on the Talmud was later carried on by the *geonim* (*see* GAON).

The Babylonian Talmud became the standard text of Jewish law and religion. It was considered divine. The *halakot* in the Talmud were authoritative; no rabbi could differ from them, he could only interpret. However, the *geonim* conceded that *haggadah* was not authoritative.

Character of the Talmud.—The Talmud consists of *halaka* and *haggadah* (Aramaic *agada*, "legend" or "narrative"), the former being far more numerous. The *amoraim* used the same method in their interpretation of the Mishnah as the *tannaim* had in their interpretation of the Holy Scriptures. They employed exegetical principles in elucidating the *halakot* and expressions in the Mishnah. Where there was tautology of words or expressions, the *amoraim* used them as basis for the deduction of new *halakot*. When there was a clash between two tannaitic principles, the *amoraim*, in their endeavour to harmonize conflicting opinions, introduced a new law.

Halakot.—The *halakot* of the Babylonian Talmud are the production of sages who lived in the Diaspora. This is reflected in the many laws that had to be adjusted to the life of Jews who no longer had their own state and government. Mar Samuel in 219 introduced a law that was far reaching in the life of the Jews throughout the ages—that the civil laws of the country were binding on the Jews. Many *halakot* reveal the influence of the laws of the country in which the Jews lived: the Palestinian Talmud reveals a marked influence of the Roman law, whereas in the Babylonian Talmud there is evidence of the effect of Parthian customs. On the other hand, many *halakot* in both Talmuds are academic, showing no reflection of the prevailing laws nor of the life of the Jews of the time.

Some *halakot*, particularly those of an academic nature, are characterized by dialectics and sometimes by quibbling and hair-splitting. Only a few *amoraim* indulged in this sort of discussion, however; the Palestinian *amoraim* protested against it, and the majority of Babylonian *amoraim* also condemned it.

Haggadoth.—A considerable part of the Talmuds consists of *haggadoth* of varying types. Those whose purpose is biblical exegesis are homiletical and explain the Scriptures allegorically. Others are stories, legends and proverbs whose purpose is to convey a moral message. Some are theological, dealing with rewards and punishments in the future world, the messianic age and the resurrection. There are *haggadoth* dealing with mysticism, angelology, demonology, astrology, dreams and their interpretation. Others are political, describing in legendary form the lives of the Romans and the Persians. There are historical legends about the lives of biblical and postbiblical people and accounts of the sufferings endured by the Jews. There are also aphorisms and maxims, humour and gaiety.

The aphorisms often reveal psychological insight and a humanitarian spirit: "He who gives charity in secret is greater than Moses himself." "The house that does not open to the poor shall open to the physician." "Let the honour of your neighbour be to you like your own." "There are three crowns: of the Torah, the priesthood, the kingship; but the crown of the good name is greater than all." "Whoever runs after greatness, greatness runs away from him; he who runs from greatness, greatness follows him." "No man is to be made responsible for words he utters in his grief." "The soldiers fight and the kings are the heroes." "When the thief has no opportunity for stealing, he considers himself an honest man." "One eats, another says grace." "The thief invokes God while he breaks into the house."

Superstition.—Superstition figures prominently in the Talmuds, particularly the Babylonian. Though some *amoraim* opposed preoccupation with magic and demonology, regarding them as idolatrous, others succumbed. The sages who deplored superstition were forced by public pressure to permit the wearing of amulets on the sabbath in order to avert the evil eye and protect the wearer from evil spirits. Dreams and their interpretation occupied the minds of men of the ancient world; references in the Bible to dreams and to those who had the wisdom to interpret them rightly are familiar. In the Talmud lengthy passages are devoted to dream interpretation, some sages believing that dreams are prophetic, others that they are of no significance. Belief in omens and astrology was widespread even among intellectuals, hence

astrological speculations are found in the Talmud. Most *tannaim* and *amoraim*, however, condemned these beliefs as idolatrous. Other *amoraim*, regarded by the people as holy men capable of performing miracles, indulged in speculations in theosophy and mysticism. But mysticism was not integral to normative Judaism, and halakists opposed it.

Science.—The sages of the Talmud were learned in geometry, astronomy and medicine, sciences that were of great importance in halakic decisions: geometry in connection with the rules and matters of agriculture, astronomy in relation to the intercalation of the months and years, medicine in the matter of the laws of ritual purity and impurity.

Other Religions.—The attitude of the sages toward other religions was antagonistic. As monotheists, who maintained that God is universal—the God of the entire human race—they opposed idol worship. They forbade the partaking of bread and wine with pagans, since they used wine in their libations to the idols and bread in their worship, and to the Jews the table was sacred. Their attitude toward non-Jews was analogous to that of the Hellenes toward non-Hellenes, but it was based not on belief in racial superiority but on religion, and the sages welcomed proselytes—one who accepted Judaism was regarded more highly than a native Jew. No laws were directed against gentile Christians, but certain strictures were passed upon Jewish Christians, who were regarded as heretics and apostates and suspected besides of being informers to the Romans. The rabbis, since they considered the Jewish Christians to be the destroyers of Judaism, showed great animosity toward them. When Christianity became a gentile religion their animosity disappeared, since the rabbis regarded Christianity as a different religion, one that brought the idea of the universality of God to the pagans. Passages about Jesus in the Babylonian Talmud show that the Babylonian *amoraim* had no conception of the historical Jesus and of the time when he lived; all they knew about him was compounded from hearsay. It is worthy of note that there is no mention of Christianity and its founders in the Palestinian Talmud. Either the rabbis were indifferent, or they were afraid to speak candidly because at that time Christianity was the dominating religion of Rome, or, perhaps, passages concerning Jesus and Christianity have been deleted.

CRITICISM AND COMMENTARY

Attacks.—The first attack against the Mishnah was made by the Church Fathers, who construed it as a menace to Christianity. The Fathers interpreted biblical passages so as to support the truth of Christianity, while the *tannaim* employed the same passages to support their own tradition. There is no word in the Mishnah about the founders of Christianity, hence the Church Fathers inferred that it denied the truth of Christianity. As a new, young religion, Christianity had to fight for its very existence, and the Mishnah was regarded as a threat. Unable to read the Talmud in the original, church authorities relied occasionally on tendentious translations made by Jewish apostates. In the 6th century the emperor Justin prohibited the reading of the Mishnah.

Throughout the middle ages the Talmud continued to be attacked by Christians. A public disputation was held in Paris in the 13th century, and later there were disputations elsewhere. In the early part of the 16th century the apostate Johan Pfefferkorn wrote a bitter attack against the Talmud, on the basis of which the emperor Maximilian ordered it to be burned; the Christian humanist Johann Reuchlin, however, defended the work against the accusations of the apostate, and the emperor rescinded his decree. The Talmud was, however, in fact frequently burned, and it is for this reason that so few early manuscripts survive. As late as the mid-18th century copies were burned in Poland, though after the middle ages the work was subject more to censorship than to destruction.

During the middle ages the Jewish Karaites, motivated chiefly by resentment against the exilarch and the *geonim*, attacked the Talmud as a compilation of casuistic arguments, myths and superstitions, denying also the divine origin of its laws (*see* KARAISM). During the period of decline in Jewish intellectual life during the late middle ages and Renaissance the Talmud came to occupy an exaggeratedly exalted position, being studied even to the exclusion of the Bible. The enlightenment that came in the 18th century, particularly as exemplified by Moses Mendelssohn, brought a reaction. Liberal Jews in modern days do not regard the Talmud as divinely inspired but as a portion of the sacred tradition and a monumental source of Jewish history. In modern *yeshibot* it is still the major, sometimes the sole subject of study.

Commentaries.—The best commentary on the Babylonian Talmud was written by Rashi (*q.v.*), but there are numerous other commentaries and novellae compiled by rabbis throughout the ages. Among translations may be mentioned Herbert Danby's of the Mishnah and the Soncino translation of the Talmud. However they are not perfect, and a critical edition of the Talmud is yet to be made. Many passages are corrupt, and the work was never well edited. While its language is chiefly Aramaic, it is a mixture of many dialects. It may be compared to a primeval, dense forest in which a person will lose his way without a skilful guide. (S. Z.)

Source Criticism.—The Talmud is not immune to the manifold problems besetting all ancient texts, particularly those that were orally transmitted (*see* TEXTUAL CRITICISM). Indeed textual criticism of the Talmud is as old as the Talmud itself, though not until modern times did it become a separate scholarly concern.

The rabbis of antiquity often emended rabbinic texts, inserting readings (*e.g., Talmud Babli, Berakot* 13b), substituting one reading for another (*e.g., T.B. Yebamot* 41a) or rejecting outright Baraitot (*e.g., T.B. Gittin* 73a) or statements of *amoraim* (*e.g., T.B. Pesahim* 11a) because they considered them to be hopelessly corrupt. Nor did later scholars of the Talmud desist from correcting one text on the basis of another that seemed to them to be more reliable. Sometimes they even emended a text without support from another text. Talmudic text criticism, unlike biblical criticism, which had to contend with a long and revered tradition, was always acceptable even to the religiously conservative scholar; it is more often than not the modern liberal scholar who protests against excessive emendation (even when sanctioned by tradition) on the ground that it is inconsonant with what is known about the natural development of texts in general.

That is not to say that the traditional attitude set no limits to talmudic textual criticism. A distinction was always drawn between a reading substantiated by internal evidence in the Talmud itself and a reading not so substantiated. The former was always taken to be valid, while the latter was acknowledged to have been open to the possibility of scribal error. With the advent of modern scholarship, this distinction was no longer compelling.

Source criticism seeks to differentiate between the original statements as they were enunciated by their authors and the forms they took as a consequence of being orally transmitted; that is, between the sources and their later traditions. It is not to be confused with the kind of analysis—frequently carried out by the rabbis of the Talmud—which merely traces the historical sources of a given passage without judging whether or not the passage faithfully reflects these sources. Source criticism claims that the transmission of the Talmud was not, and perhaps could not have been, verbatim, and that the text became altered in transmission, with the result that many statements in the Talmud have not come down in their original form. Instead, what survives is the form assumed in the last phase of transmissional development. While such a study is pertinent to most ancient texts, it is particularly relevant to the Talmud, which primarily consists of quotations and their interpretations.

Traces of source criticism can be found in geonic literature, but credit for pioneering work belongs to the Galician-born scholar H. M. Pineles (1805–70), who in his *Darka Shel Torah* (1863) for the first time offered a systematic critical approach to the study of talmudic sources. J. N. Epstein (1879–1950), professor of Talmud at the Hebrew university, Jerusalem, made a major contribution to source criticism through his comprehensive *Introduction to the Mishna* (1948) and *Introduction to Tannaitic Literature* (1957).

See also references under "Talmud" in the Index. (D. WE.)

BIBLIOGRAPHY.—M. Mielziner, *Little Maxims and Fundamental Laws*

of the Civil and Criminal Code of the Talmud (1898), *Introduction to the Talmud* (1915); Hermann L. Strack, *Introduction to the Talmud and Midrash* (1931); S. Krauss, *Talmudische Archäologie* (1910–12); W. Bacher, *Tradition und Tradenten in den Schulen Palästinas und Babyloniens* (1914), *Die Agada der Tannaiten* (1903); Max L. Margolis, *A Manual of the Aramaic Language of the Babylonian Talmud* (1910); J. Z. Lauterbach, *Midrash and Mishnah* (1916); S. Zeitlin, "The Halaka: Introduction to Tannaitic Jurisprudence," *The Jewish Quarterly Review* (1948), "Jesus in the Early Tannaitic Literature," *Abhandlung zur Erinnerung an Hirsch Perez Chajas* (1933), *The Rise and Fall of the Judaean State*, vol. i (1962). (S. Z.)

TAM, JACOB BEN MEÏR (1100–1171), a grandson of Rashi (*q.v.*), was the most famous French glossator (*tosafist*) on the text of the Talmud. In 1147 he was attacked and injured by a disorderly band who had attached themselves to the Crusaders. He escaped to the neighbouring Troyes, where about 1160 was held the first of the Jewish synods, for which the Rhinelands became celebrated. At this meeting it was laid down that disputes between Jew and Jew were not to be carried to a Christian court, but were to be settled by fraternal arbitration. New conditions of life had arisen because of the closer terms on which Jews and Christians lived, and Jacob Tam was foremost in settling the terms which were to govern the relations, from the Jewish side. His practical ordinances (*Takkanoth*), connected with marriage and divorce, trade and proselytism, as well as with synagogue ritual, bear the stamp of enlightened independence within the limits of recognized authoritative tradition and law. Of his legal work the most important was collected in his *Sefer ha-yashar*. He was also a poet and grammarian. (I. A.)

TAMALE, the capital town of the Northern Region of Ghana, is situated 437 mi. N.N.E. from Takoradi, Ghana's chief seaport on the Gulf of Guinea, and 406 mi. north from Accra, the capital of Ghana. Tamale lies 600 ft. above sea level in a plain between the White Volta to the west and the tributaries Taha Kulibong and Pagazaa to the east. Its nearest point to the White Volta is at Dalong, 22 mi. away. The climate is hot and wet. The town covers approximately 4.5 sq.mi. Pop. (1960) 40,443.

The official residence of the chief regional officer is in Tamale, and many government departments have their offices there. In addition to the regional education office, government educational establishments include a teachers' training college and a centre for instructing artisans in various trades. The vernacular literature bureau provides newspapers and literature for mass literacy campaigns in the region; personnel for these campaigns and for work in community development are trained at the rural training centre. There is a social and cultural institute in the town, and a public park and polo ground.

Tamale is an agricultural trading centre where the main occupation is farming and crop cultivation in particular. Yams, maize, guinea corn, millet, rice, peanuts and cow peas are grown, and a small quantity of cotton is produced. There is some cotton milling and the shea nut is processed. Commodities exported, but almost exclusively within Ghana, are rice, guinea corn, peanuts, cow peas and shea butter. The principal imports are cotton prints, farm tools, hardware, salt and kola nuts.

The main road, which runs northward from the coast roughly through the centre of the country, passes through Tamale; other roads enter it from the east and west. There is an air service between Tamale and Accra. (J. S. Km.)

TAMARACK: *see* Larch.

TAMARIND, the name popularly applied to a tropical leguminous tree, *Tamarindus indica*, and its pod fruit. The tree is widely distributed in tropical countries, but it is generally considered to be native to eastern tropical Africa, from Abyssinia southward to the Zambezi river. The name (meaning in Arabic "Indian date") shows that it entered medieval commerce from India, where it is used not only for its pulp but also for its seeds, which are astringent; its leaves, which furnish a yellow or a red dye; and its timber. The tree attains a height of 70 to 80 ft. and bears elegant pinnate foliage and purplish or orange veined flowers arranged in terminal racemes.

The pods are filled with an acid juicy pulp containing sugar and various acids, such as citric and tartaric, in combination with potash. The acid pulp is used as a laxative, the pods being largely exported from the West Indies and Indonesia.

TAMARISK, common name for the genus *Tamarix*, a small group of shrubs or low trees constituting the tamarisk family (Tamaricaceae). The species of tamarisk and of an allied genus *Myricaria* (false tamarisk) grow in salt deserts, by the seashore or in other more or less sterile localities in warm-temperate, subtropical and tropical regions of the eastern hemisphere. Their long slender branches bear very numerous small appressed leaves, in which the transpiring surface is reduced to a minimum. The numerous small pinkish flowers, borne in long clusters at the ends of the branches or from the trunk, give the shrubs a feathery appearance. Each has four to five free sepals, and as many petals springing with the four to ten stamens from a fleshy disk. In *Tamarix* the stamens are free, while in *Myricaria* they are united into one parcel. The fruit is capsular and contains numerous seeds, each usually with a long tuft of hairs at one end. The value of these shrubs or trees lies in their ability to withstand the effects of drought and a saline soil; they grow where little else can flourish. On this account the common tamarisk, *T. gallica*, the salt cedar or French tamarisk, is planted on seacoasts and affords shelter where none other could be provided. *T. aphylla*, the Athel tamarisk, native to western Asia, has been introduced as a windbreak in desert areas in southern California. Several species are cultivated as ornamentals.

For tamarisk manna, *see* Manna.

TAMATAVE (known locally as Tòamàsina), chief seaport of the Malagasy Republic (Madagascar), is on the eastern coast. Pop. (1956) 46,272. A coral reef forms a harbour, entered by two openings but exposed to frequent cyclonic storms and to a tidal bore. The town is built on a sandy peninsula.

Imports to Tamatave consist principally of piece goods, farinaceous foods, and iron and steel goods, and exports of gold dust, raffia, hides, caoutchouc and live animals. Of the total foreign trade of Madagascar, 46% is through Tamatave. The town is connected with the interior of the country by a network of roads and railways. Since 1895 the native population has settled in a new village to the northwest.

TAMAULIPAS, a northern Gulf-coast state of Mexico. Area, 30,822 sq.mi. Pop. (1960) 1,024,182. The central and southern parts of the state are mountainous but in the north are fertile plains. The coastal zone, sandy and broken by lagoons, is sparsely inhabited. Inland, the foothills region is wooded and fertile and low alluvial valleys penetrate into the sierras. There are four important rivers in the state. The climate, hot and humid on the coast, is pleasant in the interior. Rainfall is abundant, especially on the mountain slopes of the south. The principal industry is agriculture. Sugar, cereals, tobacco, cotton, fruit and coffee are grown. Stock raising is carried on and istle fibre is prepared for export. Copper is mined and extensive deposits of petroleum and asphalt are being exploited.

Ciudad Victoria (pop. [1960] 50,797), the capital, is a small sierra town about 120 mi. from Tampico. Other towns include Tampico, Nuevo Laredo, Matamoros, Ciudad Madero and Reinosa.

TAMAYO Y BAUS, MANUEL (1829–1898), Spanish dramatist, came of a family connected with the theatre, his mother being the eminent actress Joaquina Baus. The earliest of his printed pieces, *Juana de Arco* (1847), is an arrangement from Schiller, as also is *Angela* (1852), while *Virginia* (1853) is a dramatic essay in Alfieri's manner. *La Locura de Amor* (1855) established Tamayo's reputation as Spain's leading playwright. *Hija y Madre* (1855) is a failure, and *La Bola de Nieve* (1856) is notable solely for its excellent workmanship. Tamayo's straitened means during the next few years forced him to put original work aside and to adapt pieces from Léon Laya, Jules Sandeau and Émile Augier. In this period he produced only one original piece, *Lances de Honor* (1863), which turned upon the immorality of dueling, and led to a warm discussion among the public.

TAMBOV, an *oblast* of the Russian Soviet Federated Socialist Republic, U.S.S.R., was formed in 1937. Area, 13,243 sq.mi. Pop. (1959) 1,549,001. It occupies the low, level Oka-Don plain on the watershed between the two river basins. Only in the east does the land rise slightly toward the Volga uplands. Drainage is north to the Oka by the Tsna and its tributaries and south to the Don by the Vorona and several smaller streams. The climate is continental, with a January average temperature of 11° C. (13° F.) and a July average of 20° C. (68° F.). The annual rainfall is 19 in., with a summer maximum. The natural vegetation is forest steppe, much of which has been cleared for agriculture since the soils are rich chernozems. Only along the right bank of the Tsna and along the Vorona are large areas of pine forest on sandy soils.

In 1959 only a small proportion, 408,210 or 26% of the population, was urban. These people live in seven towns and five urban districts, of which the largest are Tambov (172,171) and Michurinsk (80,653). Most of the industry is concentrated in these two towns, the other centres being engaged in processing local farm produce.

Agriculturally the *oblast* is important, with grains occupying first place. Winter rye, spring wheat, maize (corn), oats and millet are grown. Sunflowers are the chief industrial crop, but there is some sugar beet, hemp and tobacco. Potatoes, melons and other vegetables are widespread. Beef and dairy cattle and poultry are the main livestock.

Other principal towns are Morshansk, founded in the 17th century on the Tsna (manufactures of glass, cloth and makhorka tobacco); and Kirsanov (founded 18th century) on the Tambov-Saratov railway (smelting works). (R. A. F.)

TAMBOV, a town and *oblast* administrative centre of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the left bank of the upper Tsna, on the Oka-Don plain, 260 mi. S.E. of Moscow. Pop. (1959) 172,171. It was founded in 1636 as a fortress on the Belgorod defensive line. In 1779 it became centre of a *guberniya* (pre-Revolutionary province). Development came slowly, chiefly in the late 19th century after the construction of the Moscow-Saratov railway through the town. Tambov's chief industry is engineering, chiefly the making of parts for vehicles and farm machinery, and the manufacture of electric motors and turbines, and machinery for the chemical and food industries. The last two are represented by the manufacture of synthetic rubber, dyes and paints and the processing of local agricultural produce. There are also rolling stock repair yards. Apart from the main line, there is a branch railway to Balashov and to Kamyshin on the Volga, and the Tsna is navigable downstream as far as Morshansk for pleasure launches and small barges.

In 1940 the southeastern part and adjacent suburb were constituted as the separate town of Kotovsk. Cultural institutions at Tambov include a pedagogic institute and a school of music. (R. A. F.)

TAMERLANE: *see* TIMUR.

TAMIL, peoples from south India who have perhaps the oldest cultivated Dravidian language and literature. Numbering about 28,000,000 in the 1960s, Tamil speakers (*see* DRAVIDIAN LANGUAGES) inhabited chiefly Madras state and parts of Kerala, Mysore, Andhra Pradesh and northern Ceylon (*q.v.*). Emigrant Tamil may be found in Madagascar, Malaya, Burma, Indochina, Thailand, south and east Africa, the Fiji and Mauritius islands, and the West Indies. The term Tamil was in very ancient times applied to the language and peoples of a wider area of south India, including those now speaking Malayalam. The term is used also to designate the culture and literature of the people; it is a linguistic and geographical word and includes persons of varied social status and customs.

The Tamil area in India is a centre of traditional Hinduism. The largest occupational group are cultivators, and although most are Hindu, there are Christian and Muslim Tamil. They are predominantly patrilineal with patrilocal residence; in the kinship system of the lower castes, however, matrilateral and affinal bonds are also found. Traditional marriage alliances are maintained between clanlike exogamous patrilineal subdivisions of endogamous castes and tribes. Among non-Brahmans divorce is common, as is remarriage and widow remarriage; Brahmans and some non-Brahmans disapprove of these and place greater emphasis on the lineage and on respect for elders in the extended family.

The Tamil have a long history of achievement; sea travel, city life and commerce seem to have developed early. Tamil trade with Greeks and Romans is verified by literary, linguistic and archaeological evidence. Their rich literature reaches to the early Christian era, and Tamil schools of personal religious devotion (bhakti) have been important in Hinduism (*q.v.*). South Indian classical music and the celebrated classical dance Bharata Natyam (*q.v.*) owe much to the Tamil who are masters and teachers of these arts.

See also INDIA; CEYLON; DRAVIDIAN.

BIBLIOGRAPHY.—D. T. Srinivas Iyengar, *History of the Tamils* (1929); K. A. A. Nilakanta Sastri, *A History of South India,* 2nd ed. (1958); E. K. Gough, "Brahman Kinship in a Tamil Village," *American Anthropologist,* vol. 58 (1956); L. Dumont, *Une Sous-caste de l'Inde du sud* (1957); E. R. Leach (ed.), *Aspects of Caste in South India* (1960); M. N. Srinivas, *Caste in Modern India* (1962).
(M. B. S.; R. RA.)

TAMIL LANGUAGE, one of the principal Dravidian languages (*q.v.*) spoken in South India and perhaps the only example of an ancient classical tongue which has survived as a spoken language for more than 2,500 years with its basic structure almost unchanged. The official language of Madras (*q.v.*), it is one of the 14 major languages recognized by the constitution of India (1949). At the 1951 census, Tamil was the mother tongue of 26,546,764 persons, 7.4% of the total population of India. Moreover, it is also spoken in the north of Ceylon and by many settlers in South and East Africa, Mauritius, Malaya and Fiji.

The Tamil script can be traced to Grantha and Vaṭṭēluttu epigraphical sources of the 3rd century A.D. onward (*see* TAMIL LITERATURE). That it had reached something akin to its present form by the 11th century may be seen from the inscription of Rājarāja Cola in the Great Temple in Thanjavur. Tamil script differs from others of South India mainly in excluding separate characters for the voiced velar, palatal, retroflex, dental and labial plosives, the voiced and voiceless aspirated plosives of these series, and the sibilants. However, the character of the palatal plosive *c* is realized as a dental or palatal sibilant in certain positions and is used to represent sibilants in loanwords. Also, Grantha characters are borrowed to represent the sibilants *ṣ, sh,* the voiced plosive *j* and the voiced spirant *h*. Other plosives are represented by the corresponding voiceless unaspirated plosive.

Thus, the character for *k* represents also *kh, g* and *gh* in, for instance, Sankskrit loanwords. Moreover, the plosives are realized as voiced in certain positions according to fairly consistent rules, the principal being that a medial single plosive is realized as voiced, and a medial doubled plosive as voiceless with tense utterance. Tamil includes three consonants not found in the syllabaries of the Indo-Iranian languages of India and not now fully represented in other Dravidian languages: the trilled *r,* the alveolar nasal *n* and the retroflex fricative *l*.

Like other Dravidian languages, Tamil is regarded as agglutinative in formation. While its close affinity with other south Dravidian languages, especially Malayāḷam and Kanarese, is undisputed, the affiliations of the whole family have not been satisfactorily explained. The classical view, as stated by R. A. Caldwell, is that they are connected with the Scythian languages, but many of the similarities of structure and forms he cites may be accidental.

Tamil exhibits inflexional features also, wherein postpositional words are reduced to suffixes, sometimes monosyllabic. It agrees generally with other languages in broadly classifying words into verb roots, nouns and adjectives. In common with other languages of the same family, there is evidence, although vestigial, of a separate negative verbal conjugation and, in the sense of the Spanish *ser,* the verb for "to be" is unexpressed except in the negative. Case terminations are added to the plural noun-suffix and are unaffected by it, being identical with those used for the singular.

BIBLIOGRAPHY.—T. Burrow, *Dravidian Studies,* I–VI, in *Bulletin of the School of Oriental and African Studies* (1938–48), two papers in *Transactions of the Philological Society* (1945, 1946), and, with M. B. Emeneau, *A Dravidian Etymological Dictionary* (1961); R. A. Caldwell, *A Comparative Grammar of the Dravidian Languages,* 3rd ed. (1913; reprinted 1956); T. P. Meenakshisundaram, "Tamil," in *Kalai Kkalanciyam,* vol. v (1958); K. V. Subrahmanya Ayyar, "The Earliest Monuments of the Pandya Country," in *Proceedings of the 3rd All-India Oriental Congress* (1924); A. Subbiah, "Voiced and Voiceless Stops in Tamil," in *Tamil Culture,* vol. v, no. 2 (April 1956) and "Tamil Language and Literature," in *The Languages of India* (1958).
(Jo. R. M.)

TAMIL LITERATURE. Tamil is perhaps the best-known of the Dravidian languages (*q.v.*) spoken in South India. Apart from Sanskrit, its literature is the oldest in India and, if epigraph-

ical sources are included, may be said to commence in the 3rd century B.C. with some fragmentary cave graffiti found near Madurai. (*See* also TAMIL; TAMIL LANGUAGE.)

Period of the "Sangam".—The earliest Tamil writings that may properly be termed literature are those of the *sangam,* or "literary academy." Tradition, much of it doubtless fanciful, tells of a *sangam* at Madurai which grew up during the middle ages, where a number of anthologies of secular court poetry were compiled—the Eight Anthologies, *Eṭṭuttōkai,* and the Ten Songs, *Pattuppāṭṭu.* Possibly unique in early Indian literature, which is almost entirely religious, these poems are concerned with two main topics, love and the praise of kings and their deeds. Many, especially on the latter subject, display great freshness and vigour and are singularly free from the literary conceits of much of the other early and medieval literatures of India (*see* INDIAN LITERATURE).

Since they are secular, these poems are also free from the complex mythical allusions which are such an outstanding feature of most Indian art forms. *Sangam* poetry seems to be at least partly based on personal experience; in this respect it is curiously modern and unlike anything else to be found in India before the late 19th century. From references to trade with the west and other internal evidence, it seems likely that these poems belong to the first four centuries A.D., when trade with Rome via the middle east was still active.

Paripāṭal, possibly the last of these anthologies, marks the transition to the next, more formal, stereotyped style, for it includes both religious poetry, abounding in the usual allusions, and, still in the older vein, some fine descriptive poetry, which depicts the course of the Vaigai river, on which Madurai stands.

Sixth to Ninth Centuries.—To the 6th century and after belongs that great collection of Śaiva religious poetry (*see* HINDUISM: *Gods*) called, collectively, *Tirumuṟai.* Among its 12 books are the *Tevāram,* consisting of the hymns of Appar (Tirunāvukkaracu), Tiruñāṉacampantar and Cuntarar; and Māṇikkavācakar's mystical poem *Tiruvācakam* (Eng. trans. by G. U. Pope). The 12th book, the 12th-century *Pēriyapurāṇam* of Cekkiḻār, describes the lives of the 63 Śaiva *nāyaṉmār* ("saints").

Of possibly the same period as the first few books of *Tirumuṟai* are the "Eighteen Minor Works." Largely didactic, these include the well-known collections of aphorisms, *Tirukkuṟaḷ,* by Tiruvaḷḷuvar. This work is highly esteemed by the Tamilians, and there are several translations. Another similar work is the *Nālaṭiyār.*

Similar in subject matter to the *Tevāram,* though somewhat later in date, are the hymns of the *Nālāyirappiraṉantam,* which differ only in that they praise the god Viṣṇu (Vishnu). They are the work of the 12 Vaiṣṇava devotees or Āḻvārs, the most interesting of whom is, perhaps, Āṇṭāḷ, the girl-foundling reared by a Vaiṣṇava priest.

The epics *Cilappatikāram* and *Maṇimēkalai,* written by Iḷaṅkovaṭikaḷ and Cāttaṉār respectively, may probably be assigned to the 6th or 7th century. The former tells of a merchant Kovalaṉ who neglects his wife Kaṇṇaki, lavishing his wealth on a courtesan called Mātavi. Later, impoverished as a result of his excesses, Kovalaṉ and his faithful wife travel to Madurai, where they try to start a new life, selling Kaṇṇaki's golden anklets to raise capital. Wrongfully accused of stealing these from the queen, Kovalaṉ is executed. Kaṇṇaki fully exonerates her husband before the king, who dies of remorse, and she herself, in her frenzied grief, tears off her left breast and curses the city, which is consumed by fire. She then makes her way to another royal city, Vañci, where she dies, and is received into heaven to join Kovalaṉ.

Though a Buddhist work, *Maṇimēkalai* is a sequel to the *Cilappatikāram* in that its heroine, Maṇimēkalai, is the daughter of Mātavi by Kovalaṉ. Mother and daughter have both become Buddhist nuns; the prince, Utayakumāraṉ, falls in love with Maṇimēkalai, and the pair finally realize that they were married in a former birth.

Ninth to Twelfth Centuries.—Three famous Tamil poets, Kampaṉ, Oṭṭakkūttaṉ and Pukaḻenti, belong to this period, but it is not possible to date them with any exactness. Kampaṉ was author of the principal Tamil version of the *Rāmāyaṇa.* In his version, given a South Indian setting, Rāma is regarded not only as hero of the story but as a god and incarnation of Viṣṇu—as is usually the case with vernacular *Rāmāyaṇas* in India. The work, begun by Kampaṉ, was completed by Oṭṭakkūttan.

Cayaṅkoṇṭāṉ's *Kaliṅkattupparaṇi,* an 11th-century work, describes the victorious war waged by the Coḻa king against the neighbouring ruler of Kaliṅga. It opens with a grandiose genealogy of the Tamil king, and closes with a gruesome description of the carnage of battle and of the corpses being devoured by shedevils.

A number of commentators and grammarians belong to this period. Among the former may be mentioned Aṭiyārkkunallār, who wrote the commentary on *Cilappatikāram,* and Naccinārkkiṉiyar, who, in addition to commentaries on *Pattuppaṭṭu* and some of the *Eṭṭuttōkai,* wrote one on part of the early "grammar" *Tōlkāppiyam.*

Twelfth to Sixteenth Centuries.—Tamil literature became increasingly concerned with philosophical treatises and with creating *Purāṇas* (anthologies of religious legends), which, in this case, describe the myths about the various shrines of South India. Most important among the former are perhaps those (for example Mēykaṇṭatevar's *Civañāṇapōtam*) which expound Śaiva-siddhānta philosophy.

Among the *Purāṇas* mention may be made of the *Tiruviḷaiyāṭaṟpurāṇam* of Parañcotimuṉivar, which recounts the 64 divine sports of the god Śiva (Shiva) in connection with the founding of the great temple at Madurai.

European Influence and the Modern Period.—After the collapse of the Vijayanagar empire in 1565, South India was split into a number of petty states, but some of these, such as Thanjavur (Tanjore), continued to patronize writers. Missionaries from the west were also active in the Tamil area. The earliest of these, apart possibly from St. Francis Xavier (*q.v.*), were Roberto de Nobili in the 17th century and another Italian Jesuit, Joseph Beschi, born near Mantua in 1680. Both lived as Hindu ascetics, and Beschi studied not only grammar but composition. He produced an epic, the *Tempāvaṇi,* which tells the story of St. Joseph and is based on an earlier model, *Cīvakacintāmaṇi.* In 1738 he produced a grammar of Tamil in Latin, probably the first to be written in a European language.

British Tamil studies were initiated by the manuscript collections of Col. Colin Mackenzie, who arrived in India in 1783. Later scholars of Tamil from the west included Robert Caldwell, whose *Comparative Grammar of the Dravidian . . . Languages* appeared in 1856; and the Rev. G. U. Pope who, besides writing a number of grammars, translated several Tamil works, including the *Tirukkuṟaḷ.*

In the 19th century, Tamil increasingly availed itself of western literary forms and ideas, and prose, hitherto reserved for inscriptions and commentaries, became accepted as a literary medium. The first novel was *Piratāpa Mutaliyār Carittiram,* by Vedanāyakam Piḷḷai. The most famous poet of modern times is C. Subrahmanya Bharati (d. 1922). Perhaps his most lasting work is *Kaṇṇan Pāṭṭu,* a touching and lyrical cycle on the god Kṛṣṇa (*see* VISHNUISM). His work combines the best of the techniques of east and west. More recent is *Pāratidācan* whose poems about children and the family are especially attractive.

Much modern Tamil writing is devoted to regional politics and as such can hardly be said to be of permanent value; however, such writers as Aṇṇāṯturai and Karuṇāniti have proved their worth in other work; for example, in the writing of filmscripts, and their contribution to Tamil literature may become more lasting. The economic lot of South Indian writers in the mid-20th century was a hard one, and a number of excellent literary journals, such as the humorous weekly *Āṉanta Vikaṭaṉ* and the magazine *Kalki* afforded them a most necessary medium and helped to foster good writing.

BIBLIOGRAPHY.—Translations into English, other than those mentioned in the text, include *The Lay of the Anklet,* trans. by V. R. Ramachandra Dikshitar (1939); *The Story of King Nala and Princess Damayanti,* trans. by M. Langton (1950); *The Ayodhya Canto of the Ramayana,* from Kampan's version, trans. by C. Rajagopalachari for the United Nations Educational, Scientific and Cultural Organization

(1961). V. Cronin, *A Pearl to India: the Life of Roberto de Nobili* (1959); W. C. Mackenzie, *Colonel Colin Mackenzie* (1953); V. R. Ramachandra Dikshitar, *Studies in Tamil Literature and History* (1930). (Jo. R. M.)

TAMM, IGOR EVGENYEVICH (1895–), Russian physicist, who was awarded the 1958 Nobel prize for physics jointly with P. A. Cerenkov and I. M. Frank "for the discovery and interpretation of the Cerenkov effect," was born in 1895. This effect, which was first described by Cerenkov in 1934, is concerned with the emission of light waves by electrons moving at very high speeds and is of great importance for the study of nuclear physics.

Tamm, a professor at Moscow university, and his co-worker I. M. Frank provided a theoretical explanation of the Cerenkov effect in 1937. He became one of the best-known Russian physicists and a member of the Academy of Sciences of the U.S.S.R.

In 1946 Tamm shared the Stalin prize with Cerenkov and Frank for their joint work on electron radiation. He worked on the quantum theory of diffused light in solid bodies, on the photo-effect on metals and on methods for the control of thermonuclear reactions. His method of interpreting the interaction of elementary nuclear particles is known as the Tamm method. He is the author of treatises "On the Magnetic Moment of the Neutron" (1934) and "Relativistic Interaction of Elementary Particles" (1935). (W. J. Bp.)

TAMMANN, GUSTAV HEINRICH JOHANN APOLLON (1861–1938), German physical chemist whose researches into the states and reactions of matter, especially of metals, made him one of the founders of metallurgy, was born at Yamburg (Kingisepp), near St. Petersburg (Leningrad), on May 28, 1861. He studied at the University of Dorpat (Tartu) in Estonia, and in 1892 became professor and director of the Chemical institute there. In 1903 he went to the University of Göttingen in Germany as a professor of inorganic chemistry and was head of the Institute of Physical Chemistry there from 1907 to 1930. He died at Göttingen on Dec. 17, 1938. Tammann made fundamental investigations of the changes in the condition of matter and their dependence on temperature and pressure and also studied the reactions of matter in a hard and brittle condition. Especially valuable were his introduction of thermal analysis in the construction of constitutional diagrams of metals, and his researches into the nature and structure of alloys.

Among Tammann's publications were *Lehrbuch der Metallographie* (1914; 4th ed., 1932); *Die chemischen und galvanischen Eigenschaften von Mischkristallreihen und ihre Atomverteilung* (1919). He was an editor of *Zeitschrift für anorganische und allgemeine Chemie.* (O. W. K.)

TAMMANY HALL. The name popularly used in connection with the executive committee of the Democratic party of New York county; sometimes applied to the entire party structure in the county or in the city. From about 1800, with only occasional intermissions, as a political organization it played an important role in the city and exercised considerable influence in state administration and national affairs.

Origin.—Before the American Revolution groups proclaiming fealty to King George III were organized in societies bearing the names of St. George, St. Andrew and St. David. As a countermove, the revolutionists formed associations called the Sons of Liberty or the Sons of St. Tammany, the latter after Tammanend, a Delaware Indian chief, famed for wisdom, benevolence and love of freedom, with the "saint" thrown in as a form of ridicule. After the achievement of American independence, in 1776, these associations dissolved.

During the 1780s the leaders of the aristocratic and propertied elements of New York city and state successfully managed to limit the suffrage to freeholders and to strengthen the Order of the Cincinnati, a group of former officers of the Revolutionary army with centralist and monarchial tendencies. To resist these influences, William Mooney, an upholsterer in New York city, founded the Society of St. Tammany or Columbian Order, on May 12, 1789, a fortnight after the establishment of the national government. His purpose was the creation of a national society native in character and democratic in principle and action. Its officers were given Indian titles; the society's chief was grand sachem and his fellow chiefs, sachems.

For more than three decades after its organization, Tammany represented the middle class opposition to the pretentions and power of the aristocratic party. Its "democracy" did not incorporate the aspirations of the lower economic groups. In that period it also lost its national and nonpolitical character and became intimately identified with politics in New York city.

Under the control of Aaron Burr until his political downfall after killing Alexander Hamilton in a duel in 1804, the society had an influential share in bringing about victories of the Republican (later called the Democratic) party and was richly rewarded, particularly by Thomas Jefferson after his election as president in 1800.

Organization.—Criticisms made by the opposition that a private society was engaging in politics resulted in a separation of the hall's social and political functions. In 1805 the Society of St. Tammany obtained from the legislature a charter of incorporation as a benevolent and charitable body to give relief to members and others. Politically, the Democratic party was organized as an apparently distinct body; in reality, however, the society's sachems controlled the political mechanism and by preventing hostile factions from meeting in the society building, Tammany hall, claimed the right of party regularity.

The political organization initiated at that time consisted of general, nominating, corresponding and ward committees. In the general committee was vested the power of convening the party's meetings and of making all necessary arrangements for elections. Composed originally of 30 members, three from each of the city's ten wards, this committee was gradually expanded until it had many thousands of members penetrating every section of the city. The real power consequently passed into the hands of the ward leaders, later organized as the executive committee of the party.

Boss Rule.—As early as 1806–07, revelations of widespread corruption of Tammany city officials resulted in the removal of the controller, the superintendent of the almshouse, the inspector of bread and others. Despite such proved charges, many of the removed officials, including the society's founder, remained powerful Tammany sachems.

The bitterest opponents of Tammany were the Irish immigrants who were ineligible to be members of the "native-born patriots." As a protest against Tammany bigotry, hundreds of Irish migrants broke into a general committee meeting on the evening of April 24, 1817. Within a few years, the propertied leaders of Tammany were forced for their own preservation to take in the immigrants, naturalize and vote them, and join in the fight for manhood suffrage.

In the 1830s the pressures exerted by the Workingmen's party and its successor, the Equal Rights party, forced the general committee to oust the banking and merchant leaders. The election of its grand sachem, Martin Van Buren, as president of the U.S. in 1836 redounded to Tammany's prestige. Within a few years, however, the immigrant groups, organized into gangs, came under the control of the astute, unscrupulous and engaging Fernando Wood, several times mayor of New York, who used them to break with and later control Tammany. Grand Sachem William M. Tweed initiated complete boss domination of the hall in 1868. Corruption reached a climax under Tweed, when New York city was plundered of more than $200,000,000. Tweed died in jail but most of his confederates retained their wealth. Throughout the world, Tammany became synonymous with corruption.

John Kelly, who succeeded Tweed, induced the leading reformers like Samuel J. Tilden, August Belmont and Horatio Seymour to serve as sachems. The real resuscitating factor, however, was the attachment of the tenement house masses to the district leaders. The helping hand outweighed all of the denunciations. Richard Croker in 1886 and his successor in 1902, Charles F. Murphy, carried on the façade of making liberal avowals and supporting progressive candidates for the top of the ticket but failed to curb corruption within the administrative machinery.

Reform.—After Murphy's death in 1924, Judge George W. Olvany became county leader and with the assistance of Gov. Alfred

E. Smith sought to alter the character of the hall. On March 16, 1929, Judge Olvany resigned and was succeeded by a leader of the old school, John F. Curry. Revelations of corruption in Mayor James Walker's administration, as shown in the Seabury report, discredited Curry but he remained in power until successive defeats of Tammany candidates led to his replacement by James J. Dooling in July 1934.

During the 12 years of Fiorello LaGuardia's reform administration, Tammany declined in power and was eclipsed by the more powerful Brooklyn and Bronx Democratic leaders. Many Democratic party candidates repudiated Tammany. The organization found itself in dire financial condition and had to give up its headquarters in the building maintained by the Tammany society. A succession of county leaders—Timothy J. Sullivan, Michael J. Kennedy and Edward Loughlin—promised the elimination of corruption and pledged leadership for good government but failed to induce public confidence in the hall. In the election of 1945, however, Tammany once more established its ability to turn out a sizable vote. The newly elected mayor, William O'Dwyer of Brooklyn, sought to control Tammany by ousting Loughlin and forcing the acceptance in March 1947 of Frank J. Sampson as county Democratic leader.

Lacking the support of the district leaders, Sampson functioned as a figurehead and was soon replaced by Hugo Rogers, Manhattan's borough president. An embittered struggle between city hall and Tammany ensued, reflected in a series of district leadership fights in the 1949 primaries in which the regular organization defeated a group of insurgent reformers. Rogers, however, was ousted in favour of Carmine G. De Sapio.

Pressures from within and outside the organization brought about a reconciliation of Mayor O'Dwyer and the new Tammany leader but charges of underworld control of some district leaders and of others high in Tammany councils led to insistent demands for reform. Acting Mayor Vincent R. Impellitteri, who had been denied the regular nomination, ran as an independent, anti-Tammany candidate in the special election in 1950 and beat the organization decisively.

The hall lost other important local and city elections to independent Democrats, moreover. De Sapio managed, however, to stave off threats to his leadership in the primary campaigns of 1951 and 1953. In 1953, he joined with Edward Flynn of the Bronx to support Manhattan Borough Pres. Robert F. Wagner, Jr., for the Democratic mayoralty nomination. After Flynn's death and Wagner's nomination and election, De Sapio was recognized as an important leader. He succeeded Flynn as Democratic national committeeman and also became secretary of state for New York.

See Gustavus Myers, *The History of Tammany Hall* (New York, 1917); M. R. Werner, *Tammany Hall* (Garden City, L.I., 1928); W. Riordon, *Plunkitt of Tammany Hall* (New York, 1948).

(M. R. W.; M. B. D.)

TAMMERFORS: *see* TAMPERE.

TAMMUZ, Sumerian, Babylonian and Assyrian god, who died and rose annually with dying and reviving vegetation, originally *Dumu-zi*, "the son who rises, goes forth (from the nether world)," but generally interpreted "faithful son." Philologically both interpretations are correct, and no Accadian commentary exists to explain which meaning was accepted by them. The interpretation as "the son who rises," "the resurrected child," accepted in this article, is new and differs from all views held in the standard works on this cult. It is clear, however, that the main principle of this cult is the resurrection of the dying god, and the verb *zi* stands for the root *zig* to rise up, not *zid*, to be faithful. (*See* S. Langdon, *Sumerian Liturgies and Psalms*, p. 287, 17, *ušubba-za uziga-za*, "In thy fall and in thy resurrection.")

There are many titles of the youthful god, loved by his sister, the earth and heaven goddess, Innini (Ishtar), who descends yearly into Aralû (under-world) at the time of his death to bring him back to earth in her bosom. Since he represents the mystery of life and death, as seen in the withering vegetation of the hot Mesopotamian summer, and the rapid renewal of its life at the season of the spring rains, Tammuz is the patron of flocks and irrigation as well as of vegetation. Titles such as god *Sipa*, the shepherd, *umun mirsi*, lord of the flood, god *Ab-ú*, father of vegetation, god *Zulumma*, god of the date palm, are not so old as those which reveal theological speculation concerning the brother, lover and husband of the earth mother. Theology gave rise to this, the most widely spread and profoundly religious aspect of West Asiatic and Egyptian religion (where it appears as the cult of Osiris and Isis). *Dumu-zi* appears in the oldest texts without a divine title, *c.* 3200 B.C., and in Sumerian texts of Eannatum of Lagash (*c.* 2850 B.C.), as god *Dumuzi-apsu*, "risen child of the deep," describing him as the son of the god of the nether sea of fresh waters, *Ea*, *Enki*, a title confined apparently to early Lagash, where he appears more often under the title Ningišzida, "Lord of the faithful tree," a title which developed into an independent deity. In the legend of Eridu, in which Adapa is sent to the gates of heaven for judgment because he had broken the wings of the south-wind, he found Tammuz and Ningišzida at the gate of heaven; they offered him bread and water of life, which he, on the advice of the water god of Eridu, refused and thus lost immortality. Tammuz and Ningišzida appear to have been identified with the stars Castor and Pollux or, perhaps, Procyon and Sirius.

The liturgical wailings for Tammuz during the period of his sojourn in Aralû are numerous and describe every aspect of the theological doctrines concerning him. They are invariably composed in Sumerian, rarely with Accadian interlinear translation. These wailings occurred at midsummer, and the sixth month of the calendar at Lagash from the 28th to the 24th centuries is named the "month of the festival of Tammuz." There is a variant name for this month at Nippur, called "month of the mission of Innini," referring to her descent to Aralû in search of her departed lover.

The Babylonians retained the old name of the fourth month and for some reason described it as the "month Tammuz," corresponding to July, retaining the name "month of the mission of Innini" for the sixth month (Elul). In this month (fourth) Tammuz is said to have been bound, and the liturgies speak of his having been drowned among flowers which were thrown upon him as he sank beneath the waves of the Euphrates. He is described as the shepherd who left his flocks, as the shepherds sat in the fields wailing for Tammuz.

There is a strange inconsistency in the hymns of these wailings concerning the relation of the mother goddess to her lover, Tammuz. In the early Sumerian texts she is his sister, but soon the Semitic view that she is his mother prevails. The two theories appear inconsistently together throughout the entire history of the cult. He is, however, invariably the husband and lover of the otherwise consistently described virgin goddess of love, Innini, Gestinanna, *Bêlit-sêri*, and the cult is particularly associated with the great city Erech, home of the cult of Anu, the heaven god, and Innini. In all ceremonies connected with his cult his mother or sister is invariably associated with him, and it is the one aspect of Sumerian and Babylonian religion which may be described as universal. Badtibira, near Erech, seems to have been the original seat of the cult, and in the Sumerian tradition of the ten prediluvian kings, the name of one at Badtibira is Dumuzisipa, "Tammuz the shepherd"; one of the names of Tammuz is Tibira. No great temple was built to him, and a few casual references to a temple of Tammuz clearly refer to shrines in the temple of the local deity reserved for the wailings and mystic ceremonies of the cult. Worship of Tammuz spread far beyond the lands to which the Sumerian religion was principally confined. Ezekiel speaks of it as firmly installed at Jerusalem in his time; it is mentioned in the Christian era in Mandean and Syriac literature, and it survived among the Sabeans at Harran as late as the middle ages.

In astrology Tammuz was identified with Aries; in the magic rituals he is symbolized by a white kid, and he is also connected with the ram, which led to this astral identification. Under the title *Sibzianna*, "faithful shepherd of heaven," he was identified with Orion. During the period of deified king worship in the Dungi period of Ur, and in the time of the Isin dynasty, the deified kings habitually identified themselves with Tammuz and were worshipped as husbands of the mother goddess. For this

reason it has been argued that the cult arose in prehistoric times when a king was put to death as a sacrifice to the earth mother in order to secure the good will of the deity of vegetation. There is, however, no evidence for this in the earlier texts, and so far as Sumerian religion is concerned, the theory must be dismissed as unproved. He was held to be a god of healing, bestower of health and one who, like all other deities, had power over the demons.

It is obvious that a cult which is based upon the death and resurrection of a propitiating god and upon the love of a divine mother who wails for her son, has direct connection with the facts and the theological views based upon them, which gave birth to Christianity. But the form of the cult which apparently most directly affected the origins of Christianity is that in which Marduk of Babylon was identified with Tammuz. At the Nisan or new year festival at Babylon, Bêl (Marduk) was said to have been imprisoned in the lower world, and a priestess weeps at his sepulchre. A malefactor was slain with Bêl and they descend together to the land of darkness. Beltis, his wife, descends to hell to seek him, and Bêl's garments are given to Ishtar (mother of Tammuz). Bêl was laid in a sepulchre, from which he soon comes forth. This Marduk transformation of the national Tammuz cult is only another effort of the priesthood of the capital to enlarge the worship and importance of the local cult. It obtained nowhere else in Babylonia and Assyria, and may be regarded as a true interpretation of what transpired regularly in the Tammuz cult itself. That the cult had direct influence upon the origins of Christianity cannot be denied, and the Greek cult of Adonis owes its essential content to the Sumerian Tammuz.

BIBLIOGRAPHY.—S. Langdon, *Tammuz and Ishtar* (1914); H. Zimmern, *Der Babylonische Gott Tamûz* (1909); W. W. G. Baudissin, *Adonis und Esmun* (1911); S. Langdon, *Sumerian and Babylonian Psalms* (1909); H. Zimmern, *Sumerisch-babylonische Tamûzlieder* (1909); S. Langdon, *Revue d'Assyriologie*, vol. xii., pp. 33–45; *Babylonian Liturgies* (1913); *Sumerian Liturgies and Psalms* (1919); Hugo Radau, *Sumerian Hymns and Prayers to God Dumu-zi* (1913).

(S. L.)

TAMPA, the seat of Hillsborough county, Fla., U.S., the principal manufacturing city and agricultural centre on the west coast of the state, is located at the mouth of the Hillsborough river as it enters an arm of Tampa bay. The population in 1960 was 274,970, an increase of 120.5% in the decade; that of the Tampa-St. Petersburg standard metropolitan statistical area (Hillsborough and Pinellas counties) was 772,453, an increase of 88.8%. (For comparative population figures for the two cities *see* table in FLORIDA: *Population.*) (*See* also SAINT PETERSBURG.)

History.—Tampa was the name of an Indian settlement recorded by historians in the 16th century after Spanish explorers, including Hernando de Soto, and missionaries landed in its vicinity. Fort Brooke was established on the site in 1824 when the U.S. army began the removal of Seminole Indians from north to south Florida. The fort was also active in the seven years of warfare (1835–42) brought on by attempts to remove the Indians to the west, and during the American Civil War it was bombarded and the port blockaded to prevent the shipment of cattle from Tampa to Cuba. Tampa was incorporated in 1855. In 1886 Vincente Martinez Ybor induced several cigar manufacturers to move there after they had been driven out of Havana, Cuba, by revolutionary disorders and from Key West because of labour disturbances and a disastrous fire. Tampa became a world leader in the production of clear Havana handmade cigars.

The greatest developer of Tampa was Henry B. Plant, who made it an international port and tourist resort. In the mid-1880s he connected the city by a standard-gauge railway with lines to the north; he also inaugurated steamship service to Key West and Havana. In 1891 Plant opened the Tampa Bay hotel, a dark red brick castle of Moorish architecture topped by silvery domes and 13 minarets on which he had lavished $3,000,000. Five stories in height and extending over a 6-ac. tract on the Hillsborough river, it became one of the popular luxury hotels of the world.

During the Spanish-American War about 30,000 soldiers were camped in and near the city before embarking for Cuba. Cigar makers supported the Cuban independence movement, their aid being organized by the Cuban patriot José Martí; the cottage in Tampa where he made his headquarters is now a shrine.

The Latin quarter, known as Ybor City, which accounts for about 15% of the population, is the home of Spaniards and Cubans connected primarily with the cigar industry. They maintain their own imposing social clubs and hospitals. Their restaurants have an international reputation.

In the mid-1920s the man-made Davis Islands were created in the bay to provide sites for hotels, residences, apartments, a hospital and an airport. A landscaped 6-mi. boulevard bordering the crescent-shaped bay forms a scenic residential section.

Commerce, Industry and Transportation.—Tampa is a port of entry; after the opening of the Panama canal in 1914 it became one of the most important ports on the Gulf of Mexico. The location of two-thirds of the state's citrus groves in the Tampa area has made the city a centre for processing fruit. A large percentage of all phosphate mined in the U.S. is extracted from the adjoining county; much of it is shipped from Tampa. In addition to the numerous cigar factories there are more than 700 manufacturing plants. The tourist business is also important in the economy. The city is well served by rail and air lines as well as by harbour facilities.

Education and Recreation.—The University of Tampa, a private institution chartered in 1931, in 1933 was given a 99-year lease on the Tampa Bay hotel building by the city. Florida Christian college is a private junior college chartered in 1944. The state-supported University of South Florida was established in 1958 and opened in 1960. Features of the winter season in Tampa are the Gasparilla and Latin-American pageants and festivals, baseball, golf, tennis and sailing. (A. J. H.)

TAMPERE, formerly Tammerfors, an industrial city of Finland in 61° 28′ N., 23° 45′ E., on the rapids connecting Lakes Näsi-järvi and Pyhä-järvi, and at a railway junction. Pop. (1960) 126,573. It manufactures textiles, paper and leather goods.

TAMPICO, a city and leading port of Mexico, in the state of Tamaulipas, on the north bank of the Pánuco river, about 6 mi. from the Gulf of Mexico. In summer the climate is hot and humid, although a sea breeze modifies the temperature somewhat. In winter the temperature falls to freezing on occasional days when northers blow down along the Gulf coast. Tampico is almost surrounded by swampy lands and lagoons.

The eastern and poorer part of the town stands on low ground only two or three feet above the river, and is subject to inundations. However, a modern sewer system, up-to-date street paving and a better water supply, constructed after the beginning of the 20th century, greatly improved the sanitary condition and reduced the death rate from epidemics. The western part rises about 150 ft., consists largely of residential districts and is provided with still better sanitary equipment. The business section is well built, largely of stone and brick, while many of the newer structures are of reinforced concrete and rise to six and seven stories. The city is well supplied with gas and electric light.

Tampico has excellent railway facilities, road connections with the Pan-American highway, and air services are available to Mexico City, Veracruz, Brownsville, Tex., and other points. Although a bar exists at the mouth of the Pánuco river, jetties were built and the depth increased by dredging, so that vessels drawing up to 33 ft. can approach the water front of the city, while scows and other boats of light draught can navigate the river for more than 100 mi. upstream.

The Chijol canal, begun in 1901, affords a waterway 6 ft. deep and 25 ft. wide for about 75 mi. southward through the oil fields to Túxpan. Modern port works, spacious enough to accommodate at the wharves 14 vessels at a time, steel sheds and warehouses, a union railway station within easy reach of the water front and excellent equipment for loading oil tankers make Tampico the most up-to-date harbour in Mexico.

Tampico owes its importance to four of the most productive oil fields in the country (the Ebano, Pánuco, Huasteca and Túxpan) which are situated within about 100 mi. of its site. Until 1901 it was a second-rate port, outlet of the fertile but relatively

underdeveloped hinterland, with a reputation for unhealthful and unsanitary conditions. The rapid exploitation of petroleum resources resulted in a marked increase of population from about 10,000 in 1900 to 24,980 in 1921, to 82,475 in 1940, to 94,342 in 1950 and to 134,204 in 1958. For several years Tampico ranked as the greatest oil port in the world. Pipelines lead from the nearby fields, while fleets of scows bring oil from farther up the river. Pipelines and barges together are capable of transporting approximately 1,200,000 bbl. of oil daily from the fields to the port of Tampico.

Up river from the port agricultural activities were stimulated, fruits, vegetables and grains being grown to supply the local food demands. Modern commercial establishments are stocked with U.S. and European goods. Hotels, clubs, restaurants and places of amusement are numerous and well served. Because of high prevailing wage rates, the standard of living is generally higher than in most Mexican cities. Tampico is also noted as a vacation resort with excellent bathing, golfing, hunting and fishing.

Besides those connected directly with the oil business, the industrial establishments of Tampico include many for the handling and repair of oil well machinery, yards for the building of river boats, an electric light and power plant, factories for making ice, clothing and fruit preserves, sawmills, etc. In addition to petroleum, the exports include silver bullion (from San Luis Potosí, Aguascalientes, Torreón and Monterrey), istle fibre, sugar, hides, live cattle, cottonseed cake, honey, fustic, sarsaparilla, coffee and copper ores.

Tampico grew around a monastery which was founded in the ruins of an Aztec village by Andrés de Olmos, a Franciscan friar, in 1532. It was destroyed in 1683 by pirates and not reoccupied until 1823 when Gen. Lopéz de Santa Anna ordered that the site should be settled.

In 1829 a Spanish force under Isidro de Barredas was defeated at Tampico by Santa Anna. It was briefly occupied by U.S. forces during the Mexican War and again in 1862 by the French. The arrest of some American sailors in Tampico on April 9, 1914, led to the bombardment and occupation of Veracruz on April 22, 1914, by the U.S. navy. (J. W. Mw.; R. B. McCk.)

TAMWORTH, city of New South Wales, Austr., on the Peel river, 280 mi. by rail N. of Sydney. Pop. (1961) 19,064. Tamworth is the main provincial city of north and northwestern N.S.W. and the market of a pastoral and agricultural district. Flour and saw milling, the production of glucose, starch and kindred wheat products, egg processing, joinery and the manufacture of furniture are its main industries. Tamworth was the first city in Australia to provide electric street lighting, in 1888.

TAMWORTH, a market town and municipal borough in the Lichfield and Tamworth parliamentary division of Staffordshire, Eng., 15 mi. N.E. of Birmingham by road. Pop. (1961) 13,555. Area 4.2 sq.mi. It lies at the junction of the Tame and Anker rivers, in open country with wooded uplands. Offa, overlord of Mercia, built a palace there in the 8th century and surrounded the burgh by the great entrenchment known as Offa's dyke. During the 9th century Tamworth was destroyed by the Danes, but in A.D. 913 Aethelflaed, daughter of Alfred the Great, defeated them and built a timbered stockade at the junction of the rivers where the present castle stands. William the Conqueror bestowed the manor upon Robert de Marmion, who built the Norman castle, the walls and keep of which remain. It was largely restored in the Jacobean period. Tamworth, a borough since Saxon times, once possessed a royal mint, dating from Aethelstan's reign. Originally half the borough was in Staffordshire, the other half in Warwickshire, but in 1319 Edward III granted the Warwickshire half to the burgesses by letters patent. The town was incorporated by Elizabeth I in 1560 and a further charter was granted by Charles II in 1663. The whole borough was transferred to Staffordshire in 1889. The church of St. Editha (8th century) was rebuilt, after the Danes had burned it, by king Edgar who made it collegiate in A.D. 963. The existing Decorated building was erected after a fire in 1345. The grammar school, granted a charter by Elizabeth I in 1588, once formed part of the collegiate church. Guy's almshouses (1678) were endowed by Thomas Guy, founder of Guy's hospital, London, who also built the town hall over the open market in 1701. An ancient fair in honour of St. Swithin is still held (July 26). Two members of parliament were returned by the borough from 1562 to 1885. Sir Robert Peel represented it during the first half of the 19th century. Tamworth, the centre of an agricultural and coal-mining area, specializes in agricultural engineering, textile weaving, paper making, the manufacture of asbestos and aluminum ware, and the making of terra cotta bricks from local clay.

TANA, a lake of northeast Africa, chief reservoir of the Abbai or Blue Nile. Tana lies between 11° 36′ and 12° 16′ N. and 37° 2′ and 37° 40′ E., about 6,000 ft. above the sea, on the northern portion of the Abyssinian plateau. Its greatest length is 53 mi., its greatest breadth 41 mi., and it covers 1,418 sq.mi. The shores are well defined, generally flat, but at places the mountains descend somewhat abruptly into the water. Elsewhere the land rises in gentle undulations, except at the mouths of the larger tributary streams, where are alluvial plains of considerable size. At the southeast end the lake forms a bay about 11 mi. long, and from 3 to 8 mi. across. From this bay the Abbai issues. The island of Dek (8 mi. long by 4 broad) is in the southwestern part, near the smaller island of Dega.

Lake Tana is fed by three large rivers and by many petty streams. The chief tributary is the Little Abbai, which enters the lake at its southwest corner. This river, and the Abbai itself or Blue Nile which issues from the lake, are regarded as one and the same stream and a current is observable from the inlet to the outlet. Next in importance of the affluents are the Reb and Gumara, which enter the lake on its eastern side. The outlet of the lake is marked by openings in a rocky ledge, through which the water pours by two or three channels, uniting within a couple of miles into one river—the Abbai with a width of 650 ft.

Tana has been identified with the *Coloe Palus* of the ancients, which was described by Ptolemy as a chief reservoir of the Egyptian Nile and the source of the *Astapos*, which was certainly the Blue Nile. In 1625 it was visited by the Portuguese priest Jeronimo Lobo, and in 1771 by James Bruce. It was formerly known by the name of Dembea. (C. F. R.)

TANA, a river 440 mi. long, in Kenya colony, which forms the northern boundary of Kenya protectorate (a coastal strip 10 mi. wide). Its sources are along the watershed close to the eastern wall of the eastern rift valley and it enters the Indian ocean in Formosa bay, at Kipini, in 2° 30′ S., about 120 mi. N. of Mombasa. One series of its numerous headstreams traverses the Kikuya plateau north of the Athi, others flow from Mount Kenya, all following a southeasterly direction until intercepted by the main stream which flows northeast until, when nearing the equator, it turns eastward, afterwards sweeping southward to the sea. It receives tributaries only along its upper course but on entering the plain it flows along a tortuous course, has low banks, and is in part forested and inundated at high water, but away from the river the country consists of dry plains covered with mimosa scrub. Five miles from the sea it divides in two, the southerly branch being the Sheriko. The Tana formerly entered the sea at Ozi creek, 20 mi. to the south of its present mouth. (S. M. C.)

TANAGER, the name for members of the songbird family Thraupidae (formerly Tanagridae) containing 197 species. They are small to medium-sized birds of rather generalized form, with medium-sized wings, feet and tails. The beak is heavy and somewhat similar to that of their nearest relatives, the sparrows. As a group tanagers are noted for the brilliancy of their plumage, though they have developed no ornamental tufts or plumes. Reds, yellows, greens, blues and blacks in solid colours are common and sometimes arranged in striking patterns. The sexes may be coloured similarly, or the female may be much duller than the male. In some species, however, both sexes are dull-coloured. No tanagers are notable songsters, but some have simple songs of pleasing quality.

The family is entirely confined to the new world, with its centre of abundance in the tropics. Only four species range north to the United States, and two regularly to Canada, and these are all migratory. Tanagers are arboreal birds, living in tree tops,

HAL H. HARRISON FROM NATIONAL AUDUBON SOCIETY

MALE SUMMER TANAGER (PIRANGA RUBRA) AND YOUNG

undergrowth or shrubbery. Fruit is an important part of the diet of many species, and some feed almost entirely on mistletoe berries. Other species eat some insects.

The nest is a shallow, open structure. The eggs, two to four in number, may be white or bluish, immaculate or spotted. Monogamous pairing seems the rule and, in some species at least, while the female alone incubates, both sexes care for the young. In the tropics some species remain in pairs throughout the year.

When feathers were popular in millinery, brilliant tropical tanagers were often used as hat trimmings.

The best-known species are probably those of the United States, especially the scarlet tanager (*Piranga olivacea*), the male scarlet with black wings and tail, the female olive and yellowish; the summer tanager (*Piranga rubra*), with an all-red male; and the western tanager (*Piranga ludoviciana*), with the male yellow with reddish head and black wings and mantle. (A. L. Rd.)

TANAKA GIICHI, Baron (1863–1929), Japanese statesman and general, born in Yamaguchi in 1863, graduated at the military cadet's school in 1886, and the staff college in 1892. He was minister of war in the Hara and Takahashi cabinets in 1918–21, and 1920. He became leader of the Seiyukwai Party in 1926, and formed a government in April 1927, himself taking the ministry of foreign affairs. He died on Sept. 28, 1929.

TANAQUIL, the Etruscan name of the wife of Tarquinius Priscus, or of one of his sons. After her immigration to Rome she is said to have received the name Gaia Caecilia. She was famous for her prophetic gifts. There was a statue of her as Gaia Caecilia in the temple of Sancus, which possessed magical powers. She was celebrated as a spinner of wool, and was supposed to exercise influence over Roman brides. Tanaquil and Gaia Caecilia are, however, really distinct personalities. The anecdotes told of Gaia Caecilia are etiological myths intended to explain certain usages at Roman marriages.

TANBARK OAK (*Lithocarpus densiflora*), an ornamental evergreen allied to the oaks, both of which belong to the beech family, Fagaceae. The tanbark oak, native to southern Oregon and northern California, is a tall tree of 75 to 100 ft., with alternate, leathery, toothed, more-or-less oblong three- to five-inch leaves covered on the underside with densely matted hairs that disappear in time.

The fruit is an acorn about one inch long. The tree is valued chiefly for its bark which is an important source of tannin, but it is also an attractive ornamental. (J. M. Bl.)

TANCRED (d. 1112), nephew of Bohemund I, took the Cross with Bohemund in 1096 and marched with him to Constantinople. Here he refused to take an oath to Alexius, escaping across the Bosporus in the disguise of a peasant; but after the capture of Nicaea he followed the example of the other princes and became the man of Alexius. At Heraclea, in the centre of Asia Minor, he left the main body of the crusaders and struck into Cilicia, closely followed by Baldwin of Lorraine. He made himself master of Tarsus and, when he was evicted from it by the superior forces of Baldwin, pushed further onward and took the towns of Adana and Mamistra. He joined the main army before Antioch and was prominent in the siege. He took part also in the siege and capture of Jerusalem (July 1099), after which he went to Nablus and began to found a principality of his own. He fought in the battle of Ascalon in August; and after it he was invested by Godfrey with Tiberias and the principality of Galilee, to the north of Nablus. In 1100 he attempted, without success, to prevent Baldwin of Lorraine (his old enemy in Cilicia) from acquiring the throne of Jerusalem. In the same year Bohemund I was captured by Moslems, and Tancred, relinquishing his lands in Galilee, went to act as regent in Antioch. In this capacity he regained the Cilician towns (1101) and recaptured Laodicea (1103). When Bohemund was set free, Tancred had to surrender Antioch to him. In 1104 he joined with Bohemund and Baldwin de Burg (now count of Edessa in succession to Baldwin of Lorraine) in an expedition against Harran, in which they were heavily defeated, and Baldwin was taken prisoner. Tancred, however, profited doubly by the defeat. He took over the government of Edessa in Baldwin's place, and in 1105 Bohemund surrendered to him the government of Antioch and went to Europe for reinforcements. Ruler of the two northern principalities, Tancred carried on vigorous hostilities against his Mohammedan neighbours, especially Ridwan of Aleppo; and in 1106 he captured Apamea. In 1107, while Bohemund was beginning his last expedition against Alexius, he again wrested the whole of Cilicia and Laodicea from the Greeks; and he steadfastly refused, after Bohemund's humiliating treaty at Deabolis in 1108, to agree to any of its stipulations with regard to Antioch and Cilicia. When Baldwin de Burg regained his liberty in 1108, it was only with difficulty that Tancred was induced to restore Edessa to him. But it was against the amirs of northern Syria that his arms were chiefly directed. Tancred died in 1112.

TANCRED (d. 1194), king of Sicily, an illegitimate son of Roger the eldest son of King Roger II, was crowned in January 1190 in succession to William II (*q.v.*). He was supported by the chancellor Matthew d'Ajello and the official class, while the rival claims of Roger II's daughter Constance and her husband, Henry VI, king of the Romans and emperor, were supported by most of the nobles. Tancred was a good soldier; but he was ill-supported in his task of maintaining the Norman kingdom, faced with general apathy, threatened by a baronial revolt, and menaced by Richard Coeur-de-Lion, at Messina, 1190. Henry, skilfully winning over Pisa, Genoa and the Roman Commune, isolated Tancred and intimidated Celestine III, who, on April 14, 1191, crowned him emperor at Rome. He, however, failed to capture Naples in August and retired north, leaving garrisons along the frontiers of the Regno. Tancred now sought to win over the towns by extensive grants of privileges, and at Gravina (June 1192) was recognized by the pope, whose ineffectual support he gained by surrendering the royal legateship over Sicily. In 1192 and 1193 he successfully opposed the Apulian barons, but his death at Palermo (Feb. 20, 1194) a few days after that of Roger, his son and joint king, made Henry's path clear.

TANDY, JAMES NAPPER (1740–1803), United Irishman, born in Dublin in 1740, was a prominent reformer and supporter of Lucas and Grattan. He first won popularity by his attacks on municipal corruption and his proposal to boycott English goods in retaliation for the restrictions imposed on Irish commerce. In 1780 Tandy was expelled from the Dublin volunteers (*see* Flood, Henry), but he continued his revolutionary agitation, and in 1791 co-operated with Wolfe Tone in founding the Society of United Irishmen, of which he became the first secretary. A challenge sent to the attorney general resulted in his imprisonment (1792), but on his release Tandy assisted in raising two battalions of a "national guard," which were suppressed. He then fled to America, and in 1798 proceeded to France, where he joined a number of Irish refugees. Having gained the ear of the French government, Tandy was given command of a cor-

vette, a small force of men, and a quantity of arms for distribution in Ireland. Arriving at Donegal, however, he learned that Humbert's expedition had failed, and he thereupon fled to Hamburg. The refugees were surrendered on a peremptory demand from the British government. Condemned to death, he was finally reprieved due, it is said, to Bonaparte's intercession.

Tandy's exploits completely captured the popular imagination, and he is still remembered as the hero of the "Wearing of the Green." He died in France on Aug. 24, 1803.

TANEIEV, SERGIUS (1856–1915), Russian composer and theorist, was born on Nov. 13, 1856, in the Vladimir government. He studied under Nicholas Rubinstein and Tchaikovsky at the Moscow conservatory. After a short career as a concert pianist he went back to Moscow as professor of instrumentation, then first professor of pianoforte and, in 1885, director. He died at Moscow on June 15, 1915. His published works include a symphony; five string quartets; a dramatic work, *Orestes* (performed in St. Petersburg 1895); a cantata, *John of Damascus;* and choruses for male voices and for mixed voices. He also wrote a treatise on counterpoint.

TANEY, ROGER BROOKE (1777–1864), fifth chief justice of the United States, whose decision in the Dred Scott case in 1857 became a centre of violent controversy, was born March 17, 1777, in Calvert county, Md. His father was a member of a family of planters and he inherited the conservative tradition of the landed aristocracy of the south. His education in rural schools was supplemented by a family tutor and by three years of instruction at Dickinson college, from which he graduated in 1795. Since as a second son Taney would not inherit the family plantation, he was trained for a legal and political career. He studied law at Annapolis under Judge Jeremiah T. Chase of the Maryland general court and was admitted to the bar in 1799.

In 1801 Taney settled in Frederick, Md., where he established himself as an able lawyer, achieved a position of prominence in political affairs and raised a family. His marriage in 1806 to Anne Key, sister of Francis Scott Key, was a happy one and resulted in six daughters. Although his practice was demanding and his health frail, Taney found time for political activity, serving several terms in the Maryland legislature and gradually working his way into the leadership of the Federalist party in the state.

When the War of 1812 split the Federalist party (*q.v.*), Taney, a representative of inland agrarian interests, became the leader of the group, derisively referred to as "Coodies," which supported the conduct of the war. Although the former Federalists cooperated after the war on certain matters of state politics, the breakdown of the Federalist organization left Taney without party affiliation. In the 1820s he aligned himself with Andrew Jackson and thereafter was found in the ranks of the Democrats. This transition was not unnatural since Jackson, like Taney, was a slaveowner and a member of the landed aristocracy, and they shared many of the same views.

In 1823 Taney moved to Baltimore, where professional opportunities were greater, and was soon recognized as a leader of the Maryland bar. His extensive practice broadened his knowledge of the law and gave him wide acquaintance with the economic developments of his time. In 1827 he was appointed attorney general of Maryland, an appointment which honoured his legal attainment while allowing him to continue his private practice.

Taney became a national figure as a result of his participation in the bank war of the 1830s. When Jackson reorganized his cabinet in 1831, Taney accepted an appointment as attorney general. At that time the Bank of the United States, chartered by congress in 1816, had attained a dominant position in American finance and sought to exercise its influence in order to force immediately the renewal of its charter which was to expire in 1836. Although Taney had once regarded a national bank as a desirable instrument for regulating the currency, his experience with the activities of the Bank of the United States convinced him that the bank had abused its powers. When in 1832 congress enacted a bill to renew the charter, Taney persuaded Jackson to veto it and wrote much of Jackson's veto message.

In the ensuing presidential campaign the bank took an active part, financing its supporters and attempting to frighten its opponents by manipulating the credit structure. The revelation of these events after Jackson's re-election in 1832 led Taney and others to advise Jackson to withdraw government deposits from the Bank of the United States and place them in selected state banks. Jackson delayed action for a number of months but eventually in 1833 ordered William J. Duane, then secretary of the treasury, to remove the deposits. When Duane refused, Jackson dismissed him and gave Taney a recess appointment as secretary of the treasury. Taney's subsequent action in removing government deposits to a number of state banks was subjected to virulent partisan attack, much of which was undeserved since the bank had engaged in political and financial activities of questionable character. Nevertheless, Taney's attempt to improve and regulate the currency through a system of state banks proved ineffective, in part because congress failed to enact implementing legislation. Although the bank was not rechartered, opposition to Taney and his financial program was so strong that his appointment as secretary of the treasury was rejected by the senate when it was presented for confirmation. Three years of cabinet service, during nine months of which he held a recess appointment as secretary of the treasury, thus ended, and Taney returned to private life and attempted to rebuild his law practice.

In 1836, at the age of 59, Taney began the last and greatest phase of his career. As a reward for his political loyalty, Jackson appointed him chief justice of the United States to succeed the celebrated Federalist John Marshall. Despite Whig opposition, Taney was confirmed by the senate and took the seat he was to hold for nearly three decades (1836–64). His appointment coincided with change and expansion in the personnel of the court, with the result that by 1837 all but two of the nine judges were appointees of Jackson or Van Buren. This change alarmed those who feared that property rights would be at the mercy of state legislatures. Although the Taney court did give greater scope to the exercise of state power than had the Marshall court, it was no less vigilant in maintaining federal supremacy until it foundered on the constitutional questions presented by Negro slavery.

In his first great judicial utterance, *Charles River Bridge* v. *Warren Bridge*, 11 Pet. 420 (1837), Taney expressed his deeply-held conviction that the powers conferred by corporate charters should be given a narrow interpretation in order to preserve the rights of the community. Speaking for a majority of the court, he rejected the claim of a bridge company that the subsequent grant by the state legislature of a charter to another bridge company impaired the obligation of contracts. The same emphasis on the rights of the community pervaded Taney's opinions dealing with the impact of the commerce clause on state regulatory legislation. He was unwilling to use the commerce clause as an instrument to defeat state regulation; *e.g., License Cases*, 5 How. 504 (1847), *Passenger Cases*, 7 How. 283 (1849), dissenting opinion. While he concurred in several opinions upholding state regulation on a police power rationale, his own view appears to have been that the states, in the absence of federal legislation, remained free to regulate commerce.

Despite his willingness to sustain state authority against constitutional attack, Taney's court cannot be characterized as interested only in "states' rights." Whenever state authorities threatened or interfered with the exertion of federal power, Taney was firm in upholding federal supremacy. His opinion in *Ableman* v. *Booth*, 21 How. 506 (1858), denying state power to obstruct the processes of the federal courts, remains a magnificent statement of constitutional federalism. Under Taney's leadership, federal judicial power was expanded over corporations; the admiralty jurisdiction was extended to inland waters; the federal government was held to have a paramount and exclusive authority over foreign relations; and congressional authority over U.S. property and territory was vigorously upheld.

Until the mid-1850s the court avoided embroilment in the controversy over slavery which was constantly increasing in bitter-

ness. As late as 1850, in *Strader* v. *Graham*, 10 How. 82, Taney succeeded in disposing of important constitutional questions relating to slavery on a narrow jurisdictional ground. However, by 1856 the prestige of the court was thought so great and the issue so momentous that the court was urged to settle the question whether congress had power to restrict slavery in the territories. In the Dred Scott case, 19 How. 393 (1856), one of the great "self-inflicted wounds" of the supreme court, the court yielded to the temptation to solve this fundamental political controversy by constitutional decision. Taney, writing for a majority of the court, held that Negroes could not possess rights of citizenship entitling them to sue in federal courts; that the removal of a slave to a territory in which congress, by the Missouri compromise, had prohibited the holding and owning of slaves did not result in freedom because congress lacked such power; and, finally, that the return of a Negro to a state recognizing him as a slave was conclusive as to his status. On the facts of the Dred Scott case, the third and narrowest ground was sufficient to dispose of the case. Taney's further holdings, that congress had no power to exclude slavery from the territories and that Negroes could not be citizens, were disputed by vigorous dissenting opinions and were bitterly attacked by the northern press. The decision became an issue in national politics and, with the victory of the Republicans in 1860, the prestige of the court reached the lowest point in its history. Taney's attachment for southern culture thus resulted in an unsuccessful attempt to safeguard that culture from the legislative power of the northern majority.

Taney's last years were spent in futility and despair. His background led him to sympathize with the south and he hoped that the south would be allowed to secede peacefully. During the war the court, though it continued to function as part of the government, was suspected of disloyalty. Taney, believing that war issues might come before the court, prepared in the privacy of his study a series of opinions in which he found various war measures unconstitutional. *Ex parte Merryman*, Fed. Cases no. 9,487 (1861), containing a vigorous defense of the rights of civilians in war time, is the only one of those opinions to be found in the reports. Taney's death in Washington on Oct. 12, 1864, went largely unnoticed and unregretted.

Taney's relations with his family and his associates were warm and pleasant. Not an effective platform speaker, he was at his best in personal association, where his warmth, kindness and quiet charm quickly resulted in intimacy. He presided over the supreme court with dignity and courtesy, and his personal qualities, so his colleagues have testified, led to greater unity on the court than otherwise would have occurred. His legal abilities were of the highest order, as his concise, persuasive opinions indicate. Although a devout Roman Catholic, he refused to allow members of his faith to prosyletize among his wife and daughters, who with one exception were Protestants.

Any evaluation of Taney is complicated by the fact that his actions and attitudes affecting slavery and national unity are part of the tradition defeated in the Civil War. Yet his outstanding abilities had a lasting influence on the content and development of American constitutional law.

BIBLIOGRAPHY.—The leading biography is C. B. Swisher, *Roger B. Taney* (1935). *See* also Samuel Tyler, *Memoir of Roger Brooke Taney, LL.D.* (1872); B. C. Steiner, *Life of Roger Brooke Taney* (1922); Charles Warren, *The Supreme Court in United States History*, 2 vol., rev. ed. (1926); Charles W. Smith, *Roger B. Taney: Jacksonian Jurist* (1936). (R. C. C.)

T'ANG, the dynasty that ruled China from A.D. 618 to 907. The first emperor Li Yuan (or Kao-tsu, 618–626) was a member of the aristocracy of part-barbarian, part-Chinese origin, that had formed the governing class in the Northern Chou state, centred on Ch'ang-an in the 6th century, as well as in the Sui dynasty (A.D. 589–618). Li Yuan's success in obtaining the throne at the end of the Sui dynasty was largely due to his second son, Li Shih-min (or T'ai-tsung, 626–649), who succeeded him.

T'ai-tsung was succeeded by his son Kao-tsung (649–683), who soon fell under the domination of one of the most remarkable women in Chinese history, Wu Tse-t'ien or Empress Wu. She became the real power in the land even during her weak husband's lifetime. After his death she deposed her two sons in succession and ruled in her own name, a thing never achieved before or since in China. She changed the dynastic name to Chou but insured the eventual restoration of T'ang when invasions of the Khitans and Turks in the northeast necessitated an appeal to the people's loyalty.

During the reign of Kao-tsung the T'ang empire completed, at great cost, the conquest of northern Korea. But after *c.* 682 the northern Turks revolted, following 50 years of submission, and the Four Garrisons in Sinkiang were lost to the Tibetans. Also the Korean conquests were swallowed up by T'ang's erstwhile ally, the southern Korean state of Silla, which united the peninsula for the first time.

The aged empress's forced abdication in 705 was followed by an unstable period during which her two sons, Chung-tsung (705–710) and Jui-tsung (710–712), followed one another on the throne but were dominated by cliques centred around the empress of the former, surnamed Wei, and their sister, the T'ai-p'ing Princess. Successive *coups d'état* finally led to Jui-tsung's abdication in favour of his son, known as Ming Huang the Brilliant Emperor (alias Hsuan-tsung, 712–756). His reign, the longest in the dynasty, was also its high point of prosperity. With the help of able ministers he corrected administrative abuses that had grown up in the previous decade but efforts at fiscal reform, mainly carried out by special commissioners of aristocratic background, conflicted with the interests of the rising new landowning class. A political struggle ensued which resulted in the virtual dictatorship of the imperial kinsman Li Lin-fu, from 736 onward, after which the emperor withdrew into a life of pleasure.

Under Ming Huang the Chinese once more expanded on all fronts but at the cost of establishing large standing armies composed to a dangerous extent of non-Chinese troops. In the years 751–754 the Chinese suffered defeats in Manchuria; at the battle of Talas against the Arabs in central Asia; and especially against the state of Nan Chao in Yunnan. Real disaster struck in 755 with the revolt of the commander of the northeast sector of the frontier. This was An Lu-shan, of part-Sogdian and part-Turkish origin, who owed his promotion, like the barbarian generals in other sectors, to Li Lin-fu's fear of Chinese rivals.

The rebels captured the capital and drove the emperor to flight. The crown prince (Su-tsung, 756–762), who assumed the throne in the northwest succeeded, with the aid of a contingent of Uighur troops, in recapturing Ch'ang-an and Lo-yang. But the T'ang forces lacked strength to complete their victory and the war dragged on inconclusively until 762 when the last rebel pretender was killed. The northeastern provinces remained in control of former rebel generals who were acknowledged as T'ang governors; much of the rest of the country was in the hands of military leaders who were often as insubordinate as the former rebels. In the west the Tibetans, now at the height of their power, occupied large areas and in 763 they even briefly occupied Ch'ang-an. In the north, T'ang's allies, the Uighurs, were almost as oppressive as the Tibetans. The disruption of government by the rebellion led the emperors to place increasing reliance on their personal attendants, the palace eunuchs, who were given control of the palace armies.

During the reigns of Tai-tsung (762–779) and Te-tsung (779–805) the policy of the T'ang court was mainly a negative one of trying to control the militarists by diplomacy while the country slowly recovered from the effects of war. During the brief interlude of Shun-tsung's reign in 805 a reform party tried unsuccessfully to take the palace armies out of the hands of the eunuchs, who thereafter controlled the person of the sovereign, setting up and removing emperors as they pleased.

Under Hsien-tsung (805–820) some success was achieved in asserting central authority over the provinces but it was not lasting, and the northeast remained effectively autonomous to the end of the dynasty.

Hsien-tsung was murdered by eunuchs and the short reigns of Mu-tsung (820–824) and Ching-tsung (824–827) followed. Wen-tsung (827–840) plotted with his ministers to deprive the eunuchs of their military power but the plan misfired and resulted in the

Sweet Dew incident of 833 in which three chief ministers and many other officials were slaughtered. Under Wu-tsung (840–846) the simultaneous decline of the Tibetan and Uighur empires eased the pressure on the northern and western frontiers. This was followed by the persecution of the Manichaean church in China, which had enjoyed Uighur protection, and this in turn led on to an unparalleled suppression of Buddhist monasteries and laicization of monks and nuns, a blow from which Chinese Buddhism never fully recovered.

The reign of Hsuan-tsung (846–859) was outwardly calm except for wars with Nan Chao in the south, but the demoralization of the central government and the lawlessness of the provincial armies and their leaders made the dynasty quite incapable of dealing with the disturbances, provoked to a large extent by maladministration and corruption, which broke out in an ever-rising crescendo under I-tsung (859–873) and Hsi-tsung (873–888). In 875 began the great peasant uprising known as the Huang Ch'ao rebellion, led by former salt smugglers. After roaming up and down the country for years the rebels occupied Ch'ang-an in 881. They were unable to organize the food supply of the capital and this, together with the intervention of Sha-t'o Turks, forced Huang Ch'ao to flee east in 883, where he was caught and killed.

From then on, while lawless bands terrorized the countryside, the T'ang court could only survive by playing off one military governor against another. In 903 Chu Ch'üan-chung, one of Huang Ch'ao's officers who had surrendered opportunely and obtained the governorship of the strategic region of K'ai-feng, gained control of the person of the T'ang emperor Chao-tsung (888–904). He settled the eunuch problem by a wholesale massacre. After his men had murdered Chao-tsung in 904 he placed a boy of thirteen (Chao-hsuan-ti) on the throne who was presently put through the motions of abdication. Chu Ch'üan-tsung ruled over part of north China as T'ai-tsu of Later Liang (907–915). China fell apart into a number of independent kingdoms, and unity was not restored until after the founding of Sung (*q.v.*) in 960: *See* also CHINA: *History*.

Social and Economic Life.—During T'ang the economic centre of gravity of China shifted from the Yellow River plain to the Yangtze. The population of the latter region increased enormously, swollen by influxes of refugees from the north, especially after the rebellion of An Lu-shan. The canal system built by Yang-ti of Sui became the lifeline of the capital, especially after the efficiency of the transport system had been improved in the reign of Ming Huang and again under Tai-tsung. Trade, both internal and external, was stimulated by the unification of the country and its far-flung conquests. Much of this trade was in the hands of foreigners, particularly Iranians, who settled in large communities in the major cities. They came both by land through central Asia and by sea to Canton. Especially during the first half of the dynasty China was very receptive to foreign ideas. Foreign music and dancing were popular, and foreign types are common among the familiar T'ang pottery figurines.

Migration and economic development led to a breakdown in the system of state registration of population and distribution of land and to the growth of private estates. After the An Lu-shan rebellion the old system was finally abandoned and direct taxation was based on actual land under cultivation and on an assessment of the wealth of the household. The salt monopoly, introduced as an emergency war measure during the rebellion, became a major source of government revenue and continued into the 20th century. The increase in trade led to a need for coinage; as the government mints were unable to produce an adequate supply, counterfeiting was a chronic problem. The shortage of cash and the difficulty of transporting the bulky copper coins encouraged the development of credit instruments such as the so-called "flying money," a kind of bank draft. True paper money did not appear until the 10th century.

Society remained decidedly aristocratic but economic expansion led to the creation of a more broadly based landowning class. These new men began to push their way into the government by means of the examination system, which, although a development from earlier institutions, really began in Sui and T'ang. Empress Wu, especially, used it to undermine the position of the aristocratic clans. The competition between aristocratic privilege and the ideal of selection by merit continued throughout the dynasty.

Religion and the Arts.—The T'ang emperors officially supported Taoism because of their claim to be descended from Lao-tzu, but Buddhism continued to enjoy great favour through most of the period, especially under Empress Wu. The famous pilgrim, Hsuan-tsang, who went to India in 629 and returned in 645, was the most learned of Chinese monks and introduced new standards of exactness in his many translations from Sanskrit. In spite of such direct Indian influences through Indian teachers in China as well as Chinese pilgrims to India, the most significant development in T'ang was the growth of new indigenous schools which adapted Buddhism to Chinese ways of thinking. Most prominent were the syncretic T'ien-t'ai school which sought to embrace all other schools in a single hierarchical system, even reaching out to include Confucianism as well, and the radically antitextual, antimetaphysical Southern Ch'an (Zen) school which had strong roots in Taoism. The popular preaching of the salvationist Pure Land sect was also important. After the rebellion of An Lu-shan a nationalistic movement in favour of Confucianism appeared, merging with the efforts of T'ien-t'ai Buddhism to graft Buddhist metaphysics on to classical doctrine and lay the groundwork for the Neo-Confucianism of Sung.

In literature the greatest glory of the T'ang period was its poetry. By the 8th century poets had broken away from the artificial diction and matter of the court poetry of the Southern Dynasties and achieved a new directness and naturalism. The great flowering came in the reign of Ming Huang, the time of the great figures of Li Po (701–762), Wang Wei (698–759) and Tu Fu (713–770), besides a host of lesser names. The rebellion of An Lu-shan and his own bitter experiences in it brought a new note of social awareness into the later poetry of Tu Fu. This appears again in the work of Po Chü-i (772–846) who wrote verse in clear and simple language. Toward the end of the dynasty a new poetic form, the *tz'u*, in a less regular metre than the five-word and seven-word *shih*, and meant to be sung, made its appearance. The *ku-wen* or "ancient style" movement grew up after the rebellion of An Lu-shan, seeking to replace the euphuistic "parallel prose" then dominant. It was closely associated with the movement for a Confucian revival. The most prominent figures in it were Han Yü (768–824) and Liu Tsung-yuan (773–819). At the same time came the first serious attempts to write fiction, the so-called *ch'uan-ch'i* or "tales of marvels." Many of these T'ang stories later provided themes for the Chinese drama.

The patronage of the T'ang emperors and the general wealth and prosperity of the period encouraged the development of the visual arts. Though few T'ang buildings remain standing, contemporary descriptions give some idea of the magnificence of T'ang palaces and religious edifices and the houses of the wealthy. Buddhist sculpture shows a greater naturalism than in the previous period but there is some loss of spirituality. Few genuine originals survive to show the work of T'ang master painters such as Wu Tao-hsuan (*c.* 700–760) who worked at Ming-huang's court. As a landscape painter the poet Wang Wei was a forerunner of the *wen-jen* or "literary man's" school of mystical nature painting of later times. The minor arts of T'ang, ceramics, metalwork, textiles, etc., give expression to the colour and vitality of the life of the period. Printing appeared for the first time during T'ang. Apparently invented to multiply Buddhist scriptures, it was being used, by the end of the dynasty, for such things as calendars, almanacs and dictionaries. The first printing of the Confucian classics came after the end of the dynasty in 932 in Szechwan.

BIBLIOGRAPHY.—C. P. Fitzgerald, *China: a Short Cultural History* (1935); K. S. Latourette, *Chinese, Their History and Culture* (1946); A. Waley, *Life and Times of Po Chü-i* (1949); E. G. Pulleyblank, *The Background of the Rebellion of An Lu-shan* (1955); E. O. Reischauer, *Ennin's Travels in T'ang China* (1955). (E. G. PU.)

TANGANYIKA, a republic in east central Africa within the Commonwealth of Nations that merged with Zanzibar in April 1964 to form a single sovereign state, the United Republic of Tanzania (*q.v.*). It is bounded east by the Indian Ocean, south and southwest by Mozambique (Portuguese East Africa), Malawi,

and Zambia, west by the Democratic Republic of the Congo, Rwanda, and Burundi, and north by Uganda and Kenya. Mafia Island is included in Tanganyika. The total area of 361,800 sq.mi. (937,052 sq.km.) includes 20,650 sq.mi. (53,484 sq.km.) of inland water. The capital is Dar es Salaam (*q.v.*).

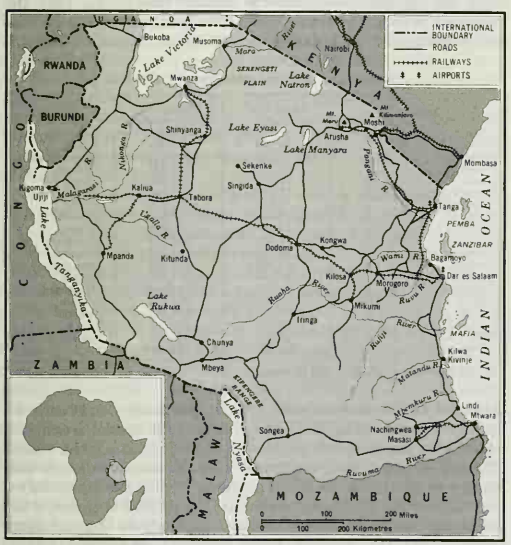
PHYSICAL GEOGRAPHY

Geology.—Most of Tanganyika is composed of ancient crystalline rocks (Archean and Proterozoic). The oldest, those of the Dodoman System of the Central Plateau, probably form a rigid shield under rocks of later age. Banded ironstones and other volcanics of the Nyanzian and Kavirondian systems with extensive granite emplacements are found to the north, while to the west and east are metamorphosed sedimentary and volcanic rocks of the Ubendean and Usagaran systems, respectively, which, with the Dodoman System, were formerly grouped under the term Basement. In the extreme west are found metamorphic rocks of sedimentary origin (Karagwe-Ankolean), intrusive granites with associated tin and wolfram, and the sediments and lavas of the Bukoban System.

Karroo beds (Upper Carboniferous—Lower Jurassic) preserved in faulted basins in the south and southwest contain Coal Measures. Younger sedimentary rocks (marls, limestones, sandstones, and shales) deposited from Jurassic to Recent times form a belt between the plateau and the coast, while lake and river sediments have been laid down inland since Pleistocene times. What are believed to be the earliest known remains of man (*Zinjanthropus boisei*) were in 1959 found in such deposits in Olduwai Gorge.

Volcanic activity associated with rift faulting occurred from Late Cretaceous times, in the southwest around Mt. Rungwe in the Rukwa-Nyasa Rift, and in the north resulting in the immense volcanic massifs and lava flows from Mt. Hanang to Kilimanjaro. Ol Doinyo (Oldongo) Lengai Volcano, near the Kenya border, was active in the 1950s.

Relief and Drainage.—The greater part of Tanganyika consists of an immense plateau varying in altitude from 3,500 to 4,500 ft. (1,070 to 1,370 m.). To the west it is bounded by steep escarpments of the western arm of the Great Rift Valley, which contains Lakes Tanganyika, Rukwa, and, outside the territory, Lakes Kivu, Edward, and Albert. The valley forks north of Lake Nyasa, and the eastern arm, which passes through the centre of Tanganyika, is less distinct; only the western wall is seen, as a cliff of varying height, until the troughlike characteristics of the Gregory Rift become noticeable near Kenya.



TOWNS AND TRANSPORTATION ROUTES IN TANGANYIKA

The mountain systems are grouped in the northeast and southwest, with a ridge of highland between the two forming the eastern margin of the plateau. In the north are the volcanic peaks of Kilimanjaro (19,340 ft. [5,895 m.]) and Meru (14,979 ft. [4,566 m.]), and the Crater Highlands, with the famous game-filled Ngorongoro caldera. Running southeast from Kilimanjaro to the coast are the Pare and Usambara ranges, while the Nguru and Uluguru mountains rise behind the coastal plain farther south. At the northern end of Lake Nyasa are the Livingstone Mountains, the Kipengere Range, the Poroto Mountains, Mt. Rungwe, and, farther west, the Ufipa Plateau.

Although appearing well watered on the map, Tanganyika has few permanent rivers, and most of the Central Plateau is devoid of running water throughout the dry season; only half the country is supplied with natural water throughout the year. The main rivers flowing to the Indian Ocean are: the Pangani, or Ruvu, which rises on Kilimanjaro; the Wami, whose tributary the Kinyasungwe-Mukondokwa has cut back into the plateau edge to capture some of the drainage of the interior; the Kingoni, or Mto Ruvu; the Rufiji, whose many tributaries include the Ruaha, which has also cut back through the plateau edge, the Kilombero, the Luwegu, and the Mbarangandu; and, farther south, the Matandu, Mbemkuru, Lukuledi, and the Ruvuma, which forms the southern boundary of the country.

The Mara, Kagera, and other rivers flow into Lake Victoria, the Malagarasi into Lake Tanganyika, and the Songwe and Ruhuhu into Lake Nyasa. Thus, Tanganyika forms the divide separating the headwaters of the three greatest rivers of Africa—the Nile, Congo, and Zambezi—flowing, respectively, into the Mediterranean, the Atlantic, and the Indian oceans. Between these systems are two great dischargeless basins. The Wembere, Manonga, and Sibiti, which have a seasonal flow only, drain much of the eastern side of the Central Plateau into Lake Eyasi, a salt lake in a subsidiary trough of the eastern Great Rift Valley, where Lakes Natron and Manyara are also salt. In the southwest the Rungwa River drains into another dischargeless salt lake, Rukwa. About half of the great lakes Tanganyika and Victoria lie within the Tanganyikan borders, but Lake Nyasa is wholly outside the country.

Climate.—The climate is characterized by alternating wet and dry seasons. The small seasonal variation in temperature is much exceeded by the diurnal range, which is greatest over the Central Plateau (as much as 20° C [36° F]) and least in coastal and lake regions (about 8° C [14.4° F]). Afternoon temperatures above 32° C (89.6° F) may prevail during the hottest months, from October to May. The equatorial climate is modified by decreased temperatures over the high plateaus and mountains, giving a range of climates from the hot, humid coastal belt to regions of snow and ice on Kilimanjaro.

The rainy season is from November or December to April or May, when humidities may be uncomfortably high near the coast, with drought for the rest of the year. But the Kenya pattern of long and short rains prevails in northeastern Tanganyika, where the main rains are during March–May and the short rains in November–December. Rainfall is broadly associated with seasonal oscillations of the convergence zone between the southeast monsoon, prevailing at the coast from April to October, and the northeast monsoon, prevailing from November to March.

Much of the territory is semiarid. Three regions can be distinguished: a coastal belt with a mean annual rainfall of 30–40 in. (760–1,000 mm.); a dry zone across the Masai Steppe and Eastern Plateau with less than 30 in.; and a wetter region with more than 30 in. in the west, but where seasonal drought is still severe. Increasingly heavy orographic rainfall occurs on windward slopes up to about 8,000 ft. (2,400 m.) (more than 70 in. [1,800 mm.] on Kilimanjaro). Very heavy rain falls at the northern end of Lake Nyasa (more than 100 in. [2,500 mm.]) and on the western shore of Lake Victoria (more than 80 in. [2,000 mm.]), where storms off the lake are concentrated by the prevailing winds. High evaporation rates limit the effectiveness of the rainfall.

Annual variations are considerable, and because of unreliable rainfall extensive areas are suited only to pastoral development.

Soil and Vegetation.—Soil types include those of the plateaus, with a horizon of concretionary ironstone (murram), deep red earths, volcanic soils, calcareous black soils, sandy soils, and alluvial soils. With some exceptions, soils are poor, deficient in plant nutrients, and easily leached. Accelerated soil erosion is a serious problem.

Miombo woodland (*Isoberlinia-Brachystegia*) occupies extensive areas in the west and south, and on open steppes grow scattered *Acacia* and myrrh tree (*Commiphora*). The grotesque baobab and symmetrical *Euphorbia candelabra* are also characteristic. Evergreen forest, in which *Podocarpus* (yellowwood), *Chlorophora excelsa* (muvule, muule, or iroko), and various cedars are of economic value, occurs on mountain slopes. Above the forest, bamboo sometimes occurs, and open woodland with giant heathers. Higher still, Afro-alpine tree communities include giant species of *Senecio* (groundsel) and *Lobelia*. Coconut palms and casuarinas abound near the coast, and extensive mangrove swamps south from the Rufiji Delta.

Animal Life.—The country is unsurpassed in the number and variety of its wildlife: elephant, giraffe, zebra, rhinoceros, buffalo, together with lion, leopard, cheeta, hyena, and other carnivores; numerous varieties of antelope, including impala, gazelle, and wildebeest; snakes, lizards, chameleons, crocodiles, and other reptiles; and more than 1,000 varieties of birds, including game birds. Chimpanzees are found in the mountains east of Lake Tanganyika. The sitatunga is found in certain swamps, and the gerenuk (another rare antelope) in the Arusha, Kilimanjaro, and Tanga regions.

The Serengeti National Park (4,450 sq.mi. [11,526 sq.km.]) is the most important reservation for plains game in Africa; it no longer includes the Ngorongoro caldera, which has become a separate conservation unit. Other, smaller national parks are those of Ngurdoto Crater, near Arusha, and Lake Manyara. Of the game reservations the largest are the Selous and Rungwa River; smaller ones include the whole areas of Mt. Kilimanjaro and Mt. Meru and others at Lake Natron and Katavi Swamp. The Rukwa Valley is one of the swarming areas of the locust, and game conservation there is administered by the Red Locust Research Association. Some of the game reservations have suffered from poaching and insufficient supervision, but at a conference on wildlife conservation held at Arusha in 1961 the Tanganyikan government announced a conservation policy more advanced than that adopted by any other African state. The tsetse fly (chiefly *Glossina morsitans*) infests wide areas, spreading sleeping sickness, but is being steadily eradicated by tree clearance and the establishment of permanent farms. Rainbow and brown trout were introduced into rivers of the northern and southern highlands, and Nile perch and tiger fish are caught in Lake Tanganyika. From November to March coastal big-game fishing is popular off Tanga and Mafia. (J. M. Ke.)

THE PEOPLE

Most of the peoples of Tanganyika are Bantu-speaking Negroes. They are among the northernmost Bantu peoples and were subjected to non-Bantu invasions from the north and to Arab contact from the coast. In the north and west are the southerly Interlacustrine Bantu kingdoms of Ha (numbering about 290,000), Haya (325,000), Fipa (80,000), and the smaller Kerewe, Zinza, Subi, and others. (*See* Bantu [Interlacustrine].) There are about 120 different tribes, and several of the large ones have been affected by the Interlacustrine invasions, notably the Sukuma south of Lake Victoria (more than 1,000,000) and the Nyamwezi in west Tanganyika (363,000). In the south, tribes related to those of Zambia and Malawi include the Nyakyusa (220,000), Bena (196,000), and the Ngoni and Yao (*qq.v.*). In the centre and north most Bantu tribes have been influenced culturally by the Masai (*q.v.*); they include the Gogo in the Dodoma Region (299,000), Hehe (*q.v.*; 220,000), Arusha (68,000), Kuria (65,000), and Nyaturu (185,000). In the northeast dwell several tribes closely related to the Kikuyu and Kamba of Kenya, including the Chagga around Mt. Kilimanjaro (more than 300,000), Sambaa (194,000), and Pare (126,000). In the east and southeast are tribes who, unlike the others, reckon descent matrilineally and have been affected by Arab influence; among these are the Makonde of the Mtwara Region (334,000), Luguru (202,000), Zaramo (183,000), Mwera (138,000), Zigua (134,000), Makua (123,000), and Rufiji (75,000).

The Great Rift Valley area in the north seems to have been a refuge for a cluster of non-Bantu tribes who were left unaffected by the many migrations of Bantu Negroes. The Hadzapi or Kindiga, hunters and gatherers reputedly related to the Bushmen of southern Africa, live on the Lake Eyasi flats; the pastoral Sandawe are said to be related to the Hottentots. These two groups are probably remnants of earlier populations. The Iraqw (134,000) and some small outlying groups are linguistically unrelated to any other known peoples; their language may have Cushitic affinities. There are also in the north about 50,000 Masai and 30,000 Barabaig, Nilo-Hamitic speakers, who migrated from Kenya, and who, as warrior tribes, greatly influenced their neighbours.

Apart from the Hadzapi, Sandawe, Masai, and Barabaig, all the tribes are agriculturalists, traditionally growing maize (corn) and millets. Save for the few kingdoms or chiefdoms (Ha, Fipa, Chagga), these peoples lack centralized political authority; their density of population and political cohesion are low. Except on the coast and in some older inland towns, where Islam predominates, the population is generally animist, although Christianity through missionary activity has made some headway. Arab contact on the coast dates from about the 5th century A.D.; many small Arab city-states were founded on trade in slaves, ivory, and gold. The slave trade, later based on Zanzibar, had devastating effects throughout central and southern Tanganyika, much of which remains thinly populated. (J. F. M. M.)

Population.—The population is unevenly distributed. One-quarter of the African peoples dwell around the shores of Lake Victoria; other large concentrations are in the highlands (especially Mt. Kilimanjaro) and the coastal towns. In 1963 the estimated population totaled 9,683,700, including 90,500 Indians and Pakistanis, 27,600 Arabs, and 22,700 Europeans. The Indians and Pakistanis are chiefly artisans and traders and nearly all live in the towns. The Arabs live mostly in the coastal areas, with a few in the Kigoma, Tabora, Dodoma, and Singida regions. Most of the Europeans are government officials, missionaries, or employees on plantations or in industrial or trading concerns; only a few hundred are permanent settlers. The Swahili language is the lingua franca. The principal townships apart from Dar es Salaam (pop. [1957] 128,742) are Tanga (38,053), Mwanza (19,877), Tabora (15,361), Morogoro (14,507), Moshi (13,726), and Dodoma (13,435).

Tanganyika: Regions and Regional Capitals

Region	Land Area (sq.mi.)	Population (1957 census)	Capital
Arusha	31,900	407,483	Arusha
Coast	13,050	537,012	Dar es Salaam
Dodoma	16,350	514,425	Dodoma
Iringa	21,750	489,291	Iringa
Kigoma	17,100	382,505	Kigoma
Kilimanjaro	5,150	473,859	Moshi
Mara	11,000	347,587	Musoma
Mbeya	23,650	540,978	Mbeya
Morogoro	28,400	547,558	Morogoro
Mtwara	32,050	747,232	Mtwara
Mwanza	13,300	836,341	Mwanza
Ruvuma	23,550	267,033	Songea
Shinyanga	19,850	660,441	Shinyanga
Singida	18,850	372,587	Singida
Tabora	59,650	567,518	Tabora
Tanga	10,850	579,382	Tanga
West Lake	15,350	514,431	Bukoba
Total	361,800	8,788,466*	

*Total population includes 2,853 in transit.

In the 1960s the African population was increasing at an average rate of nearly 2% annually, though in some of the highland tribes the increase was more rapid. The Asian communities had a higher rate of natural increase (about 2.5%). The growth of the European population was largely determined by net immigration. (J. P. Mt.; J. M. Ke.)

HISTORY

Early History and Exploration.—Most of the known history of Tanganyika before the 19th century concerns the coastal area, although the interior has a number of important prehistoric sites (*see* the East African subsections of *History* and *Archaeology* in AFRICA). Trading contacts between Arabia and the East African coastline existed by the 1st century A.D., and there are indications of connections with India. The coastal trading centres were mainly Arab settlements, and relations between the Arabs and their African neighbours appear to have been fairly friendly. After the arrival of the Portuguese in the late 15th century, the position of the Arabs was gradually undermined, but the Portuguese made little attempt to penetrate into the interior. They lost their foothold north of the Ruvuma (Rovuma) River early in the 18th century as a result of an alliance between the coastal Arabs and the ruler of Muscat (on the Arabian Peninsula). This link remained extremely tenuous, however, until French interest in the slave trade from Kilwa revived the trade in 1776 and also aroused the sultan of Muscat's interest in the economic possibilities of the East African coast, where a new Omani governor was appointed at Kilwa. For some time most of the slaves came from the Kilwa hinterland, and until the 19th century such contacts as existed between the coast and the interior were due mainly to African caravans from the interior.

In their constant search for slaves, Arab traders began to penetrate farther into the interior, more particularly in the southeast toward Lake Nyasa. Farther north two Khojas (merchants) from Surat, in India, followed the tribal trade routes to reach the country of the Nyamwezi about 1825. Along this route ivory appears to have been as great an attraction as slaves, and after the transfer of his capital from Muscat to Zanzibar about 1830, Sultan Seyyid Said himself gave every encouragement to the Arabs to pursue these trading possibilities. From the Nyamwezi country the Arabs pressed on to Lake Tanganyika in the early 1840s and in the same decade reached the kingdom of Buganda, to the north of Lake Victoria. Tabora (or Kazé, as it was then called) and Ujiji, on Lake Tanganyika, became important trading centres, and a number of Arabs made their homes there. They did not annex these territories but occasionally ejected hostile chieftains. Manwa Sera, *mtemi* (ruler) of the Nyamwezi, was deposed in the late 1850s. Mirambo, an African chief who built for himself a temporary empire to the west of Tabora in the 1860s and 1870s, effectively blocked the Arab trade routes when they refused to pay him tribute. His empire was purely a personal one, however, and collapsed on his death in 1884.

The first Europeans to show an interest in Tanganyika in the 19th century were missionaries of the Church Missionary Society, J. L. Krapf and J. Rebmann, who in the late 1840s discovered Mt. Kilimanjaro. It was a fellow missionary, J. Erhardt, whose famous "slug" map (showing, on Arab information, a vast, shapeless, inland lake) helped stimulate the interest of the British explorers Richard (later Sir Richard) Burton and John Hanning Speke. They traveled from Bagamoyo to Lake Tanganyika in 1857–58, and Speke also saw Lake Victoria. This expedition was followed by Speke's second journey in 1860, in the company of J. A. Grant, to justify the former's claim that the Nile rose in Lake Victoria. These primarily geographical explorations were followed by the activities of David Livingstone, who in 1866 set out on his last journey for Lake Nyasa. Livingstone's object was to expose the horrors of the slave trade and, by opening up legitimate trade with the interior, to destroy the slave trade at its roots. Livingstone's journey led to the later expeditions of H. M. Stanley and V. L. Cameron, both of whom went in search of the famous explorer of whom there had been no news for some years. Stanley traveled as a newspaper reporter, while Cameron also had the additional objective of exploration. As these travelers followed the slave-trade routes, considerable areas of country remained unknown to Arabs and Europeans alike. Spurred on by Livingstone's work and example, a number of missionary societies began to take an interest in East Africa after 1860.

German Ascendancy.—The main threat to the Arab slave trade, however, came from a different quarter: Leopold II of Belgium formed the International Association for the Exploration and Civilization of Africa in 1876 to open up tropical Africa. After the failure of his East African expeditions he turned to the west coast. It was then left to Germany with its newly awakened interest in colonial expansion to open up the country to European influences. The first agent of German imperialism was Karl Peters, who, with Count Joachim Pfeil and Karl Juhlke, evaded the sultan of Zanzibar late in 1884 to land on the mainland. He made a number of "contracts" in the Usambara area by which several chiefs were said to have surrendered their territory to him. Peters' activities were confirmed by Bismarck, a convert to imperialism, who had paved the way for German intervention in Africa during the Berlin Conference (1884–85). By the Anglo-German Agreement of 1886 the sultan of Zanzibar's vaguely substantiated claims to dominion on the mainland were limited to a ten-mile-wide coastal strip, and Britain and Germany divided the hinterland between them as spheres of influence, Germany taking the region later known as Tanganyika. Following the example of the British to the north, the Germans obtained a lease of the coastal strip from the sultan in 1888, but their tactlessness, and fear of commercial competition, led to the outbreak of a Muslim rising in August 1888. The rebellion was put down only after the intervention of the Imperial German government and with the assistance of the British Navy.

Recognizing the administrative inability of the German East Africa Company, which had hitherto ruled the country, the German government declared a protectorate over its sphere of influence in 1891 and over the coastal strip where the company had bought out the sultan's rights. Germany was anxious to exploit the resources of its new dependency. Between 1894 and 1906 about £4,500,000 was voted by the German government to assist in the development of East Africa, while numerous private financiers invested substantial sums in the country. Lack of communications at first restricted development to the coastal area. The introduction of sisal from Florida in 1892 marked the beginning of the territory's most valuable industry, which was encouraged by the development of a railway from the new capital of Dar es Salaam to Lake Tanganyika. In 1896 work began on the construction of a railway running northwestward from Tanga to Moshi, which it reached in 1912. This successfully encouraged the pioneer coffee-growing activities on the slopes of Mt. Kilimanjaro. Wild rubber tapped by Africans, together with plantation-grown rubber, helped swell the country's economy. The government also supplied good-quality cotton seed free to African growers and also sold it cheaply to European planters. In 1902 the Amami Biological and Agricultural Institute was opened in the Usambara Mountains to investigate some of the scientific aspects of increased production. The administration also tried to make good the lack of clerks and minor craftsmen by encouraging the development of schools, an activity in which various missionary societies were already engaged.

The enforcement of German overlordship was strongly resisted, but control was established by the beginning of the 20th century. Almost at once, came a reaction to German methods of administration with the outbreak of the Maji Maji rising in 1905. This was a revolt against the manner in which the Germans ignored local traditions, employed forced labour, and compelled people to grow cotton. Although there was little organization behind it, the rising spread over a considerable portion of southeastern Tanganyika and was not finally suppressed until 1907. It led to a reappraisal of German policy in East Africa. The Imperial government had attempted to protect African land rights in 1895 but had failed in its objective in the Kilimanjaro area. Similarly, labour legislation of a liberal character had not been properly implemented. The German government set up a separate Colonial Department in 1907, and more money was invested in East Africa. A more liberal form of administration rapidly replaced the semimilitary system that had previously existed.

World War I put an end to all German experiments. Blockaded by the British Navy, the country could neither export produce nor get help from Germany. The British advance into German territory continued steadily from 1916 until the whole country was

eventually occupied. The effects of the war upon Germany's achievements in East Africa were disastrous; the administration and economy were completely disrupted. In these circumstances, the Africans reverted to their old social systems and their old form of subsistence farming. Under the Treaty of Versailles (1919) the U.K. received a League of Nations mandate to administer the territory except for Ruanda-Urundi, which came under Belgian administration, and the Kionga triangle, which went to Portugal.

British Mandate.—Sir Horace Byatt, administrator of the captured territory and, from 1920 to 1924, first British governor and commander in chief of Tanganyika (as it was then renamed), proposed a period of recuperation before new development schemes were set on foot. The handful of Europeans in the country resented his tardiness, but Byatt sheltered behind the fact that the terms of the League of Nations mandate under which Britain was to administer the former German East African dependency were not implemented until 1922. The first important result of the mandate was the Land Ordinance (1923), which ensured that African land rights were secure.

Sir Donald Cameron, who was governor from 1925 to 1931, infused new vigour into the country. He reorganized the system of native administration by the Native Authority Ordinance (1926) and the Native Courts Ordinance (1929). His object was to build up local government on the basis of traditional authorities, an aim that he pursued with doctrinaire enthusiasm and success. He attempted to silence the criticisms by Europeans that had been leveled against his predecessor by urging the creation of a Legislative Council in 1926 with a reasonable number of nonofficial members, both European and Asian. After an investigation of Tanganyika's labour problems, a Labour Department was also set up in 1926. In his campaign to develop the country's economy, Cameron won a victory over opposition from Kenya by gaining the British government's approval for an extension of the Central Railway Line from Tabora to Mwanza (1928). His attitude toward European settlers was dependent upon their potential contribution to the country's economy. He was, therefore, surprised by the British government's reluctance to permit settlement in Tanganyika. Over the issue of closer union with the other East African territories, he stood in even stronger opposition to the British government, since he regarded the suggestion as being detrimental to the growth of his own system of native administration. His opposition played an important part in defeating the proposal. The economic depression after 1929 resulted in the curtailment of many of Cameron's development proposals; the Labour Department was abolished, and money for education was restricted. In the 1930s, too, Tanganyika was hampered by fears that it might be handed back to Germany in response to Hitler's demands for overseas possessions.

At the outbreak of World War II Tanganyika's main task was to make itself as independent as possible of imported goods. Inevitably the retrenchment evident in the 1930s became still more severe, and, while prices for primary products soared, the value of money depreciated proportionately. Tanganyika's main objective after the war was to ensure that its program for economic recovery and development should go ahead. The continuing demand for primary produce strengthened the country's financial position. The chief item in the country's development program was a scheme to devote 3,000,000 ac. (1,200,000 ha.) of land to the production of peanuts by the most modern agricultural methods. The scheme, which was to be financed by the British government, was to cost £25,000,000, and, in addition, a further £4,500,000 would be required for the construction of a railway in southern Tanganyika. It failed because of the lack of adequate preliminary investigations and was subsequently carried out on a greatly reduced scale.

Constitutional Progress and Independence.—Constitutionally, the most important immediate postwar development was the British government's decision to place Tanganyika under UN trusteeship in 1947. Under the terms of the trusteeship agreement, Britain was called upon to develop the political life of the territory, which, however, only gradually began to take shape in the 1950s with the growth of the Tanganyika African National Union (TANU). The first two African members had been nominated to the Legislative Council in December 1945. This number was subsequently increased to four, while there were three Asian nonofficial members and four Europeans. An official majority was retained. There was an important advance in 1955 when the three races were given parity of representation on the unofficial side of the council with ten nominated members each, and for a time it seemed as if this basis would persist. The first elections to the unofficial side of the council, however, enabled TANU to show its strength, for even among the European and Asian candidates only those supported by TANU were elected. A constitutional committee in 1959 unanimously recommended that after the elections, in 1960, a large majority of the members of both sides of the council should be Africans and that elected members should form the basis of the government. The approval of the colonial secretary was obtained for these proposals in December 1959, and in September 1960 a predominantly TANU government took office. The emergence of this party and its triumph over the political apathy of the people were largely due to the leadership of Julius Nyerere, and his success gave the lie to those who had insisted that British concentration on the development of local government on a traditional basis had acted as a barrier to the emergence of a unified state. Tanganyika became independent in December 1961, with Nyerere as its first prime minister. The November 1962 elections for the country's first president under a republican constitution resulted in an overwhelming victory for Nyerere. On Dec. 9, 1962, the anniversary of its independence, Tanganyika became a republic within the Commonwealth.

In April 1964 the presidents of Tanganyika and Zanzibar signed an act of union of their two countries, which would thenceforth be known as the United Republic of Tanzania. Nyerere was appointed the first president of the new republic, with Abeid Amani Karume of Zanzibar as first vice-president. (K. I.)

ADMINISTRATION AND SOCIAL CONDITIONS

Constitution and Government.—Under the 1962 constitution Tanganyika became a republic within the Commonwealth of Nations, with an executive president, who is both head of state and head of government. The first president was elected directly under a system of universal suffrage, but in the future the election of the president was to be directly linked with the election of the members of parliament (the National Assembly), to ensure that the presidency would, as far as possible, be closely associated with the majority party in the National Assembly. A presidential election is held whenever parliament is dissolved. The assembly consists of 71 elected members and up to 10 members nominated by the president, who has no power to legislate without recourse to parliament. His assent, however, is required before any bill passed by the assembly can become law. If he withholds his assent and the bill is again passed by the assembly by a two-thirds majority, he is required by law to give his assent within 21 days, unless, before that time, he has dissolved the assembly. To assist him in carrying out his duties he is empowered to appoint a vice-president (who must be an elected member of the assembly), who is leader of government business in parliament. The vice-president and other ministers comprise the Cabinet, which is presided over by the president. He is not bound to act on the Cabinet's advice. After the union of Tanganyika with Zanzibar in April 1964, the new state was to be governed in accordance with the constitution of Tanganyika until a new constitution had been approved.

Local Government.—For the purposes of local government the country is divided into 17 regions (*see* table); these are subdivided into districts under district commissioners responsible to the regional commissioners. Local government is concerned with such matters as primary education, agriculture, marketing, forestry, and health.

Common Services.—Tanganyika is a member of the East African Common Services Organization (which replaced the East African High Commission), an interterritorial body administering certain services common to Kenya, Uganda, and Tanganyika (tariffs and customs, railways, harbour services, postal services and telecom-

munications, civil aviation, commercial and industrial co-ordination). Policy decisions (which must be unanimous) of the common services authority are reached in concert by the heads of government of the three countries; the various services are dealt with similarly by a triumvirate of ministers.

Living Conditions.—The cost of living increased by more than 40% in the decade following 1950. The customs and outlook of the people vary greatly in different areas, but generally standards of living are higher in zones where cash-crop cultivation supplements subsistence agriculture. The reported number of persons in paid employment in the mid-1960s was more than 400,000. The average per capita income in the same period was about £24 ($67) annually. There were more than 40 trade unions, with an estimated membership of more than 65,000. The increase in the population of urban areas after World War II caused a housing shortage, and in the early 1950s the government began a low-cost housing program.

Health and Welfare.—In the mid-1960s the government, missions, and local authorities maintained more than 800 maternity and child-health clinics, about 250 hospitals with 15,000 beds, about 20 major leprosariums, and more than 1,000 out-patient dispensaries. Expenditure on medical and allied services (including public-health measures by local authorities) amounted to more than £2,000,000 annually. The most prevalent diseases included malaria, sleeping sickness, smallpox, and leprosy.

Justice.—There is a High Court with final jurisdiction in both criminal and civil cases, composed of a chief justice appointed by the president and of lesser judges appointed by the president after consultation with the chief justice. Subordinate courts have limited jurisdiction in each district.

Education.—Schooling in Tanganyika was pioneered by the missions and supplemented by the central government and local authorities. The schools (primary, middle, and secondary) were developed on a racial basis but were integrated from 1961. By the mid-1960s about 500,000 children (or about half of those of school age) entered primary schools, and a large proportion of these left after the four-year primary course, while after the four-year middle-school course about 20,000 pupils proceeded to secondary schools. Swahili is the language of instruction in primary schools, though English is introduced in the third year and becomes the medium in middle schools. School broadcasting is used to supplement formal work and to provide a common cultural background. In the mid-1960s it was planned to increase the primary-school course to eight years, abolish the middle schools, and expand secondary education. Trade schools include large residential establishments (mainly for building and engineering) at Ifunda and Moshi; there is a well-equipped technical institute at Dar es Salaam. The University College, also at Dar es Salaam, has faculties of law and arts. In the mid-1960s many Tanganyikans were pursuing studies abroad; these included students attending Makerere University College, Kampala, Uganda, and the University College, Nairobi, Kenya. About 2,000 teachers were being trained annually at more than 20 training centres, including a college at Changombe. Current expenditure by the government on education rose from just under £300,000 in 1947 to more than £4,000,000 in the mid-1960s (more than one-tenth of total government current expenditure).

Defense.—Tanganyika's forces comprise the two battalions of the Tanganyika Rifles, which before independence were attached to the King's African Rifles. (J. P. Mt.; X.)

THE ECONOMY

Agriculture dominates the economy, accounting for about 80% of the exports by value. There is an important livestock industry, and mining production is considerable.

Agriculture and Forestry.—Sisal is the most important product. Plants of *Agave sisalana* were introduced from Florida in 1892 by the German agronomist Richard Hindorff. By 1914 production reached 20,000 tons of fibre annually and by the 1960s usually exceeded 200,000 tons, making Tanganyika the leading world producer. Coffee, tea, pyrethrum, and tobacco are grown in the upland areas, grains (maize [corn], millet, sorghum, rice), cotton, cocoa, pulses, peanuts, and oilseeds in the higher rainfall areas of the plains, and coconuts, cassava, cashew nuts, and citrus fruits in the coastal belt. The principal subsistence crops are sugar, maize, millet, cassava, bananas, and pulses. In the 1960s irrigation projects were developed in central Tanganyika and in the Rufiji and Mbarali river basins. In the drier plains pastoral tribes herd cattle, sheep, and goats, but tsetse infestation hinders the livestock industry. Areas of true forest are limited, but valuable timbers include muninga, camphor, muvule, and African mahoganies and blackwood. The co-operative movement is widespread, and the value of produce handled is about 25% of the total value of exports.

Mining.—Diamonds account for about two-thirds by value of Tanganyika's mineral output. Diamondiferous deposits around Shinyanga, south of Lake Victoria, were discovered in 1940 by J. T. Williamson, and by the mid-1960s the mine at Mwadui (in which the government acquired a half share) was producing more than 600,000 k. annually. Gold (about £1,200,000 annually), salt, silver, mica, and tin are among mining products. Immense deposits of iron ore and coal exist in the southwest near Lake Nyasa but are difficult of access.

Power and Industries.—Electric power generation and supply are carried out by a company with an exclusive franchise. Thermal power stations exist at Dar es Salaam, Lindi, Mtwara, Iringa, Mbeya, Dodoma, Tabora, Kigoma, Morogoro, Mwanza, Arusha, and Moshi. Major hydroelectric installations include one at Pangani Falls (17.5 megawatts) and the Hale Station on the Pangani River (21.5 megawatts), completed in 1964, with a transmission line to serve Dar es Salaam and the adjacent area along the Central Railway Line.

Manufacturing industries are mostly connected with agricultural and forest products (food, flour and sisal processing, oil, and sawmilling) but include the production of textiles, clothing, cans, paint, and soap for local consumption, as well as soft drinks, beer, cigarettes, aluminum ware, bricks and tiles, coir and sisal matting, leather, cement, industrial gases, and shoes. By the early 1960s manufacturing contributed about 4% of the gross domestic product.

Trade.—About 50% of Tanganyika's gross domestic product derives from agriculture, and the economy is therefore sensitive to changes in the prices of export products, notably sisal, which fell from £100 per ton in 1950 to under £70 a ton in the early 1960s. Total exports (including those to Kenya and Uganda and re-exports) averaged about £55,000,000 annually, about one-third being taken by the U.K. and about one-fifth by other Commonwealth countries. Exports in order of value were sisal, cotton, coffee, diamonds, oilseeds and nuts, hides and skins, canned meat and meat extracts, gold, tea, beans and pulses, and tobacco. Imports (mainly textiles, vehicles and transport equipment, petroleum fuels, iron and steel, machinery, chemicals, and foodstuffs) included substantial quantities of local produce from Kenya and Uganda and totaled about £40,000,000 annually, more than one-third coming from the U.K.

Finance.—Most of Tanganyika's recurrent revenue is derived from direct sources (income and personal taxes, local government rates, company taxes, etc.) and indirect (customs and excise duties and income from government properties). Among the main items of expenditure are education, communications, power, public works, police, agriculture, health, and labour. The currency is identical with that of Kenya, being based on the East African shilling, which is legal tender for any amount and is subdivided into 100 cents. Other coins are of 1, 5, 10, and 50 cents; notes are issued of EAShs. 5, 10, 20, 100, and 1,000. EAShs. 20 = EA£1, the latter being at par with the U.K. pound sterling. Commercial banks with branches in Tanganyika include three British, two Indian, one South African, one Dutch, and one Turkish. The East African Currency Board, with headquarters at Nairobi, controls the issue and supply of currency. An ambitious development plan covering the period up to 1970, and involving the expenditure of £246,000,000, was announced in 1964.

Transport and Communications.—The railways form part of the East African Railways and Harbours System. There are

three main lines, all of metre gauge: the central line from Dar es Salaam to Kigoma (779 mi. [1,254 km.]), with branches to Mwanza on Lake Victoria from Tabora (236 mi. [380 km.]), to Mpanda from Kaliua (131 mi. [211 km.]), and to Mikumi southward from Kilosa (44 mi. [71 km.]); Tanga to Arusha (272 mi. [438 km.]), with a link eastward from Kahe to Voi (94 mi. [151 km.]) on the Mombasa-Nairobi line in Kenya and a link southward from Mnyusi to Ruvu (117 mi. [188 km.]) on the central line; and Mtwara to Masasi (154 mi. [248 km.]). Diesel traction is used.

Tanganyika has about 700 mi. (1,100 km.) of roads in towns and settlements, 3,800 mi. (6,100 km.) of territorial roads, 5,800 mi. (9,300 km.) of local roads, 10,800 mi. (17,400 km.) of district roads (usually metaled), and 8,500 mi. (13,700 km.) of village roads. The East African Railways and Harbours Administration operates passenger and freight road services, and there are widespread local bus and truck services. More than 20 shipping lines operate services to Tanganyika, some calling at Tanga and Mtwara, as well as at Dar es Salaam. There are passenger and cargo services on Lakes Victoria, Tanganyika, and Nyasa. East African Airways Corporation operates internal scheduled flights and services to the U.K., South Africa, India, and Pakistan; other airlines maintain links with adjoining territories. There is the usual range of postal services but no house-to-house delivery by postmen, the box or private bag system being used. Scheduled air services are widely used for mail transport. About half the telephone exchanges are automatic; landline and radiotelegraph services are provided, and an overseas radiotelephone service. Radio broadcasting is controlled by a corporation directed by a government-nominated board and emanates from Dar es Salaam, with a national program in Swahili and a second program in English and Asian languages.

See also references under "Tanganyika" in the Index.

(J. P. Mt.; J. M. Ke.)

Bibliography.—R. Coupland, *East Africa and Its Invaders From the Earliest Times to the Death of Seyyid Said in 1856* (1938), *The Exploitation of East Africa, 1856–1890* (1939); R. Oliver and G. Mathew (eds.), *History of East Africa*, vol. i (1963); K. Ingham, *A History of East Africa* (1962); C. Leubuscher, *Tanganyika Territory* (1944); B. T. G. Chidzero, *Tanganyika and International Trusteeship* (1961); J. C. Taylor, *The Political Development of Tanganyika* (1963); Department of Lands and Surveys, *Atlas of Tanganyika* (1956); C. Gillman, "A Vegetation-Types Map of Tanganyika Territory," Geogr. Rev., vol. xxxix (1948); Ministry of Mines and Commerce, *Summary of the Geology of Tanganyika* (1956); Bailey Willis, *East African Plateaus and Rift Valleys* (1936), D. W. Malcolm, *Sukumaland* (1953); M. F. Hill, *The Story of the Tanganyika Railways* (1960); E. W. Russell (ed.), *The Natural Resources of East Africa* (1962); J. F. R. Hill and J. P. Moffett, *Tanganyika: a Review of Its Resources and Their Development* (1955); International Bank for Reconstruction and Development, *The Economic Development of Tanganyika* (1961); Central Office of Information, *Tanganyika: the Making of a Nation* (1961). For a full bibliography *see* J. P. Moffett (ed.), *Handbook of Tanganyika*, 2nd ed. (1958). The considerable German literature includes the collected work *Deutsch-Ostafrika*, 10 vol. (1909). Current history and statistics are summarized annually in *Britannica Book of the Year*.

TANGANYIKA, LAKE, in East Africa, is the longest freshwater lake in the world (420 mi.) and the second deepest (4,708 ft.) but is comparatively narrow, varying in width from 30 to 45 mi. It covers about 12,700 sq.mi. and forms the boundary between Tanganyika and the Republic of the Congo. It occupies the southern end of the western arm of the Great Rift Valley, and for most of its length the land rises steeply from its shores. Its waters are drinkable but slightly brackish.

The lake was first visited in 1858 by R. F. Burton and J. H. Speke, who reached Ujiji, the place where in 1871 H. M. Stanley "found" David Livingstone. The southern half was first circumnavigated by V. L. Cameron in 1874, when the Lukuga outlet was discovered, and the whole lake was surveyed by Stanley in 1876.

The lake is fed by a number of rivers but is not the centre of an extensive drainage area. The largest river is the Malagarasi, which rises to the north of Kigoma. Others are the Russisi (Ruzizi), which carries the overflow from Lake Kivu, and the Kalambo, on which is situated one of the highest waterfalls in the world (over 700 ft.).

The lake is situated on the dividing line between the floral regions of east and west Africa, and the oil-palm, so characteristic of the latter, is found on its shores. Rice is grown in suitable areas on the lake littoral. An agreement was being negotiated in 1953 with the Belgian authorities to stabilize the level of the lake. Fishing for *dagaa*, a small fish not unlike whitebait, is carried on by African fishermen in dug-out canoes at night by the light of flares, and dried fish valued at over £125,000 was exported to the Congo in 1951.

Hippopotami and crocodiles abound in the waters of the lake, and in places the rare west African sharp-snouted crocodile, *crocodilus cataphractus*, occurs. The bird life is most varied, because of the lake's situation between the east and west African regions. The mountains north and south of Kigoma are the only places in Tanganyika where the chimpanzee is to be found.

There had been considerable traffic across the lake before the arrival of the European, but this had nearly all been from west to east, so that many of the numerous tribes now living on its eastern borders trace their origins to areas in the Congo. They are predominantly Bantu, with traces of Hamitic (Tusi) influence in the north and of Zulu (Angoni) in the south. The areas of both sides of the lake suffered heavily in the past from the depredations of slave traders.

The chief towns on the lake are Usumbura, Uvira, Albertville, Kigoma, Ujiji, Karema, Mtakuja, Kipili, Mpulungu, Kasanga and Moliro. Steamers ply on the lake, traversing its length in about four days. Communications across the lake are from Kigoma, the terminus of the railway line from Dar es Salaam, to Albertville, whence there is a rail link to the Lualaba river.

Bibliography.—Accounts of the discovery and early exploration of Lake Tanganyika are to be found in the works of such explorers as Sir R. F. Burton, David Livingstone, V. L. Cameron, H. M. Stanley, E. C. Hore and J. Thomson. *See* also H. Meyer, *Das Deutsche Kolonialreich* (1909–10); H. Schnee (ed.), *Deutsches Kolonial Lexikon* (1920); *Tanganyika Notes and Records* (issued twice yearly, Dar es Salaam, 1936 *et seq.*). (J. P. Mt.)

TANGERINE, a small thin-skinned variety of orange belonging to mandarin orange species (*Citrus reticulata*). The principal distinguishing characteristic of the tangerine in comparison with the rest of the mandarin species is deeper reddish-orange colour of the peel of the tangerine. *See* Orange.

TANGIER (locally Tanja) is a seaport of Morocco, on the Strait of Gibraltar, about 7 mi. E. of Cape Spartel, between two eminences at the northwest extremity of a spacious bay.

As seen from the sea, the town rises gradually in the form of an amphitheatre, with the citadel, the remainder of the English mole and York castle to the right; the commercial quarter in the central valley; and the road to Tetuan running along the beach to the left. The main thoroughfare leads from Bab el Marsa (Gate of the Port) to the Bab el Suk (Gate of the Market place) known to the English as Catherine Gate. European quarters have also been built to the north of Tangier (the Marchan quarter), to the west (on the "Mountain") and to the south (quarter of the beach and of Suani). The harbour, formed by the Bay of Tangier, is good in all weather except during a strong east wind, but large vessels have to anchor about a mile out. A deep-water harbour was begun in 1925. The population of the former international zone was 79,886 in 1940, of which 50,250 were Moslems, 7,102 Jews, 15,486 Spanish, 1,642 French, 1,389 English and 4,017 others; in 1954 the total was estimated as 175,000, including 15,000 Moroccan Jews and 42,000 Europeans (35,000 Spaniards, 8,000 French, 1,000 others). Area of the former international zone of Tangier was 135 sq.mi. Pop. of the city (1960) is 141,714.

History.—The ancient Tingis stood very near the modern site of Tangier. Under Augustus it became a free city of the Roman empire; and when Otho placed the western half of Mauretania under a procurator, he called it Mauretania Tingitana. It was held successively by Vandals, Byzantines and Arabs; and when Mulai Idris passed from Tlemçen to Fez in 788, Tangier was "the oldest and most beautiful city" of the Maghrib. After many futile attempts the Portuguese obtained possession of it in 1471; it passed with Portugal to Spain in 1580, returning to the independent Portuguese in 1656.

In 1662, as part of the dowry of Catherine of Braganza on her

marriage to Charles II, it came into the possession of the English, who defended it against Mulai Ismail in 1680 but in 1684 decided, on account of expense, to abandon it to the Moors. The sultans made Tangier the seat of foreign representation, to keep Fez free from extraneous influences; and capitulatory rights were granted to western nations (1630–1880) as well as certain powers of sanitary control. (A. BE.; X.)

A special regime for Tangier, growing out of the interest of Great Britain in the Strait of Gibraltar, was envisaged in the Moroccan arrangement of 1904 between Britain and France on the one hand and France and Spain on the other; and in the Algeciras act (1906) Tangier's international character was fully recognized. The principle was reiterated in the protectorate convention of 1912 and in the Franco-Spanish convention of the same year.

In 1913 Great Britain, France and Spain began negotiations to create a "Tangier zone" with an international regime. When, however, World War I broke out Spain withdrew, apparently in the hope of acquiring Tangier for itself.

The Statute of 1923.—It was not until 1923 that Tangier's status was at last regulated. Though France's position in Morocco had become stronger, Britain was still determined on an international form of government; and Spain supported this demand.

Negotiations culminated in the statute of Tangier, which remained in force down to 1940. It was signed by Britain and France in 1923 and by Spain in 1924. All states signatory to the Algeciras act (except the Central Powers and Russia) were immediately invited to adhere to it; and all did so except Italy and the U.S. Its main provisions were as follows:

1. The Tangier zone was to be permanently neutralized, though troops proceeding to or from the French or Spanish zone of Morocco might cross the territory after notification to the administrator. A police force, not exceeding 250 Moroccans, under a Belgian officer with French and Spanish subordinates, was to be at the disposal of the administrator.

2. Native administration and control over all Moslem institutions was to be entrusted to a mendub representing the sultan.

3. A legislative assembly of 26 members was to be instituted consisting of 17 Europeans (four French, four Spanish, three British, two Italian, one American, one Belgian, one Dutch and one Portuguese, nominated by their respective consuls) and 6 Moslem and 3 Jewish subjects of the sultan (nominated by the mendub). The mendub was to preside over the assembly.

4. A committee of control, composed of the consuls of powers signatory to the Algeciras act (former enemy countries excluded), was given a power of veto over the enactments of the legislature and charged with ensuring observance of the statute.

5. An administrator-general, head of the executive, and two assistant administrators for health and finance, were to be appointed for an initial period of six years by the sultan at the instance of the committee of control; these were to be of French, Spanish and British nationality. Appointments after this were to be assigned by the assembly to nationals of signatories of the Algeciras act.

6. An international mixed court was to administer justice in all matters affecting foreigners. It was to consist of one French, one Spanish, and two British judges, with a Frenchman and a Spaniard as public prosecutors for criminal and civil matters.

7. A port commission was also established, consisting of one member of the sultan's government, one of the legislative assembly, one of the committee of control.

France secured preponderance in the administration, helped by the fact that as the international zone was a concession from the Moroccan government, itself under French protection; Spain was allotted a secondary role. The interests of Tangier itself were in the main disregarded. An Italian request for participation and Spanish dissatisfaction with the working of the whole system led to the revision of the statute in 1928. Italy acquired a share in the administration roughly equivalent to Britain's. Spain secured a more effective suppression of arms smuggling and of political disturbances. But Spain, still aspiring to incorporate Tangier in its zone, subsequently demanded joint control with France over the native Moslem and Jewish population; for a rotation of Frenchmen and Spaniards as chief administrators; and for posts in the customs. These demands were not favourably received.

Spanish Occupation, 1940–45.—After the collapse of France and Italy's entry into World War II, Spanish troops occupied Tangier on June 14, 1940; and in Nov. 1940 the Spanish authorities incorporated it into the Spanish zone of Morocco.

Early in 1941 Britain concluded an interim agreement with Spain, which, without prejudice to British and third-party rights under the international statute, guaranteed free entrance into and departure from Tangier to British subjects and the continued functioning of British enterprises in the zone. Spain also formally undertook to keep the zone demilitarized.

Postwar Regime, 1945–56.—In August 1945 a conference on Tangier was held at Paris by representatives of Britain, France, the United States and the U.S.S.R. It was found impossible to draw up a new international statute since the U.S.S.R. refused to co-operate with General Franco's Spain; Britain, France and the United States, on the other hand, held that Spanish interests could not be ignored. Finally a compromise was reached: if Spain withdrew its troops at the request of the four powers, a provisional regime, based on the statute of 1923, as modified in 1928, was to be set up for six months; Spain was to be allowed to participate in this regime by having a Spanish consul on the control committee and a Spaniard sitting in the legislative assembly; the United States and the U.S.S.R. were to be invited to collaborate; and after six months had elapsed a new conference was to be held, to which all powers signatory to the Algeciras act (excluding former enemy countries) and the U.S.S.R. were to be invited, to draw up a permanent international statute for Tangier. Spain accepted the decision of the Paris conference; and the provisional regime came into force in Oct. 1945.

Under the Paris agreement it had been decided that the principal interested powers, France, Spain, and Britain, should no longer hold any of the most important posts in the administration; these were to be left to smaller powers. Britain renounced its former preponderant share in the administration of justice. Spain retained only its membership of the control committee and the legislative assembly. France renounced the right to have a Frenchman as administrator general but managed to insert clauses in the agreement which stipulated not only that a Frenchman should remain as second in command of the police but also that a new post, that of assistant administrator of Moroccan affairs, should be created for a Frenchman. Moreover, before anyone else had arrived, the French succeeded in recruiting enough officials to staff the administration and all the necessary police; and also, as under the statute of 1923, they continued to nominate the Moroccan members of the legislative assembly, while Moroccan interests in the control committee were represented only by a French consul general.

France thus retained a preponderant share in the administration, though the influence of the United States, preserving its capitulations there as elsewhere in Morocco, made itself increasingly felt. After six months no new conference on Tangier was called, and the provisional administration was maintained.

A new factor in the situation after World War II was the growing dissatisfaction of the Moroccans with their subordinate position. Though they had nine representatives in the legislative assembly they had no representative on the control commission, which could veto the decisions of the assembly. Italy was readmitted to the committee of control in March 1948, after signing a peace treaty with the Allies, on the invitation of France and Britain with the concurrence of all the signatories to the Algeciras act (except those excluded in 1919) and the U.S.S.R.

On March 30, 1952, the 40th anniversary of the Fez treaty which had imposed the French and Spanish protectorates on Morocco, a violent riot of the Moroccan Nationalist party took the international authorities by surprise. In June it was decided to create a motorized public corps with a Spaniard as its commander and a Frenchman as his deputy.

International Zone Abolished.—As a consequence of the recognition of Morocco's independence and sovereignty by France (March 2, 1956) and by Spain (April 7, 1956), the statute of Tangier was abolished on Oct. 29, 1956, with the signature of an

eight-power declaration and a protocol, the last-named signed also by Ahmed Balafrej, the Moroccan foreign minister.

Haj Mohammed el-Tazi, the mendub, resigned on July 8, 1956, and a governor was appointed by the king of Morocco.

On Aug. 29, 1957, a royal charter was granted to Tangier confirming the maintenance of freedom of exchanges and of foreign trade and introducing a moderate fiscal system. *See* also Morocco. (S. E. Ce.)

TANGO, a ballroom dance of Spanish American origin, popular in the Americas and Europe since about 1915. The tango is danced to music similar in rhythm to that of the Cuban habanera, though faster, and played in 4–4 time. The basic step patterns take two measures, counted slow, slow, quick, quick, quick, hold. (A. Mu.)

TANGUAY, EVA (1878–1947), U.S. vaudeville and musical comedy star, was born in Marbleton, Que., and went to the U.S. with her parents at an early age. She obtained her first footlight role at the age of eight in *Little Lord Fauntleroy* and later acted in variety, stock troupes and musical comedies. By the age of 20 she was regarded as an accomplished comedienne. At the turn of the century, Miss Tanguay shocked audiences with her daring costume—or lack of costume—and her risqué songs, and by 1910 she was an established and popular star of the theatre. Her earnings were phenomenal but she lost her fortune during the stock market crash of 1929. Her health and vision soon failed, bringing her career to an abrupt end. She spent the last two decades of her life a virtual recluse, bedridden and mostly forgotten. She died in Hollywood, Calif., on Jan. 11, 1947.

TANIMBAR ISLANDS (Timorlaut Islands), a group of islands in the Malay archipelago, southwest of the Aru Islands, between 6° 20′ and 8° 30′ S., and 130° 40′ and 132° 5′ E. Politically the islands belong to Indonesia. They form part of the outermost alpine arc which includes Timor and the Ewab Islands and recurves westward through Ceram and Buru. The group lies outside the zone of recent volcanic activity. There are about 30 islands, the largest being Jamdena (70 mi. long by 28 mi. wide). Thickly wooded hills extend along the east coast; the west coast is lower, where, opposite the island of Sera, a gulf extends far inland. In places there are precipitous cliffs, and a narrow foreshore is fringed with coconut palms and mangrove. There are many small islands off the western coast of Jamdena; off the northeastern coast is Larat (20 mi. long by 5 mi. wide), with a rocky coast and high cliffs, and with thick vegetation along the foreshore; and off the southwestern coast, Selaru, which is 32 mi. long, rather flat, and with much grassland.

There are extensive swamps on many of the islands and fever is prevalent during the southeast monsoon. The soil supports maize, rice (in small stretches), coconut and sago palms, yams, plantains, mangoes and papayas. There are practically no rivers and there is a lack of fresh water. The slight depth of soil prevents trees from attaining great height.

The vegetation includes sterculias and fig trees, leguminous trees and shrubs, myrtles, the scarlet hibiscus, pandanus, palms, a species of breadfruit tree, rare and lovely orchids of many kinds, crotons, dracaenas and many varieties of ferns. The animal life shows Australian affinities and comprises wild buffalo, wild pig, the cuscus, flying fox, many snakes, lizards and frogs. The birds include the scarlet lory, honey eater, ground thrush and oriole. Butterflies are numerous and include *Papilio aberram,* and among the beetles is a gorgeous golden-coloured buprestid (*Cyphogastra splendens*).

The islands, with two divisions, Larat and Saumlaki, are included in the province of the Moluccas. The capital of Jamdena is Saumlaki, at the southwestern end of the island, and of Larat, the town of the same name, on a narrow strait opposite Jamdena.

The isles were first visited by Capt. Owen Stanley, in 1839, and became known to Banda traders in 1877. In 1878 H. O. Forbes, with his wife, visited the islands and stayed there three months, being warned before going by the Dutch authorities of the hostility and savagery of the inhabitants; he appears to have been the first European who stayed any length of time. After his visit the islands were opened up by the Dutch, and the islanders proved responsive to outside civilizing influences.

The population (1930 census) was 31,847. The people are undoubtedly a mixed group. Some of the men are tall, others are short and thickset, the complexion of nearly all being a rich chocolate brown. The basic stock is Oceanic negroid of the Papuan variety, but there has been much mixture with mongoloid strains, first of the Indonesian type and later, especially around the coast, with Malays.

There are many pagans and a few Moslems. Christianity made some headway, and under Dutch rule cannibalism, piracy, slavery, head-hunting and intertribal fighting were suppressed. The men sometimes dye their hair a golden colour. Tattooing is known. Traces of totemism and of an age-group system survive. Their houses of wood and *atap,* sometimes built in closely packed settlements, generally on piles and entered from underneath, have carved pillars and show much artistry of construction. The men are skillful boatbuilders, use the spear and bow and arrow (with which they are adept in shooting fish), hunt, search for trepang and turtle shell, are agriculturists, keep pigs and work in iron, copper and gold. The women cultivate the fields after the men have prepared the ground, weave and dye sarongs and plait.

(E. E. L.; B. A. L. C.; T. Her.)

TANJORE, a city, with *tehsil,* and an eastern coastal district in Madras state, India. The city is on the right bank of the river Cauvery, and is an important junction on the Southern railway, 218 mi. S. of Madras. Pop. (1951): city 100,680; taluka (421 sq.mi.) 381,984. It was the last capital of the ancient dynasty of the Cholas (*q.v.*). It was the scene of the earliest labours of Protestant missionaries in India. The modern history of Tanjore begins with its conquest by the Marathas in 1674. The British first came into contact with Tanjore by an abortive expedition in 1749 to restore a deposed raja. The Marathas practically held Tanjore until 1799, when the district was ceded to the East India company. The raja retained only the capital and a small tract of country round. He died in 1833 and was succeeded by his son Sivaji, on whose death in 1855 without an heir the house became extinct. Among buildings may be mentioned the palace within the fort, and the richly ornamented temples of Rajrajeswara (*c.* A.D. 1000) and Subrahmaniya (built in the compound of the former).

The city is famous for its silk brocade, jewelry, carpets, inlaid copperwork, modeling in pith, etc.

Tanjore District has an area of 3,740 sq.mi. It is irrigated by an elaborate system of dams, cuts and canals and is exceedingly productive. The delta of the Cauvery occupies the flat northern part, which is highly cultivated, dotted over with groves of coconut trees and densely populated. The staple crop is rice, grown on about five-sevenths of the cultivated area, and rice milling has also developed. Soap and mats are made and fishing carried on. The chief seaport is Negapattinam (57,854). Other important places are Kumbakonam (91,643), also a former Chola capital, a pilgrimage place, and a manufacturing centre for metalwork, silk, cotton and sugar; and Mayuram (43,436), with an ancient temple.

The 1951 district population was 2,982,670.

TANK. A self-propelled combat vehicle embodying the military characteristics of fire power, mobility and crew protection. Tanks may be used as part of larger infantry units, as a separate striking force or for reconnaissance.

PREDECESSORS

The modern military tank is a direct outgrowth of World War I, but the idea underlying its development—protected, mobile striking power—had been used throughout the history of warfare. Many ancient peoples made use of war chariots to carry an attack to an enemy with speed and with some degree of protection. The armoured elephants of Hannibal and Kublai Khan were variations of the same idea. In the 15th century Leonardo da Vinci envisioned his covered chariots as a means of destroying the artillery just then being introduced to warfare. From time to time similar suggestions were advanced and occasionally devices of this kind were used in war. However, there is no evidence that such great

BY COURTESY OF (TOP LEFT, CENTRE LEFT) IMPERIAL WAR MUSEUM, (TOP RIGHT) NATIONAL ARCHIVES, (CENTRE RIGHT) INFORMATIONS AERONAUTIQUES, (BOTTOM LEFT) U.S. ARMY, (BOTTOM RIGHT) CHRYSLER CORP.

TANKS OF VARIOUS COUNTRIES

Top left: British Mark IV tank going through barbed wire at Cambrai, France, during World War I
Top right: French Renault tank disabled during fighting near Chaudon, France, in World War I
Centre left: British Centurion II tank, post-World War II
Centre right: French AMX-13 tank being loaded aboard transport plane, post-World War II
Bottom left: U.S. M-75 personnel carrier, post-World War II
Bottom right: U.S. M-48 tank, post-World War II

TANK

1

3

4

5

6

BY COURTESY OF (5) G. B. JARRETT. PHOTOGRAPHS, (1, 4) SOVFOTO, (2, 3, 6) BRITISH COMBINE FROM PUBLIX

TANKS OF WORLD WAR II

1. Soviet medium tank, T-34
2. Loading a U.S. "General Grant" tank with a fresh supply of ammunition after an engagement with German panzer units in the Libyan desert
3. British tank commander observing the field from a Mark IV cruiser tank during action in Libya
4. Soviet heavy tank, Stalin III
5. German "Panther" tank armed with high velocity 75-mm. gun
6. U.S. M-3 light tank moving forward to contact enemy forces in the Libyan desert, 1942

military leaders as Condé, Turenne, Frederick the Great or Napoleon made any effort to revive the idea.

The invention of the "footed wheel" by R. L. Edgeworth in 1770 led to Thomas German's invention in 1801 of the jointed track. Neither of these was of any significance until the steam engine was applied to vehicles. However, Gottlieb Daimler's patent in 1886 of the high-speed internal-combustion engine opened the way for the first really practical self-propelled fighting vehicles. While interest in track-laying vehicles and improvements in their design continued in Britain and the United States during this period, the perfection of a reasonably efficient internal-combustion engine stimulated experiments in both these countries and in Europe, with various types of armoured cars. F. R. Simms of Britain is credited with building in 1898 the first motor vehicle designed for combat purposes; it was a light four-wheeled motorcycle with an armoured shield and mounting a Maxim machine gun.

In the same year a U.S. army officer, R. P. Davidson, designed a three-wheeled automobile similarly armed and armoured. Davidson experimented also with steam-driven combat cars, as did the Russian engineer Lutski. Davidson's cars were followed in rapid succession by the Charron car in France in 1901 and the Simms war cars in Great Britain in 1902. The first turreted armoured cars were built by the firm of Charron, Girardot et Voight in France in 1904 and purchased by Russia. These were followed by the Austro-Daimler cars in Austria-Hungary, which were the first four-wheel-drive automobiles. Several other British, French, Swiss and German cars appeared. Among them was the Armstrongs-Whitworth car designed by Walter G. Wilson, who later was prominent in design and production of British tanks during World War I. During this period the track-laying vehicle, as developed by Benjamin Holt, C. L. Best and others, was making great strides agriculturally in the United States.

During the years immediately prior to World War I, tacticians of the major powers which were to engage in that struggle did not seem to envision the part such a combat vehicle could play. Lieut. Fleicz of the French army and Capt. (later Gen.) J. F. C. Fuller (*q.v.*) of the British army, as late as 1913 and 1914, respectively, predicted the machine gun would bring about static trench warfare, and suggested greater emphasis on means for overcoming it. But it was not until the soundness of their ideas was clearly demonstrated along the Aisne in 1914 that serious attention was given to armoured, trench-spanning vehicles.

WORLD WAR I TANKS

The French and Belgians were quick to improvise armoured cars for fighting delaying actions against German advance cavalry. The tactics were informal but aggressive and usually whatever the individual car commander decided. When the war stabilized on the western front in 1914, Winston Churchill, then first lord of the admiralty, authorized the stationing of a large naval air force at Calais with temporary bases as far inland as possible. Churchill quickly saw the natural affinity between aircraft and armoured vehicles. As a result, the royal naval air service under Capt. Murray Sueter improvised armoured car squadrons for protecting advanced air bases and to rescue pilots who crashed or were forced to bail out in enemy territory.

When the war had settled down to a practically continuous front, it was because artillery and the machine gun dominated the battlefield. (*See* TACTICS and WORLD WAR I.) There were no flanks. With position defense and frontal attacks the only basic tactics available to either side, defense became superior to offense. Even when salients developed which permitted a form of envelopment, the attacker's own flanks became exposed and attacks eventually ground to a standstill, with casualties reaching staggering proportions. The tremendous artillery fire preparation frequently destroyed the attacker's ability to sustain his own advance because his artillery could not be displaced forward over the ground it had helped to make impassable.

British Tanks.—In Oct. 1914, Col. (later Gen.) Sir Ernest Swinton, having seen some American Holt caterpillar tractors used as heavy artillery tractors behind the lines, was struck with the desirability of arming and armouring such vehicles as an antidote to the German machine gun. He suggested the idea to the authorities but Churchill was the only member of the committee of imperial defense who was in favour of exploring it. At the end of Feb. 1915 the war office officially abandoned the Swinton proposals but Churchill immediately created a Landships committee under the chairmanship of Eustace (later Sir Eustace) Tennyson-d'Eyncourt, a prominent naval designer.

As a result of admiralty rather than war office sponsorship, certain naval terms came into use and persisted. Parts of armoured vehicles are still known by such naval terms as hull, turret, deck, sponson, barbette, superstructure, bow, hatch and compartment.

Various experimental machines were built, ending with a large vehicle having overhead tracks, mainly the product of W. G. Wilson and Sir William Tritton. Colonel Swinton learned of these and was authorized by the commander of the British forces in France to collaborate through the war office. In June 1915 the admiralty and war office joined hands. The abortive Dardanelles landing forced Churchill's resignation in July but the committee and Capt. Murray (later Admiral Sir Murray) Sueter of the Royal Navy kept interest alive in spite of army apathy. In September, primary responsibility was turned over to the war office but the Landships committee continued to function with little change except in name.

The word "tank" was coined by Swinton. Originally, for secrecy, the hull and chassis of an experimental vehicle had been made in different shops. The hull was referred to as a water carrier for Mesopotamia and was called "that tank thing" by the workmen. The chassis was described as a demonstrational chassis for the royal marines. Swinton later recalled this and it led to his suggesting the adoption of the word "tank," which since has come into use in almost every language in the world.

The British decided to use the new tank corps to bolster a failing offensive on the Somme in spite of Swinton's arguments that it would be better to wait until more tanks were available. Out of 49 tanks, only 11 carried out the attack on Sept. 15, 1916. From then on, tanks were used in driblets and with little effect and often under unsuitable conditions because of military and political pressure. Yet their use raised Allied and lowered German morale. Not until the battle of Cambrai in Nov. 1917 were tanks used in mass. From then on they contributed greatly to reducing the previous heavy Allied casualties. (*See* WORLD WAR I.)

Production was increased largely through the organizing genius of Lieut. Albert (later Lieut. Col. Sir Albert) Stern. After the United States entered the war, Stern negotiated an Anglo-American treaty for building the components of Mark VIII tanks in both countries with assembly to take place in a factory furnished by the French.

The first British model, designated the Mark I, was produced in two types, a "male" weighing 31 tons and mounting two 6-pounders and four machine guns, and a "female" weighing 30 tons, mounting six machine guns. They were 26 ft. 5 in. long, 13 ft. 9 in. wide and 8 ft. 1 in. high, powered with a 105 h.p. Daimler water-cooled engine. Maximum speed was 3.7 m.p.h., with a fuel distance of 12 mi. They could traverse obstacles 4½ ft. high, span trenches 11½ ft. wide and climb a 22° slope. They used the rigid-type suspension on rollers, with steel tracks. Armour thickness varied from 0.2 in. to 0.4 in. A total of 150 Mark I tanks was built, and they were used in nine engagements, the last on Nov. 1, 1917. As early as April 9, 1917, at the battle of Arras, they were reinforced with the improved Mark II, and thereafter the British tank underwent constant development, the last major war model being designated the Mark V. In the last, specific improvements included an increase in speed to 4.6 m.p.h., an increase in maximum armour thickness to 0.47 in., and the substitution of a Ricardo 150-h.p. water-cooled engine, plus a better planetary-type transmission. In addition, the Mark V had twice the fuel distance possessed by the Mark I. A few Mark V Star tanks were used in combat by both British and U.S. units. These were simply the Mark V with a section built into the hull to lengthen it for greater trench crossing ability as well as to carry machine gun teams into combat. The Mark VIII was an improved version of this ve-

hicle.

During World War I the British also built a Medium Mark A or Whippet tank weighing 15.7 tons which carried a crew of only three men as compared with eight in the heavy tanks, and had greater speed and fuel distance than the heavy tanks. There later were a Medium B and a Medium C but these were available too late for combat.

French Tanks.—Though the British were the first to use tanks in battle, the French had actually preceded them in experimentation in this field. Much of this experimentation was civilian inspired. Instead of devices to defeat the German machine gun, effort initially was directed toward overcoming German barbed wire. Then, Col. (later Gen.) Eugène Estienne, like Colonel Swinton, saw an American Holt tractor and conceived an armoured fighting vehicle. In collaboration with Eugène Brillié, a Schneider engineer who had been experimenting with small Holt tractors, the Schneider C. A. or *char d'assaut* was successfully tested and placed in production.

Jules-Louis Breton, under-secretary of state for inventions, had been responsible for most of the earlier experiments. He succeeded in obtaining funds for proceeding with a larger vehicle designed by Col. Émile Rimailho of the St. Chamond firm and known as the *char St. Chamond*. Production was begun without consultation between the two groups. It was not until June 1916 that British developments in the new field were discovered. Premature British use of tanks on the Somme caused the French to rush production and the first units of *l'artillerie d'assaut* began equipping in Dec. 1916.

The first French tank attack took place in April 1917 and ended in failure. Committees bickered, and a change in high command shifted emphasis to artillery production. Nevertheless, Estienne succeeded in getting the Renault firm to design and build a light two-man tank. Another shift in high command resulted in a decision to cease building all vehicles but the Renault model. The first of these was shipped to the United States. Production was slow and it was not until the French were frightened by the German Chemin des Dames offensive that industrial production was pushed.

Although the French made many of the same mistakes as the British and there was considerable confusion over the role these new weapons of war would play, a great deal was accomplished. Jealousies existed but over-all there was less vacillation, with cancellations and reinstatements of orders, than occurred in England. The French had ordered 7,800 Renaults, of which 3,177 had been built by the end of the war, and 300 Char 2C, a 60-ton tank, of which only a few were completed. About 400 each of the Schneider and St. Chamond tanks were also built.

The Schneider and the St. Chamond were similar as to armament (one 75-mm. gun and two machine guns) except that the cannon was mounted in the right of the hull in the former and in the overhanging nose in the latter; speed was 6 m.p.h.; suspension was of the coil spring-bogie type; and tracks were of steel plate. The Schneider was lighter (14.9 tons compared with 25.3 tons) and smaller in over-all dimensions, the difference being accounted for by the bulkier power-transmission system used in the St. Chamond. This consisted of a four-cylinder 90 h.p. water-cooled engine with a dynamo and two electric motors, plus storage cells. In addition, the Schneider used a six-man crew, while the St. Chamond required nine men.

The Renault was armed with a 37-mm. gun or a machine gun, had a speed of 6 m.p.h., weighed only 7.4 tons, and was smaller in over-all dimensions than either of the other French tanks.

German Tanks.—Though the Germans had opportunities before and during World War I to take the lead in development of the tank, they failed to do so. It was not until the capture at Cambrai (Nov. 20, 1917) of a number of British tanks that the Germans possessed any serviceable armoured vehicles of this type. The captured tanks were assembled and repaired at Charleroi, where German experts studied them, but the idea of copying them was soon abandoned. Instead, reliance was placed on a model of German design which already was well along. A few of these were delivered in time to participate in the battle of St. Quentin on March 21, 1918, and in five additional actions before the Armistice in November. The Germans also used a few captured specimens in various actions, but they never had tanks in numbers at all comparable with those of the Allies.

The German model made during 1918 was built by the Daimler company. It was designed by Joseph Vollmer and known as the A-7V. It weighed 33 tons, was powered by two Mercedes-Daimler four-cylinder engines of 150 h.p. each, carried a crew of 18 men and mounted one 6-pounder gun and six machine guns. Perhaps its most interesting feature was the thickness of its armour, which varied from 0.59 in. to 1.18 in., and was designed to be proof against 37-mm. gunfire.

U.S. Tanks.—Before the United States entered the war, both the Holt Tractor company and the C. L. Best company had built some experimental tanks. As early as 1915, Cleve Shaffer proposed armouring and arming Fageol orchard tractors. At about the same time, E. M. Wheelock of Pioneer Tractor company and Norman Leeds of Automatic Machine company offered tank designs to the British and A. C. Hamilton of Oakland Motors built an experimental tank for the British at their request.

Committees created by Gen. John J. Pershing decided to obtain British and French tanks but interest lagged until the battle of Cambrai and the signing of the Anglo-American tank treaty to build the Mark VIII tank. With tremendous optimism, high production rates were forecast. The Mark VIII program failed partly because of the failure of the American aviation program since the same engine was to be used in both. The light tank program was equally bad. Pershing was notified in Feb. 1918 that these tanks would start arriving in France in April, yet at that time only a handful of people were working on the project and a contract had not even been let. By June it was evident that no American-built tanks would be available and so British and French vehicles were obtained. In return for the former, U.S. troops using British tanks had to serve in the British sector. Meanwhile, a Holt gas-electric, a Holt steam tank, an Endicott and Johnson steam tank, a Pioneer skeleton, a Ford one-man and a Ford Mark I three-man tank were being tested.

Efforts were finally concentrated on the copy of the Renault known as the 6-ton tank, on the Mark VIII and on the Ford one-man tank; but the first 6-ton tanks did not arrive in France until after the Armistice. At that time, 23,405 tanks were on order. Of these, 2,950 were Mark VIII, 4,400 were 6-ton, 1,000 were Mark I and the remainder Ford one-man. The Mark VIII was a larger version of the British Mark V, the Mark I was an improved Renault mounting both a 37-mm. gun and a machine gun. The Ford one-man tank weighed three tons, carried a machine gun and was very cheap to produce since it used many of the standard Ford Model T passenger car components.

U.S. tank units saw action in both the British and U.S. sectors and were under the over-all command of Gen. S. D. Rockenbach. Tank training in the United States was in charge of Lieut. Col. (later General of the Army) Dwight D. Eisenhower.

TANKS BETWEEN WORLD WARS I AND II

The plan for 1919, had the war continued, required a frontal assault by heavy tanks to achieve a breakthrough, with local exploitation by light tanks. This assault was to be followed by pouring fast medium tanks through the breaks, co-ordinated with aircraft attacks. Here was the real tactical lesson of World War I, but the victors retained only the memory of the war as it was fought with tanks and infantry in slow moving waves advancing under cover of a moving artillery barrage.

Voices were first raised in England against the accepted post-war tactics, principally by Col. (later Gen.) J. F. C. Fuller, author of the 1919 plan. He fostered the all-tank army and visualized land warfare resembling naval warfare. He was joined by Capt. B. H. Liddell Hart, who added the concepts of "tank marines"—tank-borne infantry co-operating with aircraft—and air-borne troops. They were joined by British Col. (later Gen.) G. LeQ. Martel, French Col. (later Gen.) Charles de Gaulle and German Col. (later Gen.) Heinz Guderian. However, the costs of the war and the quantities of tanks left over demanded economy, which

influenced subsequent vehicle development and tactics as well.

Because track life was short and there were demands for higher speeds behind the lines, there was considerable interest in combination wheel and track tanks, several of which were designed by J. Walter Christie in the United States. But eventually, as rubber bogie tires were adopted and track life increased, these were abandoned. The increasing popularity of the civilian passenger car brought about other mechanical improvements, many of which were also translated into tank design. Speeds increased as did radius of operation or fuel distance, air-cooled engines came into use, better suspensions were devised, intercommunication systems for crew members were developed, and Britain, in 1929, introduced short-wave radio for communication between tanks. Armament, however, remained light and seldom exceeded a 37-mm. gun or a heavy machine gun.

United States.—The National Defense act of 1920 eliminated the wartime tank corps and merged tanks with infantry. This action was of great influence tactically, as was peacetime economy. Interest in British tactical developments resulted in permission for the cavalry to work with "combat cars," a term coined to circumvent the law. A considerable period of bickering over appropriations and tactics resulted. A few experimental vehicles were developed which eventually resulted in an excellent and reliable M-2A4 light tank used by both infantry and cavalry, and an M-2A1 medium tank utilizing many light tank components was built for the infantry. Interest also was shown in the commercial Christie tanks, which could operate as wheeled vehicles by removing the tracks, but they were not considered sufficiently rugged for adoption especially since track life had reached 1,000 miles or more.

The M-2A4 light tank weighed 14 tons, had a crew of four and carried a 37-mm. gun and three machine guns. Its speed was 35 m.p.h. and its fuel distance was 100 miles. The M-2A1 medium tank weighed 23 tons and mounted a 37-mm. gun and light machine guns. The M-2A4 was of riveted construction; the M-2A1 medium tank was welded.

The German attack on Poland in 1939 caused a complete redesign of the M-2A1 medium tank and led to a large order for light tanks. The differences between the infantry and cavalry finally were buried in the creation of an armoured force on July 10, 1940.

Great Britain.—The British royal tank corps, though its existence was in doubt for a period immediately after World War I, continued as a separate arm. A few hundred Vickers medium tanks were built, larger versions of the French Renault but embodying many improvements. Martel and Carden Lloyd tankettes, originally intended to utilize passenger car components, came into use. These vehicles were used in a series of developmental tactical organizations along the lines suggested by Fuller and Liddell Hart. Eventually there resulted four major British vehicle types: the tankette for machine gun and mortar transport, the light tank for reconnaissance, the cruiser tank for independent long range strikes and the infantry tank for close infantry co-operation.

The first group resulted in the famous Bren carrier, the second in the light Mark VI tanks—fast, lightly armoured two-man vehicles in the five-ton class. The cruiser tanks were in the 20-ton class with speeds up to 30 m.p.h. These vehicles were developments of the U.S. Christie tanks. The infantry tanks weighed about 26 tons and had thick armour. The Christie and the infantry types were armed with a 2-pounder gun and one or more machine guns.

With the beginning of the rearmament program in 1938, the royal tank corps was consolidated with British mechanized cavalry units as the royal armoured corps, and its expansion went forward concurrently with the design of additional new tanks.

Germany.—Germany was denied the use of tanks, aircraft and heavy artillery by the victors of World War I. Thus in later years the Germans had a distinct advantage in not being saddled with a great deal of obsolescent equipment as was the case with its former enemies. The 1919 plan was studied and modified in the light of the later writings of Fuller and Liddell Hart. Where the official British and U.S. doctrine was that tanks were to be used to exploit the advantage after an enemy line was broken, and the French doctrine was slow moving, crushing fire power, the Germans adopted the view that tanks were properly used to obtain the breakthrough itself and tactics were worked out with dummy vehicles. After the Versailles treaty was abrogated, tanks were developed and openly displayed. The Germans concentrated on light and medium tanks of good design mounting slightly heavier armament than that used in other countries. Intervention in the Spanish Civil War worked out the details of German tactics and caused some modifications in design.

France.—The French experimented with a few tanks, and some modernization of wartime Renault tanks was undertaken after the war, but it was not until April 1926 that a new program for light, medium and heavy tanks was undertaken. Progress was slow until the Germans began to rearm in 1934. Even then, communist agitation and nationalization of war industries prevented as much progress as was intended. By 1939, nevertheless, approximately 3,000 modern tanks had been built. The French made notable progress in armour production; besides making considerable use of cast armour, they pioneered in work with welded armour. They were successful in combining thick, well-shaped armour with light suspension systems, thus holding down over-all weights. The French were among the first to use controlled differential steering, which permitted constant power to the track sprockets even when turning.

The principal French tanks built during the rearmament period were the 7-ton Renault AMR Type ZT armed with a light or a heavy machine gun, the 14-ton FCM armed with a 37-mm. gun and one machine gun. The former had riveted armour, the latter welded. Even greater numbers of squat Hotchkiss and Renault 13-ton tanks with cast armour were built. These mounted a 37-mm. gun and one machine gun. In the medium class, the Somua or Char 35-S mounted a 47-mm. gun and one machine gun. The heavy class included the Char B1, weighing 34 tons. The B1 had a 47-mm. cannon and a machine gun in the turret, a 75-mm. gun in the front of the hull and another machine gun in the rear of the hull. It was generally agreed that the Somua was one of the best tanks in any country at the start of the war.

U.S.S.R.—The Russian automobile corps was formed late in 1914 and many foreign automobiles were armoured in Russia and used in combat even against German trenches. British and French troops sent to Russia after the Revolution were equipped with tanks. These were given to the White Russians when the troops were withdrawn and later became the first tanks in the soviet army. A few experimental vehicles were built which copied foreign designs. Later, foreign vehicles were imported, including British commercial tanks and tankettes and a U.S. Christie. With the start of the first five-year plan in 1928 the Russians began creating designs of their own, based on these imports.

Soviet intervention in the Spanish Civil War permitted field testing of designs, but more real experience was obtained in the war with the Japanese on the Manchurian frontier in 1938–39. The poor quality of Russian gasoline led to the adoption of diesel engines of German origin, and German torsion bar suspensions also were incorporated into Russian tanks. The Russo-Finnish War showed shortcomings in design although existing light and very heavy vehicle production could not be shut off immediately.

The Russians began World War II with a light T-26 tank copied from the British Vickers-Armstrongs 6-ton tank, small amphibious tanks, BT tanks developed from the U.S. Christie, several medium tank varieties and a few heavy tanks. The numbers were greater than those possessed by the Germans but the quality was not good and they were poorly handled. The T-26 tank existed in two models, one with a single turret mounting a 47-mm. gun. The other had twin turrets, each mounting a machine gun. The light amphibious tanks carried machine guns. The 11-ton BT tanks were armed with a 47-mm. gun. Both T-26 and BT varieties had been used in Spain. The T-28 and T-32 were medium and heavy types with multiple turrets mounting machine guns and short cannon either of 47-mm. or 76-mm. calibre.

Other Nations.—Most of the smaller nations purchased tanks from other countries, principally Great Britain, France, Sweden and Czechoslovakia. The first Swedish tanks were built from parts

for German LKII light tanks shipped to Sweden after World War I and assembled under the direction of Joseph Vollmer, who had designed them originally. Spain built a few tanks of local design and the Czechs produced tanks for their own use as well as for export. The Japanese for some time used tanks purchased in Britain and France. In 1928, locally designed light, medium and amphibious tanks began to appear, some used by the army and some by the navy. Japanese tanks had considerable interchangeable parts, were thinly armoured and used diesel engines.

Poland, Austria, Australia, Belgium, Hungary, Canada and India produced some vehicles of their own design and others under license from other countries. Italy copied the French Renault near the end of World War I and built a few heavy tanks. The light Fiat copies of the Renault continued in use after the war. Later, British Carden Lloyd tankettes were built under license and evolved into a smaller vehicle, the CV 3/35. Experience gained in Ethiopia and in Spain in 1935–36 led to the design of the Italian M-39 medium tanks, lightly armed and armoured. Both were in use at the beginning of World War II.

WORLD WAR II

The Spanish Civil War had demonstrated the deadliness of a concentrated attack sustained in depth on a narrow front. It was the method used by the Germans in attacking Poland in 1939. Tanks and dive bombers were given the primary role in achieving deep, fast-moving penetrations. Motorized infantry followed close behind. The campaign was a brief dress rehearsal for invasion of the Low Countries and France the following year. French and British tanks outnumbered the German tanks but were widely dispersed in accordance with the French principle of close infantry support. German tactics of punching in mass on narrow fronts made the most of the opportunity and the campaign was soon over.

From France, the war moved to North Africa. There the tactics were similar to the naval tactics originally envisioned by General Fuller. When Germany attacked Russia, its previous tactics were not as successful. The Germans used 2,400 tanks against the Russians' 24,000. Within three months the Russians had lost 17,500 tanks and the Germans 550, but the areas were great and the Russians traded land for time. The Germans had made an organizational change which weakened their thrusts. In 1939 there had been 300 tanks per armoured division. To attain what Hitler believed would be greater mechanized fire power, more armoured divisions were created with fewer tanks per division, but these units lacked punch.

The British and U.S. forces copied German organization and followed this change also. The Russians learned by experience. When they resumed the offensive at Stalingrad, the Russians had more tanks and aircraft and applied successfully the tactics the Germans had used in France. Differences in tactical concepts were reflected in the designs produced by different armies—the Germans particularly tending as the war went on to build more heavily armed and armoured monsters, sometimes at the expense of mobility and mechanical reliability. The days when strategic surprise with large masses of armour was possible for one or another opponent gave way to fierce encounters between forces more equally matched in armour. Numbers, skill and daring were still to prove decisive in many actions, but the tank design race became one essentially between projectile and armour.

Early in the war, tanks received a great deal of publicity. They were spectacular and many people in and out of military service became conditioned to their success. Later, in certain theatres where areas for maneuver were circumscribed, people jumped to the conclusion that tanks were obsolete. Yet right up to the end of the war the issues between the Germans and Russians were largely decided by tanks and self-propelled artillery. The tank served most usefully in World War II when employed in mass on suitable occasions by commanders who fully appreciated what it could and could not do. Again, it was most effective when used as a member of an integrated team including all the other ground arms and supported by suitable aircraft.

Germany.—After the Germans overran Czechoslovakia, they took over production of Czech TNH tanks, calling them 38 (t). German production was fairly well centralized and between 1939 and 1942 no overtime was worked. After the United States entered the war the first real attempt at economic mobilization was made, but it was not until 1944 that Albert Speer, then chief of production, was given power to forbid design changes which would interfere with production.

Hitler, either directly or indirectly, made many engineering decisions, sometimes on obscure details. In spite of the German totalitarian government, there was much extravagance in the number of vehicle types and in design complications, in addition to the tremendous use made of modified captured vehicles. Allied bombing caused some dislocation of production but, from 1940 to 1944, German tank production increased tenfold, then it fell, but only partly due to bombing. Other factors were oil shortage, use of slave labour, transportation difficulties and extravagance in man hours due to lack of knowledge of mass production methods.

Between 1939 and 1944, Germany produced 42,932 armoured vehicles. As the war forced the Germans more and more on the defensive, the trend was toward the armoured assault gun and huge tank destroyers rather than tanks because more of the former could be produced for the same effort. The Germans began the war with tanks mounting heavier armament than those of the Allies and during the war they constantly increased these calibres. Armour was also constantly thickened, it being a principle that the armour of a tank should be proof against the weapons it carried. Although many vehicle types were produced, the best of the German tanks probably was the Panther, which was developed after encountering the Russian T-34. The German vehicle was superior to the Russian except in mechanical reliability. The Panther had thick, well-sloped armour of welded construction. The main armament was a high velocity 75-mm. gun. This tank weighed 47 tons and had a speed of nearly 30 m.p.h. The Tiger, a heavier and somewhat slower tank, mounted an 88-mm. cannon. Late in the war, still heavier tanks mounting high velocity cannon up to 128 mm. were in use and even larger tank guns and new tanks were being tested.

United States.—After the German invasion of Poland, rearmament started in the U.S. but there was considerable confusion. The extent and need for arming were as vague as were the possible theatres of war should the United States become involved. Existing mobilization plans had to be scrapped, and lend lease added to the confusion. After Pearl Harbor on Dec. 7, 1941, the president set up production goals so great that U.S. industry was aghast but willing to try to meet them. In spite of the president's demands that heavy tanks be produced and although the heavy M-6 tank design was available, the project was canceled because the new armoured force did not want a heavy tank and preferred to use available shipping space to ship twice as many medium tanks.

There was a great deal of bad blood between army service forces and the civilian War Production board but, in spite of the conflict between them, production was prodigious. There were 287,425 combat vehicles built during the war, together with enough spare parts to have built half again as many more.

The original M-3 medium tank had a 75-mm. gun mounted on the right front of the hull with machine guns in the bow and a 37-mm. gun and machine guns in the turret. This tank was known as the General Lee. When it was fitted with a more squat British turret, it was called the General Grant. It was produced with riveted, cast and welded armour and with different types of engines because of production limitations. Many of its limitations were overcome in a design which followed.

The M-4 medium tank, known as the General Sherman, with added armour and the 75-mm. gun in the turret, was in production by April 1942, and it reached Africa in time for the battle of El Alamein. Undergoing continual development during the war with 76-mm. guns on a 105-mm. howitzer on some models, the M-4 became by far the most widely used Allied tank in the war. A total of 49,234 was produced. Its components featured simplicity and ruggedness. Of particular interest were the volute spring bogie-type suspension, the highly developed V-8 engine and the rubber-block, rubber-jointed tracks. A life of 3,000 mi. was not unusual

for these tracks, compared to about 600 mi. for the metal-jointed German tracks. Aside from the fact that the numbers of the Sherman tank supplied to Allied armies made possible the massing of superior tank forces in critical areas, its speed, maneuverability and capacity for sustained operation under severe conditions made the Sherman tank a vital factor in Allied victory.

Late in the war a compromise between the medium Sherman and the heavy M-6 was made in the 45-ton General Pershing. It mounted a 90-mm. gun, used torsion bar suspension and a torquematic stepless transmission. A few reached Europe before the war ended and some later also were landed on Okinawa. Light tanks went through a similar process of gun calibre increase and the application of automatic transmissions adapted from motorcar design. The culmination of these was the M-24 or Chaffee 23-ton light tank, named after Brig. Gen. A. R. Chaffee.

The navy and marine corps had tested Christie amphibious tanks a few years after World War I. This led to a new concept in amphibious landings which received another impetus when the Roebling rescue machine came to their attention in 1939. As the war went on, those vehicles were refined. Beginning with the marine corps landing on Betio (Tarawa), they were used more and more by the marine corps and eventually by the army. They were propelled in water by the motion of their tracks. Finally they were armoured and later turreted. Some were used as personnel carriers and some as accompanying artillery in landing operations. About 18,620 were built during the war.

Great Britain.—Immediately following the German invasion of Poland some members of the original World War I tank committee received permission to form a new committee at the ministry of supply. They produced a heavy gas-electric TOG tank, its initials representing "The Old Gang." The ministry meanwhile released for production a heavy Vauxhall tank. Disputes followed between Winston Churchill, then prime minister, and the committee with the result that the committee was dissolved and the Vauxhall tank became the 40-ton Churchill heavy tank.

Almost all existing tanks were lost at Dunkirk. To catch up, vehicles were released for production with little testing. By the time the "bugs" were worked out, the tanks were obsolete. British tanks had a poor reputation because of this and because of lack of production inspection and control. Parliament investigated but was unable to fix responsibility. Meanwhile, production continued. The British purchasing commission in the United States contracted to build tanks similar to the American medium M-3 designed after the invasion of Poland. Near the end of the war, British tank quality had improved and weapon calibres were more nearly the equal of the Germans.

Final models of the Churchill and Cromwell mounted 75-mm. guns. The faster and lighter Comets mounted 77-mm. guns. The Cromwell and Comet were in the cruiser class. In addition to combat tanks, the British devised several self-propelled guns and many special tanks such as the A.V.R.E. for throwing high explosive petards, "flails" for detonating land mines, carpet-laying tanks or "bobbins," flame throwers and others. About 24,000 tanks were produced, 3,600 were received from Canada and the United States furnished 25,600.

U.S.S.R.—The models laid down in 1938–39 and produced by 1940 revolutionized Soviet armour. The 26-ton T-34 medium was one of these. It was of the Christie type with wide (24 in.) tracks, had a low silhouette and a 30 m.p.h. speed provided by a good 500 h.p. liquid-cooled diesel V-12 engine. It was armed with a 76-mm. gun and later with an 85-mm. gun.

Another 1940 model was the KV heavy (named for Kliment Voroshilov, soviet minister of defense), which weighed 52 tons and mounted a 76-mm. gun. The KV was much more heavily armoured than the T-34, which accounted for its greater weight, since in other respects it was much like the T-34. With these early wartime models the Soviets came into their own as designers, producers and users of fighting tanks. Before the war's end another heavy tank, the Joseph Stalin, was to appear. It was a logical step from the T-34 and the KV, in the progression toward thicker armour and more fire power, with the 51-ton JS-2 model mounting a 122-mm. gun.

In addition to Russian production of about 100,000 vehicles, the United States furnished Russia 13,303 combat vehicles (tanks, self-propelled weapons and scout cars) under lend lease. The British furnished 4,260 and Canada 1,220 tanks in addition to huge numbers of Bren carriers from both. Figures for Russian tank losses in the war have never been released but are believed to have been in the neighbourhood of 100,000. Russian equipment was designed to do the job with the least possible expenditure of man-hours to produce, with simplicity and with absolutely no emphasis on finish or crew comfort.

TANKS AFTER WORLD WAR II

Satellite nations and those friendly to the U.S.S.R. were supplied with Stalin III heavy tanks. These were developments of the original KV but with sloping armour and a dome-shaped turret mounting a 122-mm. gun. T-34/85 tanks mounting an 85-mm. gun in place of the original 76-mm. gun also were furnished to these nations. Russian troops were rearmed with the T-54, similar to the T-34 but with a domed turret and a 100-mm. gun. A new amphibious tank also appeared in 1956, followed by air-borne and rocket projector tanks in 1957. In addition to an estimated 50,000 tanks, the Russians had at least as many self-propelled guns which had been brought to a high state of development by the end of World War II.

Great Britain dropped the basic subdivision between cruiser and infantry tanks in favour of a single-purpose medium tank, the Centurion. This tank was also furnished to Sweden, Switzerland and some members of the North Atlantic Treaty organization (NATO). A still larger Conqueror was also in use in Britain. The United States continued development of the General Pershing class with a cross-drive transmission. This culminated in the M-48A2 Patton mounting a 90-mm. gun. Rounded cast armour and a new air-cooled engine with direct fuel injection were features. The navy and marine corps also continued development of LVT's, or tracked landing vehicles. The United States gave many tanks to friendly nations.

France produced a light AMX 13-ton tank as well as a few heavier vehicles. The former were also sold to Sweden, Switzerland and other countries. Czechoslovakia began producing tanks, selling them to Egypt and other countries. Sweden continued to produce tanks for use at home and for export. Yugoslavia built tanks resembling the Russian T-54. Communist China, furnished tanks by Russia, began producing tank parts.

Tanks were used in combat in the Arab-Israeli wars of 1948 and 1956 and also in Korea in 1950. In the first case they were badly handled. In Korea, terrain dictated greater use as self-propelled artillery than as tanks.

TANK DESIGN

Great advances have been made in tank design since the World War I period. Principal among these improvements were track life, fire power and sighting equipment, suspensions, armour and armour shape, speed, reliability, standardized components and the development of vehicle "families."

It once was the custom to refer to tanks as light, medium or heavy. In the late 1950s it would be more accurate to classify them as light gun, medium gun or heavy gun tanks. The effectiveness of a tank is measured by its ability to bring fire to bear on an enemy and to make such fire mobile. A self-propelled gun is classed the same way, but it usually lacks complete armour protection. Tank guns are fitted with muzzle brakes which tend to shorten the recoil. Bore evacuators also are used to suck the combustion gases out of the muzzle, thus keeping the gases out of the crew compartment. High velocity guns erode after firing relatively few rounds. United States gun designers found a way to replace liners in the field with but a few simple tools, which was a great step forward.

Tank guns were augmented by gyroscopic gun or turret stabilizers which kept the gun on target regardless of vehicle motion; by automatic range finders; by turret "baskets" which revolved with the weapon and crew; and by vision cupolas and periscopes to offset blind spots.

If armour protection were absolute, the weight would be so great the vehicle could not move. Armour thickness, therefore, is arranged to be greatest at the most vulnerable points; or rounded or angular surfaces are used to deflect fire. As it is, armour weight represents over one-third of total weight. Thus design becomes a compromise between intended armament, vehicle weight and engine horsepower. Tanks were designed to be capable of operating in extremes of temperature, to be completely waterproof, fungusproof, with low noise level, high crew comfort and capable of being maintained even though the crew members wore mittens. Additional equipment included gyrocompasses, infrared sighting equipment and electronic navigational equipment which required adequate electrical generating and storing equipment.

It can, therefore, be seen that design of a tank is no simple engineering

task. The modern tank is expensive and complex and requires over 40,000 blueprints. Experiments indicate that new metals or alloys will permit an armour to be produced which will be 40% lighter in weight with no sacrifice in strength. Smaller power units also may lead to reaching the military goal of air transportability.

FUTURE OF THE TANK

With the advent of atomic missiles the concept of war underwent a complete change. Following a period of confused thinking, military concepts began to crystallize after 1950. Land warfare in the atomic age was believed most likely to take the form of thin screens of guerrillas or skirmishers with sudden concentrations and attacks and equally quick dispersions occurring constantly.

Quick exploitation before an enemy recovers his balance after an atomic blast was considered to be of great importance. Tanks were thought to be well suited for use in atomic warfare because of the protection offered their crews from heat and flash. Their mobility would enable them to disperse and to reconcentrate rapidly as changes took place in a tactical situation either by the threat of or the use of nuclear missiles. Infantry might be too slow and too vulnerable for such activity, and probably would be carried in armoured personnel carriers accompanying tank formations.

Co-ordinated with aircraft and airborne troops, armoured troops would continue to be forces to be reckoned with in land warfare, whether or not atomic missiles were used.

See also Index references under "Tank" in the Index volume.

BIBLIOGRAPHY.—R. J. Icks, *Tanks and Armored Vehicles* (1945); M. M. Postan, *British War Production* (1952); British House of Commons, Report by Select Committee, *War Time Tank Production* (1953); C. M. Green, H. C. Thomson and P. J. Roots, *The Ordnance Department: Planning Munitions for War* (1955); F. O. Miksche, *Atomic Weapons and Armies* (1955); Royal Armoured Corps Centre, *Tank Museum Guide* (1951–55); articles in the U.S. periodicals, *Armor* and *Ordnance*. (C. E. Br.; R. J. I.)

TANKER, a ship designed to carry liquid cargo in bulk. Its cargo is usually a petroleum product, either crude oil being carried from oil fields to refineries or gasoline being carried from refineries to distribution centres. The liquid is piped into the cargo space of the ship and is transported without use of barrels or other containers. Special tankers carry other liquids such as

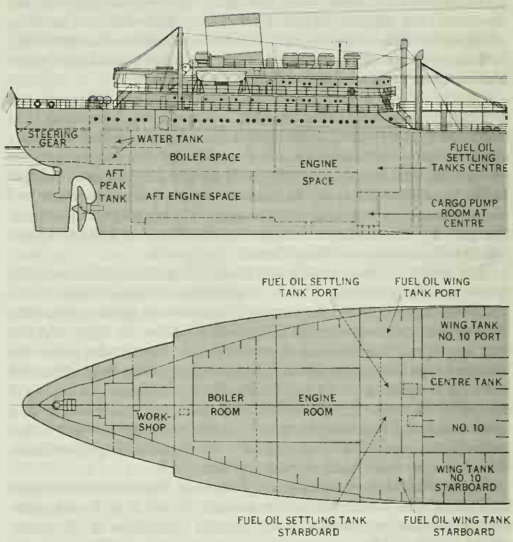
BY COURTESY OF NEWPORT NEWS SHIPBUILDING AND DRY DOCK CO.

FIG. 1.—SMALL TANKER (2,750-TON DEAD WEIGHT) USED ALONG THE EAST COAST OF THE U.S.

molasses, asphalt or even wine, but the references in this article are all to oil tankers. Tankers differ from general-cargo ships in that they normally carry a full load of cargo in one direction and return without cargo.

Tankers vary in size from small coastwise vessels about 200 ft. long, carrying from 1,500 to 2,000 tons dead weight, up to huge vessels which are the largest ships afloat, approaching 1,000 ft. in length and carrying over 100,000 tons dead weight. ("Dead weight" is the total weight that a ship can carry, consisting primarily of cargo but including fuel, water, crew, stores, etc.) Up to 1940 the upper limit in size was about 15,000 tons dead weight and about 500 ft. in length. In 1948, 28,000-ton "supertankers" appeared, followed in quick succession by tankers of 40,000, 60,000 and (in 1958) 100,000 tons. By the latter date there were more than 3,000 tankers of over 1,000 tons dead weight in the world, with a combined carrying load of more than 48,000,000 tons.

Experience with supertankers has shown that the direct cost of transporting oil goes down as the size of the tanker increases, apparently without limit; the only reason for not building larger tankers is the lack of suitable shore facilities for them. As deep-water loading and unloading facilities are created, larger and larger tankers will probably become economically feasible.



BY COURTESY OF NEWPORT NEWS SHIPBUILDING AND DRY DOCK CO.

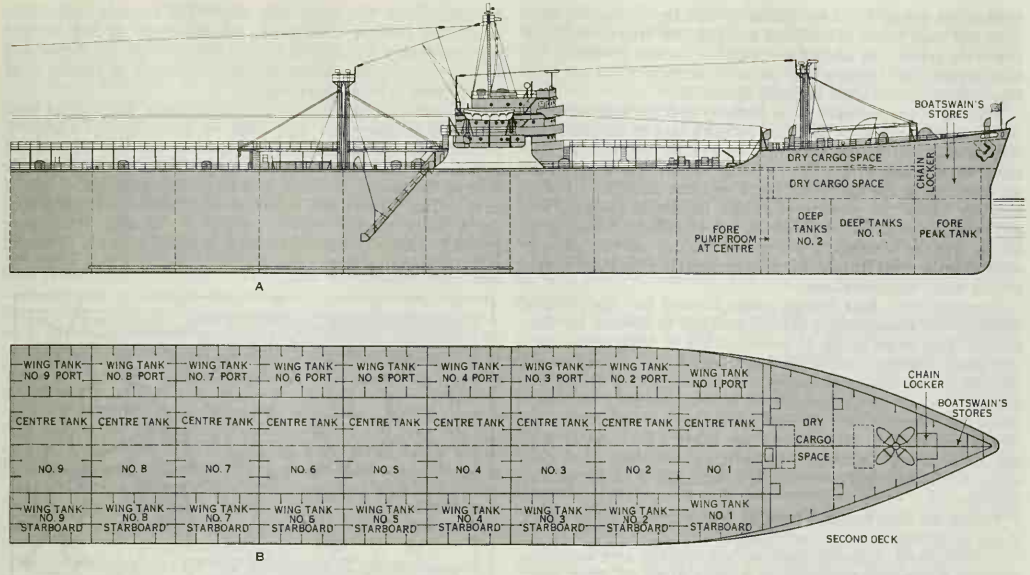
FIG. 2.—BELOW-

The speed of tankers has increased along with their size. In 1940 most tankers were powered with not over 5,000 shaft horsepower and made from 12 to 14 knots. In the 1950s most large tankers were given from 15,000 to 25,000 shaft horsepower and made from 15 to 18 knots.

Arrangement.—Tankers almost invariably have their propelling machinery located in the stern. (*See* fig. 2.) The cargo-oil space occupies about 60% of the ship's length; usually there is a general-cargo hold in the bow, along with ballast or fuel tanks. The cargo space is isolated from the forward and after portions by cofferdams, watertight compartments which are kept empty. A pump room is located just forward of the machinery space, with sometimes an additional pump room forward or amidships. There is a "forecastle" (additional topside deck at the bow) and a house amidships, containing the pilothouse, other navigational spaces and usually quarters for a few of the crew. A "poop" (additional topside deck at the stern) extends over the machinery space and, together with higher levels of deckhouses, quarters the rest of the crew. The cargo space is usually divided into two, three or four rows of cargo tanks by longitudinal bulkheads; these are further divided into individual tanks (from 35 to 40 ft. long in large vessels) by transverse bulkheads. Access to the tanks is through oiltight hatches on the deck. Tankers differ from most seagoing vessels in that they do not have a double bottom in the cargo spaces.

BY COURTESY OF NEWPORT NEWS SHIPBUILDING AND DRY DOCK CO

FIG. 3.—MEDIUM-SIZED OCEAN-GOING TANKER (27,000-TON DEAD WEIGHT) OPERATING BETWEEN VENEZUELA AND NEW YORK



DECK ARRANGEMENT OF A TANKER: (A) PROFILE; (B) PLAN VIEW SHOWING CARGO TANKS

Tanker crew accommodations are more livable than those of most vessels, because of the limited leave-time, which is often only about one day per passage. Single-person rooms, recreation lounges, and even swimming pools are fitted on the long-voyage tankers to enhance crew morale.

The density of the cargo carried by tankers varies from about 40 to over 60 lb. per cubic foot. A tanker with enough capacity for a full dead-weight cargo of light gasoline will have some empty tanks when carrying a full cargo of heavy crude oil.

Tankers are permitted by law to load more deeply, in proportion to their depth, than ordinary cargo ships. To be permitted this greater draft, they must provide special arrangements for fore-and-aft access for the crew in heavy weather, and for the quick freeing of the deck of seas which wash over it. The draft of tankers plying between the Arabian oil fields and Europe or the Atlantic coast of North America is limited by the Suez canal to 35 ft. For tankers of usual proportions this draft corresponds to about 30,000 tons dead weight. The extremely large tankers may be designed for drafts of 40 to 45 ft., and could not use canals if fully loaded. A relatively high block coefficient is used to obtain maximum dead weight. The block coefficient is the ratio of submerged volume to the product of length, beam and draft. The coefficient is frequently as high as 0.80, compared with about 0.60 to 0.70 for many passenger vessels and about 0.70 to 0.75 for most general-cargo ships.

BY COURTESY OF NEWPORT NEWS SHIPBUILDING AND DRY DOCK CO.

FIG. 4.—LARGE TANKER (60,000-TON DEAD WEIGHT) USED IN TRANS-ATLANTIC SERVICE BETWEEN THE PERSIAN GULF AND PHILADELPHIA

Construction.—The internal bulkheads, and the stiffening members, are usually entirely welded. The outside plating (shell and deck) is also welded but in large tankers a few fore-and-aft seams are riveted. It has been found that a riveted seam will usually stop a crack in the plating, thus preventing a minor crack from becoming a major fracture. Small tankers may be completely welded.

The upper deck and the bottom shell of tankers are invariably framed longitudinally; that is, the stiffening members which support the plating are run fore-and-aft instead of transversely as in other types of ships. In many tankers the side shell plating is also longitudinally framed. The thickness of the shell plating will vary from about $\frac{3}{4}$ in. for a small tanker to about $1\frac{1}{2}$ in. for a very large tanker.

Tankers are shorter-lived than most steel ships because of the corrosive nature of their cargo. Crude-oil cargoes deposit corrosive sludges on the bottom and gasoline cargoes emit corrosive vapours against the deck overhead. As a result, a tanker may last only 12 years instead of the 20-year life of a cargo ship.

This deterioration takes place mainly in the cargo spaces, leaving the bow, the stern and the machinery with years of useful life. It is, therefore, not unusual for an owner to scrap the cargo-tank portion of an old tanker and have a new one built and fitted into the old bow and stern. Since the new part is usually made larger than the discarded section, this procedure is sometimes called "jumboizing."

Protective coatings have been developed that will withstand the corrosive effects of petroleum products, but they are in general so expensive that many operators leave the interior of cargo spaces in tankers unpainted. For the tanks that are used for salt-water ballast on the return voyage, electrolytic preservative systems using sacrificial anodes of magnesium are sometimes employed.

BIBLIOGRAPHY.—Society of Naval Architects and Marine Engineers, "History of Tankers," *Transactions*, p. 135 (1943), *The Design and Construction of Steel Merchant Ships*, ed. by D. Arnott, ch. 3(A), section 8, p. 78 and ch. 6, section 16, p. 231 (1955) and "Structural Design of Tankers," *Transactions*, p. 444 (1949). (J. P. Ck.)

TANNENBERG, BATTLE OF. No portion of the earth's surface has been more fought over than East Prussia. It was in 1410, on the field of Tannenberg, that the Teutonic knights were utterly routed by their enemies the Poles and Lithuanians, and the German victory of 500 years later was in some degree the reversal of this defeat.

The Tannenberg of 1914 was the greatest victory achieved by any of the combatants at the beginning of World War I, and its strategic and moral effect upon the struggle was far-reaching.

France and Russia had been joined by military convention since 1892 and their chiefs of staff had made careful arrangements for concerted action. In case of war with the Central Powers it had been agreed that Germany was the main enemy and that the principal military effort must be made against it.

Because of the vastness of its territory and the comparatively undeveloped state of its railways, Russia's rate of concentration was considerably slower than that of France, so much so that there was a danger of the latter's being crushed before the attack on Germany in the east could begin to exert its effect. Unfortunately, under the pressure of French insistence, the good will and optimism of Russia's military leaders exceeded practical possibilities and the outbreak of war found Russian military plans in some disorder. All Russia's initial operations in East Prussia were in fact hasty improvisations.

Geographically, East Prussia points forward into the heart of Slavdom and constituted a difficult problem of defense for Germany. True to its policy of concentration of force at the decisive point, Germany had decided to advance against France with its main forces while maintaining only a minimum of strength in the east. Not wishing to abandon national territory without a blow, the Germans had decided to concentrate this minimum strength in East Prussia, despite the dangers involved in such action. Germany's strength in the east was to consist of 11 divisions, 6 first-line and 5 second-line, with but a single cavalry division.

Taking the East Prussian front as it appeared to the Russians, it can be divided militarily as follows:

1. *Königsberg Area.*—A more or less fortified area extending over the Königsberg peninsula, of a depth of 19 mi. from north to south. In the centre lay the obsolete fortress, strong against anything but heavy artillery. The area could be reinforced by sea through the canal from Pillau. It formed a strong flanking position from which to attack any force to the south.

2. *Insterburg Gap.*—A gap of 43 mi. of open country, down which ran the main Insterburg-Allenstein railway. The River Angerapp running north and south blocks this gap. It formed an excellent line of defense with the right flank resting against the impassable Lake Mauer.

3. *Masurian Lakes.*—A line of continuous lakes from Angerburg to the Russian frontier at Johannisburg, a distance of 50 mi. The narrow gaps between the lakes could be held with few troops, and the whole line presented an almost impregnable front behind which the Germans could operate to either flank.

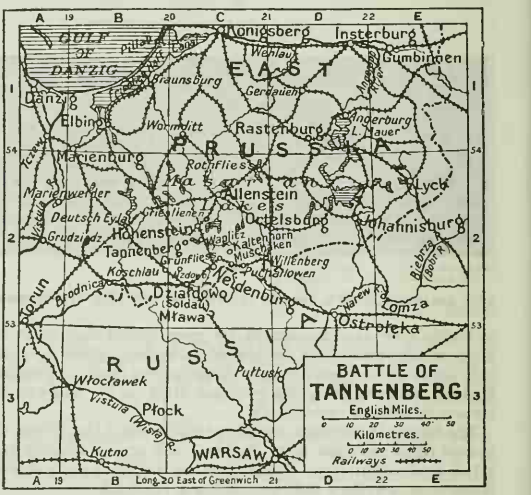
4. *Johannisburg-Dzialdowo (Soldau) Front.*—A distance of 75 mi. running east and west. On the Russian side the country had been left a desert as a defense against Germany.

The Russian Plan of Invasion.—The Russian plan for the invasion of East Prussia was to advance with two armies north and south of the line of lakes, with the object of enveloping the German forces supposedly concentrating behind this line. The northern army under Paul Rennenkampf was to advance first, toward the Insterburg gap, with the object of drawing the Germans as far to the east as possible. The southern army under Alexander Samsonov was to advance to the line Rastenburg-Allenstein, with the object of striking the rear of the Germans opposing Rennenkampf.

Rennenkampf's army crossed the frontier on Aug. 17 with a strength of six and one-half divisions and five and one-half cavalry divisions. On Aug. 20 an action was fought with the German 8th army at Gumbinnen. Max von Prittwitz und Gaffron, the German commander in the east, had concentrated his main strength to meet Rennenkampf, six divisions and a cavalry division, while leaving about two divisions to face Samsonov's army, which was concentrating on the River Narew.

Rennenkampf had moved off without a great portion of his transport and found himself forced to halt on Aug. 20. Prittwitz advanced to attack him. There were serious miscalculations in the German movements, with the result that the action was indecisive, both sides suffering heavy losses. That evening the news arrived that Samsonov had not only started to advance, but that two of his corps were already close to the frontier and that he had at least five corps with him. The German commander abandoned all hope of defeating Rennenkampf on Aug. 21 and contemplated a withdrawal to the Vistula. On communicating this on the telephone to the supreme command, he and his chief of staff were at once superseded.

The new commander, Paul von Hindenburg, had not yet been employed in the war. His chief of staff, Erich Ludendorff, reached supreme headquarters at Coblenz on Aug. 22 and the situation in the east was explained to him. He immediately ordered the troops retreating in front of Rennenkampf to halt at once and to prevent him from pursuing, and issued instructions for a concentration against Samsonov. General Hindenburg joined Ludendorff's train at Hanover and took over the command at Marienburg on Aug. 23.



Luckily for the new command, Max Hoffmann, the chief of the operations section under Prittwitz, had already carried out many steps facilitating the concentration desired by Ludendorff, and the situation, though alarming, had not grown worse since the supersession of Prittwitz. Samsonov's right was indeed well over the frontier near Ortelsburg and his left between Neidenburg and Dzialdowo (Soldau), but the Russian advance appeared to be slowing up. The German 20th corps had taken up a position north of Neidenburg and was awaiting attack. The leading units of the 1st corps were detraining south of Deutsch Eylau and were moving up on the right of the 20th corps. The 3rd reserve division had detrained at Allenstein and was moving up on the left of the 20th corps. A landwehr division and units from the Vistula garrison troops were also coming up. Thus, a strength of about six or seven divisions had been collected to oppose ten divisions and three cavalry divisions of Samsonov's army.

No decisive results could be expected from such strength, and decisive operations were the Germans' only hope. The most important feature of the situation, however, was that Rennenkampf had made no forward move since Gumbinnen and had lost all touch with the enemy in front of him. Ludendorff therefore conceived the idea of withdrawing all the troops from Rennenkampf's front, with the exception of the cavalry division, and of directing them against the right flank of Samsonov's army. Another four divisions thus obtained would give him an actual superiority over Samsonov and the direction of their attack might indeed be decisive.

An Account of The Battle.—Samsonov, who in peacetime was the governor general of Turkistan, had arrived from sick leave in the Caucasus on Aug. 12. He had seen neither his subordinate generals nor his staff before, and his immediate superior, Jilinsky,

commander of the northwestern front, was urging him forward to cross the frontier on Aug. 19. The original concentration areas of his army behind the Bobr and the Narew had been considered to be too far to the east and a general move of all corps to the west was in progress. By Aug. 20, the right wing (6th and 13th corps) had reached the frontier, with the 4th cavalry division well behind on the right. The left wing (15th and 1st corps) was just short of the frontier, with the 6th and 15th cavalry divisions covering it. Of the 23rd corps, one division was two days' march behind the left wing and the other had not yet detrained at Warsaw. The troops had been marching for a week across the desert area north of Ostroleka, over execrable roads in very hot weather. Much of the transport had been left behind, the men were tired and unfed and numbers had been much reduced by sickness. No contact had been gained with the enemy and the army was already more than 24 hours late in crossing the frontier. Reports from Rennenkampf of his victory at Gumbinnen and, during his advance westward on Aug. 24 and 25, of the signs of a hurried retreat on the part of the Germans, brought further urgings from Jilinsky to Samsonov to hurry. Samsonov obeyed to the best of his ability, and by the morning of Aug. 26 his position was as follows:

His centre of two corps (13th and 15th) was advancing northwest and was close on the line Allenstein-Hohenstein. On his right was one corps (6th) and the 4th cavalry division at Rothfliess, at two days' march from the centre. On his left was a corps (1st) and the 6th and 15th cavalry divisions near Koschlau, at one day's march from his centre. The 23rd corps troops which were up were moving into the gap between the centre and the left. His forces were thus much dispersed.

Reconnaissance had been bad and Samsonov had little information as to the enemy's dispositions. He seems to have thought that the Germans flying in front of Rennenkampf had already crossed his front in their march to the Vistula and that he was too late to cut them off. He had deviated from his instructions to advance with his left on Allenstein and now had practically the whole of his force to the west of that town. He had no knowledge of the whereabouts of Rennenkampf's army and did not realize that he was increasing the space between the two Russian armies and thereby facilitating the German operations. His supply system had broken down utterly and his men were half famished and worn out with the fatigue of excessive marching. The signal service had also broken down and orders had been sent out to corps by wireless. The Russian signals were easily read by the Germans, who thus had absolutely accurate information as to their opponents' dispositions and intentions.

Germans Attack Samsonov.—The new German command had galvanized the dispirited 8th army from the first and hope of success ran high. The attacks on Samsonov's two flanks were to be launched simultaneously on Aug. 26. On the right, the German 1st corps, strengthened by the Vistula garrisons, attacked only halfheartedly. Hermann von François, the commander of the 1st corps, was short of some of his heavy artillery and he contented himself with driving in some of the Russian outposts. On Aug. 27, however, he launched his attack in earnest. At dawn a hurricane bombardment was opened on the Russian 1st corps's positions on either side of Uzdowo. The famished Russians did not wait for the infantry attack but broke and ran. By 10 P.M. that night there was nothing left of the Russian left wing except a small rear guard just north of Dzialdowo; and that too had disappeared before the next morning.

The rear of the Russian centre was now completely exposed and François directed his march on Neidenburg. Opposition there was none, and by the night of Aug. 28 his leading troops had reached Muschaken, eight miles east of Neidenburg. By the night of Aug. 29 the whole length of the road from Neidenburg to Willenberg was held by a line of entrenched pickets. The Russian retreat to the south was completely blocked. During Aug. 30 François was attacked at Neidenburg by fresh Russian troops coming up from Mlawa and actually lost the town for a time. He held stoutly on to the line of his pickets, however, and on Aug. 31 Neidenburg was reoccupied.

On the left, the German 17th and 1st reserve corps, having turned their backs on Rennenkampf and marched due south, had arrived within striking distance of the Russian right at Rothfliess on the night of Aug. 25. Their attack on the morning of Aug. 26 came as a complete surprise. The Russian 6th corps withdrew in confusion, losing 6,000 men and 16 guns. By the night of Aug. 27 the Russian right was 32 mi. S. of Rothfliess and in a state of hopeless demoralization. The rear of the Russian centre was now exposed on its other flank. The German 17th corps continued its march to the south, while the 1st reserve corps moved in closer toward Allenstein. The story of the fighting in the Russian centre is not so inglorious as that on the two flanks. In accordance with Samsonov's plan the centre (13th and 15th corps) continued their advance on Aug. 26 toward the line Allenstein-Hohenstein. The 13th corps encountered no opposition, but the 15th had to fight its way steadily forward. For Aug. 27, the advance was to be continued.

Second German Blow.—Ludendorff, with his accurate information of the events on either flank, now decided that the German centre had retired enough. The reinforced 20th corps was therefore ordered to attack on Aug. 27. The battle which ensued between the Russian 15th corps and the German 20th corps was particularly fierce, the Russians holding their own and even advancing a little on their left. The Russian 13th corps rendered no help to its neighbour. On Aug. 28 the Germans renewed their attacks and this time the hardly tried Russian 15th corps commenced to break. The 13th corps, moving down to assist, found itself assailed in the rear by troops arriving from Allenstein and for the second time brought no help.

When night fell the two Russian corps were on a line from Grieslienen to Waplitz and in close touch with the Germans. Hoping to escape from pursuit, they began to withdraw during the night. The withdrawal soon became a rout, and a mass of troops and transport, both unfed and unwatered, became inextricably mixed in the depth of the Forest of Grünfliess. Determined attempts were made during Aug. 29 to break through to the south of Muschaken and to the north at Kaltenborn, but the German ring held fast. Only about 2,000 men effected their escape. The greater portion of these two corps surrendered to the German parties clearing the forest during Aug. 30 and 31.

Samsonov himself had moved up to Neidenburg on Aug. 27 in order better to control the battle. He was greeted with the news of the disaster to both his flanks and met streams of stragglers coming into the town. He decided to go to the headquarters of the 15th corps and take personal charge, and left Neidenburg on the morning of Aug. 28 with a small staff, all mounted on horses. At once he became mixed in the mass of transport in the forest and could exercise no effective command. He then decided to ride to the south, but during the night of Aug. 28–29 he became exhausted after having walked for some distance. He drew aside in the darkness and his staff heard a single shot. A careful search brought no result and the staff continued its way, certain that its unfortunate commander had taken his own life in his despair at the turn which events had taken.

Results.—The practical result of the fighting from Aug. 26 to Aug. 30 was the total annihilation of two Russian corps and the reduction of three other corps to half their strength in numbers. The whole Russian 2nd army was so demoralized and was so deficient in essential material, such as artillery and transport, that it was no longer of any fighting value. The Germans claimed 125,000 prisoners and 500 guns. It is certain that their booty was very great, for 60 train loads of captured material left Puchallowen after the battle and the captured horses were driven in herds to improvised camps at Neidenburg. Although greatly inferior in the whole eastern theatre, the Germans had yet managed to concentrate a superiority on the field of battle. The Russians had 132 battalions, 96 squadrons and 620 guns, including 12 heavy, while the Germans had 155 battalions, 48 squadrons and 818 guns, including 128 heavy.

BIBLIOGRAPHY.—W. von Stephani, *Mit Hindenburg bei Tannenberg* (1919); H. von François, *Marneschlacht und Tannenberg* (1920); P. von Hindenburg, *Out of my Life* (1920); E. Ludendorff, *My War Memories* (1922); J. Daniloff, *Russland im Weltkrieg, 1914–15* (1925), French trans., *La Guerre Mondiale* (1927); M. Hoffmann, *The War of*

Lost Opportunities, trans. by A. E. Chamot (1925); Eolorin, *Istoriya Voini 1914* (Prague, 1925); Ironside, *Tannenberg* (1925); Camon, *Ludendorff sur le Front Russe 1914–15* (1926); H. von François, *Tannenberg* (1926); M. Hoffmann, *Tannenberg* (1927). (*See* also WORLD WAR I: *Bibliography.*) (W. E. I.)

TANNER, HENRY OSSAWA (1859–1937), U.S. Negro painter, whose style was naturalistic and whose subject matter was most often biblical, was born at Pittsburgh, Pa., on June 21, 1859. He was the son of BENJAMIN TUCKER TANNER (1835–1923), bishop of the African Methodist Episcopal Church from 1888 to 1908 (when he retired at his own request) and author of a number of books, including *An Apology for African Methodism* (1867) and *The Negro's Origin* (1869). Henry Tanner was a pupil of Thomas Eakins, in Philadelphia, and of J. P. Laurens and Benjamin Constant in Paris, where he lived thereafter. He first exhibited at the Salon in 1895. Examples of his work may be seen at the Philadelphia (Pa.) Museum of Art; Los Angeles county museums; Art Institute of Chicago (Ill.); Metropolitan Museum of Art, New York city; and other public collections. He was a knight of the Legion of Honour and a member of several painters' societies. Tanner died in Paris on May 25, 1937.

TANNHÄUSER (TANHUSER), German minnesinger of the 13th century, who lived for a time at the court of Frederick II, duke of Austria. After Frederick's death he was received at the court of Otto II, duke of Bavaria, but he spent much time in wandering about Germany. He also went as a crusader to the Holy Land. His poems belong to the decadence of the minnesong, and combine a didactic display of learning with descriptions of peasant life. His adventurous life led him to be identified, in the popular imagination, with the knight Tannhäuser who, after many wanderings, comes to the Venusberg, or Hörselberg, near Eisenach. He enters the cave where the Lady Venus—the Frau Hulda of German folklore—holds her court, and abandons himself to a life of sensual pleasure. By and by he is overcome by remorse and, invoking the aid of the Virgin Mary, he obtains permission to return for a while to the outer world. He then goes as a pilgrim to Rome and entreats Pope Urban to secure for him the forgiveness of his sins. The pope declares it is as impossible for him to be pardoned as for the staff he has in his hand to blossom. Tannhäuser departs in despair and returns to the Venusberg. In three days the staff begins to put forth green leaves, and the pope sends messengers in all directions in search of the penitent, but he is never seen again.

This legend was at one time widely known in Germany, and as late as 1830 it survived in a popular song at Entlebuch in Switzerland, a version of which was given by Ludwig Uhland in his *Alten hoch- und niederdeutschen Volkslieder* (1844). Among the attendants of Hulda was the faithful Eckhart, and in the preface to the *Heldenbuch* he is said to sit before the Venusberg, and to warn passers-by of the dangers to which they may be exposed if they linger in the neighbourhood. The legend was reproduced by several modern German poets, and by Richard Wagner in the famous music drama.

TANNIN (TANNIC ACID), the generic name for a widely disseminated group of plant products, so named from their property of converting rawhide into leather (*q.v.*). They are soluble in water, and their solutions, which have an acid reaction and an astringent taste, are coloured dark blue or green by iron salts, a property utilized in the manufacture of ink (*q.v.*). Tannins occur in the roots, wood, bark, leaves, or fruit of many plants as normal physiological products. They also occur to the extent of 45%–65% in galls (*q.v.*), pathological growths on species of oak and sumac resulting from the attack of an insect.

Tannins may be classified into two main groups depending upon the products formed by boiling with mineral acids. Hydrolyzable tannins yield water-soluble products such as gallic acid $(HO)_3C_6H_2.CO_2H$, protocatechuic acid $(HO)_2C_6H_3.CO_2H$, and sugars. Under similar treatment, condensed tannins, by far the larger group, form insoluble precipitates called "tanners reds" or phlobophenes.

Gallotannin or common tannic acid is the best known of the hydrolyzable tannins. It is obtained from Chinese or Aleppo galls by extraction with water or organic solvents. When pure, it is a slightly yellow or tan amorphous powder or mass. It is very soluble in water, less so in alcohol and practically insoluble in ether. Hydrolysis produces gallic acid and glucose. Aleppo galls yield, in addition to gallic acid, some ellagic acid, a constituent also of myrobalan, divi-divi, algarobilla and valonia tannins. Tara, the pods from *Caesalpinia tinctoria,* a plant indigenous to Peru, contains 40%–55% of a gallotannin similar to that from galls and has become an important source for refined tannin and gallic acid. The chestnut oak (*Quercus prinus*) yields a hydrolyzable tannin important in leather manufacture.

Among the important condensed tannins are the extracts from the wood or bark of quebracho, mangrove and wattle. The structures of these tannins are uncertain, but many workers believe them to be related to the flavan catechin found in gambier. Refined analytical methods have shown that many natural tannins are mixtures of several closely related substances.

Uses.—Besides their principal application in the manufacture of leather, tannins find use as a constituent to reduce the viscosity of drilling mud for oil wells, in boiler water to prevent scale formation and in the clarification of wine and beer. Tannic acid is employed as a mordant in the dyeing of cotton with basic dyes. Tannic acid from galls is listed in the *British Pharmacopoeia* (1953) and in the *National Formulary of the United States* (1955). Because of its styptic and astringent properties, it has been used in the treatment of tonsillitis, pharyngitis, hemorrhoids and various skin eruptions. The use of tannic acid in the treatment of burns, formerly widely accepted, has become less popular. Tannic acid has been administered internally to check diarrhea and intestinal bleeding, and also as an antidote for metallic, alkaloidal and glycocidal poisons with which it forms insoluble precipitates.

(V. H. W.)

TANNING, a term describing the conversion of natural skins into leather. Modern tannage may be (1) vegetable tanning; (2) mineral tanning, principally with compounds of chromium or aluminum; (3) combination mineral and vegetable tanning; (4) oil tanning; *i.e.*, chamois leather. *See* LEATHER.

TANNU TUVA: *see* TUVA.

TANSY, the name applied to the species of the genus *Tanacetum* of the Compositae family, but most frequently to the common tansy (*T. vulgare*), a perennial herb, native to Europe, that has been widely grown both as a garden ornamental and for medicinal purposes. It long ago escaped from cultivation and is now a common weed over much of the United States. Formerly the leaves were used extensively as a seasoning in various dishes and by the medical profession for its tonic and stimulant properties. By mid-20th century it was little used although it could still be found in gardens as an ornamental plant. (J. M. BL.)

TANTA, a town of lower Egypt, in a central position nearly midway between the two main branches of the Nile, and converging point of several railways traversing the delta in all directions. Pop. (1956 est.) 164,800. It is the capital of the rich province of Gharbia, and is noted for its fairs and Moslem festivals.

TANTALUM (symbol Ta, atomic number 73, atomic weight 180.95) was discovered by A. G. Ekeberg in 1802. One stable form, Ta^{181}, and radioactive isotopes of mass numbers 180 and 182 are known. The electron arrangement in the unfilled outer orbits (O and P) is: $5s^2$, $5p^6$, $5d^3$, $6s^2$. Tantalum is always found associated with columbium (*q.v.*) in its ores and this resulted in confusion on the part of the early investigators until J. C. G. de Marignac, in 1866, developed a method for the separation of the two elements, a method which has become classical and which is of great commercial value. The important minerals are tantalite, $Fe(TaO_3)_2$, and columbite, $Fe(CbO_3)_2$, found in commercial quantities in Australia, Brazil, Belgian Congo and South Dakota. Of less importance is the mineral microlite $(6CaO.3Ta_2O_5.CbOF_3)$ found in New Mexico. These minerals may carry other elements such as titanium, tin, tungsten and manganese.

Separation.—The pulverized ore is fused in iron pots with sodium hydroxide, which converts the tantalum and columbium into sodium tantalate and sodium columbate, compounds which are insoluble in water if some free alkali is present. At the same time compounds of iron, manganese, etc., which are soluble in

hydrochloric acid are formed. The fused mass is crushed and treated with water and most of the free alkali is removed by decantation. The black residue is digested with hot hydrochloric acid to dissolve the iron and manganese compounds and to react with the sodium tantalate and columbate to form tantalic and columbic acids. These are washed by decantation and dissolved in hydrofluoric acid. Sufficient potassium fluoride is added to the solution to form the double fluorides K_2TaF_7 and $K_2CbOF_5.H_2O$. The concentration of the solution is adjusted so that it is nearly saturated with the tantalum salt when hot (90° C.). On cooling this salt crystallizes out in the form of fine needles, leaving the columbium and most of the impurities in solution. This method of separation was first used by Marignac and its efficiency is the result of the relative solubilities of the potassium double fluorides of tantalum, titanium and columbium. In the order given the solubilities are 5 g., 12 g. and 77 g. per litre at room temperature. The tantalum salt can be purified further by recrystallization from water containing a small amount of hydrofluoric acid, and this potassium tantalum fluoride is used in the preparation of all other compounds of tantalum and the metal itself.

Pure ductile tantalum was made by Werner von Bolton in 1905. The double fluoride was heated with sodium in evacuated tubes to bring about the reaction $K_2TaF_7+5Na=2KF+5NaF+Ta$. The washed, finely divided metal powder was pressed into sticks which were melted down by drawing an arc between them and a water-cooled table within a vacuum furnace. The fused buttons were quite limited in size but could be worked into rod and fine wire. In 1922 C. W. Balke perfected a process whereby tantalum bars weighing up to ten pounds might be produced. The double fluoride is electrolyzed in iron pots provided with graphite anodes. The current keeps the bath liquid and deposits the metal in the form of small crystals. When the pot is full the electrolytic mass is pulverized and the salt is separated from the heavy tantalum powder with an air blast. Final cleaning is done chemically to give a tantalum powder of high purity. This is pressed into bars which are sintered in vacuum furnaces at a temperature not far below the melting point of the metal (2,996° C.). The bar is heated by its own resistance to the passage of an electric current. These bars are very ductile and can be worked into sheet, rod and wire. This is done at room temperature, since the hot metal absorbs large volumes of the common gases and becomes brittle. Tantalum resembles but is somewhat darker than platinum in colour and its specific gravity is 16.6. It is outstanding in its resistance to the action of acids and corrosive chemicals except hydrofluoric acid, concentrated sulphuric and phosphoric acids above 175° C., and solutions of the strong alkalies.

Applications.—Tantalum was first used as a filament in electric lamps, but for this purpose it was soon replaced by tungsten. Its first extensive use was in the manufacture of electrolytic rectifiers. These (trade name "Balkite") utilize the electrolytic valve action of tantalum, which allows an electric current to pass in one direction only when immersed in acid. It is used in electronic tubes. Large quantities of sheet tantalum are used in the construction of chemical equipment for use in handling corrosive gases and liquids. Heat exchangers, evaporators, condensers and acid absorbers are built up from sheet by a special process of electric welding. Tantalum produces minimal foreign-body reactions in the human body and is, therefore, used in nerve and bone surgery with great success.

Compounds of Tantalum.—Tantalum is usually pentavalent but in some compounds its valence can be two, three or four. The pentoxide, Ta_2O_5, is white in colour and is obtained by heating the metal in air or oxygen or by the ignition of tantalic acid, $HTaO_3.xH_2O$. The latter is precipitated by the addition of ammonia to solutions of the alkali double fluorides, by the addition of hydrochloric or nitric acid to solutions of the soluble tantalates or by the action of water on tantalum pentachloride. This acid gives rise to a large number of tantalates such as $4K_2O.3Ta_2O_5.16H_2O$ and $7Cs_2O.6Ta_2O_5.38H_2O$. Tantalates are most easily made by fusion of the oxide with the carbonates or hydroxides of the alkali metals. The addition of hydrogen peroxide to solutions of the tantalates produces pertantalates of which potassium pertantalate, $K_3TaO_8.\frac{1}{2}H_2O$, is typical.

Anhydrous tantalum pentafluoride, TaF_5, can be made by the direct action of fluorine on tantalum in the form of hydroscopic prisms (melting point 97° C., boiling point 229° C.). The tantalum fluoride, formed by the solution of the metal or acid in hydrofluoric acid, is really the acid H_2TaF_7, which forms many double fluorides with the fluorides of other metals. Tantalum pentachloride, $TaCl_5$, is made by the action of chlorine on heated tantalum or by the action of chlorine and the vapours of sulphur monochloride on the heated oxide. It is a white crystalline solid (melting point 211° C., boiling point 241.6° C.). Tantalum carbide, TaC, made by the direct union of the elements or by the action of carbon on the oxide, is a very hard yellow crystalline mass which is used in the manufacture of hard carbide tools for the machining of steel (*see* CARBIDES).

BIBLIOGRAPHY.—J. A. Newton Friend (ed.), *A Text-Book of Inorganic Chemistry*, vol. vi, part III (1929); C. W. Balke, "Recent Developments in Tantalum and Columbium," *Ind. Eng. Chem.*, 30:3, pp. 251–254 (1938), and "Columbium and Tantalum," *Ind. Eng. Chem.*, 27:10, pp. 1166–69 (1935); Donald F. Taylor, "Acid Corrosion Resistance of Tantalum, Columbium, Zirconium, and Titanium," *Ind. Eng. Chem.*, vol. 42, p. 639 (April 1950). (C. W. BE.)

TANTALUS, in Greek legend, son of Zeus or Tmolus and Pluto (Wealth), daughter of Himantes, the father of Pelops and Niobe. He was the traditional king of Sipylus in Lydia (or of Phrygia), and was the intimate friend of Zeus and the other gods, to whose table he was admitted. But he abused the divine favour by revealing to mankind the secrets he had learned in heaven (according to the historian Diodorus Siculus), or by killing his son Pelops and serving him up to the gods at table, in order to test their power of observation (according to Ovid). Another story, by Pindar, was that he stole nectar and ambrosia from heaven and gave them to men. According to others, Pandareus stole a golden dog which guarded the temple of Zeus in Crete, and gave it to Tantalus to take care of. But, when Pandareus demanded the dog back, Tantalus denied that he had received it. Therefore, Zeus turned Pandareus into a stone, and flung down Tantalus with Mt. Sipylus on top of him.

The punishment of Tantalus in the lower world was famous. He stood up to his neck in water, which flowed from him when he tried to drink it; and over his head hung fruits which the wind wafted away whenever he tried to grasp them (*Odyssey*). This myth is the origin of the English word "tantalize." Another story is that a rock hung over his head ready to fall and crush him (Euripides in *Orestes*). The sins of Tantalus were visited upon his descendants, the Pelopidae. Ancient historical reminiscences and natural phenomena, especially volcanic catastrophes, are at the bottom of the legend. The tomb of Tantalus on Mt. Sipylus was pointed out in antiquity, and was identified by C. F. Texier with the great cairn beneath Old Magnesia; but Sir W. M. Ramsay inclined to a remarkable rock-cut tomb beside Magnesia.

The story of Tantalus is an echo of a semi-Greek kingdom, which had its seat at Sipylus, the oldest and holiest city of Lydia, the remains of which are still visible. There was a tradition in antiquity that the city of Tantalus had been swallowed up in a lake on the mountain; but the legend may, as Ramsay thought, have been suggested by the vast ravine which yawns beneath the acropolis. J. Hylén considered the story of the thirst of Tantalus in the underworld to be the result of the Orphic interpolator in the Nekyia of the *Odyssey*, and the Pandareus story to be an innovation of the Alexandrine poets. The essay contains a copious list of authorities and a history of the legend. According to V. Henry, Tantalus is the sun. His betrayal of the secrets of the gods refers to the sun's unveiling the secrets of heaven; the slaying of Pelops denotes the going down of the sun. (*See* PELOPS.)

TANTIA TOPI (*c.* 1819–1859), rebel leader during the Indian Mutiny, was a Mahratta Brahman in the service of Nana Sahib. He instigated the massacre of Cawnpore, and commanded at the battle of Bithur, where he was defeated by Gen. H. Havelock. With the aid of the Gwalior contingent he pressed Gen. C. A. Windham hard at Cawnpore on Nov. 27–28, 1857, but was defeated by Sir Colin Campbell on Dec. 6. Together with the rani of Jhansi he was besieged by Sir Hugh Rose in the Jhansi

fort, but escaped and collected a force of 20,000 men which Sir Hugh defeated without relaxing the siege. This was the decisive action of the campaign in central India, and Tantia Topi was obliged to seek refuge in the jungles of Rajputana and Bundelkhand, where he was taken by Major Meade, condemned and executed on April 18, 1859. He was the only rebel leader in the Mutiny who showed any conspicuous military ability. *See* INDIAN MUTINY.

TANZANIA, UNITED REPUBLIC OF. An independent sovereign state of east central Africa, within the Commonwealth of Nations, consisting of the republics of Tanganyika and Zanzibar. Pop. (1963 est.) 10,123,000 (Tanganyika 9,798,000, Zanzibar 325,000). Of the total area of 362,844 sq.mi. (939,767 sq.km.), Tanganyika comprises 361,800 sq.mi. and Zanzibar and Pemba islands 1,044 sq.mi.; Zanzibar Island lies about 20 mi. (32 km.) off the coast of Tanganyika. The new state came into existence in April 1964 under an act of union signed by the presidents of Tanganyika and Zanzibar and later ratified by the National Assembly of Tanganyika and the Revolutionary Council of Zanzibar. It was subsequently announced that until a constituent assembly had adopted a constitution (proposals for which were to be made by a special commission), the new republic would be governed in accordance with the constitution of Tanganyika, which would be modified to provide Zanzibar with a separate local legislature and executive. The powers exercised by these bodies would, however, be limited to certain internal matters; *i.e.,* those not reserved to the Parliament of the united republic and its executive. These reserved powers would include foreign affairs, defense, police, citizenship, immigration, and external trade. The president of Tanganyika became head of the united republic, and the president of Zanzibar was appointed first vice-president and executive head of Zanzibar. It was further agreed that the existing laws of Tanganyika would have validity in both countries, subject to such amendments as might be found desirable. *See* also TANGANYIKA; ZANZIBAR.

T'AO CH'IEN (T'AO YUAN-MING) (*c.* 365–427), Chinese poet and recluse, was born into a gentry family that had fallen on hard times. When he was about 20, he took a minor official post and for the next ten years was in the service of one or another of the ambitious generals who were threatening the moribund Chin dynasty. In 405, after a brief term as magistrate of P'engts'e, he retired from public life to make a precarious living as a farmer. On more than one occasion crop failure brought him and his family to the verge of starvation. On leaving his last post he had written, "My instinct is all for freedom and will not brook discipline or restraint. Hunger and cold may be sharp, but this going against myself really sickens me. Whenever I have been involved in official life I was mortgaging myself to my mouth and belly." He had no more dealings with officialdom but stuck to the life of poverty and hardship, which he preferred to the rigours of respectability. At the time of his death he was known chiefly as a recluse and an eccentric. His poetry was not much appreciated before the 8th century, but modern critics regard him as the most important poet of his period. The taste of his Six Dynasties contemporaries was for an elaborate and artificial poetry; T'ao Ch'ien's is simple and straightforward, with a minimum of artifice. It deals with country pursuits and his essentially Taoist philosophy of life and death.

T'ao the Hermit: Sixty Poems was translated and annotated with an introduction by William Acker (1952); *Poems* was translated by Lily Pao-hu Chang and Marjorie Sinclair (1953).

See J. R. Hightower, "The *fu* of T'ao Ch'ien," *Harvard Journal of Asiatic Studies,* 17:169–230 (June 1954). (J. R. HR.)

TAOISM has been an integral part of Chinese life and thought for 2,000 years. Like Confucianism (*q.v.*), it has influenced every aspect of Chinese culture, whether philosophy, religion, art, literature, or government. Throughout Chinese history, it has paralleled Confucianism, reinforcing, supplementing, and criticizing it, and it helped shape the development and transformation of Buddhism in China. In a true sense every Chinese is a Taoist. While Confucianism emphasizes social order and an active life, Taoism concentrates on individual life and tranquillity. Though this suggests that Taoism plays a secondary role, the fact is that in its advocacy of individuality, spiritual freedom, naturalness, simplicity, religious mysticism, *laissez faire* in government, and the transcendental ideal in art, it is as important as Confucianism in Chinese tradition.

The name "Taoist School" was not mentioned until the 1st century B.C., but the movement must have been developing for centuries before that. Tradition says that ancient philosophical schools emerged from government offices, and Taoism in particular from that of the historian. This theory may have been influenced by the fact that Lao-tzu (*q.v.*), traditionally considered to be the founder of Taoism, was a curator of historical documents. The deeper meaning of the theory, however, is that Taoism arose in response to actual historical situations. In the 6th century B.C. feudalism in China was beginning to crumble. In the transition to a new age, society was beset with political disorder, economic chaos, and moral decay. For several centuries each of the "100 schools" of philosophy offered its own *tao* (literally, "way") as a solution. The *tao* offered by Lao-tzu and his followers was so unusual and so challenging that their movement alone came to be known as the Taoist School, *Tao-chia.*

Taoism is not a system of purely quiet contemplation or idle speculation. It is a philosophy offering a practical way of life. Later its teachings came to be utilized in the popular religion called *Tao-chiao.* In the Chinese tradition the two have been separate, but in the West they have often been confused under the name Taoism.

PHILOSOPHICAL TAOISM

The origin of Taoism as a philosophy is obscure. The term *tao* can be traced to the earliest Chinese classics, and it is a basic term in all ancient Chinese philosophical schools, including Confucianism. But it was in Lao-tzu, Yang Chu, Chuang-tzu, the *Huai-nan-tzu,* the Neo-Taoists, and the *Lieh-tzu* that the Taoist philosophy developed.

Lao-tzu.—The fundamental doctrines of early Taoism are embodied in the *Lao-tzu,* also called the *Tao-te ching* (*q.v.*). Whether or not the book is an anthology of the 4th century B.C. (the point is disputed), its basic ideas and passages probably go back to Lao-tzu, two centuries earlier. As the second title indicates, it is a classic on Tao and *te,* the "Way" and its "virtue." Previously the connotation of Tao had been social and moral; in the *Lao-tzu* it connotes for the first time the metaphysical. Tao "exists before heaven and earth." It is the "mother" and "ancestor" of all things. It is their "storehouse." It is at once their principle of being and their substance. "All things depend on it for life." It is One, everlasting, and unchangeable; it "operates everywhere and is free from danger." It is "invisible," "inaudible," "vague and elusive," but there is in it the "form" and the "essence." It is so absolute that it cannot be described; it "has no name." Furthermore, it is "nonbeing." The last idea, which appears to be nihilistic, simply means that Tao is prior to and above all things.

In its operation Tao is characterized by *wu-wei* (literally "no action"), which really means taking no unnatural action. It means spontaneity, noninterference, letting things take their own course. Tao "takes no action" but "supports all things in their natural state," and in this way "all things will transform spontaneously." As things arise, "Tao does not turn away from them." "It benefits all things but does not compete with them." "Tao invariably takes no action, and yet there is nothing left undone."

When man follows this natural way, he will abide with the one, the eternal, and the whole and will achieve a life of peace, harmony, and enlightenment. In dealing with things, he will "produce them but not take possession of them." In government he will adopt the policy of *laissez faire,* for "the best rulers are those whose existence is merely known by the people. . . . They complete their work. Nevertheless their people say that they simply follow Nature." Thus the natural way stands in direct opposition to all artificial regulation, organization, and ceremonies. This is the reason why the *Lao-tzu* vigorously attacks all formalities and artificialities. For the same reason it condemns war, taxation, punishment, superficial knowledge, and conventional morality.

The natural way is compared with the ways of water, the female, and the infant; that is, the ways of the weak. Along with weakness, the *Lao-tzu* also advocates humility, contentment, ignorance, submission, knowing where to stop, doing away with desires, accepting disgrace, and the like. This seemingly negative morality is deceptive. After all, "the greatest skill seems to be clumsy, and the greatest eloquence seems to stutter." "There is nothing softer and weaker than water, and yet there is nothing better for attacking hard and strong things."

While the weak seems to be glorified, the strongest stress really is on simplicity. A simple life is one of plainness in which profit is discarded, cleverness abandoned, selfishness eliminated, and desires reduced. It is the life of "perfection which seems to be incomplete" and of "fullness which seems to be empty." It is the life which is "as bright as light but does not dazzle." In short, it is the life of "superior virtue." This life requires all the important moral qualities taught in almost all ethical systems. To Lao-tzu, deep love, frugality, and not daring to be ahead of the world are "three treasures," and because of them one becomes courageous, generous, and a leader of the world. He urges us to love the earth in our dwelling, love what is profound in our hearts, love humanity in our associations, love faithfulness in our words, love order in government, love competence in handling affairs, and love timeliness in our activities. He wants us to maintain steadfast quietude and to be tranquil, enlightened, all embracing, impartial, one with Nature, and in accord with Tao. He teaches us to "benefit all things," to "treat those who are good with goodness" and "also to treat those who are not good with goodness," and to "repay hatred with virtue." He admonishes us not to have ulterior motives, to show or justify ourselves, or to boast, to be proud, to hoard things, or to be extravagant. He advises us to know the subtle and the eternal, but when we do not know anything, to know that we do not know. Virtually all the ingredients of a virtuous life, including the Golden Rule, are included.

In short, the life of superior virtue is one of goodness, enlightenment, peace, and long life.

Yang Chu.—The note of life preservation is particularly strong in Yang Chu (*c.* 440–360 B.C.), who may or may not have been an early Taoist. According to surviving isolated passages of his writings, he would not enter an endangered city, would not join an army, and would not care to exchange a single hair for the profits of the entire world. His main purpose in life was to "preserve life and keep the essence of his being intact." Mencius (*q.v.*) attacked him as an "egoist who denied the special relationship with the ruler." Whether Yang's doctrine actually degenerated to an irresponsible egoism or whether Mencius distorted it is a moot point. Yang Chu's teachings are certainly twisted in the "Yang Chu" chapter of the *Lieh-tzu* (see *Later Developments,* below), where his doctrine of self-preservation is coupled with pessimism on the one hand and hedonism on the other.

Chuang-tzu.—All the basic ideas in Lao-tzu are elaborated in the teachings of Chuang-tzu (*q.v.*). But his system is more than an elaboration; in it Taoism was extended to new dimensions and to a higher degree. To him, existence is not only spontaneity but the state of constant flux and incessant transformation. In Lao-tzu, the major notes are constancy and eternity; change is a minor one. In Chuang-tzu, change is the main theme. He conceives of the universe as a great current in which one state succeeds another in endless procession. No state can be retained, and time cannot be arrested. Life goes on "like a galloping horse." In these constant changes, things transform themselves by themselves, for this is the nature of things. There is no directing agency in this dynamic drama: "if there is a true director, there is no evidence of him."

In this process of rapid change, everything must follow its own nature. It is the nature of the eagle to fly high in the clouds, the nature of the dove to hop from tree to tree. A cypress will last for a thousand years, a mushroom for only a day. Each should live according to its nature. Similarly a man should be contented with his fate and "should not attempt to replace the way of heaven with the way of man." This is the true meaning of *wu-wei.* Though there is a much stronger element of fatalism in Chuang-tzu than in Lao-tzu, following Nature to him does not mean blind submission but rather a realistic recognition of one's own limits.

In the universe, then, there are these individual and continuous transformations, each in its own natural way. Such a universal process of rising and falling, of life and death, inevitably results in a world of differences and inequality, but all these are superficial, relative, often products of subjective points of view. Life and death cause each other, and neither right nor wrong is absolute. This and that, and possibility and impossibility, Chuang-tzu says, mutually produce each other. In reality, "all things are one," for Tao embraces all of them and combines them into a unity. A man of small knowledge will cut things up, discriminate, and make distinctions; a man of great knowledge will be comprehensive, impartial, and see Tao in all its unity. The ideal man will be a "companion with Nature" and a "friend of both life and death." He makes oneness his eternal abode. He abandons selfishness of all descriptions, be it fame, wealth, bias, or subjectivity. He knows his nature, nourishes it, and adapts it to the universal process of transformation. In this way he can "roam the universe" and be "emancipated." Having attained enlightenment through "the light of Nature," he moves in the realm of "great knowledge" and "profound virtue." He is then spiritually free. As the *Chuang-tzu* itself says of him,

> Alone he associates with heaven and earth and spirit, without abandoning or despising things of the world. He does not quarrel over right or wrong and mingles with conventional society. . . . Above, he roams with the Creator, and below he makes friends with those who transcend life and death and beginning and end. In regard to the essential, he is broad and comprehensive, profound and unrestrained. In regard to the fundamental, he may be said to have harmonized all things and penetrated the highest level. However, in his response to change and his understanding of things, his principle is inexhaustible, traceless, dark and obscure, and unfathomable.

In Chuang-tzu, Taoism reached the transcendental and mystical level, but its dynamism and realism became at the same time much stronger. Generally speaking, Chuang-tzu's universe is much more complex than that of Lao-tzu. It involves not only the states of being and nonbeing but also that of neither being nor nonbeing. Lao-tzu's way is directed chiefly to handling human affairs; that of Chuang-tzu aims chiefly at dealing with the universe. Lao-tzu's goal is social and political reform; that of Chuang-tzu is to "travel beyond the mundane world." Every passage of the *Lao-tzu* can be understood in terms of ordinary human experience, but what Chuang-tzu wants is pure experience—that is, identification with the universe. He speaks of "fasting of the mind," in which "the mind is empty to receive all things," and of "sitting down and forgetting everything," in which "the body is abandoned, the intelligence is discarded, one is separated from the body and free from knowledge, and one becomes identical with a great penetration." At the same time, in his emphasis on being true to one's own nature, his philosophy is much more individualistic than that of Lao-tzu. While the notes of mysticism, quietism, and fatalism are stronger, the ultimate goal is preservation of one's original nature and through it absolute spiritual freedom.

From the time of Chuang-tzu till the 2nd century B.C., the Taoist School remained influential. It was the chief rival of both Confucianism and Moism. Followers of Lao-tzu, Yang Chu, and Chuang-tzu probably traveled in slightly different directions, but the main tenets remained prominent and clear. Later they were synthesized in the *Huai-nan-tzu.*

The *Huai-nan-tzu.*—The *Huai-nan-tzu* is an anthology of treatises by Liu An (d. 122 B.C.), prince of Huai-nan, and scholars who gathered around him as his retainers. It represents a combination of Taoist teachings, and as such its originality is negligible. In it, however, the concept of Tao becomes more concrete. Using water as its analogy, the writer describes Tao as a concrete reality, all pervasive, the cause of all things, perfectly weak, taking no unnatural action, and yet bringing all things into completion. Furthermore, cosmogony becomes more elaborate and more definite. "There was a time before the time which was before the beginning," and "there was a time before the time which was before

nonbeing." "Before heaven and earth took shape, there was only undifferentiated formlessness," which was called "the great beginning." In cosmological evolution, Tao originated from vacuity, and vacuity produced the universe, which in turn produced the material forces. The material forces of heaven and earth combined to form *yin* and *yang*, which in their turn give rise to the myriad things. In its broad outline this cosmogony has remained orthodox doctrine not only among Taoist philosophers but also among the later Confucianists.

In its attempt at synthesis, the *Huai-nan-tzu* combines certain non-Taoist elements, such as the Confucian emphasis on learning as a method of self-cultivation and the Legalist emphasis on law in government. In its discussion of the spirit, it even introduces such ideas as immortality on earth and the techniques of breathing, etc., to achieve it, thus paving the way for the later confusion of Taoist philosophy with Taoist popular religion. In spite of this fact, it maintained Taoism at a time when Confucianism had assumed the dominant role in government and thought. In addition, by approaching metaphysics and cosmology in an essentially rational spirit, it prepared for the emergence of rationalistic Neo-Taoism.

Later Developments.—In the 3rd and 4th centuries A.D., Taoism developed in three directions. The first was Neo-Taoism, the most important philosophical development in the history of Taoism, which finds expression in the commentaries on the *Lao-tzu* and the *Book of Changes* by Wang Pi (226–249) and the commentary on the *Chuang-tzu* by Kuo Hsiang (d. 312), who probably appropriated ideas from Hsiang Hsiu (fl. 250). In them, nonbeing is no longer essentially in contrast to being but is the ultimate of all, or pure being (*pen-wu*), the One and the undifferentiated. According to Wang Pi, original nonbeing transcends all distinctions and descriptions. It is the pure being, original substance (*pen-t'i*), and the One in which substance and function are identified. It is always correct because it is in accord with principle. Where Lao-tzu had destiny, Wang would substitute principle, thus anticipating the development in Neo-Confucianism.

Just as Wang Pi went beyond Lao-tzu, so Kuo Hsiang went beyond Chuang-tzu. The major concept is no longer Tao but Nature (*tzu-jan*). Things exist and transform themselves spontaneously, and there is no other reality or agent to cause them. Things exist and transform according to principle, but each and every thing has its own principle. Compared with Wang Pi, Kuo Hsiang emphasizes being rather than nonbeing and the many rather than the one. To Wang Pi, principle transcends things; to Kuo Hsiang, it is immanent in them. To neither of them, however, is the sage one who withdraws into a hermit's life; he lives a life of social and political accomplishment, through taking no unnatural action. Significantly, the greatest sage to them was not Lao-tzu but Confucius.

The second movement was that of the "pure conversation," a way of life expressed in elegant, refined, and carefree conversation, devoid of the vulgarity of politics, earthly ambitions, and worldly values. It lasted from the middle of the 2nd century through the 6th, with Juan Chi (210–263), Hsi K'ang (223–262), and Hsiang Hsiu as outstanding representatives.

The third movement was fatalistic naturalism, represented by the *Lieh-tzu*, a product probably of the 3rd century. Its ideas are all original ingredients of Taoism, but the Taoist doctrine of taking no action has degenerated into complete abandonment of effort, spontaneity is confused with resignation, and having no desire is replaced by hedonism—all negative developments. On the positive side, Tao as the self-caused ultimate reality is much more strongly affirmed, and vacuity in the sense of eliminating desires and renouncing egoism is emphasized. In both of these the *Lieh-tzu* goes beyond Lao-tzu. In being indifferent to life and death and in ignoring the distinction of right and wrong, it pushes skepticism to the extreme; and in considering unceasing transformation as going in cycles according to fate, so that life and death, by their very nature, immediately succeed each other, it gives the doctrine of transformation a peculiar form. In these the *Lieh-tzu* goes beyond Chuang-tzu. In teaching that one should do away with the distinction between oneself and others, avoid a subjective state of mind, be vacuous, pure, and tranquil, and be in accord with Nature, it is essentially Taoistic and may well reflect, or even transmit, the teachings of Lieh-tzu, or Lieh Yü-k'ao (*c.* 450–375 B.C.), whose work had disappeared centuries before.

Influence.—As a philosophy Taoism declined after the 4th century, by which time Taoism and Buddhism had begun mutually to influence each other. In the 3rd and 4th centuries, Buddhist thinkers practised "matching the concepts" of Buddhism and Taoism, as a result of which Buddhists interpreted their philosophy in Taoist terminology while Taoism assimilated Buddhist ideas. Although Taoism eventually became a handmaid of Buddhism, it exercised far greater influence on Buddhist philosophy than vice versa. The major Taoist concepts of being and nonbeing were carried over to Buddhism. Some early Buddhist schools even called themselves "School of Original Nonbeing," "School of Nonbeing of Mind," etc. Like the Neo-Taoists, the Buddhists regarded ultimate reality as transcending all being, names, and forms, and as empty and quiet in Nature.

The greatest influence on Buddhism may be seen in the Chinese development of the Buddhist Meditation School (Ch'an or Zen; *see* ZEN). The Taoist tenets that Tao is everywhere, is indescribable, and can be discovered through enlightenment had a direct bearing on the development of Zen doctrines that the Buddha mind is everywhere, is indescribable by words and inconceivable in thought, and can be realized through enlightenment. The standard ways of Zen enlightenment are "absence of thought," "forgetting one's feelings," and "letting the mind take its own course," literal borrowings from Taoism, especially from Chuang-tzu. Many early Buddhist thinkers were experts in Taoist philosophy, and throughout history many Zen masters have been personal friends of Taoists.

Just as Taoism contributed to the development of Zen in the 7th and 8th centuries, so it contributed to the development of Neo-Confucianism in the 11th (*see* CHINESE PHILOSOPHY), but because of their "way of quietness and nonactivity," the Neo-Confucianists severely attacked the Taoists, and, as a result, Taoism ceased to be a vital philosophical system in the last millennium. As a way of life, however, it has never lost its hold on Chinese culture and society. Its spirit of harmony, simplicity, and peace has been eloquently expressed in Chinese landscape painting, landscape gardening, poetry, tea drinking, etc., and its spirit of naturalism, individualism, and freedom has strongly molded Chinese life.

RELIGIOUS TAOISM

The Way of the Yellow Emperor.—While the philosophy of Taoism was unfolding, a religious current, in progress from time immemorial, was developed by priest-magicians who offered divination and magic as means to inward power, restored youth, superhuman ability, and, most importantly, immortality on earth. Somehow this movement came to be known in the 1st century B.C. as the Way of the Yellow Emperor (Huang-ti) and Lao-tzu, and enjoyed an extensive following. In A.D. 143 Chang Ling, also called Chang Tao Ling, a popular religious leader famous for healing the sick, formally organized the movement and required his many followers to donate five piculs (bushels) of rice (hence the name the Way of Five Piculs of Rice). His followers called him Heavenly Teacher, a title his direct descendants retained until the 1940s.

From the very beginning the Yellow Emperor–Lao-tzu movement was a conglomeration of sorcery, the arts of priest-magicians, and the philosophy of the Yin-Yang School, according to which all things are products of the cosmic negative and positive forces. Gradually it assimilated certain elements of Taoist philosophy. Wei Po-yang (fl. 147–167) combined all these in a coherent system and thus laid the philosophical foundation of the Taoist religion. In his *Ts'an-t'ung-ch'i* ("The Three Ways Unified and Harmonized"), he attempted to synthesize all these ideas for the sake of prolonging life through the practice of alchemy, whereby, it was believed, the two cosmic forces could be harmonized and the vital force (*ch'i*) of the universe concentrated in the human body. His book became the basis of the *Lung-hu* ("Dragon and Tiger") *Scrip-*

ture of unknown date, the *Huang-t'ing* ("Internal and External") *Scripture* (3rd century), and the *Yin-fu* ("Secret Accord") *Scripture* (8th century). Later, Ko Hung (253–333?), also called Pao-p'u-tzu, synthesized Taoist philosophy and Confucian ethics. His chief concern, however, was alchemy to prolong life, and he strongly defended the belief in immortals (*hsien*).

These two philosophers provided religious Taoism with a firm theoretical foundation and an elaborate system of practice. It was K'ou Ch'ien-chih (d. 432) who made it an organized religion, by regulating the ceremonies, fixing the names of deities, and formulating theology. His influence led to its being made a state cult in A.D. 440. State patronage reached its height in the T'ang dynasty (618–906), since when religious Taoism has existed primarily as a religion of the illiterate masses.

Features.—As a religion, Taoism is distinguished by several prominent features, outlined below.

Multiplicity of Gods.—Taoism has a god for almost everything—stars, animate and inanimate things, the whole and parts of the body, ancestors, famous historical beings, ideals, etc. Among these are the Jade Emperor, the Three Pure Ones (Heavenly Honoured Ones Without Form, With Form, and Brahma Form), the Dipper Star, the Great Unit, Gods of Literature, Wealth, Medicine, Kitchen, and Goddess of Mercy (from Buddhism). It worships numerous immortals, has 10 "Heavenly Grottoes," 36 subsidiary "Heavenly Grottoes," and 72 "Blessed Places," where immortals rule and await people to seek the "way." Besides, there are imitations of many Buddhist heavens and hells.

Search for Blessings and Longevity.—Popular Taoism is strongly characterized by piety and devotion, but its chief objectives are earthly blessings—happiness, health, wealth, the begetting of children, and, most of all, longevity. By the 7th century an elaborate system of practices was developed to achieve these blessings and to realize the Three Original Principles of Taoism, namely, Essence, Vital Force, and Spirit. These practices included breath control, baths, diet, mental concentration, sex hygiene, physical exercises, medicine, the use of charms, attempts at disappearance and change of bodily form, and alchemy, especially the effort to transmute mercury into gold. In their pseudoscientific alchemical experiments over the centuries, the Taoists made a number of scientific discoveries and perhaps contributed indirectly to the development of science in Europe.

Superstitions.—The Taoist religion is a rich reservoir of superstitions, including geomancy, fortune-telling, divination, and the use of charms. With modern reforms, these are fast disappearing but not entirely eliminated.

Sects.—Taoism is a wholesale imitation of Buddhism, notably in its clergy, temples, images, ceremonies, and canon. In organization it is generally classified under two schools, the Northern and the Southern. The Southern School, called the True Unity sect, prevailing south of the Yangtze, is traced to the so-called founder Chang Ling but more likely grew up in the 10th century. It emphasizes man's nature—that is, man's spirit or true self—and relies on charms and magic to preserve it. It depends on "self-power" and is idealistic and informal. Its priests are married, live at home, and regard as their head the Heavenly Teacher, who used to reside in the Dragon and Tiger Mountain in central China. He had no influence on the masses and should in no way be called a pope, as he sometimes has been in the West.

The Northern School, called the Complete Unity sect, probably founded in the 13th century, is centred at the White Cloud Temple in Peking. It emphasizes man's life—that is, man's vital force—and relies on medicine and diet to prolong it. It depends on "other-power" and is materialistic and formal. Its priests renounce the home, adopt vegetarianism, and live in temples.

New Societies.—The two schools have been declining since the last half of the 19th century. In their place new societies have grown up, such as the Tao Yuan (Society of the Way), founded about 1920. These are led by laymen, worship deities of all religions, shun superstition, reject the search for earthly immortality but promote charity and moral culture, and, like traditional Taoist sects, sometimes participate in political revolts.

Ethical Teachings.—To most followers, who also follow Confucianism and Buddhism, the Taoist religion is essentially a sanction of ethics. In the two most popular tracts, *The Treatise on Influence and Response* and *The Silent Way of Recompense,* which have been influential for centuries, Confucian moral values (filial piety, loyalty, love, righteousness, and good faith), Buddhist virtues (compassion, kindness to all creatures, and remonstration against all evil deeds), and Taoism's own ethics of patience, simplicity, contentment, and harmony are strongly advocated.

Modern Status.—As a religion of the masses, Taoism has been rapidly fading away, especially since the Communist government came to power in China. Without leadership, financial resources, or recruitment for priests, its future as an organized religion is dim, although private practice, including local temples and shrines in homes, will surely continue. Its superstitions, animism, and idolatry, however, have largely been replaced by social and economic reform activities. Although there is a National Taoist Association in Mainland China, practically all Taoist priests have returned to secular and productive life, and Taoist temples are being used for other purposes. But many elements of the Taoist religion—its fairy tales, magic, aesthetic imagination—are still rich in Chinese folk life, especially in festivals, while the virtuous life and spiritual serenity taught in its religious tracts will continue to influence the Chinese character for a long time to come.

See also references under "Taoism" in the Index.

BIBLIOGRAPHY.—Karl L. Reichelt, *Religion in Chinese Garment* (1951); Wing-tsit Chan, *Religious Trends in Modern China* (1953), *A Source Book in Chinese Philosophy* (1963), and (trans.) *The Way of Lao Tzu* (1963). (W. C.)

TAORMINA, a town and winter resort on the east coast of Sicily (ancient Tauromenium), in the province of Messina. Pop. (1961) 8,903 (commune). It lies on an abrupt hill 650 ft. (198 m.) above the railway station, and was occupied at least as early as the 8th century B.C. (as the discovery of a Sicel cemetery of that period shows). It was refounded by the Carthaginian Himilco in 397 B.C., after the destruction, by Dionysius the Elder of Syracuse, of neighbouring Naxos. In 392 Dionysius occupied it, and in 358 the exiles from Naxos found a home there. During the 1st Punic War it belonged to the kingdom of Hiero, and after his death it was allied to Rome. During the 1st Servile War it was occupied by Eunous and some of his followers but was taken by the consul Publius Rupilius in 132. It was one of the strongholds of Sextus Pompeius, and after defeating him Augustus made it into a *colonia.* Its marble, wine, and mullets were highly esteemed. In A.D. 902 it was taken and burned by the Saracens; it was retaken in 962, and in 1078 fell to the Normans.

The ancient town perhaps had two citadels; one the hill above the town to the west, now crowned by a medieval castle, the other the hill upon which the theatre was afterward constructed. There are some remains of the city walls. The church of San Pancrazio is built into a temple of the 3rd century B.C., dedicated to Sarapis. Other ruins belong in the main to the Roman period. The theatre, largely hewn in the rock, is of Greek origin but entirely reconstructed. The view is of exceptional beauty, Mt. Etna being seen on the southwest, while to the north are the rugged coast and the mountains of Calabria. There are remains of a smaller theatre (the Odeum), some large cisterns, and a large bath or tank known as the Naumachia. There are also remains of Roman houses, tombs, etc., and fine specimens of Romanesque and Gothic architecture (notably the Palazzo Corvaia).

TAO-TE CHING, also called LAO-TZU, is the basic Taoist writing, a work that has exerted tremendous influence in the Far East for 2,000 years. The title means the "classic of the Way and virtue" (or "and its power"). By the "Way" is meant the natural way of things, the order of Nature. *Tao* is eternal, unchangeable, all pervasive, One, indescribable, and nameless. It is "nonbeing," that is, the source from which all being has come. By "virtue" is meant naturalness, simplicity, and *wu-wei,* or "taking no [unnatural] action."

The book, consisting of from 5,227 to 5,722 words, depending on the edition, is often called the 5,000-character classic. Its 81 chapters are unrelated, repetitious, nearly all rhymed but irregular in literary style, and contain no dialogue, historical event, or

proper noun. Tradition has attributed the work to Lao-tzu (*q.v.*) of the 6th century B.C., whose private name was Tan, but doubts about its authenticity rose as early as the 5th century A.D. Most modern scholars have rejected the tradition. Some, contending that the literary style, the terminology (such as "kings and barons"), and ideas attacking ceremonies and righteousness belonged to the 4th century B.C., have attributed it to a historian of that time, whose private name was also Tan. But since the style, etc., are not confined to that period, the arguments are not conclusive. More and more scholars believe that the basic doctrines were taught by Lao-tzu but that it was a long time after his death, perhaps some time in the 4th century B.C., that the book took form.

There have been more commentaries on the *Tao-te ching* than on any other Chinese classic—about 700 in all, half of which are extant. It is also the most frequently translated; there are more than 40 versions in English. No Chinese or Japanese can be considered educated without some knowledge of this little classic. *See* also TAOISM.

See *The Way and Its Power,* Eng. trans. by Arthur Waley (1934; paperback ed. 1958); *The Way of Lao Tzu,* Eng. trans. by Wing-tsit Chan (1963). (W. C.)

T'AO YUAN-MING: *see* T'AO CH'IEN.

TAPACOLO (TAPACULO), the name given to a group of small to medium-sized birds comprising the passerine family Rhinocryptidae, peculiar to Central and South America. The 25 known species occur chiefly in the Andes, and are most numerous in Chile. A single species (*Scytalopus argentifrons*) occurs as far north as Costa Rica. The several types agree in having large, strong feet, rounded wings and rather short tails which are carried in an upright position. The bill may be either thin and awllike or decidedly thick. Certain modifications of the external nostrils are characteristic of the family. In coloration the plumage is mainly brownish or blackish and, in one form, spotted. The typical tapacolo (*Rhinocrypta lanceolata*), a native of Argentina, is crested and resembles a miniature domestic fowl. The widely distributed, wrenlike genus *Scytalopus* includes more than half of the known forms. All are notably furtive, have weak powers of flight and frequent bushes or undergrowth. Their varied calls are sometimes musical, but consist chiefly of harsh scolding notes.

Tapacolos feed on insects, larvae and related matter. Their nesting habits vary, some forms (*Scytalopus*) making nests on the ground and others in bushes. One species (*Teledromus fuscus*) digs a tunnel. Two white, unmarked eggs comprise a clutch. (E. R. BE.)

TAP DANCE. Tap dancing consists of a special agility of the feet which can be incorporated into any form of dance movement. This agility is used to produce more foot-to-floor contacts than would normally exist. In tap dancing these contacts are made in precise, rhythmical patterns with or without musical accompaniment. They are made audible by attaching small metal plates to the toe and heel of the dancing shoe.

All ethnic groups have retained some form of emphatic foot-to-floor contacts as part of their dance culture. The main sources of tap dancing are the clog dancing of northern England, the reels and flings of Ireland and Scotland, and the rhythmic foot stamping of the African Negro.

The influx to the U.S. of English and Irish labour and of African slaves brought together these basic elements. The 19th-century performers in minstrel shows and showboats stylized and fused them into a form of dance and satire which rapidly became popular. Among the most noted of these performers were Ralph Keeler and Daniel Emmett and the dance team of Lynch and Diamond.

Near the end of the 19th century tap dancing acquired a specific technique of its own. This was divided into two parts known as "buck and wing," and "soft shoe." The buck and wing was very active and fast, was danced in wooden-soled shoes and made loud sounds. Soft shoe was very relaxed and smooth and was danced in soft shoes which made soft sounds. At the turn of the century, George Primrose was the most famous of the soft shoe dancers, while Doyle and Dixon were exponents of the buck and wing.

Between 1910 and 1925 these two varieties of tap dancing merged. Wooden-soled shoes and soft soles largely disappeared, and the modern metal tap made its appearance. The technique remained very much the same and consisted principally of: striking the ball of the foot on the floor with a brushing motion forward, sideward or backward; stepping on the ball of the foot; dropping or scuffing the heel on the floor; scraping the foot over the floor; touching the toe of the foot to the floor; and making contact with any part of the shoe which would produce a sound. Handclaps and slapping of the body were also occasionally used. These steps were combined with a great variety of skillful leg movements and were influenced also by the Spanish *zapateado,* Bavarian *schuhplattler* and many other folk dances.

Some of the tap dancers who began dancing during this period were Pat Rooney, George M. Cohan, the Four Fords, Johnny Boyle, Tom Patricola and the great Bill Robinson, known as "Bojangles." In the 1930s tap was very popular as a medium for chorus dancing and in vaudeville.

Since 1940 tap dancing has added to its basic technique many of the movements of ballet and modern dancing. The footwork has also developed so that an expert tap dancer can make about 64 taps in five seconds and, standing on one leg, can make five tap sounds with a single movement of the supporting leg.

As it became more visually stimulating it was widely used in films and television. Fred Astaire, Gene Kelly, Ray Bolger and Eleanor Powell have been outstanding in these fields, and have added grace, humour and skill to this form of dance. A *Concerto for Tap Dancer* was composed by Morton Gould, to be performed with orchestra. Paul Draper introduced tap dancing in the concert field with classical and jazz music. George Tapps and Rod Strong used groups in this same medium.

There is widespread interest at all age levels in tap dancing and about 50,000 dance schools throughout the United States include it as part of their training. It is estimated that between 4,000,000 and 5,000,000 people study tap dancing. The dance teachers Ned Wayburn and Jack Manning did much to popularize this. The basic steps are comparatively simple to learn, but to become skillful requires years of concentrated study.

With all of its jumps, spins and broad movements, the essence of tap dancing remains the imaginative invention and exact execution of audible rhythmic patterns. It is an art form that has been developed principally in the U.S., and though much changed from its folk origins, it is firmly a part of American culture. (P. DR.)

TAPE RECORDING, MAGNETIC. The process of magnetic tape recording offers one of the most accurate and convenient methods of storing, and later reproducing, any type of information that can be converted into electrical signals. Sounds (*e.g.,* voice and music), electrical data from business machines and metering devices, video signals and the electrical impulses for the automatic control of machinery are examples of the material commonly stored on magnetic tape. Besides the accuracy and convenience of magnetic recording, the ability to erase the storage medium and reuse it many times gives a great economic advantage over other recording methods in many applications.

History.—The idea of storing electrical information by means of magnetized particles was expressed as early as 1888 by Oberlin Smith in an article in *The Electrical World.* However, Valdemar Poulsen, a Danish inventor, was the first to reduce the idea to practice with a recorder which stored the electrical signal on a wire of magnetic steel. He first applied for a Danish patent on this device on Dec. 1, 1898. A working model of the device created great interest at the Paris exposition in 1900, and many of the later applications of magnetic recording were envisioned at that time. Even with this enthusiasm, however, Poulsen could not find financial backers in Europe. In 1903, Poulsen and some American associates formed the American Telegraphone company for the manufacture and sale of an improved version of the device. The telegraphone would record continuously for 30 minutes on a length of piano-wire steel that moved at a speed of 84 in. per second. However, the device did not find wide application.

Following the failure of Poulsen's company, the general public heard little of magnetic recording for several years, but interest continued in many laboratories. W. L. Carlson and G. W. Car-

penter of the U.S. Naval Research laboratory applied for a patent on alternating-current bias for magnetic recording in 1921, and in 1927 J. A. O'Neill received the first U.S. patent for a magnetic tape made by drying a liquid containing magnetic particles on the surface of a strip of paper. A German patent by Fritz Pfleumer of Dresden was issued in 1928 for a similar process. Until this time, the recording medium had been a wire or tape, entirely of or plated with magnetic metals.

During the 1920s Kurt Stille of Germany acquired a number of patents in this field through which he licensed others. One such licensee was Louis Blattner of England, who built a magnetic recording machine to record and synchronize sound with motion pictures. The "Blattnerphone" was used by the British Broadcasting corporation, and later it was bought by the Marconi company. Work on this method of recording progressed in Germany, England and the U.S. during the 1930s. In particular, the Magnetophon company of Germany succeeded in producing a magnetic, film-coated tape, based on Pfleumer's patent, that constituted a low-cost recording medium. During World War II the Magnetophon, using the coated tape, served the Nazi propaganda organizations well, and following the war its principles were copied and improved upon in many countries, with Germany, England and the U.S. the leaders in this work.

Recording Media.—Magnetic materials in almost every possible form are used in one magnetic recording application or another. Drums of nonmagnetic material have their surfaces coated with a magnetic material to record short bursts of information, such as those used in electronic computers and some types of data recording. Cylindrical surfaces of rubber or plastic, impregnated with magnetic particles, are used to record short messages in telephone answering and recording machines. Disks, coated or impregnated with magnetic material and recorded in parallel or spiral tracks, are used in such applications as office dictating machines and electronic computers.

Wire, the first magnetic recording medium, is still used for certain applications, such as portable recorders. Despite improvements in tape, drums, etc., more information can be stored in a given space on wire than on any other medium. However, the fidelity of reproduction, the signal-to-noise ratio, and cross talk (resulting from adjacent recorded layers being wound together on the same spool) are generally poorer than those of tape.

A variety of magnetic tapes have been developed. Some of the early machines used a solid metal tape of magnetic material. A magnetic stainless steel, Vicalloy, consisting of 52% cobalt, 11% vanadium and 37% iron, is particularly useful for both wire and tape. However, a paper or plastic tape coated with a magnetic powder proved to be the most useful form of magnetic recording material and far outstripped the others in variety of application as well as volume produced.

The magnetic material most often used in coating tape is the red form of iron oxide, Fe_2O_3. The way in which this powder is prepared, the impurities deliberately introduced or accidentally present, and the physical size and shape of the particles have much to do with the characteristics obtained. Red iron oxide particles, needle-shaped and roughly one micron ($\frac{1}{1.000.000}$ m.) in length, are widely used. The black oxide, Fe_3O_4, has also been used as a magnetic coating material; it produces a higher signal strength, but is more difficult to erase.

The magnetic powder is thoroughly mixed with a binder, the nature of the latter depending partly on the base material to which it will be applied. The polymeric vinyl chlorides, vinyl acetate, cellulose acetate and ethyl cellulose are commonly used. The solvent employed must not attack the base material or cause adjacent layers of tape to stick to each other when the tape is tightly wound and stored at relatively high temperatures. The object is to obtain a very thin but dense layer of iron oxide on the tape in a very uniform distribution. It must adhere tightly to the base tape, but not to an adjacent layer; it must be flexible, and preferably its characteristics should not change appreciably over many thousands of usages or many years of storage.

The base material may be paper or any one of several plastics. Paper is generally lower in cost and some papers are more stable than plastics in applications where relative humidity and temperature vary widely. However, paper tears easily under tension, and since any irregularity in the thickness of the magnetic layer results in unwanted variations and noise in the recording, even the smoothest paper has a serious disadvantage in performance as compared with plastic, which can be prepared with very smooth surfaces.

Cellulose acetate film can be manufactured at moderate cost with a smooth surface, and it has reasonably good tensile strength, tear resistance and dimensional stability. It is widely used for magnetic tape in thicknesses of about 0.0015 in. Mylar or Terylene (a polyester made from polyethylene terephthalate) is more expensive than cellulose acetate, but has superior characteristics. It is stronger, and more stable under high humidity and temperature conditions. It is also more stable with respect to time, making it possible to store recorded material for many years. It has sufficient strength and stability to be used in films as thin as 0.0005 in., thus greatly reducing the storage space required for a given length of tape. However, a smooth surface is somewhat more difficult to obtain on Mylar than on cellulose acetate.

A typical method of manufacturing tape begins with combining the iron oxide and the binder. The mixture is milled in a large rotating cylinder loaded with steel balls, a process that may require from one to several days. The resulting mixture is an extremely uniform, thick fluid resembling printer's ink. This is filtered to remove any lumps of more than about one micron in diameter.

The paper or plastic film is usually processed in webs up to 20 in. wide. The coating process is done in very clean rooms where only filtered air can enter. The web passes from its roll through a cleaning stage in which both surfaces are cleaned of every trace of dust, lint and other foreign matter; this is accomplished with vacuum suction action, assisted by static removal equipment. The web is then coated with a thin layer of solvent to make the surface receptive to the coating material. When this has had sufficient time to act on the surface, the magnetic coating is applied. This may be done in a fashion similar to engraved-roll printing, in which the amount of material to be applied is metered by the depth of the pattern engraved in the surface of a printing plate. The magnetic mix is applied to the engraved roll and the surface rubbed or scraped to remove the excess. The amount remaining in the engraved pattern is then transferred to the tape web by a rolling action, and the web continues through a drying chamber. When it is dry, the surface may receive a polishing, after which the tape web is ready to be slit and reeled.

The coating remaining on the web is 0.0001 to 0.0004 in. thick when dry, and tolerances of ± 0.000012 in. may be held on the nominal thickness to assure proper performance. For tapes that are used for computers, automatic inspection devices detect any voids in the coating, since these could cause an omission in the stored information, resulting in a serious error.

The tape is slit and wound on reels varying from 3 to 14 in. or more in diameter, depending on the intended application. A 7-in. (outside diameter) reel is common in sound recording work and holds 1,200 to 1,800 ft. depending on the tape thickness.

A 10½-in. reel (2,400 to 3,600 ft.) is widely used by the broadcast industry, and larger reels are frequently used in technical data recording and business-machine operations. Nominal tape widths most commonly used are ¼, ½ and 1 in., with the ¼-in. size having the widest variety of applications.

The Recording Process.—The electrical signal to be recorded is usually applied to the tape by means of a "head" consisting of a coil wound around a core of magnetic iron which has a gap at the point where the tape moves across its surface. The current in the coil produces a magnetomotive force across the gap, magnetizing the particles in the tape. The strength and direction of final magnetization of a particular point on the tape will be determined by the strength and direction of the magnetic field at the moment that point crosses the gap. Thus the tape receives a magnetic record containing direction (north or south pole), amplitude, and linear dimension (along the tape) corresponding to the direction (plus or minus), amplitude and time of the original elec-

trical signal. The exactness of this record depends on a variety of conditions, one of the most important being the original magnetic condition of the tape. There are two conditions which give predictable results. Either the tape should be completely demagnetized before recording, or it should be completely magnetized in one direction or the other. In early wire and tape recording, the latter condition was used. The tape was carried past a permanent magnet so arranged that it brought the magnetic particles to saturation, leaving the tape fully and uniformly magnetized in one direction. The recording head then had a direct-current component added to the signal; this direct-current bias was of the proper direction and amplitude to bring the tape just back to a neutral state. The signal, adding and subtracting from this bias, left the tape magnetized by an amount proportional to its own strength. This is called D.C.-biased recording, and gives comparatively low distortion and good amplitude. However, it has a higher inherent noise level and more distortion than A.C.-biased recording.

In A.C.-biased recording the tape is first completely demagnetized and comes to the recording head in a neutral state. There an alternating current is added to the signal current; it is much higher in amplitude than the signal current, and several times greater than the frequency of the highest-frequency signal to be recorded. In crossing the gap of the recording head the magnetic particles on the tape are subjected to many cycles of the bias current, and leave the head with a magnetization which is proportional to the signal. Very low distortion and high signal-to-noise ratios can be obtained with A.C.-biased recording. For sound and general signal recording, this type of tape recording gives performance superior to that of any other recording method.

For pulse recording, in electronic computers and other digital information devices, distortion is of no importance. It is simply desired to record and later detect the presence of a pulse. In most of these recorders the tape comes to the recording head in a neutral state.

In one system the head may normally carry a current that magnetizes the tape to saturation in one direction. A pulse then carries the current to the other extreme and back again, leaving a short area on the tape magnetized in the opposite direction. This is called two-level recording.

In the other commonly used system the head normally carries no current, and pulses may carry the magnetization of the tape to saturation in either a plus or minus direction. This is called three-level recording.

Losses are encountered in recording through eddy current and resistive and capacitive losses in the windings and core of the head. Space between the head and tape causes very high losses, particularly at high frequencies (short wave lengths). The gap must be very straight and true, and the surfaces in contact with the tape must be very smooth in order to obtain the best possible transfer of the signal pattern onto the tape. Optical polishing methods are commonly used for finishing the surfaces of the gaps and faces of the head that will be in contact with the tape. The gaps used in recording heads vary in length (direction of tape motion) from 0.0002 in., when used also as a reproducing head, to 0.010 in. or more for specialized applications. The tolerances on the trueness of the faces and edges of the gap are usually held to a few millionths of an inch. The width of the cores, and the resulting recorded tracks on the tape, vary from approximately 0.025 in. for certain types of data recording to 0.250 in. for professional audio recording.

Reproduction.—In reproduction, it is usually desired to use the recorded tape to reconstruct the signal, obtaining the same polarity, amplitude and time relationship that originally existed in the recorded signal. To accomplish this, the tape is passed over the same head used in recording, or one similar to it. To reproduce the signal in the proper time relationship, the tape speed must be identical during recording and reproduction.

The magnetized portions of the tape passing over the gap in the reproducing head cause the magnetic flux in the core to change, generating a voltage in the coil. This voltage is proportional to the speed of the tape and the amplitude and frequency of recorded signal. Assuming that the tape speed is constant and that a constant-amplitude signal has been recorded, the reproduced signal will have an amplitude proportional to the frequency of the reproduced signal. This gives a response which rises at the rate of 6 decibels per octave. However, as the length of the waves recorded on the tape become smaller, the gap in the reproducing head spans more and more of a wave length, and a loss in output is encountered which is known as "gap loss." At the frequency where the gap just spans one wave length of the recorded signal, the output is zero. In order to use the tape at as slow a speed as practical and still record high frequencies (short wave lengths of recorded signal), the reproducing head has a gap made as short as the technical progress of the art, and economics, will allow.

The amplifier that obtains its signal from the reproducing head must have a response curve which compensates for the special characteristic of that head, as well as for other losses in the heads and recording process. Some of these losses can be anticipated and compensated for in the recording amplifier. Thus it is necessary to have standards established before tapes can be freely exchanged from one recording machine to another reproducing machine. This has been done to some extent in the audio field, standards being established by industry committees.

In television and data recording and in machine control the frequency response problems are quite specialized and the tapes are usually recorded and reproduced on the same machine or a similar model. Few standards have been established since they are not seriously needed for tape interchanges. In business machine applications the problems are even more specialized. Here, it is not necessary to reproduce exactly what was recorded, but merely to detect and interpret each "bit" or pulse of recorded information. The reproduced signal generally does not resemble the recorded pulse, but does bear a definite relationship to it.

Erasure.—One of the features of magnetic recording which gives it an economic advantage over other recording methods is the ability to erase and reuse the recording medium thousands of times with little or no loss in the quality of recording. Erasure may be accomplished in either of two ways: (1) A reel of tape may be erased by subjecting the whole reel to a strong alternating magnetic field and slowly reducing the field to zero; this is called bulk erasing, and is similar to the method used by a jeweler when demagnetizing a watch, although the equipment must be of a different form to accommodate the large reels of tape. (2) The second method involves using a head, very similar to a recording head, to erase the tape just before it is recorded again; the tape passes over the gap in the head and is subjected to a powerful, high-frequency alternating field which diminishes as the tape passes away from the gap, leaving the tape in a demagnetized or neutral state. Professional sound-recording engineers frequently use both methods to make sure all the previous recording has been removed from a tape before it is used again.

Tape Recording Machines.—A wide variety of mechanisms has been devised to make use of tape recording in many different applications. Since only a small amount of power is required to move the light, flexible tape over the heads for recording or for playing back, it has been possible to make battery- and spring-driven portable mechanisms to record away from established power sources; the smallest of these are pocket-sized. The miniature machines, however, sacrifice some fidelity of reproduction to attain compactness. In particular, it is difficult to attain smooth motion and uniform speed, with the result that "flutter" and change of pitch can be noted in the playback of audio material. Larger portable units are available which give quite high fidelity and are often used for broadcast program recording in the field.

A large market developed in tape recorders for use in the home. These record and play tape at 1.87, 3.75 and 7.5 in. per second, usually on one-half of the width of the $\frac{1}{4}$-in. tape; this doubles the amount of tape available for recording or makes it possible to have two simultaneous channels for stereophonic recording. Prerecorded tapes are available, but the ease with which the amateur can produce acceptable recordings is a strong factor in the popularity of such recording devices. Some machines are designed to record four tracks on a $\frac{1}{4}$-in. tape, allowing even more inten-

sive use of the tape with only a slight further sacrifice in the quality of reproduction.

Sound recorders for special applications are often quite complex. Those used for professional studio recording usually have three motors: one on the supply reel, one on the take-up reel and a third to drive the tape. Extreme care is taken to provide smooth and uniform drive so that reproduction of musical notes will remain in exact pitch with no discernible flutter. Another special application is voice-monitoring service, such as the monitoring of radio channels at an airport. Here, a 2-in. tape is used and a rotating head scans across the tape while the tape moves slowly forward. This makes possible a very long recording time on a comparatively short tape, and a particular passage may be returned to with a minimum of rewinding time.

The same principle, with much elaboration and much higher scanning speeds (1,500 in. per second), is in use for recording television broadcasts. The tape recorders used in data recording are similar to those used in professional audio recording. However, they may utilize wider tapes, and 10, 20 or more tracks may be provided by special multicore heads. In addition, the machines must be capable of starting and stopping the tape in very short intervals of time, and of driving the tape in either direction. The very low inertia of a short piece of tape makes it possible to attain speeds of 15, 30 or 60 in. per second in thousandths of a second. Various fast-acting capstans and brake systems are in use to start and stop the tape, and a slack loop is usually maintained between the reels and the working area by means of servomechanisms. *See* also PHONOGRAPH.

BIBLIOGRAPHY.—Joel Tall, *Techniques of Magnetic Recording* (1958); S. J. Begun, *Magnetic Recording* (1949); H. G. M. Spratt, *Magnetic Tape Recording* (1958); N. M. Haynes, *Elements of Magnetic Tape Recording* (1957); W. Earl Stewart, *Magnetic Recording Techniques* (1958). (W. E. Sr.)

TAPESTRY. Tapestry is popularly assumed to be almost any heavy material, ordinarily wool, hand-woven, machine-woven, or even embroidered, used to cover furniture or walls; and this definition is justified in that it is the older usage, is the definition of the Latin *tapes* and is still reflected in the parallel French derivative noun *tapissier,* "upholsterer."

In the 18th and 19th centuries the definition was narrowed to mean a close-textured plain or twill weave, with the filler, or wefts, so outnumbering the warp yarns as to conceal them completely. By various devices (sword tip, reed, batten or comb) the weft was pushed together so tightly as to create a solid, weft-faced surface.

Thus, in the 20th century, Queen Matilda's Bayeux tapestry of embroidery is considered a hanging, not a tapestry. Tapestry weaving differs from embroidery in that it is freely darned on a bare warp, instead of being wrought upon an already woven material.

There is, however, a strong conviction among textile authorities that the rigid definition of a tapestry as a weft-faced cloth must be amended. They agree that tapestrylike fabrics from the past, such as the mosaiclike twill of Kashmir shawls, the weft-twining of Chilcat blankets and the weft-wrapping of Sumak rugs are, technically, tapestry. Terms like "twill tapestry" and "twined tapestry" are often used to designate these tapestrylike fabrics in weaves other than plain weave.

TECHNIQUE

Tapestry differs from other forms of patterned weaving in that no wefts are carried the full width of the web, except by a very rare accident of design. Each unit of the pattern or the background is woven with a weft of the required colour, which is inserted back and forth only over the spot or section where that colour appears in the pattern. Typically, in plain-cloth technique, the weft passes over and under the warps alternately, and on the return shoot, under, where before it was over, and vice versa. The wefts cover the warps completely, which appear only as more or less marked parallel ridges in the texture—according to their coarseness or fineness.

Where the margin of a colour area is straight and parallel to the warps, a slit is necessarily left, which may be treated in any one of five different ways. First, it may simply be left open, as in the Chinese silk tapestries, which are called *k'ssu,* "cut silk," for that reason. Second, it may be left open on the loom, but sewed up afterward, as in European tapestries from the 14th to the 17th centuries, and also in some later types. Third, the weaver may dovetail his wefts, passing from one side, and from the other, in turn over a common warp. This may be either "comb" dovetailing—single wefts alternating—or "saw-tooth" dovetailing—clusters now from one side, now the other. Dovetailing has the double disadvantage of making the fabric heavier at that point and of blurring the outline. Persian weavers of the 16th century developed a successful variant in silk tapestry rugs, whereby a black outline weft was dovetailed over two warps—one of each of the adjacent colour areas—effectively hiding the coloured wefts in the compacting of the weave, and providing strong clear drawing. The same device is found several centuries earlier in Peru, but if there was a common source, examples from it have not yet been noted.

The fourth treatment—interlocking—was introduced in the Gobelins factory in the 18th century. Here wefts of juxtaposed colour-segments are looped through each other between the two warps which mark, respectively, the margin of each colour. This gives a continuous surface of even weight and was prized by the French weavers because the resultant effect more closely approximated that of painting.

Finally, one curious variant would, in strictest definition, fall outside tapestry, though it starts with the tapestry conception: between every two rows of tapestry-woven wefts is a cloth-woven, full-width shoot which makes the fabric solid. This type could be rigorously classified as a brocaded cloth, but the idea is that of tapestry, with the cloth structure subordinate and the problem solved being one in tapestry weave. The device, rare everywhere, is found in Japan in the 7th and 8th centuries, in east Persia in the 10th century (Moore collection, Yale university), and in medieval Peru.

Haute-lisse and Basse-lisse.—European tapestry may be woven on either of two types of looms, *haute-lisse* or *basse-lisse.* In a primitive version of the former, the warps are attached to a beam at the top, and groups of warp threads are weighted at the bottom. The weft is beaten up toward the top. Such looms are pictured on Greek vases; the weaver shown is generally labeled Penelope and her fabled scheme of undoing the day's work at night becomes clear: she simply removed the weights and slipped the wefts off the warp threads. The more highly evolved vertical frame has heavy uprights (*bâti,* in French) holding a horizontal roller (*ensouple*) top and bottom, on which the warps are stretched. Each warp passes through a loop of cord (the *lisses*), and the loops encircling the warps which correspond to uneven numbers are fastened to one slender cylinder, those to the even, to a second cylinder, both being above the weaver but within reach so that he can pull forward with his left hand now one, now the other set of warps (*i.e.,* form the shed), in order to pass his bobbin (*broche*) behind them, this being a short, pointed, slim cylinder of polished wood on which his weft yarn is wound.

The *basse-lisse* loom, on the other hand, has the rollers on the same level at table height so that the warps stretched between them are horizontal, and the *lisses* are attached to two slats, each of which is connected with a treadle so that the weaver depresses the odd-numbered, or even-numbered series of warps to form his shed with movements of his feet, leaving both hands free to manipulate his bobbin. Production on this loom is nearly twice as fast as on the vertical loom, reducing cost proportionately. The cylinders in both instances serve to roll up the finished portion and unroll a further length of unwoven warps so that the section in process is always taut and in a convenient relation to the weaver. At both types of loom the weaver works from the back side and has a mirror into which he can look through the unwoven warps, and which is placed to reflect the portion in process; but while the *haute-lisse* weaver can examine his finished work directly by walking round to the other side of his loom, the *basse-lisse* worker has to tilt up his frame.

The design is rendered first as a drawing on paper, the *petit*

patron; then it is enlarged to the size which the tapestry is to be, and is painted in exactly the correct colours on linen (later often paper). This is called the *grand patron.* Up into the early 16th century the *patron,* or cartoon, was sometimes called in French the *portrait* (or *pourtraict*). The artist specializing in designing (whether for execution in tapestry, or in other techniques and media such as embroidery, wood carving and stained glass) was called a *portraitist.*

The *grand patron* was sometimes itself hung as a substitute for tapestry; and large-scale paintings similarly rendered on great sheets of heavy but flexible linen, called *toiles peintes,* or "painted linens," were made expressly for this purpose. The most interesting collection of old *toiles peintes,* dating chiefly from the 16th century, belongs to the cathedral of Reims.

The *haute-lisse* weaver has the *grand patron* which he is following hung beside him, but the *basse-lisse* worker has his laid under his warps, so he follows from immediately above his model. In both cases the main outlines are drawn with ink on the warps.

The design is executed, in all professional European work, at right angles on the loom so that in the finished hanging the warps run horizontally. This is aesthetically advantageous, since the warp ribbing tends to create a texture more or less reinforced by linear shadows, and if vertical, these sever the design, but if horizontal they bind it into continuity. Practically, however, horizontal warps are disadvantageous since they mean horizontal slits, with weight from the heavy wool fabric pulling on them as the tapestry hangs, and thus accelerating deterioration. That the aesthetic consideration should have prevailed over the practical is noteworthy.

The extra weft ends are left hanging loose on European tapestries, but they can be fixed with an additional twist and cut short, making a reversible fabric. This is a practice especially common in far eastern tapestry weaving, which is notably fine (*k'ssu* may count 60 warp threads per inch). All European tapestries are relatively coarse up to the 18th century with, in general, from 10 to 18 warps to the inch; transition from shade to shade is effected with irregular "comb-tooth" hatching.

Twill Tapestries.—Instead of the plain-cloth binding commonly found in tapestries, a twill can be used; *i.e.,* a weave in which the weft is floated over two or more warps, then under one or more warps, with this underpass shifted always one to the right or left making diagonal ribbing. Such a weave appears in tapestry for the first time, as far as can be determined, in Persia in a unique silk and metal-thread square, the design comprising a royal portrait, figures of angels, animals and plant motives; and from the 17th century on (Moore collection, Yale university; Metropolitan museum) it is used in goat's hair or wool, especially for shawls, in Khurasan and Kerman. The type is perpetuated in the well-known Kashmir shawls, and was probably introduced into Kashmir from Persia, along with other crafts, in the 16th century.

HISTORY

Early Examples.—The earliest known specimens of tapestry weaving of linen can be dated between 1483 and 1411 B.C. One bears the cartouche of Thutmose III, another of Amenhotep II, and the third, hieroglyphics; they were found in the tomb of Thutmose IV. The Amenhotep piece has the tablet bearing the king's name and titles upheld by uraei, set in a powdered pattern of lotus and papyrus flowers, with a narrow border on each side, one with lotus flowers and buds, the other with disks. Only the designs—in red, yellow, green, blue and brown—are tapestry-woven, the ground being plain cloth; in the tapestry areas the wefts are not kept straight, but curve to follow the drawing. Thus the work is in effect bobbin-darning on warps left bare for the purpose.

Just at this period numerous Syrian ideas, techniques and perhaps actual craftsmen were entering Egypt. That the tapestry craft was not indigenous to Egypt can be assumed from the fact that, while countless mummy wrappings of the succeeding centuries have been unwound, no further instances of tapestry weaving are known to have been found until those of the 4th century A.D., a lapse of approximately 1,800 years.

Fragments of wool tapestries in graves on the Kerch peninsula, the 4th or 3rd century B.C., carry stag heads and ducks in the Hellenistic style which was widely diffused but especially strong in Syria, and about 200 to 500 years thereafter tapestry is represented again in a silk fragment, which though found at Loulan in central Turkistan has close Syrian connections. This is a band showing in seven colours (three browns, one evidently originally purple, two greens, light blue and crimson). In a central zone appears a scrolling stylized tree, in the crown of which appears a bird's head; the whole designed to present alternatively a conventional bird flanked by confronted griffins (composed of a crested bird with deer or horse feet) alternating with a symmetrical scroll. The border motive is a rosebud rendered as a heart in a deep calyx, and some of the multiple guard stripes are in continuously graduated polychromy. The bird nesting atop a cosmological tree, and the double image are both west Asiatic; fantastic griffin variations are a persistent feature of Syrian design; the heart-rose is the standard border theme on later Syrian textiles, where also compound borders with variant guard stripes are frequent, while the unusual graduated polychromy appears on approximately contemporary wool tapestries at both Palmyra and Dura-Europos. Thus this tattered strip represents the cultural linkage between the Mediterranean coast of Asia and its very heart.

The wool tapestry with graded polychromy found at Palmyra is ornamented only with groups of square dots, but that from Dura bears stylized lilies, and this figure, better drawn, appears also on wool tapestries found at Loulan which likewise may be assigned either to Syria or Persia. A closely related piece is known from south Russia. Tapestry weaving may well have been common in western Asia and probably especially in Syria at the turn of the Christian era, and had climatic conditions been less unfavourable a whole phase of the art now irretrievably lost might have survived.

Egypt.—The favourable conditions did exist, on the other hand, in Egypt, and consequently a mass of tapestry fragments has been recovered there, dating from the 4th to about the 10th century A.D., and especially numerous from the 5th to the 7th century. These are mostly garment trimmings, woven into or (in inferior specimens) sewed onto tunics and cloaks. Tunic trimmings (*see* VESTMENTS, ECCLESIASTICAL) when complete consist of stripes over the shoulders (*clavi*) down front and back, varying in width from about 1 to 6 in., ranging in length from just below the belt to the very bottom; a neck piece; panels on the shoulders and near the knees which may be round (*orbiculus*) or rectangular (*tablion*); and cuffs. Parts of pillows also are quite common, wall hangings far less frequent, and when in solid tapestry weave—in contrast to tapestry motives inwoven in cloth—rare and proportionately important. The last, notably a series of fragments in the Berlin museum representing dancers, in polychromy and silhouette, and sometimes the pillow panels, are naturalistic illustrations in a delayed Hellenistic style or subsequently in the early Byzantine style (best examples, Dumbarton Oaks, Washington, D.C.).

The garment trimmings fall into two main classes: indigenous designs rendered either in polychromy, or more often in silhouette, most typically in some tone of purple (dyed with madder plus indigo); and designs copied, with more or less distortions, from Syrian shuttle-woven silk patterns (a distinction first established by Josef Strzygowski, and repeatedly noted by Oskar Wulff and Wolfgang Volbach). In the former, constantly recurring themes are: the vine, or less often ivy; the amphora, frequently with the vine growing out of it; nude women, nude youth or boys, including dancers often armed; and animals, above all the lion and the hare —all explicable as descended from the Osiris cult and appropriate to burial robes because of their relevance to revival in a life after death.

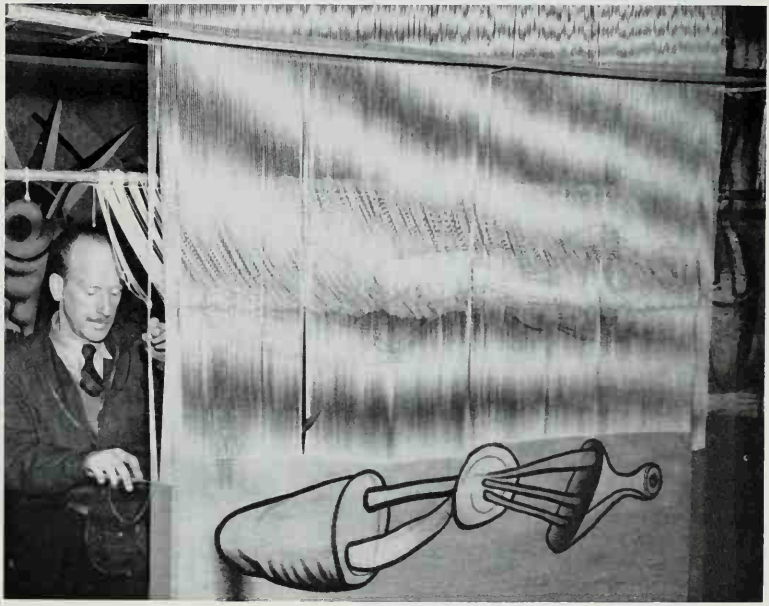
In the patterns copied from Syrian models, the styles of the different great textile-producing cities there can be recognized, and subjects from the local cults, so that, analyzed in their wider relations, they throw considerable light on the history of religion, as well as of art.

All types are woven with woolen wefts on linen warps, though

Weaving on a *haute-lisse* loom, one of the two types used by European tapestry makers. This is a vertical frame with the warps stretched on a horizontal roller

The *basse-lisse* loom, a horizontal frame with a foot-operated treadle for manipulating the warps. Production is faster with the *basse-lisse* than with the *haute-lisse*

PRINCIPAL TYPES OF LOOMS

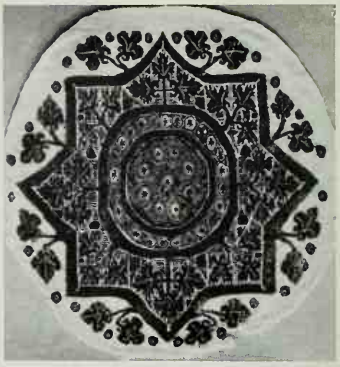


A 20th-century master weaver, Ronald Cruickshank, weaving a design by Henry Moore on a hand loom. The inked outline of the design on the warps may be seen directly above the area being woven

BY COURTESY OF (TOP LEFT, TOP RIGHT) JEAN LURÇAT, PHOTOS BY JULIETTE LASSERRE, (BOTTOM), RONALD CRUICKSHANK

TAPESTRY

Coptic, a violet silhouette showing a shepherd milking a goat; 3rd or 4th century A.D. The designs of the border (doves flanking vase of water-of-life holding vine), and motives of tapestries shown below and at right, were derived from the immortality symbolism of the Osiris cult



Violet silhouette, an abstract design with star and vine motive. This and the other Coptic tapestries on this page were found in Egyptian graves

Fragment of a polychrome portrait medallion. Probably of Syrian origin

Violet silhouette with polychrome details, showing a hare eating grapes. Coptic, 4th–7th century A.D.

Rug made by Inca weavers of Peru during the colonial period. Made of undyed cotton and wool

COPTIC TAPESTRIES AND A PERUVIAN RUG

BY COURTESY OF (TOP LEFT) CITY ART MUSEUM, ST. LOUIS, (TOP RIGHT) TEXTILE MUSEUM OF THE DISTRICT OF COLUMBIA, (CENTRE RIGHT) DETROIT INSTITUTE OF ARTS (BOTTOM LEFT) MUSEUM OF FINE ARTS, BOSTON, (BOTTOM RIGHT) THE BROOKLYN MUSEUM

A CHINESE
SILK TAPESTRY

Peony flowers and foliage in *k'ssu* (silk with gold thread used in the weft). Probably Sung dynasty (960–1279). In the Freer Gallery of Art, Washington, D.C.

BY COURTESY OF THE SMITHSONIAN INSTITUTION

TAPESTRY

BY COURTESY OF THE METROPOLITAN MUSEUM OF ART, N.Y

THE UNICORN IN CAPTIVITY

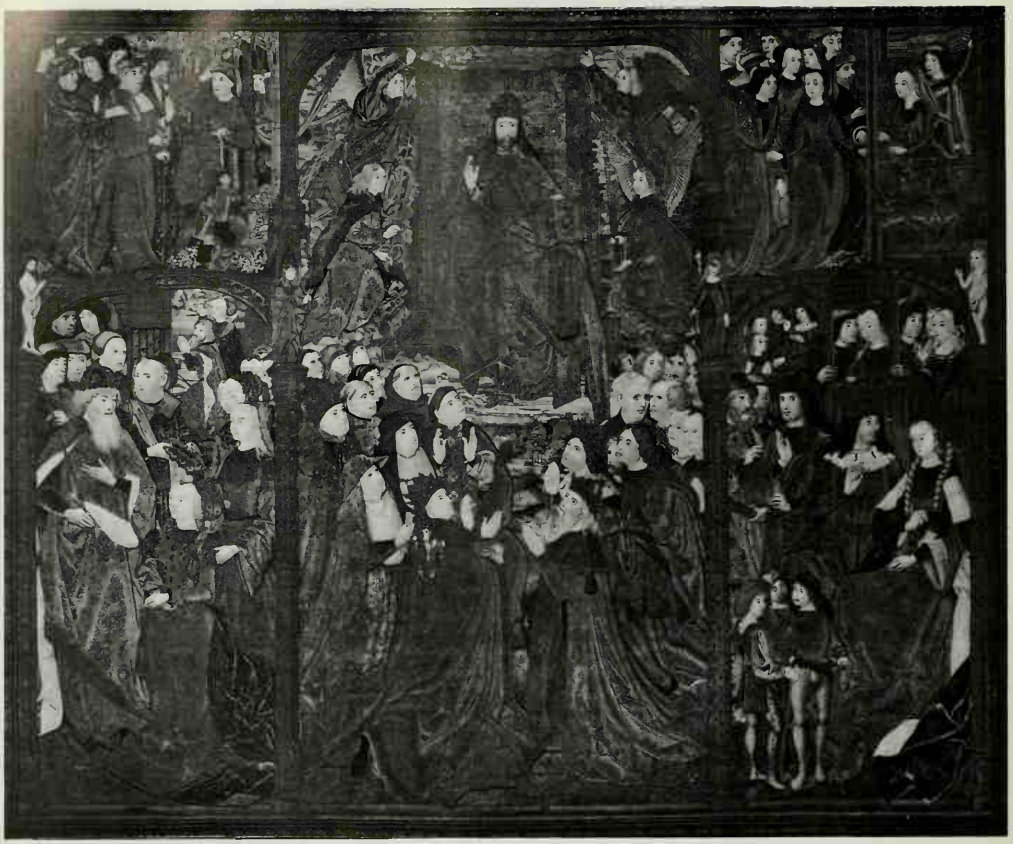
One of a series of Gothic tapestries showing the "Hunt of the Unicorn." French or Flemish. A date of about 1515 has been assigned to the series, but it has also been suggested that the set, or part of it, was prepared for the wedding of Anne of Brittany to Louis XII of France in 1499. In The Cloisters, New York city; a gift of John D. Rockefeller. 1937

BY COURTESY OF THE METROPOLITAN MUSEUM OF ART, NEW YORK

FRANCO-FLEMISH TAPESTRY

"Courtiers with Roses," the second of a set of three tapestries made towards the middle of the 15th century, perhaps at the order of Charles VII. It is woven of wool in a coarse texture of about 12 warps to the inch. The tapestry, which recalls the paintings of Jean Foucquet, has great decorative qualities. Since its exhibition at the Louvre in 1904 it has been recognized as one of the most important tapestries of its period. It was purchased in 1909 by the Metropolitan Museum of Art, New York

TAPESTRY



BY COURTESY OF THE NATIONAL GALLERY OF ART, WASHINGTON, D.C.

FLEMISH TAPESTRY OF THE GOTHIC PERIOD

"Mazarin Tapestry With the Triumph of Christ," in the Widener collection, National Gallery of Art, Washington, D.C. The tapestry was probably made in Brussels about 1500

SCANDINAVIAN TAPESTRIES

"Korna" (Cows) by Marta Mäas-Fjetterström (d. 1941), Swedish, one of the principal northern tapestry artists of the 20th century

"The Feast of Herod," a mid-17th century Norwegian tapestry in a style typical of the period. Owned by Thor Furuholmen, Oslo

BY COURTESY OF (TOP) STUDIO OF MARTA MÄAS-FJETTERSTRÖM, (BOTTOM) SMITHSONIAN INSTITUTION

TAPESTRY

PHOTOGRAPH BY COURTESY OF SIDNEY JANIS GALLERY, NEW YORK

A 20TH-CENTURY TAPESTRY

"Femme, Fleur, Etoile" (Woman, Flower, Star) by Joan Miró (1893–), Spanish. The tapestry, in the Seagram building, New York city, is a reproduction of a painting. The tapestry work was supervised by Marie Cuttoli, Paris. From the collection of Joseph E. Seagram & Sons

a few with silk wefts have appeared, and cotton wefts are occasionally introduced to give a brighter white. Primarily in the 7th, and perhaps also the 8th century, the tapestry technique may be supplemented with embroidery, as in border margins, usually in split stitch. In a special variant, which is not true tapestry, the characteristic meanders or other geometric repeats are executed with a free bobbin, "free" because it follows the design without regard to consistency of weft direction, carrying a linen thread which stands out on a violet wool tapestry-woven (or repp) ground. The multiplicity of tones in polychrome pieces is noteworthy, especially from the 7th century on, and includes clear, delicate pastels.

Egyptian tapestry weaving deteriorated sharply from the 7th to about the 11th century, when it was somewhat improved within a very limited range. Into white (or dark green) linen garments were woven tapestry trimming bands in wool, or often in silk, occasionally with metal-thread enrichments, in groups ranging from a half-inch to a foot and more in width, with patterns of geometric enlacements, frequently enclosing small quadrupeds or birds (ducks) highly stylized, and inscriptions. Many of these merely simulate writing, but many are legible, in handsome angular Kufic scripts on earlier pieces, in cursive scripts later, giving religious phrases, or the names or titles of rulers, occasionally with the place name also (*tiraz*). A characteristic colour combination in silk is golden yellow and scarlet.

Syria.—From the 6th to 8th century and doubtless from then on, striking wool tapestries were being made in Syria corresponding in style to the contemporary shuttle-woven silks, with animals or birds (commonly cocks) in energetic heraldic stylization, framed in roundels, almost always on a red ground. And later, from the 11th to the 13th century, silk and gold-thread tapestries were produced in Syria, highly distinctive in style, illustrating one of the great pagan religions still active in the old Phoenician cities when the crusaders took them over.

Persia.—Fewer specimens of the Persian tapestries have survived, but one notable fragment (Moore collection at Yale university) bears an ibex in the impressive Sassanian style, and a single piece from the Seljuk period (11th century) establishes the maintenance of the technique. This reappears in the 16th century (intermediate examples having all apparently been destroyed) as the medium for rich silk and metal-thread rugs of which only three are known still to exist (Moore collection, N.Y.), though others are illustrated in Persian miniatures.

The modern descendants of these, occasionally silk but more often wool, in slit tapestry, with simpler patterns, are the familiar *kilims;* and the type extends, with variations in decorative style, through the Caucasus and Asia Minor into eastern Europe.

Far East.—Tapestry had long since been in use in China, and there silk was the typical material, in a fine, light texture, the weave finished perfectly on both sides so that the panels are reversible. The warps are vertical in relation to the pattern. The "metal" thread, which is quite common, is the thinnest possible gold leaf applied to tough paper, cut in narrow strips and spirally wound on a silk core.

The earliest surviving examples, dating from the T'ang and early Sung periods (8th to 10th century), show only simple repeating patterns—flowerets, vine elements, ducks, lions—but are probably not representative of the full development of the time since they come only from remote desert stations of central Asia. Subsequent examples from the Ming period (1368–1622) are rare and exquisite. Abundant from the Ch'ing period (1644–1912), they show a severe artistic and technical decline. Especially plentiful are squares indicating rank, from official robes.

Tapestry technique traveled from China to Japan in the Ashikaga period (15th century), where it was used for characteristic motives, also in silk as a rule, and with vertical warps, but with a considerably coarser texture than Chinese work. The history of the art in Korea remains obscure, but rather coarse wool tapestry-woven rugs from there, with stylized motives, are well known in western markets.

Western Hemisphere.—How and when tapestry and a number of other complex Asiatic textile techniques were transmitted to the Americas is not known. Not all authorities agree that the connection took place. But examples of a fully established technique and style appear, perhaps as early as A.D. 600–700 in the Andean culture, while a vast amount of remains have come to light from the coastal cemetery sites of Peru, datable from about the 8th to the 12th century. Belted looms without heddles were used.

Bodies were buried characteristically in a squatting posture, fully dressed, indeed often wearing a multiplicity of garments hardly feasible in life, and padded and heavily wrapped into an approximate ovoid. The principal articles are: tunics, for men, varying in length and with or without sleeves; for women, full length, or in the Incaic period, a long skirt held under the arms with a sash; for both, enveloping mantles; breechcloths for men and boys; belts; and headbands. Besides tapestry, these Peruvian remains reveal the most varied techniques of weaving and needlework ever current in a single culture. (Excellent collections: Lima museum; Brooklyn museum; American Museum of Natural History, N.Y.; Textile museum, Washington, D.C.)

The warps of the tapestries are of undyed cotton (the cotton of the region being naturally either white or brown), the wefts of wool—llama, guanaco, alpaca or vicuña—with cotton sometimes introduced for bright white. The range of available colours—made with natural dyes—is large, for the most part in saturated tones; the compositions are usually polychrome and powerful contrasts are preferred.

The patterns all derive from local cults: a feline divinity, in some presentations fantastically adorned; an anthropomorphic deity (or more likely several distinct personalities) ordinarily with an elaborate headdress and frequently carrying a staff; birds, including the eagle and the condor; and fish. Plant motives are comparatively rare. All are angularly conventionalized, with the rectangle and especially the diagonal as substructure, the style showing a marked preference for meanders and reciprocals. The latter range from the most rudimentary, the serration, with its section the Z or its magnification the chevron, to complex developments of overturned interlocking figures of naturalistic origin, notably condor heads.

In the later middle ages, the characteristic principles of conventionalization appear: elimination—as, omission of the mouth; fusion—as, nose and mouth merged into one spot; exaggeration—as, human profile reduced to a huge nose; and part-for-whole—as, the feline represented only by its claw. The application of these devices is unusually extreme.

Though a great advantage of tapestry is its flexibility which makes unnecessary repeating patterns, these strongly predominate in the Peruvian use of the technique in all periods. The organization of the repeats follows basic forms relevant to shuttle weaving: stripes; bands; and rectangular checks, implicit or explicit. Schemes unrelated to the fabric structure also are used: diagonal stripes, chevron stripes, lozenge checks. Within these, the succession whether of motives or of colour changes is often exceptionally complex. The usual unit is three, or its double, six—the latter a peculiarity of textile design in this region throughout its pre-Columbian history—and the permutations thereon are numerous, varied and not infrequently remarkably ingenious.

After the Spanish conquest, European designs were supplied to highly skilled Inca weavers who executed them with their own modifications, producing an effective hybrid style (good examples, Boston Museum of Fine Arts).

Perhaps tapestry was also current in the other highly evolved pre-Columbian cultures of Central America and Mexico; but climatic conditions there have been destructive of textiles.

Europe

Gothic Period.—Tapestry weaving does not appear in Europe until the Gothic period—though the technique was probably used in classical Greece and Rome—the oldest surviving examples dating from the 12th century, of German provenance (Halberstadt cathedral). These show the "History of Abraham," "Christ and the Apostles" and "Charlemagne Amidst Four Classical Scholars," all three in crude but vigorous drawing.

From the 14th century the history is continuous, and it was

then, in the Paris shops, that the highest aesthetic level was attained, as in the superb series—however maltreated by time and restorers—illustrating the Apocalypse, ordered in 1377 by the Duc d'Anjou from Nicolas Bataille, and based on cartoons by Jean de Bondol of Bruges (Angers museum). A dignified panel shows King Arthur as one of the "Seven Heroes" (*Sept Preux*) (formerly Mackay collection, now in the Metropolitan museum, N.Y.); and a related series has been acquired by the same institution.

From the 14th century tapestry was used not only for wall hangings in the houses of the rich, churches and secular public buildings, but also for covers for cushions, chairs, benches, tables, beds, canopies, horse and mule blankets, and sometimes as rugs.

In the 15th century the production centre shifted from Paris to the Low Countries, with major "schools," more or less successive yet overlapping, in Arras, Bruges, Tournai and Brussels. Only a dorsal, signed Pierre Feré, dated 1402, bears the name of Arras, but the town can perhaps be credited with a few damaged panels of the early 15th century, notably a series in the Musée des Arts Decoratifs, Paris, with small-scale ladies and gentlemen amusing themselves outdoors, against a dark-blue ground, strewn rather sparsely with little flowering plants and bushes. This type with or without personages, but usually with at least animals and birds, continued for more than a century, and was known as *menues verdures* ("small verdures"), but is called in the modern market *mille fleurs* ("thousand flowers"). These pieces have a childlike charm, and the same is true of the contemporary remnants of a series illustrating the lives of St. Piat and St. Eleuthère made for Tournai cathedral by Pierre Feré in 1402.

Within two or three decades, a distinguished and energetic style had evolved there, exemplified by the "Passion," of St. Mark's, Venice, which lay for generations disregarded under the roof of the basilica. Tournai production continued to be important through the 15th century, with variations characterized in general by vigorous silhouette drawing, effective narrative action and the necessary combination of carrying-power with detail for closer examination. The weaving is competent, especially the exploitation of slits in facial delineation; the colours are substantial to suit the heavy texture, but clear; and most of the dyes have held well, except for purples which have usually browned.

The dukes of Burgundy, enthusiastic tapestry collectors, were good customers, employing especially the Greniers. Famous existing sets include the Beauvais cathedral "Life of St. Peter" (1460); the widely scattered "Trojan Wars" (based not on Homer but on Raoul Lefèvre's *Recuyell,* 1420); the Saragossa "Esther"; and numerous hunting and peasant scenes, part of the then current vogue of pastoral romanticism.

The identification of the cartoon painters of the 15th century has been neglected, though the possibility of deciphering their "signatures" in decorative inscriptions on the tapestries themselves has been demonstrated. Arras, Bruges, Tournai and Brussels, with Antwerp as a probable fifth had more than 60 masters, of unequal merit and productivity; about 70% are known by name, with information on them available in guild and other documents. Outstanding in Tournai were the LeQuien and LeFeire families, and their respective pupils.

But the most famous cartoon painter whose work was executed on Tournai looms was Beaudouin Bailleul, called on to design perhaps the most important series of the time, the "History of Gideon," provided by Philip the Good, duke of Burgundy, for his new Order of the Golden Fleece. Bailleul had the unique honour of being mentioned by Jean Lemaire in his *Couronne Margaritique.* No trace of the "History of Gideon" was discovered.

The most conspicuous family of weavers at the turn of the 15th century was the Poissoniers of Tournai, who produced some exotic series inspired by Vasco da Gama's trips to the Indies, and a new style of *grande verdure* suggested by Indian printed (or painted) cottons. An upper border treated as a bell-fringed valance is a recurring feature.

The Bruges "school" has been overlooked, most of its productions being merged with those of Tournai; it displays a tonic ugliness in the Doria palace (Rome) "Alexander." More elegant, but still characteristic is the "Lady With the Unicorn" series, from the first decade of the 16th century, in the Cluny museum (Paris), widely familiar because various elements from it have been utilized for countless needle-point patterns. Six allegories of the Virgin are gracefully presented, with the romantic worldliness typical of the period, against *menue verdure* on a light-red ground.

Borders are lacking on practically all earlier tapestries (the "Passion" set in St. Mark's, Venice, is a fine exception), but rouleaux bearing descriptive texts often finish the top, or less often the bottom, while the multiple scenes are divided from each other and thus framed on the sides by slender columns, frequently depicted as embossed and jeweled gold.

At the end of the 15th century, and during the first quarter of the 16th, the Brussels looms became very active, relying for cartoons, however, on Antwerp painters. Jean van Roome, who designed the tomb sculptures of Marguerite of Austria and Philippe de Savoie at Bourg-en-Bresse, was responsible for a number of tapestry *patrons* (patterns). Various other namable painters used essentially the same manner: a Maître Knoëst, known from a signatory inscription; and Maître Philippe, finally identified as Philippe de Mol.

The compositions in this group are crowded with figures in bouffant costumes commonly unpatterned, so that the faces seem to be set in a mass of heavily modeled stuffs. This effect becomes understandable when the regulations of the guild, framed in 1451, are known. All stuffs, verdures and decorative accessories were designed by the *tapissier,* but all else (faces and hands particularly) had to be designed by professional painters, under pain of fine. Mediocre examples abound but a more distinctive use of the idiom is seen in the series of the "Redemption," of which several large panels passed from the Hearst collection into the Metropolitan museum. An extensive "History of David" in the Cluny museum shows a more elegant phase of the style. Throughout this class the border, only a few inches wide, follows a formula: a garland of roses, daisies and grapes tied, perhaps, with ribbons.

A revitalized edition of this style, with more sensitive imagination, was used by the well-known painter Bernard van Orley (*c.* 1490–1540) in the set commonly known as the "Hunts of Maximilian," though erroneously, since they postdate that emperor. In a related manner, but revealing another personality, are panels illustrating the legend of Notre-Dame de Sablon, quite possibly the work of Bernard's brother, Philippe.

Designers and weavers in the Low Countries were organized in guilds, and weavers especially had careful regulations, carrying effective penalties for violation, to protect on the one hand the workers, and on the other, craft standards.

During this period, colours are predominantly dark (20, 30 or as many as 40 shades being used in a composition); the texture tends to be coarse; silk if used at all is, as a rule, limited to high lights; and metal thread is rare.

The Renaissance.—The quantity production which was bringing prosperity to Brussels guildsmen was suddenly disturbed in 1514 when Raphael executed a series illustrating the "Acts of the Apostles," to be woven by Pieter van Aelst of Brussels for Pope Leo X. The painter, secure in the acknowledged pre-eminence of his fine art, made no concession to the unfamiliar humble craft in which his design was to be executed. Air, depth and light (all alien to the conventions of a flat, rough, heavy, woven hanging) envelope large, solidly modeled figures, with little respect for silhouette or detailed elaboration. But the new manner, erroneous though it largely was for tapestry, succeeded so well that it soon supplanted the old local cartoon styles, and for the next 150 years Brussels turned out one set after another in stereotyped exaggerations of the Raphael innovations: a cluster of huge figures rendered with exaggerated high lights in a perfunctory setting. Borders became wide and elaborate, with swags, scrolls, masks, trellises, urns and sculptures.

Shades are higher, with a preponderance of tans and yellows; gradations are more gradual; strong colours like red are used chiefly for contrast (and have frequently faded); silk is lavished on the more expensive qualities; and occasional pieces are heavily brocaded with gold, characteristically "basket-woven" for maximum effect.

The characteristic *grande verdure* of the time, a specialty of Enghien, has the field filled with huge scrolling serrated leaves reminiscent of the acanthus, sometimes with flowering plants, usually with birds and insects.

Among the outstanding Brussels families in the craft were the Pannemakers, Geubels and Leyniers, the last continuing into the 18th century. From 1528 on it was obligatory to weave the Brussels city mark—an escutcheon flanked by B's—in the selvage, and the shop mark also was usually inserted.

Fortunately, William de Pannemaker, the greatest of the Brussels weavers, never forgot the first and most important duty of a tapestry, to be splendid at a distance and full of amusing surprises at close view. Pannemaker wove a magnificent set of ten large tapestries glorifying his patron, the emperor Charles V. It is now in the Spanish State collection.

The dominant influence in the Brussels industry in the 17th century was Rubens (1577–1640). His most famous set was the "Triumph of the Church" designed to the order of Archduchess Isabella of Spain (1627–28), but imitations and adaptations of his style were legion. The Renaissance qualities were pushed further: more monumentality, drama and bombast. Heavy, elaborate columns are often substituted for side borders. Additional weaving shops became important, such as those of the Hyckes and the Raes.

Flemish Weavers Abroad.—The persecution of Flemish Protestants in the first half of the 16th century drove many craftsmen, including tapestry weavers, out of the country, while throughout that century others traveled seeking their fortunes. They were welcomed everywhere as carriers of a great tradition. Such itinerant masters established more or less enduring shops, from England to Italy.

Germany and Switzerland.—From at least the middle of the 14th century, nuns in their cloisters and ladies in their castles throughout Germany and Switzerland had been busy at their looms weaving illustrations of their legends, both religious and secular. These are for the most part traditionally long, narrow strips, modest panels or cushion covers, coarse in texture, simple in shading, limited in palette but with clear strong colours effectively contrasted. Doll-like figures, monsters half fairy tale and half heraldry, undulant banderoles with naïve explanations in bold Gothic lettering, scrolling foliation and stiff flowerets often combined with patterned backgrounds to give a toyland charm.

This domestic type continued into the 17th century, but in addition, various shops were established for longer or shorter periods by wandering Flemings, who usually failed to equal the home-country style. An exception was the looms of Seeger Bombeck in the middle of the century, at Leipzig and Weimar.

Italy.—Itinerant Flemish and French weavers, setting up their looms in cities where there was temporary employment, brought tapestry weaving to Italy as early as the 15th century. First to establish a workshop was Leonello d'Este who brought Rinaldo di Gualtieri Boteram to Ferrara from Brussels. The second major factory, the Arazzeria Medici, was founded by Cosimo I in 1546. He began with two Flemish weavers from Ferrara, Nicholas Karcher and Jan Rost. The latter became the outstanding tapestry weaver in Italy. His signature on the tapestries is symbolized in a pun on his name—a chicken roasting on a spit. Well-known painters provided cartoons; *e.g.*, Bronzino, Alessandro Allori, Francesco Salviati. A noteworthy series of genre scenes was woven in 1639 from earlier designs by Francesco Ubertini. The Medici looms continued until the death of Giovan Gastone, the last Medici grand duke of Tuscany in 1737.

Spain.—Flemish tapestries came in great quantity to the churches and palaces of Spain during the 15th century; to repair and care for them, Flemish weavers were brought in. Later, one of these, Frans Tons, established the factory of Pastrana during the reign of Philip IV. Another weaver, Jacob van der Goten, was commissioned by Philip V to start the factory of Santa Bárbara near Madrid in 1720. Long active, the great contribution of this factory was the production from 1776 to 1791 of forty-five tapestries from brilliant cartoons by Goya. The factory was closed in 1835.

Northern Europe.—In Norway, fragments of tapestry with traces of human figures and trees have been found in the Oseborg ship of the early 9th century. A more complete tapestry, found between two floors of the old church of Baldishol (Hedemark), can be dated around the end of the 12th century. Identifiable figures on it are two of the months of the year, April as a bird-catcher and May, a mounted knight, each framed in an arch. The background had a decorative pattern; there is sharp, contrasting colour in the figures, and the colour areas are dovetailed.

Of the nearly 1,300 registered Norwegian tapestries, approximately 1,250 originated in the small rural communities; Norwegian tapestries may be considered a folk, or a provincial, art.

In Sweden there is no evidence of tapestry production before the middle of the 16th century when King Gustavus I introduced some Flemish weavers. The 18th century produced small coarsely woven panels, patterned with flowers and birds, attractive as a peasant craft.

England.—All through the middle ages textile art in England was practically limited to embroidery. When tapestries were needed they were imported from Flanders; but in the reign of Henry III (1216–1272), English tapestry weaving is first recorded. Edward III encouraged Flemish weavers to settle in England, and he decreed in 1370 that only the best English and Spanish wool was to be used in their manufacture. In 1374 Catherine Dutchewoman was severely punished for using linen overspun with wool.

Tapestries were also woven in the monasteries, and a few 15th-century examples remain in fragments that look like petit-point embroideries. Such were those woven at the tapestry workshops set up by William Sheldon (d. 1570), mostly small panels and cushion covers. A later specialty of these shops was a series of topographic tapestries, woven in 1588 from contemporary maps of Midland counties, which featured bird's-eye views of hills, trees and towns, surrounded, according to the custom of the period, by Flemish-styled borders of architectural and figural ornament.

In 1619 a factory of tapestry weaving was established by royal charter at Mortlake near London. The director, Sir Francis Crane, enticed about 50 Flemings to settle in England with their families. Philip de Maecht came from Paris, where he had been a leading *tapissier*, to be overseer and master weaver. The Mortlake factory flourished under the patronage of James I and Charles I. Rubens supplied cartoons and suggested to Prince Charles the purchase of the Raphael cartoons of the "Acts of the Apostles"; Van Dyke designed new borders for these. But, under Charles II, the factory's last years showed sorry deterioration; it was finally closed in 1703.

From the end of the 17th century, the Poyntz brothers had a studio in Soho, where a number of former Mortlake artisans eventually produced a distinct style based on Chinese and Indian lacquerwork. The "Sea Fight at Soleby," 1685, by Thomas Poyntz, is a renowned English tapestry set.

French State Shops.—Henry IV of France (1553–1610), a king who planned his nation's economy, emphasized the luxury production that has ever since been industrially important in France. His first step toward tapestry production was taken by establishing a shop under Gérard Laurent in the *faubourg* St. Antoine, which was later (1608) moved into the large gallery of the Louvre. There, manned primarily by *haute-lisseurs*, it continued until 1671.

Meanwhile at the turn of the 16th–17th centuries, two Flemish weavers had been taken to France by government arrangement to establish *basse-lisse* looms (*de la marche*) in Paris: François de la Planche (or Franz van den Planken) and Marc Comans. Satisfactory working conditions were found for them in an old dye factory on the outskirts of the city which belonged to the Gobelin family, and so began the establishment commonly known by that name, which has lasted ever since.

One of its first ambitious productions was an allegorical invention lauding Catherine de Medici under the guise of Artemisia; for this the *petits patrons* were chiefly by Antoine Caron (*c.* 1515–1593), who also collaborated with Laurent Guyot on a "History of Coriolanus" (first known weaving, 1600), and he in turn collaborated with Gaston Dumée on the long-repeated "Pastor Fido"

set. Meanwhile, the shop had a valuable asset in Rubens' "History of Constantine," and they revived a 16th-century pastoral of enduring popularity, the "History of Gombaut and Macée," a symbolically typical peasant couple.

De la Planche died in 1627 and was succeeded by his son who broke with the Comans family and moved to the *faubourg* St. Germain-des-Près, leaving the Comans at the Gobelins. Competition became bitter, but both continued to produce a considerable quantity, as well as good quality, until they were superseded in 1662 by the royal factory, which purchased the Gobelins works as its seat.

Colbert, as usual alert to profitable possibilities, recruited skilled personnel, not only from the de la Planche and Comans shops, but also from the old Louvre enterprise, and thus established ateliers of both horizontal and vertical looms. As director, he utilized Charles Le Brun (1619–1690), already experienced with tapestry in a brief-lived undertaking by Colbert's ill-fated predecessor, Nicolas Fouquet; the Gobelins shops were officially chartered as a royal factory in 1667. Le Brun designed the most important sets, the "Elements," the "Seasons," and above all, the "Life of Louis XIV," supplemented by the "Royal Residences," 37 large panels altogether. Le Brun's triumphant academicism dominated the work of Antoine and Charles Coypel (1661–1722, 1694–1752), and a half-dozen other designers who devoted themselves to classical and Biblical themes; a lighter spirit was introduced by Claude Audran the Younger (1638–1734), with his talent for decorative inventions, especially grotesques.

Louis XV (1710–1774), in his turn, was celebrated in a set of "Hunts" by Jean Baptiste Oudry (1686–1755). But the outstanding artist working for the factory during this reign was François Boucher (1703–1770), whose delicious sophistication and polished elegance posed new problems for the weavers. Now indeed they must learn to paint with a bobbin, and to this end hundreds of new dyes were perfected for both wool and silk, until about 10,000 hues were available, to effect almost imperceptible tonal modulations; and interlocking of the wefts was introduced to render the transitions practically invisible, while the finest textures practical were used, from 20 to occasionally as many as 40 warps to the inch.

Both Oudry and Boucher worked also for another major state-subsidized factory which had been carried on at Beauvais by two Flemings, Louis Hinart, for 20 years (1664–1684), and Philippe Béhagle for 27 more (1684–1711). Oudry made for it his famous "Fables de La Fontaine." Boucher in 1736 began to purvey Italian fêtes, Chinese scenes and various pastoral scenes with titillating overtones. The factory has been especially noted for furniture and screen panels.

Two new types of decorative panels appear at the height of production, the architectural composition and the grotesque. The former usually shows a complex arcade with massive balustrade on which stand urns of flowers and perhaps a bird or two, as a peacock with trailing tail. In the latter, architectural tracery defines a complex of panels, framing a medley of festoons, scarves, vases, musical instruments, *putti*, masks and comedy actors. For the English taste, entire wall coverings were made at the Gobelins, much larger than, but stylistically like, these decorative panels.

Meanwhile Charles Coypel produced the great "Don Quixote" set, brilliant decorations. But by the end of the century, though technical standards were maintained, artistic deterioration set in.

A third factory at Aubusson, which operated as a modest private undertaking for a century and a half, also was chartered as a royal factory in 1665, but practically all the output there was relatively coarse material for the open market. The most effective pieces are *chinoiseries*. Coarse, rather dull, verdures were made in quantities. Aubusson architectural panels either imitate those of the greater ateliers, often with more complex elements and the addition of animals, or depict a damasked wall hung with a painting or cluster of decorative objects, and garlands. The factory was especially successful in its production of carpets, with conventional or floral designs.

In all three shops the idea of tapestry as a woven picture was so predominant that the commonest border was a woven imitation of a carved and gilt frame.

Apprentices in these factories, in order to attain the rank of master, had to produce a small specimen of their work with a difficult subject. Sometimes an elaborate vase of flowers was chosen, but often copies were made of paintings in the royal collection, and similar little pieces were made by the weavers in their own time to supplement their income.

Russia.—Peter the Great, as part of his program of modernizing his country, established looms in his capital in 1716, with workers from the Gobelins. The most striking productions are portraits of Catherine the Great. (P. An.; D. W. Li.)

Modern Design in Tapestry

During the main portion of the 19th century, in tapestry as in other luxury arts past styles were reproduced as closely as possible. The sources of modern tapestry design, however, lay in folk traditions and in progressive painting. The influence of industrialization was inescapable, of course, not only in tools, materials and dyes, but in the new middle-class market and its new demands. Besides portieres and lambrequins, the times demanded small pictorial panels, woven pictures in fact.

In England programmatic designers began the revival and purification of craftsmanship; and in 1881 William Morris established a tapestry factory at Merton abbey in Surrey. For about 15 years he and his associates had been designing not only for looms but also for pictorial wall decorations and stained glass windows; they were well prepared professionally to design tapestries. Morris and Walter Crane (1845–1915) contributed sketches but most Merton tapestries were designed by Edward Burne-Jones (1833–1898), the Pre-Raphaelite painter. His greatest effort was for a set portraying the legend of King Arthur, finished in 1894 for Stanmore hall and now owned by the Duke of Westminster. Far more adventuresome than any of these were the tapestry designs made in the 1880s by Arthur Heygate Mackmurdo (1851–1942), one of the most original craftsmen of his generation.

In Norway, shortly after 1890, special efforts were made to base a modern tapestry art on native medieval weavings. The leaders were Gerhard Munthe (1849–1929), a well-known painter, and Frida Hansen (1855–1931), a weaver who studiously researched the peasant craftsmanship of Norway and evolved a particular, light and open weave. Her work was featured at the World's Columbian exposition in Chicago, 1893; her later career was centred in Paris.

Swedish tapestry designers active at this period were the painter Gustaf Adolf Fjaestad (1868–1948); Alf Wallander, after 1896 chief designer at the Rörstrand potteries (1862–1914); Gustaf Ferdinand Boberg, a leading architect (1860–1946); and Agnes Branting (1862–1930), a well-known writer.

From 1896 to 1903 a small German tapestry industry emulating Scandinavian models was set up at Scherrebek near the Danish border of that day. After crude beginnings Scherrebek soon called on leading modern designers for patterns. Otto Eckmann (1865–1902), Hans Christiansen (1866–1945) and Walter Leistikow (1865–1908) are the best known. At Kiel and Meldorf, not far away, similar enterprises were attempted, doubtless influenced like Scherrebek by the zeal of Justus Brinckmann, founder of the Hamburg Museum of Arts and Crafts. In Austria the Krainsche Arts and Crafts school at Laibach wove tapestries as early as 1898. Modern tapestry design was hindered in France by the academic administration of the state factories. Progressive artists, however, were affected by the English Arts and Crafts movement in the later 1880s. At Pont-Aven Paul Gauguin (1848–1903), Emile Bernard (1868–1941), and their friends designed in or for needle point and appliqué, the latter especially apt to their style of broad, unbroken colour areas boldly outlined. In Paris some of the artists known as the *Nabis* followed suit. Paul Ranson (1862–1909), married to a deft needlewoman, designed more skillfully for the medium than any of the painters at Pont-Aven. At this time the young Aristide Maillol (1861–1944) began to design wall hangings that, he claimed, helped him find his true style in art. His first piece, exhibited 1893, showed stitching that followed the curved or angular outlines of each colour area, a

craft elaboration that was lacking in the work of the Ransons. Maillol briefly set up a needlework shop in Banyuls (1893–1895) with a few girl workers. When he married one and moved to Paris the idyll ended, but it seems pictured in a slightly later hanging, "Girls With Guitars," now in the Arts and Crafts museum of Copenhagen. Maillol's wish to weave true tapestry was granted by a generous patron in 1898; but with the first work on the loom the artist was stricken by an eye ailment. From then on he had to devote himself to sculpture, becoming one of the modern masters of the art. Before 1900 the Gobelins at last were weaving wall hangings designed by three leading graphic artists of the day: Louis Maurice Boutet de Monvel (1851–1913); Eugène Grasset (1841–1917); and Lucien Lévy-Dhurmer (1865–1953).

In 1893 William Baumgarten brought some Aubusson weavers to the U.S.; later, the U.S. painter Arthur B. Davies had cartoons executed in France.

After World War I, French tapestry took the lead in a double way. The great modern painters of the day—Picasso, Braque, Matisse, Léger, Rouault, Miró—permitted their works to be reproduced with extraordinary fidelity under the supervision of Mme. Marie Cuttoli, a Paris connoisseur and promoter of exceptional taste. Simultaneously another tendency began under the leadership of Jean Lurçat, who designed startlingly effective wall decorations for execution on the loom; scorning the reproduction of oil paint and *gouache* effects, Lurçat's direction has continued to lead to rich results from the talents of a large group of young French artist designers.

During the 1920s and early 1930s the Bauhaus in Germany developed a branch of modern tapestry design, abstract in composition and profoundly rooted in the loom technique and in the exploitation of fibres as strong textural elements. Anni Albers, long a resident and teacher in the United States, may be considered the chief practitioner of this kind of tapestry (more often restricted in size). She was the first weaver to receive the accolade of a one-man show at the Museum of Modern Art, N.Y., and like most modern tapestry weavers, she also designed for the textile industry.

In Scotland the Edinburgh Tapestry company and the Golden Targe Tapestry studio continue, along with some others, the tradition of William Morris, influenced by more contemporary French examples. Graham Sutherland and Henry Moore are the most eminent of the modern British artists who have had cartoons executed and installed, often in important public buildings.

In Canada, Mexico and the United States, 20th-century tapestry has been much admired and emulated. The principal show place for such efforts has been the Museum of Contemporary Crafts, established by the American Craftsmen's council in 1956.

Folkcrafts, including Mexican tapestry weaving, have been revived throughout Latin America with various success. Similarly the Scandinavian lands, including Finland, have had a brilliant resurgence of tapestry inspired by folk traditions. The Northern church has been unusually receptive to this art.

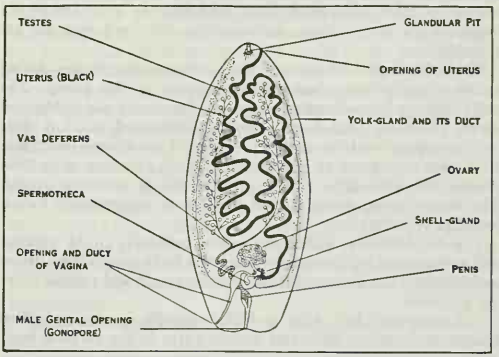
Marta Måås-Fjetterström (d. 1941) is the best-known Swedish tapestry artist, and her atelier continued to produce excellent works. Her predecessors, too, were remarkable. In Finland a freer, more colourful art, more delicately scaled, has been practised by many, among whom the best known may be Martta Taipale, Laila Karttunen and, for damask tapestry, Dora Jung. In Norway, and to a lesser degree in Denmark, similar work was being done.

See also Index references under "Tapestry" in the Index volume.

(ER. K.; D. W. LI.)

BIBLIOGRAPHY.—E. A. Barnard and A. J. B. Wace, *The Sheldon Tapestry Weavers and Their Work* (1928); H. C. Marillier, *English Tapestries of the 18th Century* (1930) and *History of the Merton Abbey Tapestry Works* (1927); W. G. Thomson, *Tapestry Weaving in England* (1912); Phyllis Ackerman, *Tapestry, the Mirror of Civilization* (1933), *Gothic Tapestries*, Catalogue of Loan Exhibition, The Arts Club of Chicago (1926) and *European Tapestries*, Catalogue, San Francisco Museum of Art (1922); J. E. Anderson, *A Short Account of the Tapestry Works at Mortlake*, manuscript in Library of Victoria & Albert Museum, n.d.; Jules Badin, *La Manufacture de Tapisseries de Beauvais Depuis ses Origines Jusqu'a nos Jours* (1909); Albert F. Calbert, *The Royal Spanish Tapestries* (1921); *Ciba Review*, "Tapestry," Basel, Switzerland: Society of Chemical Industry, no. 5 (1938); *Craft Horizons*, "Aubusson Tapestries in the U.S.," May–June 1959, "(The) Child Weavers of Egypt," Jan. 1958 and "Scandinavian Design," March–April 1958; Demotte, *La Tapisserie Gothique*, preface by Solomon Reinach (1925); Detroit Institute of Art, *Four Hundred Years of Tapestries*, Catalogue of an exhibition, Detroit (1945); E. du Somerard, "Catalogue Describing David and Unicorn Tapestries of the Cluny Museum" (1883); Charles Mills Gayley, *The Classic Myths in English Literature and Art* (1939), good background material; Jules Guiffrey, *Les Manufactures Nationales de Tapisserie les Gobelins et Beauvais* (1908); Guillaume Janneau, *L'Evolution de la Tapisserie* (1947); George Leland Hunter, *Decorative Textiles* (1918), *The Practical Book of Tapestries* (1925) and *Tapestries, Their Origin, History and Renaissance* (1912); Thor B. Kielland, *Norsk Billedvev, 1550–1800*, 3 vol. (1953–55), magnificently illustrated, English summary in each volume; André Lejard (ed.), *French Tapestry* (1947), especially for chapter on 20th century; Jean Lurçat, *Designing Tapestry* (1950); Philip Ainsworth Means, *Examples of the Pre-Incaic Period* (1930); Musée Des Arts Decoratifs, *Tapisseries, '58*, Catalogue (1958); Museum of Contemporary Crafts, *French Tapestries*, Catalogue (1959); Lila M. O'Neale in collaboration with A. L. Kroeber, "Textile Periods in Ancient Peru," Berkeley, Calif.: Univ. of Calif., *Publications in American Archaeology & Ethnology*, vol. 28 (1930); *Ornamo*, Finland: Applied Art Association, Ornamo (1950–54), for articles on work of Dora Jung, Eva Anttila, Martta Taipale and Laila Karttunen; Oslo-*Kunstindustrimuseet Gammel Norsk Vaevkunst*, Kristiana: Mittet (1913–21); Anna Marie Pope, *Tapestries by Hannah Ryggen* (1955–56); William George Thomson, *A History of Tapestry: From the Earliest Times Until the Present Day* (1930); Toledo Museum of Art, *International Exhibition of Modern Tapestries, December 7–January 25, 1930*, Catalogue, the Toledo Museum of Art (1929); "The Vatican, Its History, Its Treasures," in *Gallery of Tapestries*, pp. 231–250 (1914); Pierre Vercours (ed.), *Tapisseries de Jean Lurçat, 1939–1957* (1957); Victoria & Albert Museum, *Catalogue of Tapestries in the Victoria & Albert Museum*, by A. F. Kendrick (1914), *Brief Guide to Peruvian Textiles* (1926), *Tapestries*, Department of Textiles, vol. 3 (1913–15) and *The Raphael Cartoons* (1950, 1958); Bernard Vuilleumier, *Art of Silk Weaving in China* (1939); Adèle Coulin Weibel, foreword to *2000 Years of Tapestry Weaving: A Loan Exhibition, 1951* (1951); A. Wauters, *Les Tapisseries Bruxelloises* (1878). (D. W. LI.)

TAPEWORMS (CESTODES or CESTODA), a class of endoparasitic Platyhelminthes (*q.v.*) in which the body is usually flattened and ribbon-like, and may be unisegmental or composed of a chain (strobila) of "segments" (proglottides). In the latter case the segmentation is not comparable with that of an annelid or arthropod, but is probably to be looked upon as an adaptive device



FROM LANKESTER, "TREATISE ON ZOOLOGY" (BLACK)

FIG. 1.—DIAGRAM ILLUSTRATING REPRODUCTIVE SYSTEM OF AMPHILINA FOLIACEA, A CESTODE WORM, INFESTING THE BODY OF STURGEONS

for the multiplication and dissemination of eggs. The number of segments in a complete strobila varies, in different species, from three or four to many hundreds, and the length of the chain from less than a millimetre to several metres. Each segment contains a separate set of reproductive organs, but the nervous and excretory systems are continuous throughout the chain. Tapeworms have no mouth or alimentary canal, liquid food being presumably taken in by absorption through the body-wall. The adult worms are, with few exceptions, intestinal parasites of vertebrates. In almost all known cases the life-history involves an intermediate host and a somewhat complex metamorphosis.

General Morphology.—The body of the adult is covered ex-

ternally by a cuticle, below which there is a cellular subcuticular layer. All the internal organs are embedded in loose parenchymatous tissue. In most forms the parenchyma contains numerous "calcareous corpuscles"—small refringent nodules composed mainly of carbonate of lime. The function of these is unknown. Possibly they are merely stored-up excretory products. The musculature consists mainly of longitudinal fibres, often in definite bundles and situated, for the most part, rather deeply in the parenchyma.

In the segmented forms there is, at what is usually called the anterior end, a specialized "head" or scolex, serving as an organ of fixation. This is generally provided with muscular suckers or variously modified sucker-like organs known as bothria or bothridia, and may, in addition, bear a central proboscis-like structure, the rostellum, which is commonly armed with cuticularized hooks. The region nearest to the scolex is the zone of proliferation of new segments, which are, as a rule, continually being formed throughout the life of the worm. As they pass back along the strobila the segments become successively sexually mature and finally gravid, the most posterior usually containing little but a uterus crowded with eggs.

The nervous system consists of an intercrossing system of nerve-fibres and ganglionic cells at the anterior end (in the scolex in strobilate forms), and usually a single pair of main lateral nerve cords which are continuous throughout the body and without segmental ganglia. The excretory or "water-vascular" system is composed typically of two pairs (dorsal and ventral) of main longitudinal canals, continuous throughout the strobila and connected together anteriorly.

The animals are, with rare exceptions, hermaphrodite, each individual, or each segment in strobilate forms, being potentially male and female. The male duct and the vagina may have separate apertures or may open side by side into a common atrium. There is considerable variation in the position of these openings. In unsegmented forms they may be terminal, subterminal or ventral, while in segmented forms they may be on the ventral surface or on one of the lateral margins of the segment. The essential organs of the genital apparatus do not differ greatly from those of other groups of Platyhelminthes. The oviduct connected with the ovary and its associated glands, and with the inner end of the vagina, leads into a uterus in which the fertilized eggs are accumulated.

Classification.—There have been considerable changes during recent years in the systematic arrangement of this group. The older division into unsegmented forms (Monozoa) and segmented forms (Merozoa) has been generally abandoned, since it does not correspond with an arrangement based on internal anatomy. It is now customary to divide the Cestoda into two main subclasses, the Cestodaria and the Eucestoda or cestodes proper. The former group contains two families of unsegmented forms, parasitic in fishes:

(1) Amphilinidae, with a protrusible proboscis at the anterior end and a very long uterus traversing the body length three times and opening near the anterior end, whereas male and vaginal pores are posterior.

(2) Gyrocotylidae, with a frilled margin, posterior ruffled rosette, and uterine, male and vaginal pores in the anterior body half.

The Eucestoda include a large number of the more typical tapeworms, sometimes unsegmented but usually with a long ribbon-like body divided into a few to many segments and with a scolex. There are five main orders although others may be necessitated by recent findings of unusual species.

I. Phyllobothrioidea. These are intestinal parasites of elasmobranch fishes and are characterized by the four mobile, often large and leaflike bothridia which may be armed with hooks or subdivided by septa.

II. Diphyllidea. This order includes only the genus *Echinobothrium,* parasitic in elasmobranch fishes. These are segmented tapeworms of not more than 20 segments with a scolex having two bothridia and with a spiny "neck" region behind the scolex.

III. Tetrarhynchoidea. These tapeworms, also parasitic in the intestine of elasmobranch fishes, are distinguished by the four protrusible proboscides covered with hooks or spines that can be everted through four openings on the scolex. The scolex in addition has four bothridia or bothria.

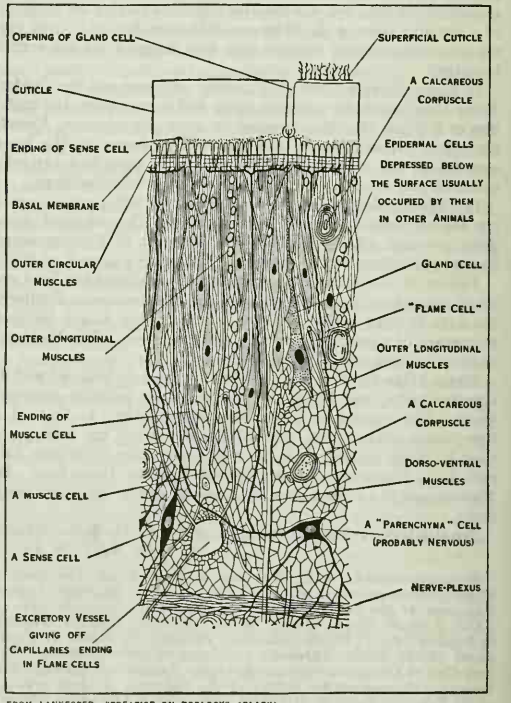
IV. Dibothriocephaloidea. This order, inhabiting the intestine of various vertebrates, includes a number of unsegmented tapeworms such as *Caryophyllaeus* and *Archigetus* and also many segmented types. The scolex is usually provided with two shallow bothridia. Here belongs *Diphyllobothrium latum,* a human parasite whose complicated life history is discussed below.

V. Taenioidea. This is the largest and most important order, including all the more common tapeworms of higher vertebrates. The body is always segmented and the scolex bears four true suckers or acetabula, often with the addition of a rostrellum armed with hooks. The life cycle includes a bladder-worm stage (*see* below).

LIFE HISTORY

Development.—The eggs of Cestodes are usually enclosed in two membranes, one of which may show cuticularized thickening, forming a hard "shell." In the Bothriocephalidea the shell is provided with an operculum at one pole, like that of a digenetic Trematode. The embryo, at the stage when it is ready to hatch, is a spherical or oval mass of cells (called an "onchosphere"), provided at one pole with cuticularized hooks, of which in most species there are three pairs—hence the name "hexacanth embryo." In certain Amphilinoinei, however, the embryo has ten hooks. The embryonic hooks are used as levers, chiefly in burrowing among the tissues of the intermediate host.

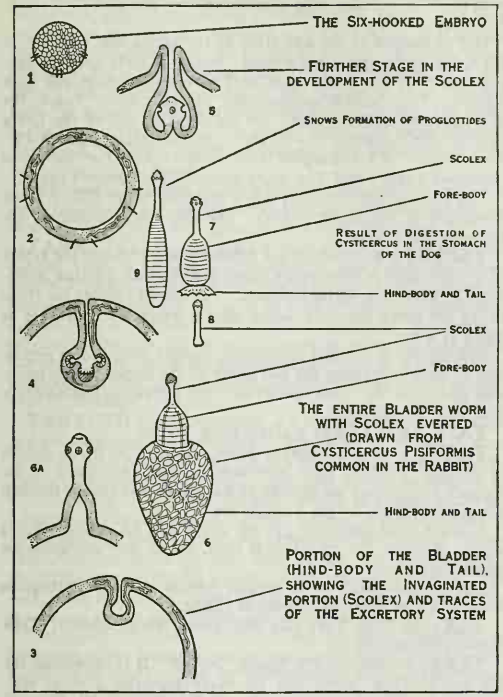
In some of the Dibothriocephaloidea the embryo is provided with an external ciliated envelope and hatches in water, where it swims about by means of its cilia. In the human parasite *Diphyllobothrium latum,* for example, the ciliated embryo is at first free



FROM LANKESTER, "TREATISE ON ZOOLOGY" (BLACK)

FIG. 2.—MICROSCOPIC STRUCTURE OF A TAPEWORM. A PORTION OF A TRANSVERSE SECTION THROUGH THE BODY-WALL OF LIGULA INTESTINALIS

in fresh water and is then swallowed by a small copepod (*Cyclops* or *Diaptomus*). Shedding its ciliated coat, the embryo penetrates into the body-cavity of the copepod, and develops into a more



FROM LANKESTER, "TREATISE ON ZOOLOGY" (BLACK)

FIG. 3.—DEVELOPMENT OF A TAENIID TAPEWORM (TAENIA PISIFORMIS) 1–5, development of the six-hooked embryo, within the intermediate host, into a bladder worm, and successive stages in the formation of the head. 6–9, evagination of the head, and further development of the larval form into an adult worm, in the final host

elongate form called the *procercoid*. If the copepod is swallowed by a suitable fish, such as a pike, perch or trout, the procercoid develops further among its tissues into a *plerocercoid*, and is then infective for the final host, infection being acquired by eating the fish in a raw or imperfectly cooked state.

For the development of this species, therefore, and of others related to it, two successive changes of host are required. The larval forms known as *Sparganum*, which occur in many land vertebrates, including man, have been shown to be the plerocercoids of *Diphyllobothrium*.

The larval forms of Tetrarhynchoidea, resembling the scolex of the adult, without the strobila, or with only an unsegmented appendage, occur encysted among the tissues of various marine animals, chiefly teleostean fishes.

Among the best known of the Taeniidea the developmental history is quite different. The human parasite *Taenia solium* may be taken as typical. If an egg of this worm be swallowed by the proper intermediate host (in this case the pig) the hexacanth embryo is liberated in the intestine and proceeds to pierce the intestinal wall, wandering about among the tissues and usually coming to rest among the muscles. Here it grows into a small bladder, in whose wall a depression or invagination appears on the side opposite to the embryonic hooks. At the bottom of this invagination five thickenings appear, destined to become the four suckers and the median rostellum. Finally the hooks are developed on the rostellum, and the bladder worm, or *cysticercus*, is fully formed. Should this be ingested by a human being with raw or imperfectly cooked pork, the "head" is evaginated and attaches itself to the lining of the intestine, and the "neck" joining it to the bladder begins to form a series of proglottides. Another human parasite, *Taenia saginata*, has a similar development, but is without hooks, and makes use of the ox as intermediate host.

Many modifications occur in the form of the bladder worm, but the essential features of the development are the same throughout the order. In some forms the original bladder worm gives rise to several or many scolices instead of the usual single scolex, a phase of asexual multiplication thus being introduced. In the gid-parasite of the sheep (*Taenia* or *Multiceps multiceps*) the bladder worm, which occurs in the brain and is often called a *Coenurus*, gives rise to numerous scolices, each capable of developing into an adult worm if ingested by a dog or other suitable carnivorous host.

In *Echinococcus granulosus* this phase of multiplication is elaborated still further, the original embryo developing in the intermediate host into a large and relatively thin-walled bladder or cyst ("hydatid"), from the inner surface of which there are budded off a large number of "brood-capsules," each containing from 10 to 30 scolices. Daughter cysts may also be budded off, and these may give rise to further scolices, so that altogether a very large number of potential adult worms may be formed from the original embryo.

This worm occurs, in the hydatid stage, in the liver and other organs of a large number of hosts, including man, sheep, ox and pig. The adult is a very small worm with only three or four segments, and occurs in the intestine of the dog, wolf and fox.

As may be noted from the examples already given, the intermediate host is usually an animal likely to be devoured by the final host. Thus the dog harbours several Taenioid tapeworms whose intermediate hosts are rabbits and sheep. The cysticercus of *Taenia taeniaeformis* of the cat occurs in rats and mice. A number of species of *Hymenolepis* parasitic, as adults, in ducks and geese make use of small fresh-water Entomostraca such as *Cyclops* and *Cypris* as intermediate hosts. Houseflies and earthworms are the vehicles of infection for several of the tapeworms of fowls. *Dipylidium caninum*, a common parasite of the dog and cat, passes its cysticercus stage in fleas and in the louse, *Trichodectes*. The louse is capable of swallowing the eggs of the worm, but the adult flea cannot do so. Fleas become infected during their larval period, the eggs, when swallowed, hatching in the hinder part of the intestine and the embryo boring through its wall into the body-cavity. Here they do not continue their development until the pupal stage has been passed, the cysticerci becoming fully formed only in the adult flea. Infection is acquired by dogs and cats, and accidentally by man, through swallowing fleas containing them.

The larvae of many of the Phyllobothrioidea, whose adult forms are found in the spiral valve region of the intestine of sharks and rays, occur among the tissues of other marine animals. One such larval form (apparently belonging to a species of *Tylocephalum*) occurs in the pearl-oyster, and is believed to be one of the principal causes of the production of pearls. This, apparently, takes place only in the event of the death of the parasite from some unknown cause.

Economic Importance.—From the fact that tapeworms often occur in enormous numbers in the intestine of an apparently perfectly healthy animal, it is clear that they are not always definitely injurious to their hosts. As parasites of man and of domestic animals, however, they have considerable importance. As a rule their ill effects are most manifest in young animals, and they are sometimes responsible for serious digestive and nervous disturbances, and even for severe losses among stock and poultry. The adult worms, besides depriving the host of a certain amount of nutrient material, secrete substances which have been shown to be definitely toxic in many instances, when injected into the blood or body-cavity of experimental animals. The same is true of the larval forms, and certain changes take place in the blood cells of infested animals which indicate that some of these poisons are normally absorbed into the system. It is probable, however, by the migrations of the larvae within the body, and the secondary invasions by pathogenic organisms

for which they prepare the way, that the most serious harm is done.

Only in exceptional cases, such as the invasion of the brain by the hydatid of *Echinococcus,* are tapeworms really dangerous to man. (H. A. B.; L. H. H.)

TAPIOCA, a food product made from cassava flour or starch (*see* CASSAVA). The terms cassava and tapioca as applied to starch are synonymous. In the manufacture of tapioca, cassava flour is moistened and heated with continuous stirring. This treatment causes the starch to agglomerate into small semitranslucent white balls.

This convenient pelleted form of cassava starch is the tapioca of the food industry. Cassava starch when cooked produces clear, tender flavourless gels that retain their clarity and tenderness when cooled. The use of tapioca as a thickening agent for puddings, fruit pies and soups is based on these aesthetic gel properties. (R. M. SA.)

TAPIR, the name for large woodland mammals, forming the family Tapiridae of the order Perissodactyla (*q.v.*), with four front and three hind toes, massively built and with the nose and upper lip produced to form a short flexible trunk. Four species inhabit South and Central America, and one, *Tapirus indicus,* the Malay region. The latter is the largest, and is readily recognized by its black fore- and hindquarters and white body. It may stand 3 ft. 6 in. at the shoulder and is solitary, shy, nocturnal and inoffensive, frequenting deep forests in the neighbourhood of water. Its diet is exclusively vegetarian. The American species are all nearly uniform dark brown when adult, but the young, like those of *T. indicus,* are spotted and striped with white. The best-known new world form is *T. terrestris* of Brazil and Paraguay. In habits it and the other three species resemble the Asiatic form. The peculiar distribution of the tapirs is explained by the presence of fossil forms in Europe, China and the United States in Miocene, Pliocene and, in the two latter localities, Pleistocene formations.

YLLA FROM RAPHO-GUILLUMETTE

FEMALE SOUTH AMERICAN TAPIR AND YOUNG (TAPIRUS TERRESTRIS)

TAPPAN, ARTHUR (1786–1865), U.S. philanthropist and abolitionist, was born on May 22, 1786, at Northampton, Mass. Beginning as a dry-goods importer in 1807, his partnership in a silk-jobbing firm established his fortune. He was of a stern and pious nature and often championed humanitarian and religious causes. Among those organizations to which he contributed were the American Bible society, the American Missionary society and the American Tract society. He donated money to establish Kenyon college, Gambier, O., and Oberlin college, Oberlin, O.

Tappan was an active supporter of various antislave movements. He joined the American Colonization society, later withdrawing in order to work for the abolition of slavery in the U.S. In 1840 he became president of the American and Foreign Anti-Slavery society and founded the *American and Foreign Anti-Slavery Reporter.*

He advocated the abolition of slavery through political means within the constitution, and in 1840 and 1844 he supported the Liberty party's candidate for president, James G. Birney.

He died in New Haven, Conn., on July 23, 1865.

TAPPAN, EVA MARCH (1854–1930), U.S. teacher and author of widely read children's books on historical and literary topics, was born Dec. 26, 1854, in Blackstone, Mass. She received her early education in seminaries where her mother taught. After graduation from Vassar college, Poughkeepsie, N.Y., she taught in Wheaton seminary, Norton, Mass. (1875–80), and later at Raymond academy, Camden, N.J., where she served as associate principal from 1884 to 1894.

She received the M.A. degree from the University of Pennsylvania, Philadelphia, in 1895 and the Ph.D. degree in 1896. From 1897 to 1904 she taught in the English high school, Worcester, Mass., and upon retirement devoted herself exclusively to writing, except for a period of service during World War I as an editor with the United States food administration. She died in Worcester, Mass., Jan. 29, 1930.

Her dedication to the education of the young was reflected in the nature and scope of her writing. Simple in style and organization, many of her books were used as texts in grade and high schools. They include among many others, *Charles Lamb, the Man and the Author* (1896), *In the Days of Alfred the Great* (1900), *Our Country's Story* (1902), *In the Days of Queen Victoria* (1903), *When Knights Were Bold* (1912), *Our European Ancestors* (1918) and *The Story of Our Constitution* (1922).

The story of her own early life is recorded in her *Ella, a Little Schoolgirl of the Sixties* (1923). She also edited numerous anthologies for children.

TAPPET is an intermediary fitment which converts the action of a rotary cam or lifter into a linear motion, with a return movement by gravity or spring pressure. The largest tappets are those in the ore stamp batteries, which raise a weight of sometimes as much as a ton.

The tappets in internal-combustion engines require nice regulation to ensure obtaining the full power of the engine and to maintain silence in the valve operation. *See* INTERNAL-COMBUSTION ENGINES; AUTOMOBILE.

TAPS AND DIES: *see* MACHINE TOOLS.

TAPTI, a river of western India. It rises in Betul district of Madhya Pradesh, flows westward between two spurs of the Satpura hills, across the plateau of Khandesh, and thence through the plain of Surat to the Arabian sea.

It has a total length of 436 mi. and drains an area of 25,174 sq.mi. For the last 32 mi. it is tidal, but is only navigable by small vessels.

The port of Swally at its mouth, famous in Anglo-Portuguese history, is now deserted because of silting.

TAR: *see* COAL TAR; GAS INDUSTRY; NAVAL STORES; TARS, LOW-TEMPERATURE.

TARA, a village of Co. Meath, Ireland. It is celebrated for the hill of Tara, which was for many centuries a royal residence and the scene of great meetings of the people. The isolated hill, upon which five highroads converged, is about 510 ft. in height. On it are six raths or circular earthworks, the largest of which, the king's rath (*rath-na-riogh*), encloses other works, including the *forradh* or meeting place, a flat-topped mound. On this (but not in its original position) stands a pillar stone, said to be the stone of destiny on which the Irish kings were crowned. An oblong enclosure, 759 ft. in length by 46 ft. in breadth, formed of earthworks, with entrances at intervals on each side, represents the banqueting hall. Extensive excavations were begun in 1953 with the assistance of the Irish government.

In the middle of the 3rd century A.D. King Cormac Mac Art is said to have founded there schools of military science, law and literature. In the time of St. Patrick, Tara was a great centre of pre-Christian religious rites, and about 560 it was abandoned as a royal residence.

There the Danes were defeated in 980; and in 1843 the hill of Tara was the scene of one of Daniel O'Connell's mass meetings to demand repeal of the legislative union (Aug. 15).

TARAFA ('Amr ibn ul- 'Abd ul-Bakrī), Arabian poet of the 6th century, who, after a wild and dissipated youth spent in Bahrein, left his native land after peace had been established between the tribes of Bakr and Taghlib and went with his uncle Mutalammis, who was also a poet, to the court of the king of Hira, 'Amr ibn Hind (d. 568–569), and there became companion to the king's brother. Having ridiculed the king in some verses, he was sent with a letter to the ruler of Bahrein, and in accordance with the instructions contained in the letter was buried alive. One of his poems is contained in the *Moallakat.*

His collected works were published in W. Ahlwardt's *The Diwans of the Six Ancient Arabic Poets* (London, 1870).

Some of his poems were translated into Latin by B. Vandenhoff (Berlin, 1895).

TARAI (TERAI) (*i.e.*, "moist land"), the name of the submontane strip of marshy jungle stretching beneath the lower ranges of the Himalayas in northern India. This strip may be said to extend roughly from the Jumna river on the west to the Brahmaputra on the east, and a large portion of it lies within Nepal. The term, however, is now officially confined to a subdivision of Naini Tal district in Uttar Pradesh. At its northern edge, where the waterless forest tract of the bhabar ends, a series of springs burst from the surface, and these, increasing and uniting in their progress, form the numerous streams that intersect the Tarai. The Gogra is the great river of the Tarai proper and is navigable from the Himalayan foothills. Elephants, tigers, bears, leopards and other wild animals are found.

Everywhere it is most unhealthful, inhabited only by tribes who seem immune to malaria, though large tracts were drained for cultivation and settled with agriculturists from the plains.

TARANTELLA, an Italian folk dance in rapid $\frac{6}{8}$ time, danced in couples and characterized by light quick hops and tapping foot movements. A disease (or form of hysteria) called tarantism, prevalent in Italy from the 15th to 17th century, was supposedly curable only by frenzied dancing of the tarantella. According to popular superstition, the dance also cured the sting of the tarantula; but probably it derives its name from the town of Taranto. The tarantella appears in many ballets, including *Napoli, La Tarentule* and *La Boutique Fantasque*. (LN. ME.)

TARANTO (anc. TARENTUM, *q.v.*), a seaport of Puglia, Italy, and the capital of the province of Taranto 50 mi. W. by N. from Lecce by road, and 68 mi. by rail (44 mi. W. by S. from Brindisi) Pop. (1961) 200,974 (comm.). The city proper is situated on a rocky island 56 ft. above sea level, which, in ancient times was a peninsula, the isthmus on the west having been cut through by Ferdinand I of Aragon. This island separates the Gulf of Taranto (Mar Grande) from the deep inlet of the Mar Piccolo, and is sheltered by two other flat islands, San Pietro and San Paolo. This rock is the site of the citadel of the ancient town; its population is confined within small houses and narrow streets.

The Strada Garibaldi along the Mar Piccolo is inhabited by fishermen whose language retains traces of Greek. The cathedral, dedicated to San Cataldo, an Irish bishop, has externally some remains of Saracenic Gothic; internally it has been completely modernized (retaining however 16 ancient columns in the nave) and the shrine of the patron saint has been termed "an orgy of rococo" (1657). It has a crypt of the 6th or 7th century. There is a museum in the former convent of San Pasquale, with an especially fine collection of vases, terra cottas, and silverware, jewelry, Greek statuary, mosaics, etc., from Taranto and other sites in Apulia.

Adjacent is the Palazzo degli Uffizi, containing various public offices. To the east, across the swing bridge over the channel connecting the Mar Grande with the Mar Piccolo is the new part of the town, on the site of the main part of the ancient city, extending as far as the arsenal and even beyond: while the Boyo, with the railway station, lies to the west of the island. There, too, is the commercial harbour. The chief industry is the cultivation of oysters in the Mar Piccolo; besides oysters Taranto carries on a large trade in *cozze*, a species of large black mussel, which is packed in barrels with a special sauce. Excellent fish abound in the Mar Piccolo, 93 different species being found.

The ebb and flow of the tide is distinctly visible there, Taranto being one of the few places in the Mediterranean where it is perceptible. In 1861 the strategic importance of Taranto was recognized, and the arsenal is the most important in Italy after Spezia. It extends for $1\frac{1}{2}$ mi. along the southern coast of the Mar Piccolo, its chief basin.

At Taranto the Italians suffered their first major naval losses of World War II. On the night of Nov. 11–12, 1940, British torpedo planes attacked a group of ships there, damaging two cruisers, two or three battleships and two fleet auxiliary vessels. In the fall of 1943 the city was occupied by British troops.

In 927 Taranto was entirely destroyed by the Saracens, but rebuilt (though henceforth restricted to the island) in 967 by Nicephorus Phocas, to whom is due the construction of the bridge over the channel to the northwest of the town, and of the aqueduct which passes over it. The town was taken by Robert Guiscard in 1063. His son Bohemund became prince of the Terra d'Otranto, with his capital there. After his death Roger II of Sicily gave it to his son William the Bad. In 1301 Philip, the son of Charles II of Anjou, became prince of Taranto. The musician Giovanni Paisiello (1741–1816) was born there. The castle, originally Byzantine, dates in its present form from 1480 and 1577.

The tarantula (*see* below), inhabits the neighbourhood of Taranto. The wild dance, called tarantella (*q.v.*), was supposed, by causing perspiration, to drive out the poison of the bite. (T. A.)

TARANTULA, strictly speaking, a large spider (*Lycosa tarantula*) with a body about one inch long which takes its name from the town of Taranto in southern Italy. Its bite was formerly believed to be the cause of a disease known as tarantism. In this malady the victim wept, skipped about and then went into a wild dance (*see* TARANTELLA). Unless assisted by special music or magic antidotes, the victim was expected to die. Investigations have shown that the bite of this spider is not dangerous to man and that the disease attributed to it must have had other causes.

Lycosa tarantula belongs to the wolf spider family (Lycosidae) and has numerous allies in various parts of the world. The tarantula spins no web and catches its prey by pursuit. It inhabits dry, well-drained ground where it digs deep burrows. In the winter, the opening of the burrow is covered with silk, and the spider becomes dormant. It lives for several years.

The name tarantula is often applied indiscriminately to large, fearsome hairy spiders of the suborder Mygalomorphae but is usually reserved for members of the family Theraphosidae, which are found in the southwestern United States, Mexico and tropical America. Many of these spiders tunnel into the soil, spending most of their life within this single burrow. They feed on large insects and occasionally on frogs, toads and mice.

In the southwestern United States, some tarantulas of the genus *Aphonopelma* attain a length of about 2 in. and a leg spread of from 4 to 5 in. These tarantulas are quite harmless and have been trained as pets; however, they can inflict painful wounds in self-defense. In South America, some tarantulas have a body length of more than 3 in. A few of these build huge webs and catch and kill large insects and even small birds for food. Most of these forms are harmless to man, except for their bite.

See also SPIDER. (C. J. Go.; M. L. G.)

TARAPACÁ, the northernmost province of Chile, bounded north by Peru, east by Bolivia, south by Antofagasta and west by the Pacific. Area 22,422 sq.mi. Pop. (1960) 123,070. It is part of the Atacama desert (*q.v.*) region of the Pacific coast of South America and is without water except at the base of the Andes, where streams flow down into the sands and are lost. In some of these places there is vegetation and water enough to support small settlements.

The wealth of Tarapacá is in its immense deposits of nitrate of soda (found on the Pampa de Tamarugal, a broad desert plateau between the coast range and the Andes, which has an elevation of about 3,000 ft.). The mining and preparation of nitrate of soda for export has provided an uncertain livelihood for the population of this province. Silver is mined near Iquique (*q.v.*), the provincial capital. Unfavourable world markets for nitrate caused a decline in population after 1940. The principal ports of the province are Pisagua, Iquique and Arica (*q.v.*), from the first two of which "nitrate railways" run inland to the deposits.

Tarapacá was ceded to Chile by Peru after the war of 1879–83 and was organized as a province in 1884.

TARASCANS, the principal native people of the southwest Mexican state of Michoacán, who term themselves and their language Purépecha. An expanding independent nation at the time of the Spanish conquest, most of the Tarascans by the 20th century were confined to the pine and oak-clad highlands dotted with volcanic cones and lakes west from Morelia and Lake Pátzcuaro and north from Uruapan and Paricutín volcano. Despite a culture that is more old world than native, there is a marked retention of racial

identity; and in the second half of the 20th century there were approximately 60,000 persons of Purépecha speech, of which perhaps one-quarter were monolingual. This language constitutes a linguistic family with no known relatives. Tarascan archaeology and history have been little studied, but artifacts and colonial accounts indicate that this group was the equal or superior of their Mexican contemporaries in the working of feathers, shell, wood, metal and stone, and in military prowess; only in architectonics were they greatly inferior.

The Tarascans have no special tribal organization, but participate in much the same political, economic and religious organizations as do their mestizo neighbours. Agriculture is dominant; maize and wheat occupy most of the acreage, followed by pulses, cucurbits, barley, cabbage and chile pepper. Livestock includes chickens, oxen, pigs, sheep, beehives, burros and turkeys. Probably the turkey was domesticated by the Tarascans. In the mountain or sierra region forest products are quite important, and there the wooden house or *troje* prevails. Among the leading cottage handicrafts are woodworking, pottery making and weaving. Many Tarascans go as *braceros* ("day labourers") to the United States, and most of the nearly 120 communities have at least one man who has lived in the United States.

Among the leading communities are the Sierra capital Paracho, strongly Indian Cherán and Santa Fe, and the sentimentally important former capitals Tzintzuntzan, Ihuatzio and Pátzcuaro.

BIBLIOGRAPHY.—J. A. Mason, "The Native Languages of Middle America," in *The Maya and Their Neighbors* (1940); L. Mendieta y Núñez *et al., Los Tarascos* (1940); D. D. Brand, "An Historical Sketch of Geography and Anthropology in the Tarascan Region," *New Mexico Anthropologist*, pp. 37–108 (June 1944); Smithsonian Institution, *Publications of the Institute of Social Anthropology*, no. 1, 2, 6, 7, 11 (1944–1951); Sociedad Mexicana de Antropología, *El Occidente de México* (1948); G. Aguirre Beltrán, *Problemas de la Población Indígena de la Cuenca del Tepalcatepec* (1952). (D. D. B.)

TARASCON (TARASCON-SUR-RHÔNE), a town of southeastern France, in the *département* of Bouches-du-Rhône, 62 mi. N.W. of Marseilles by rail. Pop. (1962) 6,475. Tarascon stands on the left bank of the Rhône opposite Beaucaire, with which it is connected by a railway bridge and a suspension bridge. The church of St. Martha, built in 1187–97 on the ruins of a Roman temple and rebuilt in 1379–1449, has a Gothic spire and interesting tombs in the crypt. Of the original building there remain a porch and a side portal flanked by marble columns with capitals like those of St. Trophimus at Arles. The former leads to the crypt, where are the tombs of St. Martha (1658), Jean de Gossa, governor of Provence under King René, and Louis II, king of Provence.

The castle, built on a rock, was begun by Count Louis II in the 14th century and finished by King René in the 15th. It contains a turret stair and a chapel entrance, both 15th-century, and fine wooden ceilings. The building is now used as a prison. The *hôtel de ville* dates from the 17th century.

The civil court of the *arrondissement* of Arles is situated at Tarascon, which also possesses a commercial court, and fine cavalry barracks. The so-called Arles sausages are made there, and there is trade in fruit and early vegetables. In *Tartarin de Tarascon* Alphonse Daudet has satirized the provincial life of Tarascon, which has a tribunal of commerce and the fair of Beaucaire. It formerly had the two fêtes of La Tarasque, the latter in celebration of St. Martha's deliverance of the town from a legendary monster of that name. King René presided in 1469, and grand exhibitions of costume and strange ceremonies take place during the two days of the festival. Tarascon was originally a settlement of the Massaliots, built on an island of the Rhône. The medieval castle, where Pope Urban II lived in 1096, was built on the ruins of a Roman camp. Tarascon preserved the municipal institutions granted it by the Romans, and of the absolute power claimed by the counts of Provence only recognized the rights of sovereignty.

TARAXACUM, a genus of weedy, perennial herbs, bearing yellow flowers, widespread in the northern hemisphere. The best known of the various species is the common dandelion (*q.v.*).

TARBERT, a fishing village at the head of East Loch Tarbert, an inlet of the sea on the western shore of the mouth of Loch Fyne, Argyllshire, Scot., 38 mi. S.S.W. of Inverary by road. Pop. (1951) 1,426.

The name (met elsewhere in Scotland) denotes an isthmus and the village is in fact on the neck of land joining Kintyre to the rest of Argyllshire. This isthmus is little more than 1 mi. wide, and boats were dragged across to the head of West Loch Tarbert, a narrow sea loch nearly 10 mi. long. There is a good harbour with a pier for passenger steamers about $\frac{3}{4}$ mi. from the village. The herring fishery, though shrunken, is the main industry together with some shipbuilding. Overlooking the harbour are the ruins of a castle built by Robert Bruce in 1326.

TARBES, a town of southwestern France, capital of the *département* of Hautes-Pyrénées, 98 mi. W.S.W. of Toulouse on the Southern railway. Pop. (1962) 46,162. Under the Roman dominion Turba, which was about 11 mi. S.E. of the present town of Tarbes, was the capital of the Bigerriones, one of the states of Novempopulania. The bishopric of Tarbes dates from the 5th century, and in feudal times its bishops held the chief temporal authority, that of the counts of Bigorre, of which Tarbes was the capital, being limited to the quarter of the town where their castle was built. The English held the town from 1360 to 1406. In 1569–70 Tarbes was twice taken by Gabriel, count of Montgomery, and the inhabitants driven out, but in Aug. 1570 the peace of St. Germain allowed them to return. Subsequently Tarbes was several times taken and retaken, and a number of the inhabitants of Bigorre were forced to take refuge in Spain, but in 1594 the members of the League were finally expelled. The English, under Wellington, gained a victory over the French near Tarbes in 1814.

Tarbes stands in a fertile plain, stretching to the Pyrenees, on the left bank of the Adour, streams from which are conducted through the town. The lines of the Southern railway from Morcenx to Bagnères-de-Bigorre and Lourdes and from Toulouse to Bayonne cross there. Chief among the many open spaces is the Jardin Massey (35 ac.), given to his native town by a director of the gardens of Versailles and containing a museum of paintings and antiquities. Near a small lake stands a cloister (15th century) brought from the abbey of St. Sever-de-Rustan, 14 mi. N.E. of Tarbes. The architecture of the cathedral, Notre-Dame-de-la-Sède, is heavy, but the cupola of the transept (14th century), and a rose window of the 13th century, in the north transept, are interesting. There is also a modernized Carmelite church of the 13th century.

Tarbes is the seat of a bishopric under the archbishop of Auch, of a prefecture, tribunals of first instance and of commerce, a chamber of commerce and a board of trade arbitrators. Tarbes has an important stud for the breeding of Anglo-Arabian horses, much used by light cavalry. The industrial establishments include tanneries, potteries, sawmills and turners' shops. There are also important fairs and markets.

TARBOOSH, the close-fitting, flat-topped and brimless cap, in shape like a truncated cone, made of felt or cloth, worn by Mohammedan men throughout the east either as separate headgear or forming the inner part of the turban. It used to be worn as the badge of a Turkish subject in Turkey and Egypt, where it was red in colour with a black or blue silk tassel. It is the same as the fez. In 1925, Turkish citizens were forbidden to wear the fez, under severe penalties, by Pres. Kemal Pasha.

TARDE, GABRIEL (1843–1904), one of the most versatile and productive of French social scientists of his day, was born March 12, 1843, at Sarlat, Dordogne. He served the French government as a provincial magistrate and as director of the bureau of criminal statistics at the ministry of justice. In 1900 he was appointed professor of modern philosophy at the Collège de France. He died at Paris, May 13, 1904. The sequence of invention, repetition, conflict and adaptation emphasized in Tarde's social philosophy bears a close resemblance to Hegel's dialectic. Tarde held that about one person in a hundred is inventive, yet that invention is the source of all progress. Inventions are imitated, but differ in degree and type of imitation, and this promotes opposition, both between varied limitations and between the new

and the old in culture. The outcome of this opposition is an ultimate adaptation. This adaptation then becomes in itself an invention.

He saw this sequence as an unending cycle constituting the process of history and social experience (*Social Laws*, 1898). He treated repetition in his most famous work, *The Laws of Imitation* (1890); opposition in *Universal Opposition* (1897); and adaptation in *Social Logic* (1895). In *Comparative Criminality* (1886) and other works on criminology he attacked the extreme biological theories of C. Lombroso and his school, pointing out the environmental influences on behaviour. He also contributed to the studies of law and political science, crowd psychology and economics. His *Economic Psychology*, 2 vol. (1902), later stimulated J. A. Hobson in England and Thorstein Veblen and his followers in the United States.

Bibliography.—Henri Bergson *et al., Gabriel Tarde* (1909); M. M. Davis, *Psychological Interpretations of Society* (1909); H. E. Barnes, "The Philosophy of the State in the Writings of Gabriel Tarde," *Phil. Rev.*, vol. xxviii, pp. 248–79 (1919). (H. E. Bar.)

TARDIEU, ANDRÉ PIERRE GABRIEL AMÉDÉE (1876–1945), French politician and writer, was born in Paris on Sept. 22, 1876. Educated at the École Normale Supérieure, he was *chef de cabinet* to Waldeck-Rousseau from 1899 to 1902. From 1902 to 1914 he lectured at the École des Sciences Politiques and the École Supérieure de la Guerre, and was foreign news editor on *Le Temps*. Entering the chamber of deputies in 1914 he acted as special commissioner in the United States (1917–18), member of the Peace conference (1918–19), and minister of the liberated regions (1919–20). As captain of Chasseurs between 1914 and 1916, he was wounded and three times mentioned in dispatches. At the Peace conference (1919) he had a considerable share in drafting the political and territorial clauses. He presided over the Committee of Five which drafted the Allied reply to the German observations on the draft terms of the peace treaty. Tardieu also presided over the Alsace-Lorraine committee, the Committee of the Saar and that of the execution of the treaty.

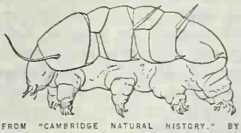
Tardieu's political attitude was characterized by his inflexible attachment to the policy of Clemenceau, with whom he was associated in the closest collaboration for war and for peace. Three times in succession, between 1919 and 1924, he refused to enter the government, in order to maintain his independent defense of this policy. As director of the *Echo National*, the daily paper founded by Clemenceau and himself, he waged a fierce war in foreign affairs, against the successive revisions of the treaty of Versailles; and in internal affairs, against the steps which paved the way for the victory of the *cartel des gauches*. Between 1924 and 1926 he had no seat in the chamber, but was again a member from 1926 to 1936. He was premier, 1929–30, and again in 1932. Tardieu died Sept. 17, 1945, at Menton, France.

Tardieu was the author of the following historical works: *Questions Diplomatiques* (1905); *France and the Alliances* (1908); *La conférence d'Algésiras* (1909); *Le Prince de Bulow* (1909); *Fürst v. Bulow* (1910); *La France et les alliances* (1910); *Le mystère d'Agadir* (1912); *Notes sur les Etats-Unis* (1917); *L'Amérique en armes* (1919); *La Paix* (1921); *Le Slesvig et la Paix* (1925).

TARDIGRADA, a class of minute animals doubtfully placed among the arthropods (*q.v.*). The largest of the bear animalcules or water bears, as these animals have been commonly called, are not much more than 1 mm. in length, while most of them are very much smaller. They are free-living forms found in variable habitats, in damp moss, on flowering plants, in sand, fresh water and even in the sea; and in adaptation to this wide range of external conditions a large number of genera and species have been evolved.

The water bear has a well-developed head region and a short body composed of four fused segments, each represented externally by a pair of short, stout, unjointed limbs generally terminated by sharp claws, which seem to range in number from four to nine, but sometimes, as in *Batillipes*, ending in three pairs of racquet-shaped, digitiform outgrowths. The limbs of the last pair, project backward from the posterior end of the body on each side of the anus. The cuticle, which is not chitinized, may be smooth or sculptured in various ways and is sometimes strengthened with segmentally arranged plates (as in *Echiniscus*) and sometimes provided with long paired hairlike cirri. An interesting feature of the structure of Tardigrada is the more or less fixed number of cells of which their body is composed. Thus the median dorsal strip of epidermis is composed of only 24 pairs of cells.

No special organs of circulation or respiration are known. The alimentary canal traverses the body from end to end. Within the fore gut is a pair of protrusible stylets, large glands that open into the esophagus, a muscular pharynx and large median and lateral glands that open into the intestine posteriorly. The nervous system, which is remarkably well-developed, consists of a bilobate cerebral ganglion, frequently provided with a pair of eye-spots, and of a ventral chain of five large ganglia connected by lateral commissures. The sexes are not distinct and the generative products are discharged either into the posterior end of the alimentary canal or directly to the exterior through a median pore in front of the anus. The eggs are large and in some cases at least are enclosed in the cuticle of the parent which is cast to form a case for them. The young when hatched are about one-third the size of the parent which they closely resemble except for the occasional absence of one pair of limbs which develop later.



FROM "CAMBRIDGE NATURAL HISTORY." BY PERMISSION OF THE MACMILLAN COMPANY, PUBLISHERS

ECHINISCUS SPINULOSUS, AFTER DOYÈRE

The Tardigrada are divided into two orders: (1) the Heterotardigrada containing two families with nine genera (*Echiniscus, Batillipes*, etc.) in which the head is provided with two pairs of cirri in front, a pair of lateral cirri and a so-called *clava* or spatulate process, one on each side; and (2) the Eutardigrada containing two families and three genera (*Macrobiotus, Hypsibius* and *Milnesium*) in which cephalic cirri are wanting.

The most remarkable feature of the tardigrades is their ability to withstand desiccation and low temperature, a phenomenon known as anabiosis. Specimens kept for eight days in a vacuum, transferred for three days into helium gas at room temperature, then exposed for several hours to a temperature of −272° C. came to life again when they were brought to normal room temperature; further, 60% of specimens kept for 21 months in liquid air at a temperature of −190° C. came to life. Specimens allowed to dry at room temperature and kept dry for 18 months returned to life in about two hours after having been transferred into water.

The Tardigrada cannot be assigned with complete confidence to the Arthropoda until it is proved that their oral stylets are modified appendages. Assuming that they belong to the Arthropoda, they must rank as a class by themselves. Some authors give Tardigrada phylum rank; a few scholars have suggested that the group is closer to Annelida.

Bibliography.—E. Marcus, "Tardigrada," in Ward and Whipple's *Fresh-Water Biology*, 2nd ed. by W. T. Edmondson (1959) and "Tardigrada," in Bronn's *Klassen und Ordnungen* (1929); R. W. Pennak, "Tardigrada. Ecology of the Microscopic Metazoa, etc.," *Ecol. Monographs*, 10:572–579.

TARDIVEAU, RENÉ: *see* Boylesve, René.

TARENTUM (Taras), an ancient Greek city of southern Italy, situated on the north coast of the gulf of the same name, on a rocky peninsula at the entrance to the only secure harbour in it. The entrance was defended by the two islands called the Choerades (now S. Pietro and S. Paolo). It was a Spartan colony, founded by Phalanthus about 708 B.C. (*See* Sparta.) Taras was a mythical hero, son of Neptune. Situated in a fertile district, especially famous for olives and sheep, with an admirable harbour, great fisheries and prosperous manufactures of wool, purple and pottery, Tarentum grew in power and wealth and extended its domain inland. A great defeat by the natives in 473 B.C., led to a change of government from aristocracy to democracy. A feud with the Thurians about the district of the Siris was settled in 432 by the joint foundation of Heraclea.

In the 4th century Tarentum was the first city of Magna Graecia, and its wealth and artistic culture at that time are amply attested by its coins. (*See* Numismatics.) In the second half

of the century Tarentum was at constant war with the Lucanians, and did not hold its ground without the aid of Spartan and Epirote *condottieri*. Then followed war with Rome (281), the expedition of Pyrrhus, whom Tarentum summoned to its aid, and at length, in 272, the surrender of the city by its Epirote garrison. Tarentum retained nominal liberty as an ally of Rome. In the Second Punic War it went over to Hannibal in 212, and suffered severely when it was retaken and plundered by Fabius (209), who sold 30,000 citizens as slaves. It revived after receiving a colony in 123 B.C., which received the name of Neptunia. In the time of Augustus it was essentially Greek and a favourite place of resort (Horace, *Od.*, iii, 5, 53). Belisarius ordered it to be refortified, but it was soon taken by Totila, who made it his treasure store. After his defeat by Narses, it was sold to the Byzantine empire by its Gothic governor.

One of the most interesting discoveries of recent years has been that of a terramara on the so-called Scoglio del Tonno on the northwest of the town, which in its type and in the character of the objects found there, is exactly identical with the terremare of the Po valley. It seems, however, to be an isolated colony, and not to prove a parallel development in north and south Italy. The only relic of any building of the Greek city is a part of a Doric temple on the island—two fluted columns, with a lower diameter of 6½ ft., and a height of 28 ft., and some fragments of the entablature, belonging probably to the beginning of the 6th century B.C. The rock occupied by the modern town was the citadel, but was connected with the land to the west by an isthmus, which was only cut through by Ferdinand I of Aragon. The line of the walls which defended the city on the east (land) side has been traced, and a few remains of well cut blocks, with Greek masons' marks, have been found. In the centre of the Agora was the huge bronze Zeus by Lysippus, and facing on to it a painted portico, with pictorial representations of the life of Phalanthus, and the foundation of the city, and the museum. There was also a fine gymnasium and other buildings mentioned by classical writers. Strabo's description of the site (vi, 3, 1) is a good one. The Roman amphitheatre, on the other hand, and remains of Roman baths by the seashore, have been found; the former perhaps occupies the site of the ancient theatre, in which the Roman ambassador was received in 281 B.C. Tarentum was the birthplace of Archytas and Aristoxenus (*qq.v.*). *See* also TARANTO.

TÂRGOVIŞTE (TÎRGOVIŞTE), a city of Rumania, in the region of Ploeşti, situated at the foot of the Carpathians, on the right bank of the Jalomita river, 48 mi. N.N.W. of Bucharest. Pop. (1963 est.) 27,900. A branch line connects Târgovişte with the main Walachian system, and is prolonged northward into the hills, where there are rich deposits of petroleum, salt and lignite. Coal is also found but not worked. Apart from the scanty ruins of a 14th-century palace, the most interesting building in the town is the Metropolitan church, still one of the finest in the country, with its nine towers and monuments of the princely house of Cantacuzino. It was founded in 1515 by Neagoe Basarab, builder of the famous cathedral of Curtea de Argeş. Târgovişte is a garrison town, with a cavalry training school and an artillery depot and repairing arsenal.

Târgovişte was the capital of Walachia from 1383 to 1698. In the 15th century it was sacked by the Szeklers. Michael the Brave defeated the Turks under its walls in 1597. In the 16th century it had a population of 60,000 and contained 70 churches and 40 convents. Its importance and population decreased when the capital was moved to Bucharest.

TARGUM. The Targums are Aramaic translations—often interpretations or paraphrases—of the books of the Old Testament. In their earliest form they date from the time when Aramaic superseded Hebrew as the spoken language of large numbers of the Jews. (*See* HEBREW LANGUAGE; SEMITIC LANGUAGES.) In their origin they were designed to meet the needs of the unlearned among the people who had ceased to understand the Hebrew of the Old Testament. In the absence of any precise evidence on the point it is impossible to give more than a rough estimate as to the period at which Hebrew, as a spoken language, was generally displaced by Aramaic. It is certain, however, that Aramaic was firmly established in Palestine in the 1st century A.D., despite the fact that Hebrew still remained the learned and sacred tongue. Hence it may be reasonably inferred that the mass of the people had adopted Aramaic at a considerably earlier period, probably as early as the 2nd century B.C., and that the need of Aramaic translations of the sacred text made itself felt only a little later.

Just as the old Greek translation of the Pentateuch, the Septuagint (*q.v.*), came into being in Egypt about 200 B.C. to satisfy the needs of a Jewish population that increasingly favoured Greek over Hebrew as its vernacular, so, it appears, did the Targum on the Pentateuch appear at about the same time in western Asia—*i.e.*, Babylonia and Judaea—where Aramaic had become the commonest vernacular tongue. At first the Pentateuch alone was involved, apparently because only the Law was as yet canonized as Sacred Scripture. In time, however, in the case of both the Septuagint and the Targums, the books of the Prophets and of the Hagiographa (Psalms, Proverbs, Job, Ruth, Lamentations of Jeremiah, Song of Solomon, Ecclesiastes, Esther, I and II Chronicles; for Ezra, Nehemiah and Daniel, however, no Targums have been found) were turned into Greek and Aramaic.

The status and influence of the Targums became assured after the Temple was destroyed, when synagogues replaced the Temple as houses of worship. For it was in the synagogue that the practice of reading from the Law and from the Prophets became widely observed, along with the custom of providing these readings with a translation into Aramaic.

This role of the Aramaic Targum became so generally recognized that by the 2nd century A.D. the Mishna (codification of the oral tradition, incorporated in the Talmud) takes it for granted and merely inculcates certain regulations to be observed by the *Meturgeman* ("translator"), who had by this time acquired a definite status. The Mishna indicates that the *Meturgeman*, who was distinct from the reader, translated each verse of the Law into Aramaic as soon as it had been read in Hebrew; in the readings from the Prophets three verses might be read at a time (*cf.* Luke iv, 16 ff.; Acts xiii, 14 ff., 27; xv, 21).

The translator endeavoured to reproduce the original as closely as possible, but, inasmuch as his object was to give an intelligible rendering, a merely literal translation would sometimes be insufficient; hence the translators tended to take a more elastic view of their obligations. To prevent misconceptions, they expanded and explained what was obscure, adjusted the incidents of the past to the ideas of later times, emphasized the moral lessons to be learned from the national history and, finally, adapted the rules and regulations of the Holy Scriptures to the conditions and requirements of their own age. As time went on, the practice of introducing additional matter of an edifying character grew in popular favour and was gradually extended. The method by which the text was thus utilized as a vehicle for conveying homiletic discourses, traditional sayings, legends and allegories is abundantly illustrated by the later Targums, as opposed to the more literal translations of the earlier Targum of Onkelos (*see* below) and on the Prophets.

It would be incorrect to suppose, however, that the translation of the text was left entirely to the individual taste of the translator. The latter is rather to be regarded as the representative of a school of thought reflecting the spirit of the age in which he lived.

There can be little doubt that the Targums existed for a long time in oral form, and so long at least as the Targum tradition remained active there would be little temptation to commit it to writing. But it is highly probable that this procedure, in the case of the Targums, was mainly followed with respect to those parts of the Old Testament which were read in the synagogical services (*e.g.*, the Law and the Prophets) and that it was less rigidly observed in regard to the other portions of Scripture; a written translation of the latter would be of special value for the purpose of private study. Hence there is no need to reject the tradition as to the existence of a written Targum on Job in the time of Gamaliel I (1st century A.D.), especially as references to Targum manuscripts occur in the Mishna and elsewhere. But it was not these manuscripts but the living tradition of the learned that

was recognized as authoritative throughout the period that closes with the compilation of the Talmud in the early centuries of the Christian era. The official recognition of a written Targum, and therefore the final fixing of its text, belongs to the post-Talmudic period, and is not to be placed earlier than the 5th century A.D.

TARGUMS ON THE PENTATEUCH

Onkelos.—The so-called Targum of Onkelos owes its name to a mistaken reference in the Babylonian Talmud. In its original context, that of the Jerusalem Talmud, the passage refers to the Greek translation of Aquila the proselyte. Onkelos the proselyte was mistakenly identified with him and then credited with the Aramaic translation of the Pentateuch. With the exception of this one reference, the Targum is always introduced in the Babylonian Talmud by the phrase "as we translate" or "our Targum"; it is probable, therefore, that the name of the author, or authors, was unknown to the Babylonian Jews. It is first quoted under the title of the Targum of Onkelos by the gaon Sar Shalom (d. *c.* A.D. 859).

The Targum of Onkelos may be regarded as an official translation of the Law, in the Judaean Aramaic dialect, which was carried out in Babylon, probably about the 4th century A.D.; in its final form it cannot be earlier than the 5th century A.D. The translation, as a whole, is good, and adheres very closely to the Hebrew text.

Of all the extant Targums, that of Onkelos affords perhaps the most characteristic and consistent example of the exegetical methods employed in these works. Two principles may be said to have guided the translators. On the one hand, their primary object was to produce a faithful rendering of the original which at the same time would be intelligible to the people; for this purpose a purely literal translation would be insufficient. On the other hand, they regarded it as necessary to present the sacred text in such a manner as best to convey the particular form of interpretation then current. Later Jewish exegesis was especially concerned to eliminate everything in the sacred writings that might give rise to misconception with respect to God on the part of the unlearned.

Hence in the Targums various expedients were adopted for avoiding any reference to the Deity that might be misunderstood by the people or that involved apparent irreverence. Examples of this peculiarly Targumic method are: (1) the insertion of "word," "glory," or "presence" before the divine name, when God is referred to in his dealings with men; (2) the insertion of the preposition "before" when God is the object of any action; (3) the use of the passive for the active voice; (4) the use of periphrasis for the more pronounced anthropomorphisms, such as "to smell," "to taste"; (5) the use of different expressions or the insertion of a preposition before the divine name, when God is compared with man or the same action is predicated of God and man; (6) the use of distinctive terms "God" and "idol" where in the original the same characters are used to denote both God and heathen gods. Instances of this endeavour to maintain, as it were, a respectful distance in speaking of God occur on every page of the Targums; at the same time, however, and by no means infrequently, less obviously physical human actions and passions are ascribed to God, a phenomenon that requires further study.

Jerusalem Targums.—Two other Targums to the Pentateuch are cited by Jewish authorities under the titles of the Targum Jerushalmi and the Targum of Jonathan ben Uzziel. Of these, the former contains only portions of the Pentateuch and is therefore usually designated the Fragmentary (Jerusalem) Targum.

The second Jerusalem Targum owes its ascription to Jonathan ben Uzziel to the incorrect solution (probably in the 14th century A.D.) of the abbreviated form by which it was frequently cited (viz., Targum Jerushalmi), hence the term "Pseudo-Jonathan" by which it often is designated. This Targum often represents a later and more successful attempt to correct and supplement the Targum of Onkelos by the aid of variants derived from an older source. It exhibits, also, that tendency to expand the text by additions of every kind that has been already noted as characteristic of the later stages of targumic composition. Homilies, legends, traditional sayings and explanations are utilized by the targumist, so that at times his works convey the impression more of a late exegetical work than of a translation.

It is not possible to determine the sources and dates of composition of these two Targums. It is generally agreed that, though the extant manuscripts are late, they preserve elements that go back to an early period. Most of these elements derive from the text tradition that gave rise to the Targum of Onkelos, yet others are independent of Onkelos.

We must assume that a tolerably fixed Targum tradition existed in Palestine from quite early times, and that the Targum of Onkelos was the main representative of this Targum tradition that existed among the Jews down to the 7th century A.D., the period to which the internal evidence compels us to assign the Targumim. (*Cf.* B. J. Roberts, *The Old Testament Text and Versions,* pp. 197–207 [1951]).

TARGUMS ON THE PROPHETS

The official Targum on the Prophets is stated by the Babylonian Talmud to have been said by Jonathan ben Uzziel, the disciple of Hillel in Judaea, in the 1st century A.D., and is usually known, therefore, as the Targum Jonathan. Elsewhere in the Talmud, however, the quotations from this Targum are given under the name of Joseph bar Chijah, head of the school at Pumbeditha, in Babylonia, in the 4th century A.D. Both in language and style it closely resembles the Targum of Onkelos, and appears to have been modeled on that translation; in certain passages, indeed, it appears to have made use of it. Probably, like Onkelos, it did not assume its final form in Babylon before the 5th century A.D. It naturally follows from the character of the original that the rendering of this Targum is less literal than that of Onkelos, especially in the prophetic books, but, when due allowance is made for the difficulty of the Hebrew, it may be described on the whole as a faithful reproduction of the original text. Its peculiarities of rendering are because of the same principles that were noted as underlying the translation of the Pentateuch.

Anthropomorphisms, as a rule, are avoided by means of the same expedients as those employed by Onkelos, expressions derogatory to the dignity of God, or of the heroes of the nation, are softened down, while figurative language is either boldly transposed or its character clearly shown by introduction of the particle "as" or "like." There is, further, a tendency to interpret the prophetic utterances so as to make them applicable to the immediate enemies of the Jewish people. Lastly, in the obscurer passages the Haggadic method of interpretation (*i.e.*, employing legend, folklore, homiletics, etc.) is employed to its fullest extent, while the translation throughout shows a marked tendency to explanatory additions.

Of a Jerusalem Targum to the Prophets little is known, though it is hardly doubtful that such a Targum existed, if only in oral form.

TARGUMS ON THE HAGIOGRAPHA

These Targums are almost entirely the work of individuals, made in imitation of the older Targums.

Psalms and Job.—These Targums present certain features in common and may therefore be treated together. Like all the later Targums they exhibit a large amount of explanatory addition, chiefly Haggadic in character. At the same time the translation of the original is not neglected, and, when separated from the later accretions, this is found to follow the Hebrew fairly closely. Peculiar to these Targums are the double translations they give to many verses, one of which is usually Haggadic in character while the other is more literal. They cannot be earlier than the 7th century A.D. and possibly are of a considerably later date.

Proverbs.—This Targum stands apart because of the peculiarity of the language in which it is written. The influence of the Peshitta (Syriac) version of the Bible is so clearly marked that it has been described as a Jewish revision of that version. Setting aside the Syriasms due to the use of the Peshitta, however, the Targum shows affinity to the Targums on the Psalms and Job. The

translation is literal and almost entirely free from Haggadic additions.

Megilloth.—The chief characteristic of the Targums to the Megilloth (*i.e.*, the latest books of the Hagiographa, namely the Song of Solomon, Ruth, Lamentations of Jeremiah, Ecclesiastes and Esther) is their exaggerated use of paraphrase. They mark the final stage in the development of Haggadic interpretation, in which the translation of the text has frequently disappeared in a mass of fanciful matter.

The Targum of Esther is known in three recensions: (1) that of the Antwerp Polyglot (late 16th century), almost a literal translation; (2) that of the London Polyglot (1657), which gives practically the same text with many additions of a Haggadic character; (3) the so-called second (*sheni*) Targum, a much larger work, containing a collection of later Midrashim ("explanatory commentaries") to this book.

The Targum on the Song of Solomon is of a similar character to that of the "second" Targum on Esther. These have been dated halfway between the Babylonian Targums (Onkelos and that to the Prophets) and the Jerusalem Targums on the Pentateuch and those on the greater Hagiographa.

Chronicles.—This Targum was first edited from an Erfurt manuscript in 1680–83; a more complete and accurate edition from a Cambridge manuscript was published in 1715. In the translation, which at times is fairly literal, use appears to have been made of the Jerusalem Targums on the Pentateuch and of the Targums on the books of Samuel and Kings. The text represented by the Erfurt manuscript is assigned to the 8th century, that of the Cambridge manuscript to the 9th century A.D.

Other Books of the Hagiographa.—No Targums have been discovered to Daniel, Ezra and Nehemiah.

BIBLIOGRAPHY.—W. Bacher, "Targum," *Jewish Encyclopedia*, vol. xii, pp. 57–63; B. J. Roberts, "The Aramaic Targumim," *The Old Testament Text and Versions*, ch. 14, with bibliography (1951); P. Churgin, *Targum Jonathan to the Prophets* (1927); *The Targum to Hagiographa*, in Hebrew (1945); F. Rosenthal, *Die aramaistische Forschung seit T. Nöldekes Veröffentlichungen* (1939); J. F. Stenning (ed.), *The Targum of Isaiah*, with a translation (1949); A. Sperber (ed.), *The Bible in Aramaic Based on Old Manuscripts and Printed Texts*, vol. i, *The Penteteuch According to Targum Onkelos* (1959).

(J. F. S.; H. M. O.)

TÂRGU-MUREŞ, a town of Rumania, in the Magyar Autonomous Region, 79 mi. E. of Cluj by rail; ceded to Hungary in 1940 and returned after World War II. Pop. (1963 est.) 74,004. It is situated on the left bank of the Mureş, and is a well-built town, once the capital of the territory of the Szeklers. On a hill dominating the town stands the old fortress, which contains a beautiful church in Gothic style built about 1446, where in 1571 the diet was held which proclaimed the equality of the Unitarian Church with the Roman Catholic, Lutheran and Calvinistic churches. The Teleki palace contains the Teleki collections, which include a library of 70,000 volumes and several valuable manuscripts (*e.g.*, the Teleki codex), a collection of old Hungarian poems, and a manuscript of Tacitus, besides a collection of antiquities and another of minerals. Târgu-Mureş also has an interesting Szekler industrial museum. Trade is chiefly in timber, grain, wine, tobacco, fruit and other products of the neighbourhood. There are manufactures of sugar, spirits and beer.

TARIFF, a tax levied upon goods as they cross national boundaries. The words tariff, duty, and customs are generally used interchangeably to denote taxes on commodities that are traded internationally.

NATURE AND FUNCTION OF TARIFFS

Tariffs may have either a revenue-raising or a protective function, but a tariff designed primarily to raise revenue may exercise a strong protective influence and a tariff levied primarily for protection may yield revenue. Gottfried Haberler in his *Theory of International Trade* suggested that the best objective distinction between revenue duties and protective duties (disregarding the motives of the legislators) is to be found in their discriminatory effects as between domestic and foreign producers. If domestically produced goods bear the same taxation as similar imported goods, or if the goods subject to duty are not produced at home, even after the duty has been levied, and if there exist no home-produced substitutes toward which demand is diverted on account of the tariff, then the duty is not protective. A purely protective duty tends to shift production away from the export industries into the protected domestic industries and those industries producing substitutes for which demand is increased. On the other hand, a purely revenue duty will not cause resources to be invested in industries producing the taxed goods or close substitutes for such goods, but it will divert resources from the production of export goods to the production of those goods and services upon which the additional government receipts are spent. From the purely revenue standpoint, a country will select a relatively small number of articles of general consumption and subject them to low duties so that there will be no tendency to shift resources into industries producing such taxed goods (or substitutes for them). During the period when it was on a free-trade basis, Great Britain followed this practice, levying low duties on a few commodities of general consumption such as tea, sugar, tobacco, and coffee. If a country wishes to protect its home industries its list of protected commodities will be long and the tariff rates high.

Tariffs may be further classified into three groups—transit duties, export duties, and import duties.

Transit Duties.—This type of duty is levied on commodities that originate in one country, cross another, and are consigned to a third. As the name implies, transit duties are levied by the country through which the goods pass. Such duties are no longer important instruments of commercial policy. During the mercantilist period (17th and 18th centuries) and even up to the middle of the 19th century in some countries, such duties played a role in directing trade and controlling certain of its routes. The development of the German *Zollverein* (customs union) in the first half of the 19th century was partly the result of Prussia's exercise of its power to levy transit duties. The most direct and immediate effect of transit duties is to reduce the amount of commodities traded internationally and raise their cost to the importing country. In 1921, the Barcelona Statute on Freedom of Transit abolished all transit duties.

Export Duties.—Export duties are no longer used to a great extent, but they were common in the past and were significant elements of mercantilist trade policies. Their main function was to safeguard domestic supplies rather than to raise revenue. Export duties were first introduced in England by a statute of 1275 that imposed them on hides and wool. By the middle of the 17th century the list of commodities subject to export duties had increased to include more than 200 articles. With the growth of free trade in the 19th century, export duties became less appealing; they were abolished in England in 1842, in France in 1857, and in Prussia in 1865. At the beginning of the 20th century only a few countries levied export duties: Spain still levied them on coke and textile waste; Bolivia and Malaya on tin; Italy on objects of art; and Rumania on hides and forest products. Export duties are generally levied by raw-material-producing countries rather than by advanced industrial countries. The neo-mercantilist revival in the 1920s and 1930s brought about the reappearance of export duties, but only to a limited extent. Export duties were not levied by the United States and were specifically prohibited by the Constitution. This prohibition was incorporated in the Constitution mainly because of pressure from the South, which wanted no restriction on its freedom to export agricultural products.

If the primary objective of the export duty is revenue, rates are usually low, for in the last analysis the duty is borne by the domestic producer, who cannot charge prices higher than the world prices. In South and Central America practically all exported commodities are taxed, the duties ranging from 1% to 5%. Several countries depend upon export duties for a large part of their revenue. Chile, for example, at one time derived more than 75% of her total revenue from export duties.

To the extent that export duties raise the cost of raw materials to foreign manufacturers relative to the cost to manufacturers in the exporting country, export duties may act as a form of protection to domestic industries and stimulate their output. As

examples, Norwegian and Swedish duties on exports of forest products were levied chiefly to encourage milling, woodworking, and paper manufacturing at home. Similarly duties on the export from India of untanned hides after World War I were levied to stimulate the Indian tanning industry. In a number of cases, however, duties levied on exports from colonies were designed to protect the industries of the mother country and not those of the colony.

If the country imposing the export duty supplies only a small share of the world's exports and if competitive conditions prevail, the burden of an export duty will likely be borne by the domestic producer, who will receive the world price minus the duty and other charges. But if the country produces a significant fraction of the world output and if domestic supply is sensitive to lower net prices, then output will fall drastically and world prices may rise and as a consequence not only domestic producers but also foreign consumers will bear the export tax. How far a country can employ export duties to exploit its monopoly position in supplying certain raw materials depends upon the success other countries have in discovering substitutes or new sources of supply.

Import Duties.—Import duties are the most important and most common types of custom duties. As noted above, they may be levied either for revenue or protection or both. Tariffs are not a wholly satisfactory means of raising revenue, for tariff rates cannot be readily adjusted to current fiscal needs. In the United States, for example, revenues from import duties in 1808 amounted to twice the total of government expenditures, while in 1837 they were less than one third of such expenditures. Until near the end of the 19th century the customs receipts of the U.S. government made up about half of all its receipts. This share had fallen to about 6% of all receipts before the outbreak of World War II and by 1965 was down to about $1\frac{1}{2}\%$.

A tariff may be either specific, ad valorem, or compound (*i.e.*, a combination of both). A specific duty is a levy of a given amount of money per unit of the import, such as $1 per yard or per pound. An ad valorem duty, on the other hand, is calculated as a percentage of the value of the import. Ad valorem rates furnish a constant degree of protection at all levels of price (if prices change at the same rate at home and abroad), while the real burden of specific rates varies inversely with changes in the prices of the imports. A specific duty, however, penalizes more severely the lower grades of an imported commodity. This difficulty can be partly avoided by an elaborate and detailed classification of imports on the basis of the stage of finishing, but such a procedure makes for extremely long and complicated tariff schedules. Specific duties are easier to administer than ad valorem rates, for the latter often raise difficult administrative issues with respect to the valuation of imported articles.

A list of all import duties is usually known as a tariff schedule. A single tariff schedule, such as that of the United States, applies to all imports regardless of the country of origin. This is to say that a single duty is listed in the column opposite the enumerated commodities. A double-columned or multi-columned tariff provides for different rates according to the country of origin, lower rates being granted to commodities coming from countries with which tariff agreements have been negotiated. Most trade agreements are based on the most-favoured-nation clause, which extends to all nations whatever concessions are granted to the most-favoured nation. (*See* MOST-FAVOURED-NATION TREATMENT.)

Every country has a free list that includes articles admitted without duty. By looking at the free list and the value of the goods imported into the United States under it one might be led to conclude that tariff production is very limited, for more than half of all imports are exempt from duties. Such a conclusion, however, is not correct, for it ignores the fact that the higher the tariff, the less will be the quantity of dutiable imports. Attempts to measure the height of a tariff wall and make international comparisons of the degree of protection, based upon the ratio of tariff receipts to the total value of imports, are beset by difficulties and have little meaning. A better method of measuring the height of a tariff wall is to convert all duties into ad valorem figures and then estimate the weighted-average rate. The weight should reflect the relative importance of the different imports; a tariff on foodstuffs, for example, may be far more important than a tariff on luxuries consumed by a small group of people. A tariff index may be used to measure either the degree of protection afforded by a tariff schedule or the relative degree of import obstruction that it provides. If the amount of protection is under consideration then it must be conceded that even a uniform ad valorem rate of duty will provide different protection to different countries, depending in part upon demand and cost differences between countries. The degree of obstruction to imports, on the other hand, would depend not only upon the height of the tariff but also upon the responsiveness of demand and supply, the changes of which are not considered in computing the tariff index. Consequently, it is misleading to construct only one tariff index for a given country. For each country a number of tariff indexes should be computed, one for each country from which it imports, for it often happens that a tariff severely restricts imports from certain countries but has little effect on imports from other countries.

TARIFFS IN THE 20TH CENTURY

Viewing the development of tariffs and international trade since 1900, three broad periods can be distinguished: (1) from 1900 to World War I, (2) from 1919 to World War II, and (3) from World War II to the 1960s. During the first period tariffs were the main form of import control. They were moderate and relatively stable, and equality of competitive opportunity for imports from all countries was the rule, with exceptions found mainly in the trade relations between a few countries and their colonies.

The second period, from 1919 to 1939, was marked by frequent changes in duties, their rise to a higher level, the development of new control devices, and the negotiation of special trade agreements. There was some resurgence of protective tariffs immediately after World War I, caused in part by the strong national feelings aroused by the war and by difficult postwar economic adjustments. As political and economic conditions improved toward the middle of the 1920s the severity of trade restrictions was relaxed. In the late 1920s many governments negotiated commercial treaties that resulted in mutual reductions in duties. Soon after the beginning of the depression of the 1930s, however, this liberal trend was reversed. Import controls and higher tariffs were adopted by one country after another. Two outstanding examples of such legislation were the Smoot-Hawley Tariff (1930) in the United States and the British tariff of 1932 that ended nearly a century of virtual free trade for Great Britain. The world economic conference that convened in London in 1933 to deal with tariffs and international trade failed because it came too late to supersede the various national recovery programs that had already gained momentum in many countries, particularly the United States. Despite the moderating influence of the U.S. Reciprocal Trade Agreements Act (1934), which was designed to reduce tariffs and stimulate trade, the dominant trend in the trade policies of most countries during the 1930s was in the opposite direction; no effective methods of concerted action were devised for checking the deterioration of the economic relations among nations.

The third period, from World War II to the 1960s, brought radical changes in the nature and control of world trade. During the war the shortage of many types of goods for ordinary civilian use, and of the shipping to carry them, limited the volume of ordinary trade. Under these circumstances, tariffs and other means of control became subordinate to the decisions of the exporting countries as to whether the particular goods could be spared and as to whether transport was available to carry them. To a large extent, governments themselves participated in the procurement and distribution of many classes of goods. As the war progressed, the enlarged production programs of the Allies and of the countries friendly to them resulted in huge international movements of supplies. The problem of paying for these goods was virtually eliminated by the U.S. Lend-Lease Act, which authorized the president to make them available to friendly countries on a lend-lease basis, *i.e.*, without immediate compensation. A significant feature of the agreements concluded by the United States with more than 30 of the United Nations was the uniform pledge of all

the contracting countries to collaborate after the war in a program designed to promote more liberal conditions for international trading and a more prosperous expanding world economy.

Soon after the war the United States and Canada, in order to facilitate the resumption of normal trade, relaxed their controls upon exports and extended loans to many other countries. Western European countries negotiated numerous trade agreements with each other. Unlike their predecessors of the prewar era, these agreements did not deal with changes in tariffs or other trade controls. The negotiations centred rather upon the lists of goods that each country was desirous of obtaining from the other and the delivery of which the other government agreed to facilitate. Generally speaking, even after the transitional period following World War II, the course of international trade was determined only to a small degree by the height of the tariff or by licensing systems. The main determinants were the needs of the war-torn areas for products essential to the rehabilitation of their economies.

By the close of 1947, new or tightened import restrictions had been brought into operation by various countries to prevent further loss of foreign exchange. These restrictions were directed chiefly against imports from the dollar countries of North and South America. To ease the settlement of uneven balances in the trade among the European countries themselves, a special intra-European payments plan was inaugurated in 1948. Under this arrangement portions of the dollar aid allocated through the Marshall Plan to European countries were made conditional upon the setting up by these countries of special accounts of equivalent value in their own currencies, against which members with whom they had an export surplus could draw to finance their deficits. This was later replaced by a flexible system for clearing balances, known as the European Payments Union (*q.v.*).

The progressive improvement in production, monetary stability, and trade balance which many countries were making gave promise, by the beginning of 1950, of some trade normalization. Early in 1950 the western European countries that had received Marshall Plan aid agreed upon a series of steps for the liberalization of import quotas. Later, smaller groups of countries formed customs unions such as the Benelux customs union, the European Economic Community (Common Market), the European Free Trade Association (EFTA), and the Latin American Free Trade Association. (*See* ECONOMIC UNION.) On the other hand, in the major countries of Asia, progress toward resumption of normal trade relations was still largely dependent upon political developments. Among the factors hampering trade in Asia were the Communist control of China and the continued disturbances in Southeast Asia. On the favourable side were the stabilization of the Japanese yen and the transfer of Japanese trade to private channels.

General Agreement on Tariffs and Trade.—The most important expression of the postwar desire of the nations of the world to liberalize and expand international trade was the General Agreement on Tariffs and Trade (GATT). It consisted of an integrated set of bilateral trade agreements (negotiated in five international conferences) aimed at the abolition of quantitative trade restrictions and the reduction of tariff duties among the contracting parties. GATT was concluded at Geneva, Switz., in 1947 and was originally considered as an interim measure pending the formation of a United Nations agency to supersede it. For this latter purpose the United Nations Economic and Social Council called an international conference on trade and employment that met in Havana, Cuba, from November 1947 to March 1948 to draw up the charter of an agency to be known as the International Trade Organization (ITO). The ITO was designed to support national policies for economic development and full employment and contribute to the gradual liberalization of world trade. Simultaneously the members of the preparatory committee of the ITO charter met in Geneva and negotiated the General Agreement on Tariffs and Trade. Of the 56 nations represented at Havana, 53 signed the proposed ITO charter, but only one subsequently ratified it, and plans for ITO were therefore abandoned. Since the ITO did not come into being, GATT was amplified and further enlarged at Annecy, Fr., in 1949, at Torquay, Eng., in 1951, and at Geneva in 1956 and 1961. A new conference for further negotiations, known as the "Kennedy round" (in honour of Pres. John F. Kennedy during whose term the U.S. Trade Expansion Act of 1962 was passed) formally opened in Geneva in May 1964.

GATT includes a long schedule of specific tariff concessions for each contracting party, representing tariff rates which each country has agreed to extend to others. In addition it includes a set of general rules for the conduct of international trade; these rules are, in effect, a code of commercial policy. They provide for unconditional most-favoured-nation clauses, elimination of quantitative trade restrictions, uniform customs regulations, and the obligation of each contracting party to negotiate for tariff cuts upon the request of another. The negotiating procedures followed under GATT agreements are essentially multilateral. The close contacts among the country representatives who attend the negotiating sessions, and the fact that each country is affected not only by its bilateral negotiations but also by the negotiations under way among other pairs of countries, make GATT negotiations in effect multilateral.

By 1965 the membership of GATT included 64 countries, which accounted for more than four-fifths of all world trade. Trade negotiations under GATT covered scores of thousands of commodities. Tariff concessions were granted on about two-thirds of the total imports of all GATT countries.

For current developments in tariff negotiations see *Britannica Book of the Year.*

BIBLIOGRAPHY.—F. W. Taussig, *Tariff History of the United States* (1931); Gottfried Haberler, *The Theory of International Trade with its Applications to Commercial Policy* (1936); Asher Isaacs, *International Trade* (1948); John Letiche, *Reciprocal Trade Agreements* (1948); Lawrence Towle, *International Trade and Commercial Policy* (1956); Delbert Snider, *Introduction to International Economics* (1958); S. Enke and V. Salera, *International Economics* (1957); J. B. Condliffe, *The Commerce of Nations* (1950). (A. A. Ps.)

TARIJA, a department of southeastern Bolivia established in 1832 during the administration of Andrés de Santa Cruz. Area 14,526 sq.mi., pop. (1961 est.) 141,000. It is bounded south by Argentina, east by Paraguay, and west and north by the departments of Potosí and Chuquisaca. The plains provide excellent grazing lands for breeding horses, cattle and sheep. Corn, wheat, fruits and vegetables are grown.

The eastern tropical region in the Chaco (*q.v.*) lowland produces tobacco, sugar cane, rice, bananas, cassava and citrus. Mineral resources include petroleum, iron and silver. Oil refineries are at Sanandita and Fortín Campero.

Tarija, the departmental capital (pop. [1961 est.] 20,846), situated in a small valley on the banks of the Guadalquivir river, was founded by Don Luis de Fuentes y Vargas, in 1574. Its elevation, 6,250 ft., helps to temper the climate. Tarija may be reached by air, and motor road from Tupiza, and from La Quiaca, Arg.

(J. L. TR.)

TARIM, the chief river of Sinkiang province, China, which gives its name to the great basin between the Tien Shan and Kunlun mountain systems of central Asia. It is 1,360 mi. (2,190 km.) long, and the area of the basin is approximately 300,000 sq.mi., though another 50,000 sq.mi. portion of the southwest Gobi desert sometimes is included with the Tarim basin.

The Tarim river is formed by the confluence of the Kizil (Kashgar) and the Yarkand rivers in the far west; it then flows for about 230 mi. N.E., to be joined from the north by the Ak Su and shortly, from the south, by the Khotan, though the latter only reaches the Tarim for a period of about 40 days per year. Trending eastward, along the northern side of the Takla Makan desert, the Tarim divides repeatedly and has inter-dune lakes along its course. The Konche, issuing from the lake called Baghrash Köl, joins the Tarim just eastward of long. 88° E. The lower course of the Tarim is sluggish, diminishing in volume and is aggrading its bed. In 1921 it broke out of its channel and returned to an old course followed previously to A.D. 330, also moving the marshy Lop-Nor back to its classical location. The river is usually frozen from Dec. through Feb. The March thaw causes inundations, followed by a great summer flood as the snows melt in the Tien Shan and the Kunlun.

The basin properly consists of three regional divisions though

sometimes five units are distinguished. The first division is the Takla Makan desert or central, waterless sand desert. The terminal depression of Lop-Nor is the second division, which often is considered a part of the Takla Makan. Its salt-encrusted dry lake beds, old river channels and changeable marshy lake terminals of the Tarim seldom have permitted permanent occupancy. The third division comprises the belts of oases surrounding the Takla Makan on the south, west and north along the piedmont fringes north of the Kunlun, east of the Pamir and south of the Tien Shan. These oases lie between the gravel glacis of the upper piedmonts and the sands of the Takla Makan. The best-known oases on the south are Karghalik, Khotan, Keriya, Niya, Cherchen and Charkhlik. To the west are Yarkand and Kashgar. On the north are Wu-shih, Ak Su, Bai, Kuchche, Korlo and Karashahr. The southern oases have smaller water supplies and cultivable areas than those of the north, but the western oases are the largest. The two other regional units often included in the Tarim, for convenience, are: the Turfan depression, lying beyond the Kuruk Tagh to the northeast in the indeterminate zone between the Tarim basin and the Gobi country; and the Su-lo (Shuleh) basin at the western end of the Kansu corridor, north of the Nan Shan.

Problems of Climatic History.—The present-day climate of the Tarim basin is discussed in the article SINKIANG, but the problems of past climatic changes have been the subject of much study. In general the belief has increased that climate has not changed markedly since the end of the last glacial era, but that the water supplied by the melting of glacial ice left in the high mountains has been failing. The problem is connected with the supposed desiccation of central Asia. (*See* ASIA: *Exploration.*)

Regardless of cause, great changes have occurred in the Tarim basin during the last 2,500 years. Settlements away from the piedmont were abandoned, sometimes to be replaced by smaller settlements higher on the piedmont edge. Some localities formerly well watered and prosperous now are deserted and quite uninhabitable. Many streams no longer penetrate the desert fringe, and the Tarim has shifted its course repeatedly, to alter the route of the Silk road.

Historical Geography.—The chief old land route of contact between eastern and western Asia passed through the Tarim basin. A northern route along the southern foot of the Tien Shan connected the trans-Caspian lowlands, via Kashgar and the Pamir passes, with the Kansu corridor in the east; a southern route along the northern Kunlun piedmont made the same connections. The region was always essentially a corridor zone, since there never was sufficient moisture, widely distributed, to support extensive plant, animal or human life. This explains why the great nomadic tribes of Dzungaria, the Wu-sun, Sakas, Yueh-chi, Huns, Turks and Mongols raided or made tributary the Tarim oases, but never permanently crossed the Tien Shan to occupy the basin. Yet the oases provided the resource for careful irrigation agriculture to small groups of people, and provided food, animals and water for small numbers of migrants, travelers and traders.

Modern Conditions in the Tarim Basin.—Basically, the Tarim is yet untouched by modern developments, either of mineral resources or of water conservation-irrigation systems. Chinese administration, in times of strong Chinese power, has kept the routes open and free from banditry, but during the 19th century such periods were short. The oases populations chiefly are Turkic peoples, unsympathetic to the Chinese, but neither sufficiently numerous nor united to develop their regional entity. The Chinese fear of foreign exploitation has prevented both full exploration of possibilities and the development of known resources. The lack of transport development has kept connections uncertain and movement both slow and costly. The Tarim has lain too far from the centre of any modern regional development to have shared in that growth. The Communist Chinese have plans for marked development of several localities of the basin, but the comparatively small opportunities make them of relatively low priority and, by the early 1960s, little was achieved other than the political organization of the region and the settling of Chinese nuclei in most of the important oases. A few agricultural processing plants were established, and resource exploration begun. The extension of the railway through the Kansu corridor into Dzungaria, in the late 1950s, and the eventual connections with the Soviet rail system may do much to lessen the remoteness of the Tarim basin.

Oasis agriculture remains the mainstay of the scattered settlements. Grains, cotton, silk, fruits and wool are the chief agricultural products, and Khotan jades are the only other important item. Little is exported from the basin, but local surpluses find a market among travelers passing through when conditions permit. *See* also SINKIANG; KASHGAR; KHOTAN; YARKAND; LOP NOR; GOBI; TAKLA MAKAN; DZUNGARIA. (J. E. SR.)

TARKINGTON, (NEWTON) BOOTH (1869–1946), U.S. novelist and dramatist whose satirical pictures of midwesterners represent his best work, was born in Indianapolis, Ind., July 29, 1869. He studied at Phillips Exeter academy, Exeter, N.H., and at Purdue (Lafayette, Ind.) and Princeton universities. In 1902–03 he was a Republican member of the Indiana legislature. A versatile and prolific writer, Tarkington won early recognition with the melodramatic novel, *The Gentleman From Indiana* (1899). This was followed by the immensely popular romance *Monsieur Beaucaire* (1900), which he later adapted for the stage. *The Conquest of Canaan* (1905), a "problem" novel, was probably the most mature work of his early period. His humorous portrayals of boyhood and adolescence, *Penrod* (1914) and *Seventeen* (1916), became classics of their kind. He was equally successful with his portrayals of midwestern life and character: *The Turmoil* (1915), *The Magnificent Ambersons* (1918; Pulitzer prize, fiction, 1919) and *The Midlander* (1924), combined as *Growth* (1927); and *The Plutocrat* (1927). *Alice Adams* (1921; Pulitzer prize, fiction, 1922), a searching character study, is perhaps his most finished novel. He continued his celebrated delineations of feminine character in *Claire Ambler* (1928), *Mirthful Haven* (1930) and *Presenting Lily Mars* (1933), and wrote several domestic novels in his later years. His many plays include *The Man From Home* (1906), *Mister Antonio* (1916) and *Up From Nowhere* (1919), all with H. L. Wilson, and *Beauty and the Jacobin* (1912). He died on May 19, 1946, in Indianapolis.

Adopting William Dean Howells and Mark Twain as his literary models, Tarkington possessed an informal, charming style and a gift for characterization. However, sentiment often obscured his critical vision.

BIBLIOGRAPHY.—R. C. Holliday, *Booth Tarkington* (1918); James Woodress, *Booth Tarkington, Gentleman From Indiana* (1955); D. Russo and T. Sullivan, *A Bibliography of Booth Tarkington* (1949).

TARLAC, an agricultural province on the central plain of Luzon, Republic of the Philippines, in the region drained by the Agno river. Area 1,175 sq.mi., pop. (1960) 426,647. Soils are of alluvial origin and range from sandy loams, through clay loams to clays. In many areas the subsoil is a tight clay. Rice is the principal crop and in a part of the province two crops per year are possible, the second by means of controlled irrigation. Sugar cane is the second crop, and one of the most modern sugar *centrals* in the republic is at San Miguel, near the centre of the province. Other crops are corn, beans, tobacco, sweet potatoes and coconuts. Pampango is the principal vernacular, but Tagalog, Ilocano and Pangasinan are also spoken. With considerable absentee ownership, this was an area where the Hukbalahap movement was quite strong. (*See* PHILIPPINES, REPUBLIC OF THE.)

Tarlac municipality (pop. [1960] 98,285), the provincial capital, was founded in 1686 and lies on the railway and paved highway from Manila to northern Luzon. (AN. C.)

TARLTON (TARLETON), **RICHARD** (d. 1588), English actor, for several years chief comedian of the Queen's Players, was born, according to Fuller (*Worthies* iii, 139), at Condover, Shropshire. He was probably at one time an innkeeper. In 1583 he is mentioned as one of the original company of Queen's Players, and he had become an experienced actor; he remained with them until his death. Tarlton was Elizabeth I's favourite clown, and his talent for impromptu doggerel on subjects suggested by his audience is recorded in the verb "tarltonize." For the Queen's Men he wrote *The Seven Deadly Sins.* Many songs and witticisms of the day were attributed to him, and after his death *Tarlton's Jests,* many of them older than he, made several volumes. He is said to have been the Yorick of Hamlet's soliloquy.

TARN, a *département* of France, formed in 1790 of the three dioceses of Albi, Castres and Lavaur, belonging to the province of Languedoc. Pop. (1962) 319,560. Area, 2,232 sq.mi. Tarn is bounded north and east by Aveyron, southeast by Hérault, south by Aude, southwest and west by Haute-Garonne, northwest by Tarn-et-Garonne. The *département* forms the southern buttress of the Plateau Central, with the Montagne Noire (3,970 ft.) near its southern border. The general slope is from east to west, and in the east the Monts de Lacaune reach 4,134 ft. at the Pic de Montalet. The greater part of the *département* is floored by ancient rocks with some granite, especially in the north, and the general level is from 1,500 to 2,000 ft. The Aveyron runs along the northern boundary, the Tarn across the middle of the *département*, and the Agout across the south, and they join one another west of the boundary on their way to Garonne. The Aude receives the streams from a part of the Montagne Noire.

The limestone and sandstone foothills are clothed with vines and fruit trees, and are broken by deep alluvial valleys of extraordinary fertility. The eastern portion of the *département* has the climate of Auvergne, the severest in France, but that of the plain is Girondin. The winter average temperature reduced to sea level is 41° F and the summer average temperature is 70° F, but the great elevation often reduces these figures by 15° to 20° F. The rainfall, 29 or 30 in. at that place, exceeds 40 in. on the Lacaune and Montagne Noire.

The west and centre produces cereals, wheat, oats, maize and vines; the valleys around Castres provide natural pasture for cattle; market gardening is carried on in the west. There are mines of coal and iron, and quarries of lime. The industries include the manufacture of textiles, hosiery, brushes, morocco leather, hats, metal foundries, dye works and glassworks. The *département* is famous for its sparkling wines. The *département* is divided into two arrondissements (Albi and Castres) and there are 36 cantons and 325 communes. The *département* is in the 16th military region, and the académie (educational division) of Toulouse, where is its court of appeal.

The chief towns are Albi (the capital and seat of an archbishop, with suffragan bishops at Rodez, Cahors, Mende and Perpignan), Castres (*q.v.*), Gaillac, Lavaur, Mazamet and Cordes. Burlats has ruins of an old church and château; Lisle d'Albi, a *bastide* with a 14th-century church, and Penne has ruins of a medieval château.

TARN, a river of southern France, 233 mi. long, a tributary to the Garonne. It rises at 5,249 ft. on the Hercynian gneisses of Mt. Lozère, flows westward and, having received the Tarnon, enters a gorge in the Jurassic limestones, which separates the Causse de Sauveterre from the Causse Méjan. It receives the Jonte (left) and passes between the Causse Noir, the Larzac plateau and the Causse de St. Affrique (left) and the Lévezou range and the Plateau of Ségala (right). At Millau it receives the Dourbie (left) and lower down the Dourdou (left) and from there to the cascade of Sabo, above Albi, the river crosses Permian and Silurian rocks. Having entered the plain, the river has cut out a deep bed in the Tertiary strata, passes Gaillac, and, after receiving the Agoût (left), it turns northwest, receiving the Tescou (right) at Montauban and the Aveyron (right). After passing Moissac it joins the Garonne on its right bank.

TARN, a name applied in England, especially in the lake district and in Scotland to small lakes or pools in the mountainous districts. It applies especially to such as have no visible affluent streams. The term is sometimes used also in reference to a marsh or bog. *See* also LAKE.

TARN-ET-GARONNE, a *département* of southwestern France, formed in 1808 of districts formerly belonging to Guienne and Gascony (Quercy, Lomagne, Armagnac, Rouergue, Agenais), with the addition of a small piece of Languedoc. From 1790 to 1808 its territory was divided between the *départements* of Lot, Haute-Garonne, Tarn, Aveyron, Gers and Lot-et-Garonne. It is bounded north by Lot, east by Aveyron, south by Tarn and Haute-Garonne, and west by Gers and Lot-et-Garonne. Area, 1,441 sq.mi. Pop. (1962) 175,847. The *département* is the region focusing upon Montauban and includes the junction of the Aveyron and the Tarn and their union with the Garonne. These rivers separate hills of Pliocene strata, and in the extreme northeast stands out the edge of the Plateau Central and the limestone Causse de Quercy (1,634 ft.). The climate is mild and agreeable; the mean annual temperature being about 56° F. Rain falls seldom, but heavily, especially in spring, the annual rainfall being 28 or 30 in.

The wide alluvial valleys of the three large rivers are most productive. Cereals, especially wheat, maize and oats, occupy more than two-thirds of the arable land of the *département*. The vine is grown everywhere and large quantities of grapes are exported as table fruit. Potatoes are also grown. Plums and apricots are abundant. The breeding of horses, especially for cavalry purposes, is actively carried on; and the rearing of horned cattle, both for draft and for fattening, is also important. Sheep, pigs, poultry, and silkworms are also sources of profit. There are metal foundries, various kinds of silk mills, and manufactories of straw hats, wool, paper and brooms. The principal exports are fruit, wine, flour, truffles from the Rouergue, and early vegetables. The canal of the Garonne traverses the *département* for 48 mi. and the Garonne and the Tarn furnish 82 mi. of navigable waterway. The *département* is served by the Orléans and the Southern railways. The *département* forms the diocese of Montauban under the archbishop of Toulouse, and belongs to the académie (educational division) of Toulouse, where is its court of appeal, and to the district of the XVII corps d'armée (Toulouse). It has two arrondissements (Montauban, the capital, and Castelsarrasin), 24 cantons and 195 communes.

TARNOPOL (TERNOPOL), formerly a province of Poland, bounded north by the province of Volhynia, west by Lwów, south by Stanislawów and east by the U.S.S.R. Area 6,268 sq.mi. Pop. (1931) 1,604,000. Tarnopol was part of the plateau of Podolia joining the Lublin uplands, having the appearance of hills where the land slopes down to the Volhynian plain, especially in the Gology. The south was bounded by the Dniester, with its lower valley, the tributaries of which drained the bulk of the province except in the north, where the Bug and Styr have their sources. These tributaries flow from north to south and from the deep ravines which are a feature of Podolia. The plateau is sometimes flat, sometimes hilly and wooded, as on the Miodobory along the eastern border. The Strypa and the Lipa were scenes of stubborn Russian defense during World War I. Tarnopol was a fertile agricultural province. It was famous for its horses and cattle. The chief towns were Tarnopol (pop. [1959] 52,245), Zloczów and Barezany. The province was occupied by soviet troops at the start of World War II and by German troops from 1941 to 1945. It was incorporated into the U.S.S.R. after World War II as an *oblast* in the Ukrainian S.S.R. Area 5,367 sq.mi. Pop. (1959) 1,085,586.

TARNOPOL (TERNOPOL), a town formerly in Poland, in Tarnopol province, 87 mi. E.S.E. of Lvov (Lwów). Pop. (1959) 52,245. The town was founded in 1540 by Jan Tarnowski. It became a fortress and received many privileges from the Polish kings. The ancient synagogue of the reign of Casimir the Great remains. Industry consists in corn milling and the preparation of wax and honey. It is known for its trade in horses, agricultural produce and spirits. Soviet troops occupied Tarnopol in 1939, and it was captured by axis forces in 1941. It was incorporated into the U.S.S.R. after World War II.

TARNÓW, a town of Poland, in Kraków province. Pop. (1960) 71,000. It is situated on Biala river, not far from its junction with the Dunajec, and is the seat of a Roman Catholic bishop. It possesses a cathedral in Gothic style, built in the 15th century, with monuments of the Tarnowski and Ostrogski families, to which the town formerly belonged, another church, built in 1454, and a diocesan museum with Polish paintings of the 15th and 16th centuries. On a hill near the town stand the ruins of an old castle of the Tarnowski family and a small church over 800 years old. The town hall is an old and interesting building. Agricultural implements, glass and chicory are manufactured. Tarnów was occupied by German forces in World War II, but was returned to Poland in 1945.

TARNOWSKI, JAN [called MAGNUS] (1488–1561), Polish

general. After a careful education at the palace of Matthew Drzewicki, bishop of Przemysl, he occupied a conspicuous position at court in the reigns of John Albert, Alexander and Sigismund I. In 1509 Tarnowski distinguished himself in Moldavia, taking part in the victories of Wisniowiec (1512) and Orsza (1514). He then travelled in the Near East, and northern and western Europe. While in Portugal he received from King Emanuel the chief command in the war against the Moors, and was promoted by Charles V. to a count of the Empire. On the death of Nicholas Firlej in 1526 Tarnowski became grand hetman of the crown, or Polish commander-in-chief, and won his greatest victory at Obertyn (August 22, 1531) over the Moldavians, Turks and Tatars. Tarnowski took the royal side during the "Poultry War" of 1537; and also in 1548 when the *szlachta* tried to annul by force the marriage of Sigismund Augustus (*see* SIGISMUND II) with Barbara Radziwill. In 1553, indeed, he was in opposition to the young king; yet he remained emphatically an aristocrat, intensely opposed to the democratic tendencies of the *szlachta*, and working for a firm alliance between the king and the magnates. Though a devout Catholic, he was opposed to the exclusive jurisdiction of the bishops and would have limited the authority of Rome in Poland. Tarnowski invented a new system of tactics to increase the mobility and security of the armed camps within which the Poles had so often to encounter the Tatars. His principles are set forth in his *Consilium Rationis Bellicae* (best edition, Posen, 1879). As an administrator he did much to populate the vast south-eastern steppes of Poland.

TARO (*Colocasia antiquorum*), a coarse herbaceous plant of the arum family (Araceae, *q.v.*), called also eddo, probably native to southeastern Asia, whence it has spread to the Pacific islands, where it is extensively cultivated for its large, spherical, underground tubers, rich in starch; these form an important article of food, especially in Hawaii. The Polynesian poi is a thin pasty mass of taro starch.

It is closely allied to the dasheen (*C. esculenta*), native to Indonesia, various cultivated forms of which are known as elephant's-ear. The dasheen tubers have a nutty flavour and contain more carbohydrates and proteins than the potato.

TAROK, a game of cards very popular in central Europe and to some extent in France (*tarot*) and Italy (*tarocchi*). It is the oldest surviving card game (*see* CARDS, PLAYING). The modern game uses 54 cards, comprising 32 suit cards, 21 taroks (permanent trumps) and a joker. Each suit includes four court cards, instead of three as in the 52-card French pack. The taroks, originally illuminated with individual designs of esoteric significance, now are usually printed only with the Roman numerals from I to XXI.

The play of the game centres upon the manipulation of the taroks, each of which is privileged to win any suit card. Tarok includes many features that were borrowed by other games—competitive bidding, a scat or widow, melding of honour combinations, the sequence of matadors, the suitless or wild joker, the effort to capture specific counting cards in tricks and extra score for the last trick.

TARPAULIN or **TARPAULING** (as if *tarpalling*, from *tar*, and *palling*, a covering, Lat. *palla*, a mantle), is a heavy well-made, double warp plain fabric, of various materials, used chiefly in the manufacture of covers for weather protection. It is usually made from cotton duck, single filled, double filled, or Army construction. Certain tarpaulins are made from drill and flax. To render it proof against rain and other atmospheric influences, it is generally treated with waxes, petrolatums, and other waterproofing agents. Some tarpaulins are also made from jute (*q.v.*), most are equipped with grommets or eyelets in the corners, spaced around the cover. The fabric is fastened over the object to be protected.

TARPEIA, in Roman legend, daughter of the commander of the Capitol during the war with the Sabines caused by the rape of the Sabine women. According to the story, she offered to betray the citadel, if the Sabines would give her what they wore on their left arms, meaning their bracelets; instead of this, keeping to the letter of their promise, they threw their shields upon her and crushed her to death. Simylus, a Greek elegiac poet, makes Tarpeia betray the Capitol to a king of the Gauls. The story may be an attempt to account for the Tarpeian rock being chosen as the place of execution of traitors. According to S. Reinach, however, in *Revue archéologique*, xi. (1908), the story had its origin in a rite—the taboo of military spoils, which led to their being heaped up on consecrated ground that they might not be touched. Tarpeia herself is a local divinity, the manner of whose death was suggested by the tumulus or shields devoted to her cult, a crime being invented to account for the supposed punishment.

See Sir George C. Lewis, *Credibility of early Roman History*; A. Schwegler, *Römische Geschichte*, bk. ix. 10; Livy, i. 11; Dion. Halic. ii, 38–40; Plutarch, *Romulus*, 17; Propertius, iv, 4; Ovid, *Fasti*, i, 261; C. W. Müller, *Frag. Hist. Graec.*, iv, p. 367.

TARPON (*Tarpon atlanticus*), a fish allied to the herrings, but with a large mouth and very large, thick, silvery scales. It occurs in the warmer parts of the western Atlantic, and is an active fish, preying on the fry of other fishes. It reaches a length of 7 ft., and a weight of more than 200 lb.

TARQUINIA (anc. TARQUINII), a town of Italy, in the province of Viterbo, 62 mi. northwest by rail from the city of Viterbo, 490 ft. above sea level. Pop. (1951) 7,533.

It is picturesquely situated, and commands a fine view of the sea. It possesses mediaeval fortifications, and no less than 25 towers are still standing in various parts of the town, which thus has a remarkably mediaeval appearance. The castle, on the north, contains the Romanesque church of S. Maria in Castello, begun in 1121, with a fine portal of 1143, a *ciborium* of 1166 and a pulpit of 1208, both in "cosmatesque" work; the pavement in marble mosaic also is fine. There are several other Romanesque and Gothic churches and other buildings. The Gothic Palazzo Vitelleschi (1439) with remarkably rich windows, contains the fine Government museum, now including the Bruschi collection, of Etruscan antiquities from the tombs of the Etruscan city, which probably occupied the site of the Roman town, now deserted, its last remains having been destroyed by the inhabitants of Corneto (the mediaeval name by which the town was till lately known) in 1307. Scanty remains of walling and of buildings of the Roman period exist above ground.

The importance of Tarquinia to archaeologists lies mainly in its necropolis, situated to the south-east of the mediaeval town, on the hill which, from the tumuli raised above the tombs, bears the name of Monterozzi. The tombs themselves are of various kinds. The oldest are *tombe a pozzo*, or shaft graves, containing the ashes of the dead in an urn, of the Villanova period, the earliest belonging to the stage known as First Benacci (*see* ETRUSCANS) and the latest to the middle or end of the 9th century B.C. immediately after the coming of the Etruscans, a few contemporary graves of whom, containing Egyptianizing scarabs and some gold and silver, have also been found. In some of these tombs hut urns, like those of Latium, are found. Next come the various kinds of inhumation graves, the earliest of which, the so-called warrior's grave, belongs to the early "Vetulonian" period, and the Bocchoris tomb to 730–690 B.C. (MacIver, *Villanovans and Early Etruscans*, 1924, 40–56; 158–166): the most important are rock-hewn chambers, many of which contain well-preserved paintings of various periods.

Tarquinia was the chief of the 12 cities of Etruria, and appears in the early history of Rome as the home of Tarquinius Priscus and Tarquinius Superbus. The people of Tarquinia and Veii attempted to restore Tarquinius Superbus after his expulsion. In 358 B.C. the citizens of Tarquinia captured and put to death 307 Roman soldiers; the resulting war ended in 351 with a 40 years' truce, renewed for a similar period in 308. When Tarquinia came under Roman domination is uncertain, as is also the date at which it became a municipality; in 181 B.C. its port, Graviscae (mod. Porto Clementino), with Government salt-works 4 m. S.W., in an unhealthy position on the low coast, became a Roman colony. It exported wine and carried on coral fisheries.

TARQUINIUS PRISCUS, LUCIUS, fifth legendary king of Rome (616–578 B.C.). He is represented as the son of a Greek refugee, who removed from Tarquinii in Etruria to Rome, by the advice of his wife, the prophetess Tanaquil. Appointed guard-

ian to the sons of Ancus Marcius, he supplanted them on their father's death. He laid out the Circus Maximus, instituted the "great" games, built the great sewers (*cloacae*), and began the construction of the temple of Jupiter on the Capitol. He carried on war successfully against the Sabines and subjugated Latium. He is said to have raised the number of the senators to 300, and to have doubled the number of the knights.

The introduction of many of the insignia of war and of civil office is assigned to his reign, and he was the first to celebrate a Roman triumph, after the Etruscan fashion, in a robe of purple and gold, and borne on a chariot drawn by four horses. He was assassinated at the instigation of the sons of Ancus Marcius.

The legend of Tarquinius Priscus is in the main a reproduction of those of Romulus and Tullus Hostilius. There seems to have been originally only one Tarquinius; later, when a connected story of the legendary period was constructed, two (distinguished as the "Elder" and the "Proud") were introduced, separated by the reign of Servius Tullius, and the name of both was connected with the same events.

For the constitutional reforms attributed to Tarquinius, *see* ROMAN HISTORY; for a critical examination of the story, Schwegler, *Römische Geschichte*, bk. xv.; Sir George Cornewall Lewis, *Credibility of early Roman History*, ch. 11; W. Ihne, *History of Rome*, i., E. Pais, *Storia di Roma*, i. (1898), who identifies Tarquinius with Tarpeius, the eponymus of the Tarpeian rock. Ancient authorities:—Livy i. 34–41; Dion. Hal. iii. 46–73, Cic. *de Repub.*, ii. 200.

TARQUINIUS SUPERBUS, LUCIUS, son of Lucius Tarquinius Priscus and son-in-law of Servius Tullius, the seventh and last legendary king of Rome (534–510 B.C.). On his accession he proceeded at once to repeal the recent reforms in the constitution, and attempted to set up a pure despotism. Many senators were put to death, and their places remained unfilled; the lower classes were deprived of their arms and the completion of the fortress-temple on the Capitoline confirmed his authority over the city. The outrage of his son Sextus upon Lucretia (*q.v.*) precipitated a revolt, which led to the expulsion of the entire family. All Tarquinius's efforts to force his way back to the throne were vain (*see* PORSENA, LARS), and he died in exile at Cumae.

In the story certain Greek elements, probably later additions, may easily be distinguished. Tarquinius appears as a Greek "tyrant" of the ordinary kind; on the other hand, an older tradition represents him as more like Romulus. This twofold aspect of his character perhaps accounts for the making of two Tarquinii out of one (*see* TARQUINIUS PRISCUS, LUCIUS). The well-known story of Tarquinius' repeated refusal and final consent to purchase the Sibylline books has its origin in the fact that the building of the temple of Jupiter Capitolinus, in which they were kept, was ascribed to him.

For a critical examination of the story *see* Schwegler, *Römische Geschichte*, bk. xviii.; Sir G. Cornewall Lewis, *Credibility of early Roman History*, ch. 11; E. Pais, *Storia di Roma*, i. (1898); and, for the political character of his reign, ROMAN HISTORY. Ancient authorities:—Livy i. 21; Dion. Hal. v. i–vi. 21.

TARRAGON (*Artemisia dracunculus*), also called estragon, is a small but vigorous perennial herb of the family Compositae, native to southern Europe and western Asia. It is widely grown for its aromatic leaves, which are used for flavouring vinegars, sauces, pickles, salads, etc., and for its volatile oil, known as estragon oil, also employed as a flavouring agent and in perfumes.

(Q. J.)

TARRAGONA, a maritime province in the northeast of Spain, formed in 1833 from the southern part of the province of Catalonia, and bounded S.E. by the Mediterranean, N.E. by Barcelona, N. by Lérida, W. by Saragossa and Teruel, and S.W. by Castellon de la Plana. Pop. (1960) 362,679; area 2,426 sq.mi. The Ebro flows through the southern portion of the province, and below Tortosa forms a conspicuous marshy delta, but elsewhere the coast-line is unbroken. The hills are clothed with vineyards, which produce excellent wines, and in the valleys are cultivated all kinds of grain, vegetables, rice, hemp, flax and silk. Olive, orange, filbert and almond trees reach great perfection, and the mountains yield rich pastures and timber. Manufactures are well advanced, and comprise silk, cotton, linen and woolen fabrics, velvet, felt, soap, leather and spirits. There are also many potteries and cooperages and flour, paper and oil mills. Silver, copper, lead and other minerals have been found, and quarries of marble and jasper are worked in the hills.

TARRAGONA, the capital of the Spanish province of Tarragona, a flourishing seaport, and the seat of an archbishop; at the mouth of the river Francoli, 63 mi. (101 km.) by rail W.S.W. of Barcelona. Pop. (1960) 43,519 (mun.).

Tarraco, the capital of the Iberian Cessetani, many of whose coins are extant, was one of the earliest Roman strongholds in Spain. It was captured in 218 B.C. by Gnaeus and Publius Cornelius Scipio, who improved its harbours and enlarged its walls. A Roman monument on a hill 3 mi. E. is known as the Sepulcro de los Escipiones, and locally believed to be the tomb of the Scipios, who were defeated and slain by the Carthaginians under Hasdrubal Barca in 212 B.C., but there is no good reason to believe that the monument is older than the 1st century A.D. As the Colonia Triumphalis, so called to commemorate the victories of Julius Caesar, Tarraco was made the seat of one of the four assize courts (*conventus juridici*) established in Hispania Citerior. Augustus spent the winter of 26 B.C. here, and made Tarraco the capital of the whole province, which received the name of Hispania Tarraconensis. A temple was built in his honour. It was afterwards restored by Hadrian (A.D. 117–138), and the city became the Spanish headquarters of the worship of the goddess Roma and the deified emperors. Its flax trade and other industries made it one of the richest seaports of the empire.

To the Romans the Visigoths under Euric succeeded in 467, but on their expulsion by the Moors in 711 the city was plundered and burned. In 1089 the Moors were driven out by Raymond IV. of Barcelona, and in 1118 a grant of the fief was made to the Norman Robert Burdet, who made it a frontier fortress against the Moors. The British took the city in 1705; the French captured it in 1811. In the civil war of 1936–39, Tarragona remained in Loyalist possession until Jan. 15, 1939.

Tarragona is on the coast railway from Barcelona to Valencia, and is connected with the Ebro Valley Railway by a branch line to Reus. The old town, with its dark and steep alleys, occupies a rugged hill which rises abruptly from the sea to an altitude of about 550 ft. Many of the houses in this quarter are very old, and are built partly of Roman masonry; one such fragment, immured in the palace wall, is inscribed with the epitaph of a charioteer (*auriga*). Massive ruined walls encircle the old town. Their lowest course is "Cyclopean," consisting of unhewn blocks about 12 ft. long and 6 ft. wide; Roman masonry of the Augustan age is superimposed. The six gates and the square towers are also, to a great extent, "Cyclopean." Tarragona cathedral is one of the noblest examples of early Spanish art. The main body of the building dates from the end of the 12th century and the first half of the 13th, and is of transitional character. On the north-east side is a cloister contemporary with the church, with which it communicates by a very fine doorway. The cloister contains much remarkable work, and the tracery of the windows bears interesting marks of Moorish influence. Two other noteworthy churches in the city are San Pablo and Santa Tecla la Vieja, both of the 12th century.

There is a fine Roman aqueduct; the Roman amphitheatre was dismantled in 1491 to furnish stone for the eastern mole, though a few rows of seats are left near the sea-shore; and the museum contains a large collection of Roman antiquities. The Torreón de Pilatos is said to have been the palace of the Emperor Augustus. When the monks of the Grande Chartreuse were compelled to leave France, they settled at Tarragona in 1903, and established a liqueur factory. A characteristic feature of Tarragona is the number of its underground storehouses for wine (bodegas); wine is exported in large quantities. Chocolate, soap, flour, ironware, paper, pipes and salted fish are also manufactured.

See G. W. Edwards, *Spain* (1926).

TARRASA, a town of Spain, in the province of Barcelona, 6 mi. W.N.W. of Sabadell on the Barcelona-Lérida railway. Pop. (1960) 92,234 (mun.). Tarrasa was a Roman municipality, and a bishopric from the 5th century to the Moorish invasion in the

8th. Tarrasa is now mostly a modern industrial town.

TARS, LOW-TEMPERATURE, the primary condensation products resulting from the carbonization of coal and such other organic substances as wood, peat and lignite at temperatures not exceeding 700° C. The term "low-temperature tar" usually refers to a low-temperature coal tar.

Wood Tar.—A by-product of the carbonization of wood and two distinct types of wood tar are recognized: (1) hardwood tar and (2) resinous wood tar, the former derived from woods such as oak, beech, etc., and the latter from resinous stools and roots and particularly from pine wood.

Hardwood Tars.—The condensed volatile product of wood distillation is known as crude pyroligneous acid, which has the average composition: water 81%, methyl alcohol 3%–4%, acetic acid 6%–8%, tar 7%. After settling, the major portion of the tar separates from the aqueous acid although the latter always retains some tar in solution. The chemical constituents of wood tar form a complex mixture comprising the "fatty" acids, esters, ketones, alcohols, phenols (usually polyhydric) and their methyl ethers, together with waxes. The most important representatives of these groups present in the tar are formic and acetic acids, with their methyl esters, acetone, methyl and allyl alcohol, guaiacol, catechol, and esters of pyrogallol.

Beechwood tar is practically the only wood tar subjected to a complete straight distillation. Three fractions are usually collected, first the light oils and water containing acetic acid and methyl alcohol up to 180° C., then the heavy oils up to 240° C. and finally the pitch. The important fraction is the heavy oil from which beechwood creosote is obtained (*see* CREOSOTE).

The oils from wood tar are used as gum inhibitors in gasoline and as flotation oils; the pitch is used in briquetting.

Resinous Wood Tars.—These tars are represented by pine-wood tar, which is commonly termed "Stockholm" or "Archangel" tar and is made extensively in the forests of the U.S.S.R., Finland and Sweden. It is an important commercial product and differs from hardwood tar in containing the pleasant smelling mixture of terpenes known as turpentine. Pine-wood tar is the residue left after the turpentine has been distilled, usually with the aid of steam. It is used widely in the manufacture of cordage, *e.g.*, tarred ropes and twine, and for impregnation of hemp fibre for oakum. It is used to a slight extent in pharmacy as a component of some ointments and antiseptics. Distillates of pine-wood tar, particularly the creosote fraction, are used in froth flotation processes.

Peat Tar.—A jet black, semisolid oil, lighter than water, this can be dehydrated by heating at 100° C. On distillation up to 360° C. a hard pitch is obtained along with an oily distillate, the latter consisting of a neutral fraction, containing the solid waxes, phenols and only traces of basic compounds. The most striking characteristic of the neutral oils is their high degree of unsaturation as manifested by absorption of atmospheric oxygen. The waxes, which resemble the montan wax of lignite, melt at about 40° C. The acidic (phenolic) constituents of the tar consist of phenol and its homologues, the cresols and xylenols in small quantity, mixed with tar acids boiling at 250°–360° C. The latter show a carbolic acid coefficient of 31, *i.e.*, they are 31 times more powerful than phenol as bactericides. Among the phenolic constituents of peat tar, guaiacol, methylguaiacol and a methyl ether of pyrogallol have been identified. These ethers are absent from coal tar, but are found in wood-tar creosote, of which they form the major portion.

Peat tar may therefore be looked on as one of the transition steps from wood tar to coal tar. Although the products which can be isolated from peat tar are of undoubted value, the difficulty of winning and drying peat militates against the carbonization of this fuel on an industrial scale.

Lignite and Brown Coal Tars.—These tars resemble peat tar. They are usually black oils of rather buttery consistency owing to the presence of paraffin wax. Of an average density of 0.95, they retain a high percentage of water but solidify completely at 6°–8° C. In contradistinction to tars from bituminous coal, they are almost completely soluble in gasoline, a residue of about 5% only being usually left.

They are paraffinoid in nature, the wax content being as high as 10%. The crude wax melts at about 40° C. but from it a series of waxes with melting points from 46° to 73° C. has been obtained. The phenolic constituents resemble those of peat tar and are efficient bactericides. The basic components belong to the pyridine and quinoline series and are chiefly methylated derivatives. A distinctive characteristic of lignite tars is the presence of ketones in the neutral oils—a resemblance to the oils of wood tar. Lignite tars are of increasing importance as a source of fuels for internal-combustion engines, burning and lubricating oils and pitch.

Low-Temperature Coal Tars.—Vacuum tar has only been produced experimentally with the object of elucidating the composition of coal and the effect of heat upon this material. It is obtained when coal is carbonized at temperatures up to 450° C. under a pressure of 5–40 mm. and the average yield is 14 gal. (6.5% by weight) per ton of coal. It is usually lighter than water (0.99 sp.gr.) and is especially interesting in that some of its constituents are identical with substances found in petroleum oils, and extractable from coal itself by solvents such as pyridine. It consists of 40%–45% paraffins and naphthenes, 40% "ethylenic" hydrocarbons richer in carbon than mono-olefines, 7% aromatic hydrocarbons, chiefly homologues of naphthalene, and 10% phenols, mostly cresols and xylenols, with smaller amounts of tertiary alcohols and basic compounds. Naphthalene, anthracene, benzene and its homologues are all absent.

Among the more interesting constituents of vacuum tar are the naphthene, $C_{30}H_{60}$ called melene which is also found in Galician petroleum and in the distillate of beeswax, and the hydrogenated phenols (alcohols) such as hexahydro-*p*-cresol and its homologues.

Low-temperature tar is produced by the carbonization of bituminous coals, containing 28%–35% volatile matter, at temperatures not exceeding 700° C., which is an optimum temperature for the production of gas, tar and smokeless fuel (*see* CARBONIZATION, LOW-TEMPERATURE). It is essentially a mixture of those volatile products of coal which are liquid at ordinary temperatures and which have not been subjected to the secondary thermal decomposition incidental to high-temperature carbonization in horizontal retorts. It is therefore regarded as a "primary" tar. Many factors affecting the yield, composition and other characteristics of low-temperature tars have been varied in experimental plants with the result that there have been produced many types of tars to which it has been impossible to assign a common composition.

Low-temperature tar is usually a brownish black oil, much less viscous than the tar made from the same coal at high temperatures. It is obtained in an average yield of 16–18 gal. per ton of coal carbonized, or 8%–9% by weight. This is approximately double the weight obtained by distilling the coal at 900°–1,200° C. Its specific gravity is 1.02–1.07, and owing to the low temperature at which it is formed, it is mainly paraffinoid in character, whereas high-temperature tar is almost wholly aromatic. The most distinctive characteristics of low-temperature tar are the almost entire absence of benzene and its homologues, naphthalene, anthracene and phenol; the unsaturated nature of the neutral oils, from which can be isolated solid paraffins; and the presence of solid, amorphous compounds, of either basic or phenolic nature.

The crude tar is a most unstable product and begins to decompose at about 200° C., finally yielding, when distilled to 360° C., an average of 65% oily distillate and 35% pitch. The amount of pitch formed on distilling low-temperature tar is therefore only half that obtained in a similar operation with high-temperature tar. The former, too, contains much less "free carbon" (*i.e.*, matter insoluble in benzene) than the latter, the average amount being 2%–4%. By modified extractions with caustic soda and sulfuric acid (*see* COAL TAR) the crude distillate can be divided into (1) neutral oils; (2) phenols; and (3) bases.

Neutral Oils.—These consist mostly of paraffins, naphthenes and olefines, with comparatively small amounts of methylated derivatives of naphthalene and anthracene. On treatment with

concentrated sulfuric acid, the neutral oils are absorbed with the exception of some 12% consisting mainly of paraffins. Crude solid wax representing over 1% of the tar is obtained on cooling an acetone solution of the neutral oils. In the paraffin series, all the lower members have been identified, and in the higher ranges, the solid members from $C_{26}H_{54}$ to $C_{29}H_{60}$. The latter melt at 62° C. α- and β-methylnaphthalenes have been isolated, but no naphthalene. Similarly β-methylanthracene occurs in a complex mixture of hydrocarbons crystallizing on cooling the high fractions of the distillates, but anthracene is absent. Neither fluorene, nor carbazole has been identified in low-temperature tar; in addition to the hydrocarbons acenaphthene and diphenyl, there occur the fully hydrogenated products, perhydrofluorene, perhydroacenaphthene and dodecahydrodiphenyl.

Phenols.—The extraction of the phenolic constituents of the tar or its distillates is complicated by the fact that resinous and asphaltic substances accompany the phenols in solution in the alkali. These impurities, however, can be eliminated by agitation of the alkaline solution with organic solvents or by saturation with salt. The crude phenols (10%–20% of the tar) obtained on acidification as a black oil, can be further purified by extraction with light petroleum, followed by precipitation of the residue from an ethereal solution by fresh petroleum. The petroleum extract consists of low-boiling phenols including phenol itself, the cresols and 1,2,4-, 1,3,5-, 1,3,4-, and 1,4,2- xylenols. Catechol, β-naphthol and trimethylphenols have also been identified. The higher phenols of the tar are amorphous solids varying in colour from pale yellow to black.

Bases.—The crude bases, representing 2%–3% of the tar, are also composed of a low mobile fraction and a higher viscous fraction containing solid, amorphous substances of unknown constitution. The lower bases consist of pyridine and quinoline with their mono-, di- and tri- methyl derivatives. Only traces of primary bases are present in the tar. There have been identified pyridine, 2-methylpyridine, 3-methylpyridine, 4-methylpyridine, 2,5-dimethylpyridine, 3,4-dimethylpyridine, 2,4,5-trimethylpyridine, 2,3,4-trimethylpyridine, 2,4,6-trimethylpyridine, aniline, 2,3,5-trimethylpyridine, *p*-toluidine, *o*-toluidine, *m*-toluidine, 2,3,-4,5-tetramethylpyridine, quinoline, isoquinoline, and 2-methylquinoline.

Utilization of Low-Temperature Tar Products.—The commercial possibilities of low-temperature tar have not yet been fully developed, but its application in various directions has been demonstrated. The low-boiling distillates, owing to their "unsaturated" nature, form satisfactory fuel possessing certain antidetonating qualities. It has been claimed that the higher fractions of the neutral oils form an exceptionally good lubricant, whilst the "middle" or "dead" oils are suitable for combustion in the diesel engine. Special characteristics are also claimed for low-temperature tar pitch, which can be used along with or in place of high-temperature tar pitch.

See H. H. Lowry, *Chemistry of Coal Utilization* (1945).

(D. D. P.; F. E. Cl.)

TARSIER, a small arboreal mammal, *Tarsius*, typifying a suborder of Primates (*q.v.*). Its name refers to the great length of the anklebones or tarsus, but an equally distinctive character of this animal is its enormous, gogglelike eyes. The head in which these are set is rounded, the muzzle short and pointed, while the ears are large. Rather smaller than a squirrel, it has a slender body and long, thin tail which terminates in a tuft of hair and serves as a rudder and balance. The fingers and toes end in disklike pads, which assist the animal to adhere to branches and other smooth surfaces. While it is doubtful if more than a single species exists, numerous races have been described on the various islands: Sumatra, Borneo, Celebes and some of the Philippines. The tarsier feeds on insects and lizards, sleeps during the day but is active at night, moving by jumping, an action for which the structure of its hind legs is well adapted. It is rare, not more than two being found together, and brings forth only one young at a time.

Bibliography.—C. Hose, *Mammals of Borneo* (1893), *Fifty Years of Romance and Research*, pp. 95, 96 and appendix C; G. Elliot Smith *et al.*, *Proc. Zool. Soc. London* (1919); H. H. Woollard, "Monograph on Tarsius," *Proc. Zool. Soc. London* (1925); W. E. Le Gros Clark, *Proc. Zool. Soc. London* (1924); F. Wood Jones, *Man's Place among the Mammals* (1929).

(J. E. Hl.)

TARSUS (mod. *Tersous*), an ancient city in the fertile plain of Cilicia. The small Cydnus river flowed through the centre of the town, and its cool swift waters were the boast of the city. The city is first mentioned on the Black Obelisk, as captured by the Assyrians *c.* 850 B.C. It was probably an old Ionian colony, settled (like Mallus) under the direction of Clarian Apollo. Its importance was the result of: (1) its excellent and safe harbour; (2) its possession of a fertile territory; and (3) its command of the first wagonroad made across Mt. Taurus, which was cut through the Cilician Gates, a narrow gorge 100 yd. in length, originally only wide enough to carry the waters of a small affluent of the Cydnus. The greatness of Tarsus rested therefore mainly on the two great engineering works, the harbour and the road.

Tarsus is most accessible from the sea or from the east. Even after the Cilician Gates were cut, the crossing of Taurus was a difficult operation for an invading army (as Xenophon and Arrian show). Hence Tarsian history (where not determined by Greek maritime relations) has been strongly affected by Semitic influence, and Dion Chrysostom, about A.D. 112, says it was more like a Phoenician than a Hellenic city (which it claimed to be). After the Assyrian power decayed, princes, several of whom bore the name or title Syennesis, ruled Tarsus before and under Persian power. Persian satraps governed it in the 4th century B.C.; and struck coins with Aramaic legends there. The Seleucid kings of Syria for a time kept it in a state of servitude; but it was made an autonomous city with additional citizens (probably Argive Greeks and Jews) by Antiochus IV Epiphanes in 171 B.C.; and then it began to strike its own coins. It became one of the richest and greatest cities of the east under the Romans after 104 B.C., and was favoured by both Antony and Augustus: the reception there by the former of Cleopatra, who sailed up to the city in a magnificent vessel, was a striking historic event. In spite of its oriental character, it maintained a university where Greek philosophy was taught by a series of famous Tarsians, who influenced Roman history. Chief among them was Athenodorus Cananites (*q.v.*), teacher and friend of Augustus. St. Paul, a native of Tarsus, proudly describes himself as "a citizen of no mean city."

Tarsus depended for its greatness on commerce, peace and orderly government. It was not a strong fortress and could not be defended during the decay of the empire against barbarian invasion. The Arabs captured the whole of Cilicia shortly after A.D. 660, and Tarsus seems to have been a ruin for more than a century after the conquest. But Harun al-Rashid rebuilt its walls in 787, and made it the northwestern capital of the Arab power in the long wars against the Byzantine empire. All the raids, which were made in Asia Minor regularly, year by year, sometimes twice in one year, through the Cilician Gates and past the fortress Loulon, issued through the north gate of Tarsus, which was called the Gate of the Holy War. The western gate remained standing, and was misnamed St. Paul's Gate. The caliph Mamun died on such a foray in A.D. 833, having caught a chill at a great spring north of the Cilician Gates beside Ak-Keupreu. He was brought to Tarsus where (like the emperor Tacitus) he died and (like the emperor Julian) was buried. His illness recalls the fever which Alexander the Great contracted from bathing in the Cydnus. Nicephorus Phocas reconquered Tarsus and all Cilicia for the empire in A.D. 965. In the first crusade Baldwin and Tancred captured Tarsus, A.D. 1099, and there the two leaders had a serious quarrel. It formed part of the kingdom of Lesser Armenia for great part of the three centuries after A.D. 1180, and it was fortified by Leo II and Hethoum I. But Turkoman and Egyptian invaders disputed its possession with the Greek emperors and Armenian kings and with one another. Finally it passed into Ottoman hands about the beginning of the 16th century.

The ruins of the ancient city are very extensive, but they are deeply buried, and make little or no appearance above the surface except in the Dunuk Tash (popularly identified as the tomb

STEWART
Official tartan of the royal family

BLACK WATCH
A regimental tartan of the British army

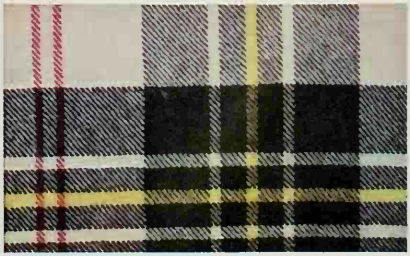
BUCHANAN

ROB ROY

GORDON (Dress)

MACLEOD OF LEWIS

MACDONALD (Ancient)

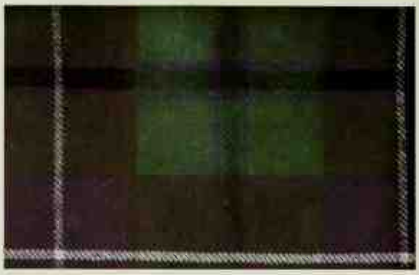


MACPHERSON (Dress)

SCOTTISH CLAN AND REGIMENTAL TARTANS

TARTAN

MACINTOSH (Hunting Ancient)



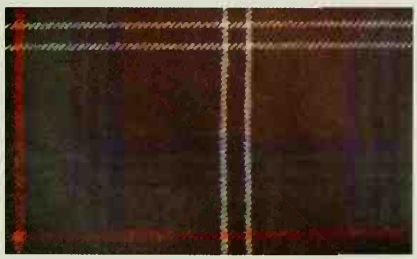
DOUGLAS

MACKENZIE (Ancient)

MACLACHLAN

SINCLAIR (Dress)

SKENE



CHISHOLM (Hunting)

MACMILLAN (Hunting)

TARTANS OF SOME SCOTTISH CLANS

of Sardanapalus, a monument which, however, was at Anchiale, not at Tarsus). This shapeless mass of concrete was probably the substructure of a Graeco-Roman temple, from which the marble coating was removed. The modern town has considerable bazaars and trade; but the climate is very oppressive, because of the proximity of vast marshes which occupy the site of the harbour and the lower part of the original Cydnus course. The river was diverted from its former course by Justinian in the 6th century. The emperor's intention was only to carry off the surplus waters in time of flood and prevent inundations in the city, not to deprive Tarsus of what was its chief pride and boast; but gradually the neglect of subsequent centuries allowed the channel in the city to become blocked by accumulation of soil, and now the whole body of water flows in the new channel east of the city, except what is drawn off by an artificial irrigation course to water the gardens on the western side of the city. The population in 1950 was 33,704.

BIBLIOGRAPHY.—The literature regarding Tarsus is scanty, and few ancient inscriptions have been published. *See* W. B. Barker, *Lares and Penates* (1853); G. F. Hill in the *British Museum Catalogue of Coins;* Six in *Numismatic Chronicle*, pp. 152 ff. (1884), pp. 329 ff. (1894); E. Babelon, "Perses Achémenides," *Catalogue Bibl. Nat.;* the numismatic works of B. V. Head, F. Imhoof Blumer, etc.; Waddington in *Bulletin de Corr. Hell.*, vii, pp. 282 ff.; Sir William Ramsay, *Cities of St. Paul*, pp. 85–245 (Garden City, N.Y., 1907), and "Cilicia, Tarsus and the Great Taurus Pass" in *Geographical Journal*, pp. 357–410 (1903); R. Heberdey and A. Wilhelm, "Reisen in Kilikien" in the *Denkschriften d. kais. Akademie Wien*, xliv, with works of other travellers, especially V. Langlois and Macdonald Kinneir. Callander in *Journal of Hellenic Studies*, pp. 58 ff. (1904), studied Dion Chrysostom's two *Tarsian Orations*.

TARTAGLIA (TARTALEA), **NICCOLÒ FONTANA** (1499–1559), Italian mathematician, was born at Brescia, where he was brought up in poverty by his mother. During the sack of Brescia by the French army (1512) the boy received sabre cuts on the head; the ensuing speech difficulties caused the nickname Tartaglia, the stutterer, which he adopted. In 1534 he settled in Venice as a teacher of mathematics. In 1535 he was challenged to a problem-solving contest by Antonio Maria Fiore, a pupil of Scipione del Ferro, professor of mathematics at Bologna. Del Ferro had discovered the solution to the cubic equation, but kept it secret; upon his death it passed to Fiore. All problems proposed by Fiore led to cubic equations and when Tartaglia rediscovered the solution his victory was assured.

In 1537 appeared Tartaglia's first work *Nova Scientia*, a treatise on gunnery, important as a pioneer attempt to establish the laws of falling bodies. Shortly afterward Tartaglia was approached by Girolamo Cardano (*q.v.*), physician and lecturer in Milan, requesting that Tartaglia publish his solution of the cubic or let Cardano include it in his forthcoming book. This Tartaglia refused, but on a later visit to Milan, hoping to become adviser on artillery to the Spanish army, he confided the solution to Cardano under promise of secrecy. In 1545 appeared Cardano's *Ars Magna*, giving the solution to the cubic as well as the quartic, the latter discovered by his pupil, Lodovico Ferrari. Tartaglia responded in *Questioni et inventioni diverse* (1546), a collection of problems proposed to him by prominent men and dedicated to Henry VIII. He berated Cardano for his breach of faith; this led to a sequence of six printed cartels and countercartels between Ferrari and Tartaglia. The feud culminated in a public dispute in Milan, Aug. 10, 1548, which Ferrari apparently won. Tartaglia lost his position as public lecturer at Brescia and returned to Venice. His troubles in Brescia are described in the *Travagliata inventione* (1551). Best known is his *Trattato di numeri et misure*, 3 vol. (1556–60), an encyclopaedic work on elementary mathematics. Tartaglia also published translations of Euclid and Archimedes. He died in Venice, Dec. 13, 1559.

His works were published as *Opere del famosissimo Nicola Tartaglia* (1606).

See also O. Ore, *Cardano, the Gambling Scholar* (1953).

(O. OE.)

TARTAN, a worsted cloth woven with alternate stripes or bands of coloured warp and weft, so as to form a chequered pattern in which the colours alternate in "sets" of definite width and sequence. The weaving of particoloured and striped cloth cannot be claimed as peculiar to any special race or country, for such checks are the simplest ornamental form into which dyed yarns can be combined in the loom. But the term tartan is specially applied to the variegated cloth used for the principal portions of the distinctive costume of the Highlanders of Scotland. For this costume, and the tartan of which it is composed, great antiquity is claimed, and it is asserted that the numerous clans into which the Highland population were divided had each from time to time a special tartan by which it was distinguished. After the rebellion of 1745 various acts of parliament were passed for disarming the Scottish Highlanders and for prohibiting the use of the Highland dress in Scotland, under severe penalties. These acts remained nominally in force till 1782, when they were formally repealed, and since that time clan tartan has, with varying fluctuations of fashion, been a popular article of dress, by no means confined in its use to Scotland alone; and many new and imaginary "sets" have been invented by manufacturers, with the result of introducing confusion in the heraldry of tartans, and of throwing doubt on the reality of the distinctive "sets" which at one time undoubtedly were more or less recognized as the badge of various clans.

Undoubtedly the term tartan was known, and the material was woven, "of one or two colours for the poor and more varied for the rich," as early as the middle of the 15th century. In the accounts of John, bishop of Glasgow, treasurer to King James III., in 1471, there occurs, with other mention of the material, the following:—"Ane elne and ane halve of blue Tartane to lyne his gowne of cloth of Gold." It is here obvious that the term is not restricted to particoloured chequered textures. In 1538 accounts were incurred for a Highland dress for King James V. on the occasion of a hunting excursion in the Highlands, in which there are charges for "variant cullorit velvet," for "ane schort Heland coit," and for "Heland tartane to be hose to the kinge's grace." Bishop John Lesley, in his *De origine, moribus, et rebus gestis Scotorum*, published in 1578, says of the ancient and still-used dress of the Highlanders and Islanders, "all, both noble and common people, wore mantles of one sort (except that the nobles preferred those of several colours)." A hint of clan tartan distinctions is given by Martin Martin in his *Western Isles of Scotland* (1703), which work also contains a minute description of the dress of the Highlanders and the manufacture of tartan.

The following lines give a brief description of the colours of the tartans of the principal clans. The kilt-tartan colour is given in each case; the plaid-tartans vary in slight particulars.

Campbell of Breadalbane, green ground with black and blue half-inch wide crossings and double yellow overcheck. *Campbell of Argyll*, green ground with wide blue and black crossings, narrow black crossings and alternate yellow and white overcheck. *Cameron*, red ground with wide green crossings, and yellow overcheck. *Forbes*, green ground, blue and black wide crossings and narrow crossings of black, white overcheck. *Fraser*, red ground with green and blue wide crossings, and white overcheck on the red. *Graham of Monteith*, green ground, wide cross bands of black with blue, and double pale blue overcheck. *Grant*, red ground with narrow and wide crossings of green and blue, pale blue overcheck. *Macdonell of Glengarry*, green ground with wide crossings and black and blue, narrow crossings of four red lines and white overcheck. *Macdonald*, green ground with black and blue wide crossings and four red overlines in two groups, and fine red overcheck. *Macdonald of Clanranald*, green ground, wide crossings of black and blue, narrow crossings of red grouped together, and white overcheck. *Macgregor*, red ground, wide and narrow crossings of green, white overcheck edged with black. *Mackintosh*, red ground, wide crossings of green, narrow crossings of blue and blue overcheck. *Mackenzie*, green ground, wide crossings of blue and black, with one white overcheck and red overcheck. *Macleod*, green ground, wide black and blue crossings, overcheck of red and overcheck of yellow. *Macpherson*, red ground with bright blue, green and black wide and narrow crossings, narrow white crossings and yellow overcheck. *Munro*, red ground, green and black wide and narrow crossings, narrow red crossings, double yellow overcheck. *Murray of Athole*, green

ground with blue and black wide crossings and treble overcheck in red. *Royal Stewart*, red ground with green and blue wide and narrow crossings, bright blue narrow crossings, and double overcheck of white and yellow.

See W. and A. Smith, *Tartans of the Clans of Scotland* (1850); J. Sobieski Stuart, *Vestiarium Scoticum* (1842); R. R. M'Ian, *Clans of the Scottish Highlands* (1845–46); J. Grant, *Tartans of the Clans of Scotland* (1885).

TARTAR: *see* TARTARIC ACID.

TARTARIC ACID is one of the most important of the organic acids. Its sparingly soluble acid salt was known to the Greeks and Romans as a deposit in wine casks or vats during the fermentation of grape juice to wine. This crude crystalline crust of argol, which contains about 75% potassium acid tartrate (bitartrate), can be partly purified by recrystallization to tartar. When further purified by removing impurities, it becomes cream of tartar. K. W. Scheele first isolated tartaric acid in 1769 by boiling tartar with chalk and decomposing the product with sulfuric acid. Tartaric acid crystallizes from water in colourless monoclinic prisms. At ordinary temperatures these crystals are free of water of crystallization, have the composition $C_4H_6O_6$ and melt at 169°–170° C.

Manufacture.—The by-products of wine fermentation, press cakes from grape juice, lees (dried slimy sediments) and argols constitute the raw materials from which tartaric acid, cream of tartar and Rochelle salt (potassium sodium tartrate) are made. The raw materials containing impurities which otherwise would present filtration difficulties during the recovery process, are subjected either to steam autoclaving at 50–60 lb. pressure or dry roasting at 150°–160° C. The crude tartar is indirectly heated in the dry state, suspended in water, then neutralized with calcium hydroxide; the precipitated calcium tartrate is filtered. The filtrate is evaporated for recovery of potassium sulfate, and the cake is treated with sulfuric acid. Sparingly soluble calcium sulfate remains largely undissolved, whereas the tartaric acid passes into solution. The dilute liquor is concentrated to the crystallizing point in vacuum pans and cooled. The crude product is redissolved in water, treated with chemicals to remove impurities and decolorized by heating with charcoal. The filtered colourless solution is concentrated and allowed to crystallize. After the granular material is removed by centrifugation and dried in a rotary drier, the pure tartaric acid meets the requirements specified in the *United States Pharmacopeia*. Rochelle salt is prepared from argols blended to contain 75%–85% potassium bitartrate and less than 6% calcium tartrate. The roasted product is neutralized with sodium carbonate, clarified and evaporated to crystallize the Rochelle salt. Cream of tartar (potassium hydrogen tartrate) is usually produced by mixing filtrates from the manufacture of tartaric acid, Rochelle salt and potassium sulfate. After purification and decolorization with carbon, the material is allowed to precipitate under acid conditions. Tartar emetic (antimony potassium tartrate), a drug known for centuries, is prepared by heating in water, 3.4 parts of antimony oxide and 4.3 parts of potassium bitartrate. The product separates in colourless transparent crystals which gradually lose water of crystallization, becoming opaque.

Applications.—Tartaric acid is widely used in carbonated drinks, effervescent tablets and powders and as an acidulent in gelatin, dessert and fruit jellies. It has many industrial applications; *e.g.*, in the cleaning and polishing of metals, in calico printing for controlling the liberation of chlorine from bleaching powder, as a mordant in wool dyeing, in certain photographic processes for printing and developing, and as ferric tartrate in blueprinting. Rochelle salt is used extensively in the silvering of mirrors, as a component of crystal-controlled electronic oscillators, in the processing of cheese and as a mild cathartic in drug preparations. Cream of tartar is employed in baking powders, in hard candies and taffies for its ability to invert cane sugar, as a cleaner of brass, in the electrolytic tinning of iron and steel and in gold and silver coating of metals. Tartar emetic is used extensively in therapeutics for nauseant, expectorant and emetic effects. In tropical infections such as trypanosomiasis and kala azar, it is administered intravenously. Tartar emetic finds use in other varied fields; *e.g.*, in the control of insects on farms and as a mordant in fixing basic colours on cotton, leather and fur.

Chemical Reactions.—Tartaric acid is oxidized by hydrogen peroxide in the presence of ferrous salts to dihydroxymaleic acid. Reduction with hydrogen iodide forms first *d*-malic and finally succinic acids. Heating forms pyruvic and pyrotartaric acids and finally a black residue. Ammoniacal silver solution is reduced with the formation of a silver mirror. Tartaric acid has the ability to form complex metal ions with aluminum, iron, copper and lead, and prevents their precipitation on the addition of bases.

Stereoisomeric Tartaric Acids.—Four varieties of tartaric acid are recognized: (1) the ordinary dextrorotatory tartaric acid found either free or as potassium and calcium tartrates in the juices of tamarinds, mulberries, pineapples, unripe beetroot and especially in grapes; (2) levorotatory tartaric acid, which occurs naturally in the fruit and leaves of *Bauhinia reticulata*, a native tree of the French Sudan, but which originally was obtained by the resolutions of the salts of racemic acid. Except as regards its interactions with other optically active substances, its chemical properties are identical with those of the dextro-acid and in all its physical properties it resembles this acid except that it turns the plane of polarization of light to the left and its crystalline salts show hemihedral faces like those of the dextro-acid, but oppositely situated, as object to mirror image; (3) racemic acid, $(C_4H_6O_6)_2.2H_2O$, which is obtained from the tartar mother liquor and dried at 110° C. to the anhydrous form, melts at 206° C. It is produced synthetically by oxidation of fumaric acid, by hydroxylation of fumaric or maleic acids, or by mixing molecular proportions of the *d*- and *l*- acids; (4) mesotartaric acid is obtained when cinchonine tartrate or *d*-tartaric acid or its alkali salts are heated for prolonged periods at 170° C. The anhydrous form of the acid melts at 159–160° C. The meso form can be synthesized by hydroxylation of fumaric or maleic acids. It is inactive like racemic acid but rather for the reason that the molecule has a plane of symmetry. Unlike the racemic mixture, it is not resolvable into optically active varieties. The chemical study of the optical properties of these four tartaric acids by Louis Pasteur forms the foundation of modern conceptions of stereoisomerism. *See* STEREOCHEMISTRY: *Early History of Stereochemistry.* (P. P. R.)

TARTARS: *see* TATARS.

TARTARUS, in Greek mythology, the son of Aether and Gē, father of Typhoeus and the giants. In the *Iliad* the word denotes an underground prison, as far below Hades as earth is below heaven, in which those who rebelled against the will of Zeus were confined. In later commentaries Tartarus is the place of punishment of the wicked after death. *See* ABYSS.

TARTINI, GIUSEPPE (1692–1770), Italian violinist, composer and musical theorist, was born at Pirano, Istria, on April 8, 1692. In early life he studied, with equal want of success, for the church, the law courts and the profession of arms. As a young man he was wild and irregular, and he crowned his improprieties by clandestinely marrying the niece (or the daughter of a dependent) of Cardinal Cornaro, archbishop of Padua. The cardinal resented the marriage and Tartini, thinking his life in danger, fled for safety to a monastery at Assisi, where his character underwent a complete change. He studied the theory of music under Padre Boemo, the organist of the monastery, and learned to play the violin in so masterly a style that his performances in the church became the wonder of the neighbourhood. For more than two years his identity remained undiscovered. Then his retreat was revealed to the cardinal, who, hearing of his changed character, readmitted him to favour and restored him to his wife.

Tartini next removed to Venice, where he repaired, by the aid of good instruction, the shortcomings of his own self-taught method. After a period of study at Ancona, he returned to Padua, where he was appointed solo violinist at the church of San Antonio. From 1723 to 1725 he acted as conductor of Count Kinsky's private band in Prague. In 1728 he founded a school for the violin in Padua. The date of his presence in Rome does not seem to be clearly established, but he was in Bologna in 1739.

Afterward he returned to Padua, where he died Feb. 16, 1770.

Tartini's numerous compositions illustrate his passionate and masterly style of execution, in which he surpassed all his contemporaries. He frequently headed his pieces with an explanatory poetical motto, such as "Ombra cara," or "Volgete il riso in pianto o mie pupille." He told Lalande in 1766 that the sonata known as *Il Trillo del Diavolo*, or *The Devil's Sonata*, was the fruit of a dream, in which the devil played an exquisite sonata. Tartini's first book of sonatas was published at Amsterdam in 1734, the second at Rome in 1745, and many chamber works appeared during his lifetime.

Tartini contributed to the science of acoustics by his discovery (independently of Sorge, 1740, to whom the primary credit is now given) of what are still called "Tartini's tones" or differential tones. When any two notes are produced steadily and with great intensity, a third note is heard, whose vibration number is the difference of those of the two primary notes. It follows from this that any two consecutive members of a harmonic series have the fundamental of that series for their difference tone—thus, $\frac{E}{C}$, the fourth and fifth harmonic, produce C, the prime or generator, at the interval of two octaves under the lower of those two notes; $\frac{E}{G}$, the third and fifth harmonic, produce C, the second harmonic, at the interval of a fifth under the lower of those two notes. The discoverer was wont to tell his pupils that their double-stopping was not in tune unless they could hear the third note; and Henry Blagrove (1811–72) gave the same admonition. Tartini made his observations the basis of a theoretical system which he set forth in his *Trattato di musica, secondo la vera scienzia dell' armonia* (Padua, 1754) and *Dei Principij dell' armonia musicale* (Padua, 1767; Eng. trans., 1771).

He also wrote a *Trattato delle appogiature*, posthumously printed in French, and an unpublished work, *Delle Ragioni e delle proporzioni*, the manuscript of which was lost.

TARTU (Dorpat), a city of Estonia, situated on the Ema (Embach) river, in 58° 23′ N. and 26° 23′ E. Est. pop. (1959) 74,000.

The name "Tartu," adopted after Estonian independence in 1918, means lowland of Tar or Taara, a god of Estonian mythology probably related to the Scandinavian Thor. The principal part of the town lies south of the river, and the more important buildings are clustered around the two eminences known as the Domberg (Cathedral hill) and the Schlossberg (Castle hill), which in the middle ages were occupied by the citadel, the cathedral and episcopal palace.

Following a great fire in 1777 the town was almost entirely rebuilt, and its fortifications were transformed into tree-lined promenades. In addition to the 13th-century church of St. John, one of the finest and greatest brick buildings in mediaeval Livland, a university with an observatory, an art museum, a botanical garden and a library of several thousand volumes housed in a restored portion of the cathedral were established there. The Estonian National museum in its dome-shaped building was located just outside the town.

The university, chartered by Gustavus Adolphus of Sweden in 1632, moved to Parnu on the advance of the Russians in 1699 and again moved during the occupation by Peter the Great. Not until 1802, under the patronage of Alexander I, was the university restored. German influence predominated until the Russification of 1896.

In 1919 the institution was reopened as the Estonian university. The technical school was moved to Tallinn.

The astronomy department of the university became famous partly as a result of the labours of F. G. W. von Struve (1820–39) and partly because of the work of Joseph von Fraunhofer (*q.v.*), who developed a great refracting telescope, presented by Alexander I.

The foundation of Dorpat is ascribed to Yaroslav, prince of Kiev, and is dated 1030. In 1224 the town was seized by the Teutonic Knights, and in the following year Bishop Hermann erected a cathedral on the Domberg. From that date till about 1558 the town enjoyed great prosperity, and the population reached 50,000.

In 1558 it was captured by the Russians, but in 1582 was yielded to Stephen Bathory, king of Poland. In 1600 it fell into the hands of the Swedes, in 1603 reverted to the Poles and in 1625 was seized by Gustavus Adolphus of Sweden. The Russians again obtained temporary possession in 1666, but did not effect a permanent occupation till 1704.

In 1708 the bulk of the population was moved to the interior of Russia.

The town passed from Russian to Estonian rule in 1918, but suffered during German occupation and the later struggle with the bolshevists. In 1940 the town again fell under soviet control, and from 1941 to 1944 it was under German occupation.

TARWEED, the name given in the United States to various plants of the tribe Madieae of the family Compositae (*q.v.*), natives chiefly of the Pacific coast of North America. They are mostly annuals, with showy yellow or white flowers. About 25 species occur in California, Oregon and Washington. Several of these, as the yellow tarweed (*Hemizonia virgata*), the coast tarweed (*H. corymbosa*) and the clustered tarweed (*H. fasciculata*), are valuable honey plants.

The most widely known species is the Chile tarweed (*Madia sativa*), which is native to Chile and has been naturalized in California and Oregon. It is a useful forage plant, the seeds of which yield a pleasant, edible oil.

TASCHEREAU, SIR HENRI ELZEAR (1836–1911), chief justice of Canada, was born at St. Mary's in Beauce county, Que., on Oct. 7, 1836. He was educated at the seminary of Quebec and was called to the bar there in 1857, becoming a queen's counsel in 1867. He entered parliament and was elected Conservative member for Beauce county in the legislative assembly (1861–67).

He gained a reputation with his work on *Criminal Law Consolidation and Amendment Acts of 1869 for the Dominion of Canada with Notes, Commentaries, etc.* (vol. i, 1874; vol. ii, 1875, and later editions); and soon afterward published *Le Code de procédure civile du Bas-Canada* (1876). He became successively puisne judge of the superior court of Quebec (1871), judge of the supreme court of Canada (1878) and chief justice of Canada (1902–06). He was knighted in 1902.

In 1895 Taschereau had become dean of the faculty of law at Ottawa university, where he had previously held a chair. He became a member of the judicial committee of the privy council in 1904.

Taschereau died at Ottawa on April 14, 1911.

TASHKENT, city and *oblast* of the Uzbek Soviet Socialist Republic, U.S.S.R., situated in a loess oasis, watered by the interlocking Chirchik and Keles tributaries of the Syr-Darya.

Irrigation cultivation is very ancient in the oasis; the town itself dates back at least to the 7th century A.D. In 582 the Turks of Transoxiana divided into two khanates, the western one having two headquarters, one at Urumchi and one north of Tashkent, and it is recorded that the Chinese in 659 claimed the territory of the western Turks, including Tashkent. It was captured in 1865 by the Russians and was made the administrative centre of the former government of Russian Turkestan. The Russians built a new town nearly 8 mi. long and 4 mi. broad, considerably to the northeast of Eski-tashkent (Old Tashkent), which is now practically deserted.

The houses are low because of frequent earthquakes and are almost hidden by the poplar, willow and fruit trees which surround them and line the broad streets. Its population in 1959 was estimated to be 911,000. The population is mixed, with Russians, Sarts and Uzbeks predominating.

Tashkent has cotton-cleaning factories, *makhorka*-tobacco factories, a leather, machinery and cellulose industry and sawmills; there is municipal electricity, water and tram service. Much fruit and agricultural produce are grown in the oasis.

Under the postrevolutionary regime efforts were made to raise the status of women.

A state library, a museum including a collection of Graeco-

Bactrian coins, a teachers' training college and the Uzbek Academy of Sciences were established there. The population of Tashkent *oblast* in 1959 was estimated to be 2,263,000.

TASMAN, ABEL JANSZOON (*c.* 1603–*c.* 1659), the greatest of Dutch navigators and explorers, the discoverer of Tasmania, New Zealand, the Tonga and the Fiji Islands, was born at Lutjegast in Groningen, probably in 1603. He entered the service of the Dutch East India company in 1632 or 1633. His first exploring voyage was to Ceram as captain of the "Mocha" in 1634. He sailed in 1639 under Matthew Quast on an expedition in search of the "islands of gold and silver" in the seas east of Japan. After a series of trading voyages to Japan, Formosa, Cambodia and Palembang, he was chosen by the governor general, Anthony van Diemen, to command in 1642 the most ambitious of all Dutch voyages for the exploration of the southern hemisphere.

Tasman's Voyage 1642–43.—By 1642 Dutch navigators had discovered discontinuous stretches of the western coast of Australia, but whether these coasts were continental and connected with the hypothetical southern continent of the Pacific remained unknown. Tasman was to answer this problem, following instructions based on a memoir by Frans Jacobszoon Visscher, his chief pilot. He was to explore the Indian ocean from west to east, south of the ordinary trade route, and, proceeding eastward into the Pacific (if this proved possible), to investigate the practicability of a sea passage eastward to Chile, to rediscover the Solomon Islands of the Spaniards and to explore New Guinea.

Leaving Batavia on Aug. 14, 1642, with two ships the "Heemskerk" and "Zeehaen," Tasman sailed to Mauritius (Sept. 5–Oct. 8), then southward and eastward, reaching his most southerly latitude of 49° S. in about 94° E. Turning north he discovered land on Nov. 24 in 42° 20′ S. and he skirted its southern shores, naming it Van Diemen's Land (now Tasmania). A council of officers on Dec. 5 decided against further investigation, so he missed the opportunity of discovering Bass strait. Continuing eastward, he sighted on Dec. 13, in 42° 10′ S., the coast of South Island, New Zealand, and explored it northward, entering the strait between North Island and South Island but supposing it to be a bay. He left New Zealand on Jan. 4, 1643, at North cape under the impression that he had probably discovered the west coast of the southern continent, which might be connected with the "Staten Landt" discovered by W. C. Schouten and J. Le Maire south of South America: hence the name of "Staten Landt" which Tasman gave to his discovery in honour of the states-general.

Convinced by the swell that the passage to Chile existed, Tasman now turned northeast and on Jan. 21 he discovered Tonga and on Feb. 6 the Fiji Islands. Turning northwest, the ships reached New Guinea waters on April 1 and Batavia on June 15, 1643, completing a ten months' voyage on which only 10 men had died from illness. Tasman had circumnavigated Australia without seeing it, separating it from the hypothetical southern continent.

The council of the company decided, however, that he had been negligent in his investigation of the lands discovered and of the passage to Chile. They sent him on a new expedition to the "South Land" in 1644.

Tasman's Voyage 1644.—Tasman's instructions ordered him to establish the relationships of New Guinea, the "great known South Land" (western Australia), Van Diemen's Land and the "unknown South Land." He sailed from Batavia on Feb. 29, steering southeast into Torres strait, which he mistook for a shallow bay, coasting the Gulf of Carpentaria and then following the coast of Australia to 22° S.

Although he was rewarded with the rank of commander and was made a member of the Council of Justice of Batavia, his second voyage was also a disappointment to the company because it had failed to reveal lands of potential wealth. In 1647 Tasman commanded a trading fleet to Siam, and in 1648 a war fleet against the Spaniards in the Philippines. He had left the company's service by 1653 and died probably before Oct. 22, 1659, and certainly before Feb. 5, 1661.

BIBLIOGRAPHY.—The standard authorities are J. E. Heeres (ed.) *Abel Janszoon Tasman's Journal* (1898); R. Posthumus Meyjes (ed.) *De Reizen van Abel Janszoon Tasman en Franchoys Jacobszoon Visscher* (1919). *See* also R. H. Major, *Early Voyages to Terra Australis*, Hakluyt Soc. (1859); J. B. Walker, *Abel Janszoon Tasman; his life and voyages* (1896); J. C. Beaglehole, *The Discovery of New Zealand* (1939); G. C. Henderson, *The Discoverers of the Fiji Islands*, pp. 3–99 (1933). The first published account in England of Tasman's voyage of 1642–43 (a translation from the Dutch of Arnoldus Montanus, 1671) appeared in John Ogilby's *America*, pp. 653–661 (1671). In 1682 Robert Hooke published in his *Philosophical Collections*, no. 6, pp. 179–186, the first English translation of D. R. van Nierop's *Eenige oefeningen* (1674), comprising extracts from a journal of the voyage.
(H. M. Ws.)

TASMANIA, a state of the commonwealth of Australia comprising one large and numerous smaller islands, the largest of which, King, Flinders and Cape Barren islands, lie in Bass strait. Except for Macquarie Island (lat. 54° 35′ S.), Tasmania is located on the continental shelf of Australia, being separated from the mainland by a 150-mi. stretch of shallow sea (30–40 fathoms). The shield-shaped main island tapers southward from latitudes 40° 40′ to 43° 40′ S. and extends eastward from longitudes 144° 35′ to 148° 20′ E.; it has a maximum length of 180 mi., a maximum width of 190 mi. and an area roughly that of Scotland (26,215 sq.mi.) or 0.88% of the commonwealth of Australia.

PHYSIOGRAPHY

Geology.—Pre-Cambrian quartzites and schists (Davey group), in general striking north-south, outcrop in western Tasmania, succeeded by Late Proterozoic dolomite, slate and quartzite (Carbine group) and by at least 10,000 ft. of Middle and Upper Cambrian tuffs, lavas, graywackes, slates, conglomerates and breccias (Dundas group). The Cambrian rocks, containing trilobite and dendroid fossils, have been metamorphosed to form keratophyric "porphyroids," an important ore-bearing zone, but some intrusive porphyry may also be present. Tillite occurs on King Island (Upper Carbine). Extensive sills of basic rock, intruded during the Middle or Upper Cambrian and subsequently serpentinized, are the source of alluvial osmiridium deposits. Coarse, siliceous Lower Ordovician conglomerate (Junee group) is followed by about 5,000 ft. of richly fossiliferous Gordon limestone ranging from Lower Ordovician to Lower Silurian and this by about 5,000 ft. of richly fossiliferous quartzites and slates ranging from Lower Silurian to Lower Devonian (Eldon group). Mountain building in the Middle Devonian period was associated with the introduction of extensive granites and the formation of metallic mineral deposits (copper, lead, zinc, tin, tungsten and gold, all of which have been profitably mined). After continental glaciation in the Upper Carboniferous, marine transgression resulted in the deposition of tillites, sandstones, shales and limestones, all with glacial erratics and rich in fossils. Two fresh-water interludes added kerosene shales and coal seams. The Triassic, which follows with disconformity, comprises sandstones and plant-bearing shales (with important subbituminous coal seams) and, finally, tuffaceous feldspathic sandstones. The Permo-Triassic sequence is intruded by thick sills and dikes of dolerite, which cannot be older than Upper Triassic nor younger than Cretaceous. The land surface was peneplained in the Late Mesozoic and broken by normal faulting in the Early Tertiary, resulting in block mountains up to 5,000 ft. high and fault lowlands. The Bass strait depression was also initiated and occupied by the sea, which encroached beyond the present shores to deposit sandstones and limestones. The fault lowlands rapidly filled with sands and clays to a maximum thickness of 1,000 ft., and extensive basalts (probably Pliocene) poured over the landscape. Subsequent movements have been confined to minor upwarps. Tasmania was glaciated during the Pleistocene and, because of a cognate eustatic retreat of the sea, Bass strait was drained. Subsequently a rise in sea level again flooded Bass strait, together with the estuaries of the Tamar, Derwent, Huon and Gordon rivers, producing some excellent harbours. (S. W. Cy.; M. R. Bs.)

Physiography.—Tasmania is characterized by rugged relief and swiftly flowing rivers with irregular gradients. The only notable exception to this occurs between Launceston and Hobart, where former lake floors give rise to extensive areas of slight relief at relatively low altitudes. Elsewhere, even along the apparently flat, narrow coastlands of the west and northwest, the

landscape is sharply dissected by rivers, notably of the Gordon, Pieman and Arthur systems. In the west and southwest are mountain ranges where sharp peaks, resulting from the glacial erosion of hard quartzites and conglomerates and reminiscent of Alpine forms, rise to 5,000 ft. (Mt. Ossa, the highest in Tasmania, is 5,203 ft.) and where valleys are commonly up to 2,000 ft. deep. Central Tasmania consists of dolerite plateaus, bounded mainly by steep, often step-faulted scarps (hence Western Tiers), above which rise tablelands and flat-topped hills (mesas and buttes). Like the northeastern granitic plateaus beyond the Tamar-Esk depression, the central plateau averages 3,500 ft. but, unlike the northeast, it slopes southeastward, being drained by the Derwent, and bears witness, particularly in its many lakes, to Pleistocene glaciation. Most of Tasmania above 2,200 ft., except for the northeast, exhibits typical features of glaciated landscapes (U-shaped valleys, cirques, moraines, roches moutonnées, etc.).

The coast line is rugged and irregular because of recent subsidence by about 300 ft. Where the drowned land surface was comparatively level, however, as in the northwest, a recent emergence of about 70 ft. has produced a more mature coast line with sweeping beaches a few miles in length. Elsewhere, the development of spits, bars and tied islands, though still at an early stage, provides some evidence of a straightening coast line.

Climate.—Tasmania, nowhere far removed from the sea, lies in the belt of westerly winds and antarctic lows (cyclones). Hence its weather is markedly variable not only from day to day but also from season to the same season in ensuing years. Like southwest England and Ireland, its climate is cool, equable and humid, but because of rugged relief there are marked variations both regionally and locally. Except in upland areas, temperatures range from 60° to 65° F. in summer to 45° to 50° F. in winter, the east coast being about 4° warmer than the west coast because of the influence of warm (north) and cool (south) sea currents respectively. In general, rainfall is determined by relief and decreases eastward; it averages 50 in. a year on the west coast, 90–120 in. on the western highlands (144 in. at Lake Margaret), 40–60 in. on the northeastern highlands and 20–40 in. elsewhere. Except locally in the southeast, where southeasterly winds often bring heavy rains in spring and summer, there is a winter or late winter maximum. Tasmania enjoys a temperate summer climate which is attractive to Australians from hotter and drier regions.

Soils.—Tasmania with its rugged relief, humid climate and forest vegetation is characterized by gray, sandy podsols, interspersed with extensive tracts of shallow immature soils. In superhumid highland areas podsols are replaced by peats, and in the subhumid southeast, by brown earths and black soils. Alluvial soils occupy limited tracts of river valleys, being significant only along the middle Derwent. Fertile red loams derived from weathered basalt occur in the northwest. (G. G. F.; P. St.)

Vegetation.—Climatic, physiographic and edaphic variety gives to Tasmania a wide range of habitats for plants. Three plant formations occur: sclerophyll forest, temperate rain forest and austral-montane. Sclerophyll forest is characteristic of the regions having an annual rainfall of less than 60 in. The fine forests in the 30–60-in. rainfall areas have been extensively exploited for timber and for the paper industry; they have also suffered from destructive fires. The species of economic importance are *Eucalyptus regnans, E. obliqua, E. gigantea* and, in the south and southeast, *E. globulus*. In the drier central and eastern parts of the state a savanna woodland occurs in which *E. pauciflora* is dominant; she-oaks (*Casuarina* species) and wattles (*Acacia* species) are conspicuous. Temperate rain forest which develops in superhumid areas is characterized by myrtle (*Nothofagus cunninghamii*), a timber tree which may reach a height of 100 ft., and by sassafras (*Antheroperma moschatum*). This community contains several endemic conifers valuable as timber although no longer plentiful—the Huon pine (*Dacrydium franklinii*), King Billy pine (*Athrotaxis selaginoides*) and celery-top pine (*Phyllocladus rhomboidalis*). In certain areas, particularly in the forests of the south and southwest, an almost impenetrable thicket develops known as horizontal scrub. This is due to the growth of a small tree, *Anodopetalum biglandulosum*. The slender trunk falls over under its own weight and from it branches arise which behave in the same way so that the tangled interlacing branches spread like a platform above the ground. In both sclerophyll and rain forest formations large areas of acid waterlogged soil support a specialized plant community named after its most conspicuous member, button grass (*Gymnoschoenus sphaerocephalus*). In the wettest parts this sedge forms large tussocks 2 to 6 ft. high, but elsewhere other monocotyledons, chiefly members of the families Cyperaceae and Restionaceae, are dominant. Plants of the mountain plateaus include many which have a subantarctic distribution. Of these the perennial evergreen cushion plants are particularly interesting, Tasmania having five species including *Donatia novae-zelandiae*, found also in New Zealand, and *Phyllachne colensoi*, found in similar habitats in New Zealand and Tierra del Fuego.

See J. D. Hooker, *Botany of the Antarctic Voyage III, Flora Tasmaniae* (1860); L. Rodway, *The Tasmanian Flora* (1903); W. M. Curtis and J. Somerville, "The Vegetation (of Tasmania)" in *A.N.Z.A.A.S. Handbook of Tasmania* (1949); W. M. Curtis, *The Student's Flora of Tasmania*, Part I (1956). (W. M. C.)

HISTORY

Tasmania, or Van Diemen's Land as it was called until 1853, was discovered in 1642 by the Dutch navigator, Abel Tasman (*q.v.*). It was named after Antony Van Diemen, the governor general of the East Indies at whose command Tasman sailed. Further exploration awaited the French and British navigators, Marion Dufresne (1772), Tobias Furneaux (1773), James Cook (1777), William Bligh (1788 and 1792), Bruni d'Entrecasteaux (1792–93), John Hayes (1794), George Bass (1798), Nicolas Baudin (1800 and 1802) and Matthew Flinders (1798 and 1802). The first British settlement was made at Hobart (1804), and was followed by another which was to become Launceston (1806); most of the first settlers, both free and convict, were transferred from Norfolk Island. The settlement, under the resident lieutenant governors David Collins, Thomas Davey and William Sorell, was governed directly from New South Wales. Free immigrants were attracted by land grants and by the development of the sheep and whaling industries. Van Diemen's Land was made a colony by order in council in 1825, and henceforward its government was practically separated from that of New South Wales. Until 1850 lieutenant governors George Arthur, Sir John Franklin, Sir John E. Eardley-Wilmot and Sir William Denison governed through nominated executive and legislative councils and built the foundations of the administrative structure of the colony. In this period, despite the depredations of bushrangers and the "black war" which ended with virtual extinction of the aboriginal people, great material progress was made. In 1828 the total white population was 17,000, of whom 7,500 were convicts under restraint, though a majority were assigned as servants to free settlers. Grain was exported regularly to New South Wales, while rising wool prices brought general prosperity and stimulated much speculation in land, even, from 1835, in the Port Phillip district beyond Bass strait. British capital entered the colony, particularly through British government expenditure on the convict establishments. The adoption of the probation system for convicts from 1840 marked a change in British policy on transportation. The influx of population was rapid until 1847, growing from 46,000 to 70,000, but the ratio of free to unfree was reversed. Simultaneously the boom gave way to an economic depression and the colonial government, despite severe cuts in expenditure, faced a deficit, which local opinion attributed to the cost of police and jails. Belatedly and reluctantly the home authorities offered some relief, but not before public opinion was roused to demand representative institutions and the cessation of transportation. Representatives were first elected to the legislative council in 1851 and transportation ceased in 1853. The constitution was settled more definitely in 1856 with the election of a two-chamber parliament and the appointment of the first responsible ministry.

The gold discoveries in Victoria in 1851 caused inflation and labour shortage in Tasmania and the economy was unable to support the strain. The whaling industry declined and with it shipbuilding. The adoption of a protectionist policy by Victoria in-

jured Tasmanian manufactures. After 1857 a prolonged depression, lasting 25 years, was marked by deficits, reduction in services, opposition to all forms of taxation, political feuds and frequent changes in ministries. Under competition from more suitable wheatlands on the Australian continent, wheat production gradually declined to its present insignificance.

After 1880 economic conditions improved, largely because mining industries were opened up in the west and northeast. The hardwood forests were exploited and fruitgrowing in the south became the basis of small manufacturing and export industries. Railways, roads and harbour facilities were developed, while relative political stability was achieved, particularly during the Fysh ministry (1887–92). With the failure of the Bank of Van Diemen's Land in 1893 began a period of business recession, unemployment and government retrenchment. From 1894 to 1899 there was a slow recovery during Sir Edward Braddon's ministry. In the elections for the house of assembly in 1897, proportional representation in the form of the Hare-Clark system (*see* REPRESENTATION) was introduced and has been continued with some modifications to this day. In the 1890s a majority of Tasmanians supported the movement that ended in 1901 with Australian federation (*see* AUSTRALIA).

Excepting two periods of economic recession, 1922–26 and 1931–35, the 20th century saw continuous development. In the first decade the fruit and fruit-preserving industry grew until it became of considerable importance in the state's economy. The northern coastal districts developed potato growing and dairying, while other parts were noted for fine wool, barley and fat lamb production (see *Agricultural and Pastoral Industries,* below). To secure a balanced economy, to prevent the drift away of population and to promote decentralization, state governments supported deliberately the development of secondary industries, based on the state's hydroelectric undertakings (see *Manufacturing Industries and Power Supply,* below). Population increased from 172,000 in 1901 to 308,752 in 1954.

The political life of the state was organized by two parties, Labour and Liberal, which, under the existing proportional representation system, rarely have a large parliamentary majority, though Labour governed continuously from 1934 to 1953. The legislative council or upper house, elected on a more restricted franchise, still possesses effective constitutional powers. Undoubtedly the greatest change after 1901 (and particularly after 1933) lay in the increasing financial dependence of the state on the resources of the Australian commonwealth (federal) government (see *Finance,* below).

BIBLIOGRAPHY.—J. West, *History of Tasmania,* 2 vol. (Launceston, 1852); J. Fenton, *History of Tasmania* (Hobart, 1884); R. W. Giblin, *Early History of Tasmania,* 2 vol. (Melbourne, New York, 1939); W. D. Forsyth, *Governor Arthur's Convict System: Van Diemen's Land 1824–36* (London, New York, 1935); K. E. Fitzpatrick, *Sir John Franklin in Tasmania, 1837–1843* (Melbourne, 1949); W. A. Townsley, *The Struggle for Self-Government in Tasmania, 1842–1856* (Hobart, 1951); also articles in *Historical Studies, Australia and New Zealand,* vol. 1 no. 1, vol. 3 no. 12, vol. 4 no. 15 and vol. 5 no. 17 (Melbourne, 1940 *et seq.*); R. M. Hartwell, *The Economic Development of Van Diemen's Land, 1820–1850* (Victoria, New York, 1954). (W. A. TY.)

POPULATION AND SOCIAL CONDITIONS

The population of Tasmania in 1954 was 308,752. The state's two cities, Hobart and Launceston, were comparatively small and contained only 30% of the population, Hobart, the state capital, having a population of 54,887, and Launceston 37,627.

Lacking any great metropolitan centre, Tasmania provides no extreme contrasts of wealth or poverty. A considerable proportion of the surplus of the larger manufacturing enterprises is distributed to nonresident shareholders; consequently the more prosperous sections of the community are mostly urban retailers, owners of the larger pastoral properties and the more successful members of various professions. As in other parts of Australia, social stratification is not clearly marked and income differences are reflected in material possessions rather than in ways of life or thought. On the other hand, some families which have lived in Tasmania for several generations enjoy special prestige in the community, and this may be associated in some cases with political or social influence. The numerical importance of small and medium property owners (themselves often wage earners) is reflected in the relatively conservative tendency of both political parties. Political discussion is mainly concerned with the costs and benefits of government services and the details of development policies.

The work force of the larger industrial, mining and public utility undertakings is organized in well-established trade unions. Comparatively few major industrial disputes occurred in Tasmania during the years 1945–58. The minimum wages and conditions of Tasmanian employees are regulated either by awards of the commonwealth arbitration court, or by state wages boards, of which the determinations follow the general pattern established by the commonwealth court.

Education.—Educational services are provided by the state government, which since 1936 has developed area schools in country towns to which pupils are brought by motor transport, thus making it possible to provide a more varied curriculum for country children than was possible in very small local schools. Education is compulsory up to 16 years of age. Secondary education is provided in state high schools, and also in 13 private (fee-charging) schools situated in Hobart and Launceston. Most of the private schools are associated with various religious denominations and receive no state assistance; about one-fifth of the students undergoing secondary education go to private schools. The University of Tasmania (founded 1890) in 1957 had a teaching staff of approximately 80 with about 1,000 students, and awarded degrees in the faculties of arts, law, science, engineering and commerce.

Health.—Hospital services are provided by the state as well as by private (but subsidized) institutions, and state medical officers may be appointed in rural municipalities which cannot support a private practitioner. Child welfare services helped to reduce the number of deaths under one year per 1,000 live births from 58 in 1920–29 to 24 in 1951–55.

Social Services.—Various social service payments (*e.g.,* child endowment [family allowances], unemployment and sickness benefits, age and widows' pensions) are payable under commonwealth legislation (mainly enacted 1940–46), and together with high levels of employment during the period 1942–57 have almost entirely removed the worst symptoms of poverty.

ECONOMIC CONDITIONS

Agricultural and Pastoral Industries.—Agriculture is restricted by physical factors and accessibility to the northwestern coastlands from Smithton to Devonport, the central northern lowlands from Deloraine to Launceston and Cressey, the state highway and railway zone from Launceston to Hobart (the midlands), and the valleys of the southeast. In general, mixed farming is characteristic. Thus the north, more humid and having richer soils than other farming areas, specializes in potatoes and fodder growing, dairying, fat lamb production and the fattening of beef cattle; apples and pears are grown in the Mersey and Tamar valleys and in the Lilydale and Scottsdale districts. Beef cattle are kept also in all other farming areas, particularly on marginal lands, but in the drier areas south of Launceston, cereals, mostly oats but also soft wheat and barley, together with pulse, fodder and sheep, are the basis of mixed farming. The midlands are the principal sheep-raising region, being partly dependent on the lake plateau for summer grazing, and are noted for fine merino wool and stud stock. Dairying reappears in the southeast, though it is less important here than in the north. Hops are a specialty of the irrigated alluvial lands of the middle Derwent, berry and stone fruits of the lower Derwent, and apples and pears in the Huon valley. Elsewhere, the land provides only rough grazing for sheep, or timber for the sawmill and paper industries; substantial highland areas in the west and southwest are infertile and inaccessible.

Fisheries.—Tasmanian fisheries, in contrast with those of other Australian states, are mainly offshore and mainly dependent on export markets. They enjoy the advantage of a coastline richer in natural harbours, anchorages and protected fishing grounds than that of the mainland, but productivity is limited by a narrow continental shelf and by the low plankton content of the predominating waters of the subtropical East Australian current. Fishing is

mostly confined to the northwest and southeast. The principal species caught are barracouta (*Thyrsites atun*) and crayfish (*Jasus lalandii*), but snapper or school shark (*Notogaleus rhinophanes*) and gummy shark (*Mustelus antarcticus*), together with whitebait (*Lovettia sealii*), found mainly in the Leven and Mersey rivers but also in the Huon river, increased greatly in importance during the 1940s. Barracouta are taken mainly with jigsticks (8–10-ft. poles with two barbless hooks) and crayfish with beehive lobster pots.

Since World War II the fishing industry, hitherto dependent on a variable demand for fresh fish from the mainland, has been stabilized to some extent by the development of fish canning. Factories are situated at Stanley, Ulverstone, Launceston, Triabunna, Dunalley, Hobart, Margate and on Flinders Island. The chief markets are Melbourne and Sydney.

Mining and Metallurgy.—Mining industries, which consume 50%–60% of the state's hydroelectric power production, play an important part in the Tasmanian economy. At Mt. Lyell copper —about 8,000 tons a year—is produced from chalcopyrite by flotation and electrolytic refinement. The ore, which contains only 0.65% copper, yields gold and pyrites as by-products. Silver-lead and zinc ores are mined partly at Zeehan but also at Rosebery. There zinc, lead and copper concentrates are separated by flotation from an extremely fine-grained ore containing mainly sphalerite and galena with some pyrite and chalcopyrite. Lead and copper concentrates are exported but zinc concentrate is refined electrolytically (annual output about 80,000 tons) at Risdon, by-products being sulfuric acid and superphosphate fertilizer. Since World War II tungsten minerals have become important with the development of extensive scheelite deposits on King Island. The scheelite, which occurs in a metamorphosed limestone, is separated by flotation and the concentrate (about 1,000 tons a year) exported. Wolframite is also mined in several places. Alluvial tin deposits are common in the northeast where the most productive tin mine in Australia (about 500 tons of tin concentrate a year) is working at Aberfoyle. Tin associated with pyrrhotite is being mined at Renison Bell and is separated by gravity concentration. Gold, a by-product of several ores, is now relatively unimportant, there being few gold mines, most of these in the northeast. Osmiridium occurs in placer deposits derived from serpentine but is little mined. Although bauxite deposits are considerable, they are being held in reserve for the aluminum works at Bell bay. Triassic coal, subbituminous with a high ash and sulfur content, is mined (about 250,000 tons a year) mainly on the east coast (St. Mary's) but also in the Derwent valley (Hamilton). Mining is by the bord-and-pillar method but is hampered by small-scale faulting and large dolerite intrusions. Permian coal and oil shale, not economically significant, are found near Devonport. Limestone, widely distributed throughout the state, is quarried (about 150,000 tons a year) for the production of agricultural lime, cement and carbide. At Electrona the only carbide produced in Australia is manufactured from Ordovician limestone quarried at Ida bay.

Manufacturing Industries and Power Supply.—Industrial development, which showed considerable expansion during and after World War II partly because of state encouragement, is based mainly on the processing of such primary products as fruit, dairy produce, wool, timber and minerals.

Although shipping difficulties have hampered development, Tasmanian industry is favourably placed both for interstate and overseas export, and it has cheaper power (wholly hydroelectric) than the rest of Australia. In fact, more than two-thirds (683,700 h.p. in Aug. 1958) of Australia's hydroelectric plant and about one-half (2,500,000 h.p.) of its potential resources are in Tasmania. Development dates from 1895. Since 1911 it has received much of its impetus from the growth of electrolytic zinc refining, still the major industry consuming hydroelectric power, though the carbide works and the newsprint mills using Tasmanian hardwoods, all situated near Hobart, have become important consumers. In 1929 the Hydro-Electric commission was created by the state government to undertake further development of power resources, all of which are state-owned. The Tarraleah power station (126,000 h.p.) on the Nive river, together with the Tungatinah development, harnesses the catchment runoff west of the Great lake (over 3,000 ft. above sea level). The Shannon (14,500 h.p.) and Waddamana (132,600 h.p.) stations, on the Shannon and Ouse respectively, are supplied from the Great lake itself, which is regulated by the Miena dam. On the Derwent, the Clark dam, at the foot of which lies the Butler's Gorge station (17,100 h.p.), provides additional storage for Tarraleah. The Trevallyn development on the Tamar near Launceston, together with Tungatinah, added 252,000 h.p. to the system in 1954. Trevallyn (112,000 h.p.) is designed primarily to supply the aluminum works at Bell bay, near the Tamar mouth, where production from Malayan bauxite began early in 1955. By 1958, Tungatinah (175,000 h.p.) had been augmented by stations at Wayatinah (61,500 h.p.), on the Derwent, and at Lake Echo (45,000 h.p.).

Although Hobart and Launceston are the chief industrial centres, Burnie and Devonport are important producers of paper, paints, pigments, cement (Railton), plywood and other building materials, as well as confectionery and egg and milk products.

External Trade.—Tasmania lies within the Australian customs area, and under the commonwealth constitution no restrictions can be placed on trade between Tasmania and the five other states. Interstate trade, of which a small but increasing proportion is carried by air, accounts for about three-fifths of Tasmanian exports and four-fifths of its imports. Of total exports, primary products, such as greasy wool, fresh fruit, hides and skins, metalliferous ores, fish, potatoes, hops and peas, furnish more than one-third; secondary products include scoured wool, zinc ingots, newsprint, canned fruit, butter, sawn timber, copper, carbide, cement and, to a lesser extent, woolens, worsteds and confections. Imports are mainly producers' materials and manufactured consumers' goods, principally machinery, motor vehicles, petroleum products, hardware, textiles, clothing and foodstuffs. In addition, substantial quantities of raw wool (grades not produced in Tasmania) are imported for the Launceston textile industry.

Transport and Communications.—Transport to the Australian mainland is provided by cargo and passenger steamships and by frequent air services. Since World War II Australian coastal shipping services have encountered various difficulties (*e.g.*, industrial disputes, wharf congestion, shortage of ships), and the airlines have gained a substantial part of the passenger and import traffic.

The railway system comprises more than 800 route miles (gauge 3 ft. 6 in.) and is mainly government-operated to facilitate cheap transport for bulky products. Although less than half the main road mileage has been sealed, motor transport is widely used for the carriage of goods and passengers. Internal transport is co-ordinated by the Transport commission, which has encouraged or undertaken the development of road-passenger services on several routes. Road-freighting services are licensed by the commission, which also operates its own services to some places. In mid-1958 about 85,000 motor vehicles were registered in Tasmania, of which 24,000 were commercial. The Tasmanian telephone system has been linked to that of continental Australia since 1936. Broadcasting stations (both national and commercial) serve all the main centres and television will be introduced about 1960.

Finance.—After federation in 1901, Tasmania's fiscal problems were due to relatively low taxable capacity, the lack of benefit (especially in earlier years) from Australian protectionist policies, and the resulting difficulty of providing public services at a standard comparable to those of wealthier states. To meet the difficulties of less developed states, the constitution of the commonwealth of Australia provides in section 96 that the federal parliament may grant financial assistance to any state on such terms and conditions as it thinks fit. Tasmania first received a federal grant under this section in the financial year 1912–13, but until 1934–35 the amount of such grants depended upon the discretion of the commonwealth government of the day. Between 1925 and 1930, methods of assessing section 96 grants on an objective basis were carefully examined by Tasmanian economists, whose work laid the foundations for procedures adopted by the Commonwealth Grants commission established in 1933. In 1934–35 and subsequent years federal grants under section 96 were paid to Tasmania (in common

with Western and South Australia) in strict accordance with the recommendations made by the commission as to the amount necessary to enable these states to function at a standard not appreciably below that of the wealthier states. Since 1942 the commonwealth has also paid substantial tax reimbursement grants to each state on condition that no taxation is levied by that state on income received by persons or companies. Tasmania has benefited considerably from these tax reimbursement arrangements, which have reduced the adverse effects of Tasmania's relative disadvantage in taxable capacity, and have made possible somewhat higher standards, especially in education and health services. Tasmania is represented on the Australian loan council, which annually determines the amount and distribution of borrowing by Australian public authorities. (G. G. F.; P. St.)

Bibliography.—G. L. Wood, *The Tasmanian Environment* (Melbourne, 1923); A. Lowndes and W. Maze, *Land Utilization Regions of Tasmania* (Sydney, 1937); State Economic Planning Authority, *Regional Planning Atlas: Economic Resources of Tasmania* (Hobart, 1947); Australian and New Zealand Association for the Advancement of Science, *Handbook to Tasmania* (Hobart, 1949); T. G. Taylor, *Australia* (London, Toronto, 1951); Commonwealth Statistician (Canberra), *Statistics for the State of Tasmania, Pocket Year Book of Tasmania, Monthly Review of Business Statistics* (Tasmania); also *Year Book of the Commonwealth of Australia.*

L. F. Giblin, *Federation and Finance,* Report of the Australian and New Zealand Association for the Advancement of Science, vol. xviii (Sydney, 1926); *Report of Committee on Tasmanian Disabilities under Federation* (Hobart, 1925); *Reports of Commonwealth Grants Commission,* esp. 3rd (Canberra, 1936); H. P. Brown, "Some Aspects of Federal-State Financial Relations" in G. Sawer (ed.), *Federalism in Australia* (Melbourne, 1949; London, 1950).

TASMANIAN DEVIL (*Sarcophilus ursinus*) is a large, heavily built relative of the dasyures. In size it may be compared to a badger. The general colour of the fur is black tinged with brown, with white patches on the neck, shoulders, rump and chest. It is a burrowing animal, nocturnal and carnivorous, and commits great depredations on the sheep yards and poultry lofts of the inhabitants. (*See* Marsupialia.)

TASMANIANS. The Tasmanians, who are now extinct, were of medium height; had black to dark brown skins; woolly hair; heavy brows; longish, oval or pentagonal, flattish and small sized (cranial content) heads; short broad noses and large teeth.

Culture.—They were food gatherers, expert hunters and trackers of game, nomads, moving in small groups of the family type, with very little government but recognizing the areas belonging to other groups as reserved and not to be interfered with except on risk of war. Hereditary or permanent chiefs do not seem to have been found. No marriage ceremony is recorded and there seems ground for holding that they were polygynous. They are typical representatives of the Early Stone and Wood Age. The stone implements are very crude in form and finish (akin to Mousterian products) and were made of phthanite, a fine grained sandstone. Fire was made with the groove and saw or with the drill. Stones with black and red bands were used for tallies or mnemonics for absent friends. All have been lost and their exact significance is not quite clear. They made rough drawings on bark with charcoal, ornamented themselves, using red ochre for the hair and wore shell necklaces, fillets of gay flowers, or festoons of showy berries. Huts and windbreaks were used as shelters. Skins were thrown over the back against rain but otherwise they were naked. Scarification was practised and they rubbed themselves with powdered charcoal and red ochre. They greased their bodies against rain. They made rafts generally of the bark of some species of Eucalyptus, which was rolled up into cigar-shaped bundles, three going to a raft, broad in the middle and tapering to a point at each end. These were good fair weather craft but dangerous in storms. They used their spears for fishing but had neither bow and arrow nor boomerang.

Language.—There were five dialects, classified on a geographical basis in two main groups as eastern and western. Enough remains to show that they were all related and were dialectical variations from a common archetype. The language is said to have been musical and soft, vowels being peculiarly full and round. They had no *d, f, v, s,* or *z.* Words began generally with a consonant, *cr, pr* and *tr* being common, other combinations being rare. Words end with a vowel as a rule. Distinct forms were used for singular and for plural.

Bibliography.—H. Ling Roth, *The Aborigines of Australia,* 1890, a careful and complete collection of earlier information which has been used by later writers. *See* also W. J. Sollas, *Ancient Hunters,* 3rd ed. 1924.

TASSIE, JAMES (1735–1799), Scottish gem-engraver and modeller, was born at Pollokshaws, near Glasgow, on July 15, 1735, and died in London on June 1, 1799. During his earlier years he worked as a stonemason until he removed to Glasgow to attend the academy established by the brothers Foulis, the printers, and became one of its most distinguished pupils. Subsequently he visited Dublin and became acquainted with Dr. Quin, who had been experimenting in imitating antique engraved gems in coloured pastes. He engaged Tassie as an assistant, and together they perfected the discovery of an "enamel," admirably adapted by its hardness and beauty of texture for the formation of gems and medallions. Tassie went to London in 1766. After a hard struggle the beauty and artistic character of his productions came to be known. The empress of Russia commissioned a collection of about 15,000 examples; all the richest cabinets in Europe were thrown open to him for study and reproduction; and his copies were frequently sold by fraudulent dealers as the original gems. He exhibited in the Royal Academy from 1769 to 1791. In 1775 he published the first catalogue of his works, followed in 1791 by two volumes quarto, enumerating nearly 16,000 pieces.

Tassie also executed many large profile medallion portraits of his contemporaries. They were modelled in wax and were then cast in white enamel paste, the whole medallion being sometimes executed in this material; while in others the head only appeared in enamel, against a background of ground-glass tinted by paper placed behind. At his death, in 1799, Tassie's works numbered about 20,000 pieces.

His nephew, William Tassie (1777–1860), also a gem-engraver and modeller, succeeded to the business. His portrait of Pitt, in particular, was very popular, and circulated widely. He bequeathed to the Board of Manufactures, Edinburgh, a valuable collection of casts and medallions by his uncle and himself.

See Gray's *James and William Tassie,* 1895.

TASSO, TORQUATO (1544–1595), the greatest Italian poet of the late Renaissance and author of the epic *Gerusalemme liberata,* was born in Sorrento on March 11, 1544, the son of Bernardo, also a poet and author of the long and romantic *Amadigi* (1560), and Porzia de' Rossi. His childhood was overshadowed by family misfortunes and when, in 1552, Bernardo was banned from the kingdom of Naples, Torquato followed him into exile. Four years later, while father and son were staying in Rome, his mother died. His father's exile and the somewhat mysterious death of his mother struck the imagination of the young Torquato. In 1557 he went with his father to Pesaro and Urbino and there he had his first experience of court life. In 1559 they were in Venice and in 1560 Torquato went to Padua to study law, philosophy and rhetoric, a study which he continued in Bologna between 1563 and 1564, and then again in Padua until 1565. In Padua he met the philosopher Sperone Speroni and became aware of the subtleties of the classicist theory of poetry. He became a friend of Scipione Gonzaga, the future cardinal, and entered the Academy of the Eterei which Gonzaga had founded. In Padua he also composed his first epic, the *Rinaldo* (publ. 1562), and his lyrical poems to Lucrezia Bendidio and Laura Peperara.

Toward the end of 1565 Tasso entered the service of Luigi Cardinal d'Este and went to live in Ferrara, where, as a young poet admired by the court, he passed the happiest period of his life. In the Ferrarese academy he expounded his *Conclusioni d'Amore* (publ. 1581) according to the Renaissance tradition of love treatises. From Paris, where he went with the cardinal in 1570, he wrote an interesting letter which shows him as a subtle political observer. In this period he composed his complimentary verses for the duke's sisters, Lucrezia and Leonora d'Este, and his love lyrics for Laura Peperara and Eleonora Sanvitale. He associated with men of letters such as Guarini and Giambattista Pigna, and participated in the life of the court—his pastoral drama

Aminta (publ. 1580) was performed on the island of Belvedere in the summer of 1573. In 1574 he began his tragedy *Galealto re di Norvegia* and in April 1575 he completed the *Gerusalemme liberata* which he had begun a few years earlier. In Jan. 1572 he had entered the service of Duke Alfonso II, with a high salary and without heavy duties.

During the last months of 1575, however, this happy period of his life began to be overshadowed by the psychological disturbances which tormented him for the rest of his life. Before publishing the *Gerusalemme*, he asked many people for advice, including Gonzaga, the theologian Silvio Antoniano, Speroni and the versifier Pietro Angeli da Barga. The critical observations for which he had asked humiliated and irritated him, and he was deeply concerned with most of the moral and poetical problems raised by his advisers. At this time Tasso developed an acute egotism accompanied by a persecution mania, which showed itself in such episodes as his throwing a knife at a servant whom he suspected of spying on him, and in his frequent, restless travels. From Sorrento, where he stayed with his sister in 1577, he went to Rome, and thence, in April 1578, again to Ferrara. At the beginning of July he moved to Mantua, and then to Padua, Venice, Pesaro and, finally, to Turin, as a guest of the marquis Filippo d'Este. In Feb. 1579 he returned to Ferrara during the festivities for the wedding of the duke with Margherita Gonzaga and, thinking himself neglected, he flew into a rage. The duke had him arrested, chained and imprisoned in the Ospedale di Sant'Anna, where he remained for seven years. That this was due to his alleged love for Leonora, the duke's sister, must be considered a legend. It is more likely that the real cause of the duke's severity was the danger Tasso represented to the Este court, where, even after the exile of the duke's mother, Renée of Valois, a sympathizer of Calvin, there survived sympathy for Calvinistic reform. Accusations of heresy by Tasso could have helped the pope in his anti-Este policy.

During his seclusion in Sant'Anna the poet remained in touch with the outside world through his correspondence and the many occasional poems which he wrote, and he also composed most of his *Dialoghi* and the *Apologia* in defense of his epic. Finally, in July 1586, thanks to the intervention of Father Angelio Grillo and of Vincenzo Gonzaga, prince of Mantua, Tasso was allowed to leave Sant'Anna.

In Mantua, where he received a warm welcome, he finished his tragedy *Galealto* under the new title *Torrismondo* (publ. 1587). In Oct. 1587 he fled from Mantua and took refuge in Rome, where he spent most of his last years as the guest of Cardinal Gonzaga and then of the cardinals Pietro and Cinzio Aldobrandini, to whom he dedicated in 1593 his recast epic *Gerusalemme conquistata*. He went for short periods to Florence in 1590, to Mantua in 1591 and to Naples, where he hoped to recover his mother's dowry, in 1588, 1592 and 1594. In Naples he was the guest of G. B. Manso who eventually became the author of one of the first biographies of Tasso, and there he wrote his religious poems, the *Monte Oliveto* (publ. 1605) and the *Sette giornate del mondo creato* (publ. 1600–05), besides a fragment of the *Vita di San Benedetto* and his last dialogues. In Rome he composed *Le lagrime di Maria Vergine* (1593) and *Le lagrime di Gesù Cristo* (1593). During his last years his illness and discontent continued, despite the fact that he was then receiving many honours and proofs of friendship—in 1594, plans were being made for his poetical coronation in Rome. By then he was tired and no longer interested in fame. In March 1595 he fell ill and asked to be transferred to the convent of Sant'Onofrio in Rome, where he died on April 25, 1595.

Tasso's Works.—When Tasso composed his epic *Rinaldo* (publ. 1562), relating the youth of the hero in 12 cantos *in ottave*, he paid little attention to the Aristotelian rules: the poem reveals his technical ability, but not as yet his poetical genius. In comparison the *Aminta*, composed ten years later, is already a masterpiece. This pastoral drama reflects the ephemeral period of happiness which Tasso enjoyed at the Ferrarese court. Technically it stands halfway between the eclogue in dialogue form and the melodrama, and is full of literary reminiscences of Theocritus, Moschus, Achilles Tatius, Virgil, Ovid, Propertius, Claudian and Nemesianus, as well as of Petrarch, Politian, Jacopo Sannazaro, Ariosto, Pietro Bembo and others. The allusions to the court are of great delicacy and the result is a work of extreme literary preciosity. The *Aminta* has been defined as the idealization of court life. The inspiration of the poet is lyrical rather than dramatic and in this respect the work ranks with his sonnets, *canzoni* and madrigals.

The *Rinaldo*, the *Aminta* and the *Rime* (publ. 1581–87) are a prelude to the composition of the *Gerusalemme liberata*. The idea of an epic poem on the conquest of Jerusalem arose in Tasso's mind at a very early date (the first drafts belong to his Venetian and Paduan period), and grew during his stay at Ferrara; the work was completed in 1575. The subject and structure of the poem offered him the possibility of expressing his complex nature, while responding to a taste that was typical of his time. The *Liberata* resumes and develops a Renaissance literary tradition: the desire to renew the classical epic and to reconcile it with the new trends of the Counter-Reformation. While engaged in composing the *Liberata*, Tasso wrote his *Discorsi dell'arte poetica* (rewritten later as *Discorsi del poema eroico*, publ. 1594), in which he expounded his qualified acceptance of the Aristotelian rules (for instance, unity of action was not to exclude a variety of episodes) while in fact upholding the rights of poetry. In the *Liberata*, Tasso narrates the actions of the Christian army led by Geoffrey of Bouillon during the last months of the first crusade, culminating in the conquest of Jerusalem and the battle of Ascalon. To the principal action he adds a number of episodes, including (in deference to the d'Este family) the story of Rinaldo d'Este, his rebellion, his love for the pagan Armida and his repentance and decisive participation in the final battle. Tasso also adds the story of the love of the Italian Tancred for the beautiful Saracen Clorinda, whom Tancred kills without recognizing her; the secret passion of Erminia, princess of Antioch, for Tancred; and the intervention of supernatural forces in favour of Aladino, king of Jerusalem. These episodes give him the opportunity of expressing the different aspects of his inspiration but his lyricism predominates even in his epic work. The language is lyrical in the Petrarchan tradition, and lyrical, too, is the rhythm of his hendecasyllables and of his *ottave*.

The poetical novelty of the *Liberata* aroused violent criticism, to which Tasso replied in 1585 with his *Apologia*. Tasso's prose works (*Dialoghi* and *Lettere*) show his ability as a stylist who, while following classical models, achieves a modern form in which classical and popular patterns are admirably blended. The works of his last years have no poetical significance: the *Conquistata*, in which he tries to rewrite his epic according to strict moral and rhetorical preconceptions, is a document of the poet's decline, while the tragedy *Torrismondo* shows passages of real poetry only where the poet allows his lyrical inspiration to express itself.

Tasso's life soon became the subject of a legend concerning his alleged love for the princess Leonora and his mysterious madness. The latter is described as simulated in Alessandro Guarini's *Farnetico savio* (1610), while Tasso's love affairs are narrated in the early biographies by G. P. D'Alessandro (1604) and G. B. Manso (1621). This legendary tradition lasted until the 19th century, in spite of the doubts already expressed in the 18th century by G. Tiraboschi and P. A. Serassi. During the romantic period Tasso's life became a symbol of the poet's life and the subject of novels and dramas (of which the most famous is Goethe's *Torquato Tasso*, 1790). Later criticism, while rejecting legendary details of Tasso's life, concentrated on determining the pathological background of his genius and on underlining the moral weakness of the man. Twentieth-century criticism tends to see him as the tragic result of the civilization of the Renaissance.

Standard editions of Tasso's works are: *Le Rime*, ed. by A. Solerti (1898–1902); *La Gerusalemme liberata*, ed. by A. Solerti (1895–96), L. Caretti (1958); *Aminta*, ed. by B. T. Sozzi (1957); *La Gerusalemme Conquistata*, ed. by L. Bonfigli, 2 vol. (1934); *Dialoghi*, crit. ed. by E. Raimondi (1958); *Le Lettere* (1852–55) and *Prose diverse* (1875) ed. by C. Guasti; and *Appendice alle opere in prosa*, ed. by A. Solerti (1892). English editions include *Godfrey of Bulloigne, or the Recoverie of Hierusalem*, trans. by

E. Fairfax (1600); *Jerusalem Delivered*, trans. by J. K. James, 2 vol. (1884); L. E. Lord, *A Translation of the Orpheus of A. Politian and the Aminta of T. Tasso* (1931).

BIBLIOGRAPHY.—For a full bibliography *see* A. Tortoreto and J. G. Fucilla, *Bibliografia analitica tassiana, 1896–1930* (1935); A. Tortoreto, "Nuovi studi su T. Tasso (Bibliografia analitica: 1931–1945)," in *Aevum*, XX, pp. 14–72 (1946); and more recent, Tortoreto's annual contributions in the periodical *Studi Tassiani*. *See* also A. Solerti, *Vita di T. Tasso*, 3 vol. (1895); E. Donadoni, *T. Tasso: Saggio critico*, 2 vol. (1921); W. Boulting, *Tasso and his Times* (1907). (G. A.)

TASTE: *see* SMELL AND TASTE.

TATA, a Parsee priestly family, originally from Baroda state, India, has produced distinguished merchants, industrialists and philanthropists. Founder of their greatness was JAMSETJI NASARWANJI TATA (1839–1904), educated at Elphinstone college, Bombay. Joining his father's trading firm in 1858 and aided by a powerful imagination, he entered on an outstanding career which contributed massively to India's industrial development. He organized cotton mills in Bombay and Nagpur; founded the Tata Iron and Steel company, one of the largest integrated steelworks in the world, and planned the use of hydroelectric energy resulting in the formation, after his death, of the Tata Power companies which supply electricity to Bombay city and the surrounding areas. He introduced sericulture into India; founded the Indian Institute of Science in Bangalore; applied the findings of science to the cultivation of cotton and other crops; built the Taj Mahal hotel; and established an endowment for the advanced professional and technical training of Indians abroad. A man of high social ideals, he was a pioneer in his attitude to labour, and devoted an overwhelming proportion of his firm's profits to practical philanthropy.

His sons, SIR DORABJI J. TATA (1859–1932) and SIR RATANJI TATA (1871–1918), inherited both his fortune and his traditions. Under their guidance the firm expanded into a group of associated companies, covering basic industries. Sir Ratanji founded a department of social science and administration at the London School of Economics, and financed certain social studies at the University of London. Sir Dorabji established a trust for international research into leukemia, and endowed a Cancer Research hospital in Bombay.

In 1932 on the death of Sir Dorabji, SIR NAOROJI SAKLATVALA, one of the founder's nephews, became chairman of the group. On his death in 1938, J. R. D. TATA (1904–), whose father R. D. TATA was a cousin and partner of the founder, became chairman. By the late 1950s the group controlled the largest single aggregation of Indian industry, including among its interests not only textiles, steel and electric power, but also agricultural equipment, locomotives, diesel trucks, chemicals, cement, vegetable oils, soap and toilet products, insurance, radios and industrial investment. Eighty-five per cent of the capital of the parent company, Tata Sons Private Limited, was held by the various charitable trusts endowed by the family. (B. SH. S.)

TATARIA (TARTAR), an autonomous soviet socialist republic of the Russian Soviet Federated Socialist Republic, U.S.S.R., was created on May 27, 1920. Area: 26,255 sq.mi. Pop. (1959) 2,850,417 (urban 1,190,308).

Physiography.—The country is situated on the middle Volga and the lower Kama which divide Tataria in three parts: the smaller, on the Volga's right bank, is traversed by the Sviyaga river flowing northward to the Volga; two others are on the left bank of the Volga and are separated by the Kama flowing westward. The Volga itself turns in Tataria from east to south. The whole country is mainly a plain 330 to 600 ft. above sea level; only in the southeast it rises to about 1,000 ft. (highest point 1,246 ft., near Bugulma). The land is cultivable, 35% being black earth and about 40% podsolized soils. Known mineral resources include petroleum (in the southeast), gypsum, coal and phosphorites. The climate is continental, five months of the year having an average temperature below 32° F., with January average of 14°; the summer is hot with July averages of 60°. The rainfall averages 16 in. per annum.

The People.—The Tatars (*q.v.*) are a Turkic people settled in the middle and lower Volga basin. In 1939 the total population of the republic was 2,919,423, including 1,424,667 Tatars (48.8%), while the Russians were 42.9%, the remainder being Chuvash, Mordvinian and other minorities.

Kazan is the capital of Tataria and its population rose from 179,023 in 1926 to 646,806 in 1959. Other important towns are (pop. [1959]) Zelenodolsk (60,000) and Bugulma (60,980).

History.—In the 9th century modern Tataria was part of the eastern Bulgarian khanate that comprised the lands now forming the republics of Bashkiria, Udmurtia, Mari, Chuvashia and Mordvinia. Batu Khan, son of Juji and grandson of Genghis, destroyed this empire in 1236, founding the Golden Horde instead. About 1430 Ulugh Mohammed Khan, an army commander, broke away from the Golden Horde and founded an independent Tatar khanate with Kazan as capital. It prospered for more than a century, but was in turn destroyed in 1552 by Tsar Ivan the Terrible. The Tatars rose against the Russians in 1553–57, 1572 and 1582 but were defeated. They participated in the peasant risings led by I. I. Bolotnikov (1606–07) and by Stenka Razin (1667–71). In 1708 Peter I made Kazan the chief town of a new and vast province (*guberniya*) and in 1718 the headquarters of the Volga-Caspian "admiralty" were installed there. The policy of russification was accompanied by compulsory conversion of Muslims to Orthodoxy and in 1744 alone more than 400 mosques were destroyed. Therefore the Kazan Tatars played an important part in the peasant revolt led by E. I. Pugachev in 1773–75, supplying many of his military commanders, the most prominent being Iderkei Almetyev and Bakhtiar Kankeyev. Under Catherine II a more enlightened policy prevailed: in 1776 Tatars were permitted to be traders and in 1788 some religious freedom was restored and the great muftiate of Ufa organized. In 1804 the Kazan university was founded, but by 1914 only six Tatars had graduated from it. A similar policy applied to the Kazan Medical school founded in 1814: by 1903, of 10,975 graduates only 70 had been Tatars.

However, toward the end of the 19th century a nationally minded Tatar intelligentsia existed. One of them, Ismail Bey Gasprinsky, a Crimean Tatar, in 1883 founded at Bakhchi-Sarai the newspaper *Terciman* ("The Interpreter"). He also founded a model Muslim private primary school and by 1905 there were about 5,000 such schools among Russia's Turkic-speaking populations. The Tatars were active in the Russian 1905 Revolution. Under their leadership, at Nizhni Novgorod (now Gorki), a Muslim league was founded in Aug. 1905. In the first and second *dumas* (1906–07) there were 30 Muslim deputies, but the electoral reform of 1907 reduced their number to 10 in the third *duma*. The idea of racial, linguistic and religious solidarity among all Turkic peoples, called Pan-Turanianism (*q.v.*), became popular among the Tatar youth.

After the Revolution of 1917 the Tatars tried to form a Volga-Ural Tatar state. On March 23, 1918, the Soviets promised to create a Tatar-Bashkir autonomous republic. On Aug. 6, 1918, however, Kazan was seized by the Czechoslovak legion and became a stronghold of the "White" Russian government. It was retaken by the Red army on Sept. 10, 1918, and the Soviet government on May 27, 1920, created a Tatar A.S.S.R. (A separate Bashkir A.S.S.R. had been created on March 23, 1919 [*see* BASHKIR AUTONOMOUS SOVIET SOCIALIST REPUBLIC].)

In 1926 the Latin alphabet was introduced in Tataria and three years later printing of Turkic newspapers and books using Arabic script stopped completely. But on May 5, 1939, the Cyrillic alphabet replaced the Latin and Galei Afzaletdinovich Dinmukhametov (1892–1951), chairman of the presidium of the Tatar supreme soviet, welcomed this step toward strengthening of collaboration between Tatar and Russian cultures. Many of the new generation of Tatar intelligentsia saw no contradiction between Tatar patriotism and loyalty to the Union of Soviet Socialist Republics.

Government.—Tataria has a supreme soviet elected every four years on the basis of one deputy for every 15,000 of the population. It also elects deputies to the supreme soviet of the U.S.S.R. (one for every 300,000 of the population) and 11 deputies to the soviet of nationalities of the U.S.S.R. It has a council of ministers appointed by the Tatar supreme soviet.

Economy.—Between 1913 and 1956 the area of cultivated arable land increased from 2,491,000 to 3,333,000 ha. The area under grain remained constant (about 2,400,000 ha.) but that under potatoes, vegetables and industrial cultures increased considerably. Livestock included cattle, pigs, sheep and horses. Tataria also produces timber. Industrial products include building bricks, leather shoes and gypsum. The petroleum industry made great strides after World War II. Leninogorsk and Almetyevsk are the new oil towns. Almetyevsk is linked with Gorki and Perm by oil pipelines. There is also a Kazan-Gorki natural gas pipeline. In 1956 the republic had 238 mi. of railways, 3,351 mi. of highways and 266 mi. of waterways. (K. Sm.)

TATARS (less correctly Tartars), a name given to nearly 4,500,000 inhabitants of the U.S.S.R. of Turkic origin. The name is apparently derived from *Tata* or *Dada*, a tribe inhabiting northeast Mongolia in the 5th century. Before the Russian Revolution the term Tatars was loosely applied by the Russians to a large part of the Turkic peoples of the empire, including those of Azerbaijan and northern Caucasus. It was even used generically for all Moslems living in Russia. After the Revolution its use became largely confined to certain peoples of Tatar stock living in the Volga region, the Urals and in western Siberia.

The Volga Tatars consist of four groups: (1) the Kazan Tatars, descendants of the Kipchaks; (2) the Mishars; (3) the Kasimov Tatars; and (4) the Astrakhan or Karagash Tatars. The great majority of the Volga Tatars inhabit the Tatar Autonomous Soviet Socialist Republic and the northwestern part of the Bashkir A.S.S.R. The remainder are distributed among other peoples in the Volga region, in the Urals, in western Siberia, Kazakhstan and other central Asian republics. The Siberian Tatars are divided into the Baraba, the Chulym and the Tobolsk Tatars and are mostly settled along the middle Irtysh river and along the Ob. Some Siberian and Caucasian peoples formerly called Tatars are now no longer described, and apparently never described themselves, as such. Many so-called Tatars are to be found in Moscow, in Leningrad, in the Donbas and in the Caucasian oil fields. Tatars formerly made up 23% of the population of the Crimean Tatar A.S.S.R., but they were deported in 1946 for their alleged co-operation with the Germans and are no longer recognized as a people.

The major part of the Volga Tatars had their origin in Turkish-speaking Bulgarian tribes who came from around the Sea of Azov in the 7th and 8th centuries. Together with local tribes they made up the Golden Horde from the 13th century, and from the 15th century they formed the khanate of Kazan which was finally subdued by and annexed by Russia in 1552. The Mishars and Kasimov Tatars broke away from the Golden Horde in the 14th and 15th centuries and settled round the middle course of the river Oka.

The Tatars are Moslems and speak a Turkic language. Owing to their longer association with Russia they are more russianized than any other Moslem people. (*See* also Turks or Turkic Peoples.) (G. E. Wr.)

TATE, (JOHN ORLEY) ALLEN (1899–), U.S. poet and critic known as the author of "Ode to the Confederate Dead" and other intellectual poetry, but whose greatest influence was in criticism, was born in Clarke county, Ky., on Nov. 19, 1899. He was graduated in 1922 from Vanderbilt university, Nashville, Tenn., where he was a pupil of John Crowe Ransom and a member of the group of poets known as "Fugitives" and of the social commentators, "Agrarians" (*see* their *I'll Take My Stand*, 1930, which criticized the current faith in industrial progress).

Tate taught English at Southwestern college, Memphis, Tenn.; North Carolina College for Women, Greensboro; Princeton university; and the University of Minnesota, Minneapolis. His one novel, *The Fathers* (1938), enjoyed a success of esteem. Most of his poems are collected in *Poems, 1922–47*, and his volumes of critical essays include *Reactionary Essays on Poetry and Ideas* (1936), *Reason in Madness* (1941), *On the Limits of Poetry* (1948) and *The Forlorn Demon* (1953). Tate's "Miss Emily and the Bibliographers" challenged the domination in literary studies of the historical school of scholarship; his "Emily Dickinson" adumbrated that poet's fame. The view of the nature of poetry set forth in such essays as "Literature as Knowledge" and "Tension in Poetry" became a commonplace of modern analytic criticism.

See John M. Bradbury, *The Fugitives* (1958). (R. St.)

TATE, SIR HENRY (1st Bart.) (1819–1899), English sugar refiner, philanthropist and founder of the Tate gallery in London, was born at Chorley, Lancashire, on March 11, 1819, the son of a Unitarian minister. Starting life as a grocer's assistant in Liverpool, he eventually became a prosperous sugar refiner. By 1878 he extended his business to London where he built a second sugar refinery, with outstanding success. Among numerous benefactions he built, equipped and endowed the Hahnemann hospital, Liverpool, provided a library for Liverpool University college and another for Manchester college, Oxford, and gave three public libraries to the London districts of Brixton, Streatham and South Lambeth. Tate's offer to the nation of his private collection of 19th-century pictures, together with the money for a building, was accepted in 1892, and the National Gallery of British Art (known as the Tate gallery) was opened at Millbank, London, on July 21, 1897. Two years later, by his further generosity, the accommodation was almost doubled.

Tate became a trustee of the National gallery in 1897 and was made a baronet in the following year. He died at Streatham on Dec. 5, 1899.

See R. H. Blackburn, *Sir Henry Tate* (1940).

TATE, NAHUM (1652–1715), English poet laureate and playwright, adapter of other men's plays, collaborator with Nicholas Brady (*q.v.*) in the famous *New Version of the Psalms* (1696). He was born in Dublin, Ire., graduated from Trinity college there and moved to London. Tate wrote some plays of his own but he is best known for his adaptation of the Elizabethans. His version of Shakespeare's *King Lear*, to which he gave a happy ending in which Cordelia married Edgar, held the stage well into the 19th century.

Tate also wrote the libretto for Purcell's *Dido and Aeneas* (*c.* 1689). Some of his hymns found a lasting place in Protestant worship—"While shepherds watched"; "Through all the changing scenes of life"; and "As pants the hart for cooling streams."

Tate was commissioned by Dryden to write the second part of *Absalom and Achitophel* (1682), although Dryden added the finishing touches (probably including the portraits of Elkanah Settle and Thomas Shadwell) himself.

The best of Tate's own poems is *Panacea, a Poem on Tea* (1700). He succeeded Shadwell as poet laureate in 1692 and died in London on July 30, 1715. The belief that Tate lived until 1716 long caused it to be thought that he was superseded as poet laureate during his lifetime.

TATIAN (Tatianus) (2nd century A.D.), Christian apologist, missionary and heretic, and editor of the *Diatessaron* (*i.e.*, the four Gospels in a continuous narrative). He was born in Mesopotamia of Syrian parentage and educated in Greek learning, in which he became proficient. He was initiated into the Mysteries, though into which is not known, but after this became acquainted with the Old Testament and was converted to Christianity. Tatian then went to Rome, where he was a hearer of Justin Martyr (*q.v.*), and together with the latter incurred the enmity of a certain philosopher Crescens.

After the death of Justin, Tatian became a heretic. Among his pupils were Rhodon, and, perhaps, Apelles and Clement of Alexandria. He made a missionary journey to the east and worked in Cilicia and Pisidia, using Antioch in Syria as the centre of his efforts.

The heresy that Tatian either founded or adopted was the Gnostic heresy of the Encratites. Their main doctrines were the evil nature of matter, an absolute forbidding of marriage, abstinence from wine and perhaps from meat. It would also seem that Tatian believed in the existence of aeons, one of whom was the Demiurge of the world. He denied the salvation of Adam. It is also stated that in his celebration of the Mysteries (*i.e.*, the Eucharist), he used only water.

According to Eusebius, Tatian wrote many books; of these the names of the following survived: *On Animals, On Demons, Ad-*

dress to the Greeks, Book of Problems, an attempt to deal with contradictions found in the Bible, *Against Those Who Have Discussed Divine Things, On Perfection According to the Saviour,* the *Diatessaron,* and a recension of the Pauline epistles. Of these books only two—the *Diatessaron* and the *Address to the Greeks*—are extant. The *Address to the Greeks* belongs to Tatian's Catholic period. He has the double purpose in view of exposing the weakness of the pagan view of the universe and of commending the Christian explanation. The omissions in the *Address* are even more remarkable than its statements.

There is at the most not more than an allusion to Christ, who is never mentioned by name, and though there are frequent allusions to the regaining of life, which is accomplished by union with the Logos, there is no reference to the doctrines of the incarnation or of the atonement.

Editions of the *Address to the Greeks* are found in O. von Gebhardt and A. Harnack, *Texte und Untersuchungen zur Geschichte der altchristlichen Literatur,* vol. iv (1888), and E. J. Goodspeed, *Die ältesten Apologeten,* pp. 266–305 (1914); an English translation by B. P. Pratten was published in the "Ante-Nicene Christian Library," vol. iii, pp. 3–48 (1867). An English translation of the Arabic text of the *Diatessaron* was published by J. H. Hill as *The Earliest Life of Christ Ever Compiled from the Four Gospels* (1894).

Bibliography.—Johannes Quasten, *Patrology,* vol. i, pp. 220–228, with full bibliography (1950); Robert M. Grant, "The Date of Tatian's Oration," *Harvard Theological Review,* 46:99–101 (1953), and "The Heresy of Tatian," *Journal of Theological Studies,* N.S., 5:62–68 (1954); Arthur Vööbus, *History of Asceticism in the Syrian Orient,* vol. i, pp. 31–45 (1958). (B. M. M.)

TATRA MOUNTAINS, a name usually reserved for the highest group of the central Carpathians (the Slovak Vysoké Tatry), rising abruptly as an isolated igneous group, on the border between Poland and Czechoslovakia, from a high plateau of about 2,600 ft. to altitudes exceeding 8,400 ft.; it extends between the Váh, Poprad, Dunajec and Arva rivers.

The mountains have a west–east length of about 30 mi. and a width varying from 9 to 15 mi.; Gerlachovka (8,711 ft.), is the highest peak in the Carpathians. In everything except glaciers the range bears the Alpine imprint. Hotels, summer resorts and centres for winter sports exist. Railways lead to the most interesting districts.

South of the valley of the Váh runs a lower parallel range of similar structure, the Nízské Tatry (Low Tatra), reaching its maximum heights in the peaks of Dumbier (6,703 ft.) and Králová Hola (6,391 ft.). (W. S. L.; X.)

TATTA (Thatta; formerly also Thato), an ancient town, with modern district, now in the Hyderabad division of West Pakistan. The town is 7 mi. from the right bank of the Indus and 50 mi. E. of Karachi; pop. (1961) 12,786. Tatta was the capital of the Samma dynasty in lower Sind in the 16th century. An English factory was established there in 1758, but soon withdrawn.

Tatta District (6,933 sq.mi.) comprises the pre-partition Karachi district, less the former Federal capital area (the present Karachi district). It occupies the Indus delta, with a tongue of land in the highlands north of Karachi. Pop. (1961) 361,733.

TATTOOING is the introduction of pigment through ruptures in the skin. Sometimes the term is loosely applied also to scarification (*see* Mutilations and Deformations; Branding). Tattooing proper has a nearly world-wide distribution, except that it is rare among populations with the darkest skin colour, and was absent from most of China (at least in recent centuries). Among various peoples tattooed designs are thought to provide magical protection against sickness or misfortune, or serve to identify the wearer's rank, status or membership in a group. Decoration is perhaps the commonest motive for tattooing.

Tattooing occurs on Egyptian mummies dating from about 2000 B.C., and is mentioned by classical authors for the Thracians, Greeks, Gauls, ancient Germans and ancient Britons; the Romans tattooed criminals and slaves. After the advent of Christianity tattooing was forbidden in Europe, but persisted in the near east and in other parts of the world.

In the Americas many Indian tribes customarily tattooed the body or the face, or both. The usual technique was simple pricking, but some California tribes introduced colour into scratches, and many tribes of the arctic and subarctic, most Eskimos and some peoples of eastern Siberia, made needle punctures through which a thread coated with pigment (usually soot) was drawn underneath the skin. In Polynesia, Micronesia and parts of Malaysia pigment was pricked into the skin by tapping on an implement shaped like a miniature rake. In moko, a type of Maori (*q.v.*) tattooing, shallow coloured grooves in complex curvilinear designs were produced on the face by striking a miniature bone adze into the skin. In Japan needles set in a wooden handle are used to tattoo very elaborate multicoloured designs often covering much of the body. Burmese tattooing is done with a brass penlike implement with a slit point and a weight on the upper end. Sometimes pigment is rubbed into knife slashes (*e.g.,* in Tunisia, among the Ainu of Japan, Ibo of Nigeria and Chontal Indians of Mexico); or the skin is punctured with thorns (Pima Indians of Arizona and Senoi of Malaya) or other simple needlelike implements.

Tattooing was rediscovered by Europeans when the age of exploration brought them into contact with American Indians and Polynesians. The word tattoo itself was introduced into English and other European languages from Tahiti, where it was first recorded by James Cook's expedition in 1769. Tattooed Indians and Polynesians, and later Europeans tattooed abroad, attracted much interest at exhibits, fairs and circuses in Europe and the U.S. during the 18th and 19th centuries. Stimulated by Polynesian and Japanese examples, tattooing "parlours" where specialized "professors" applied designs on Euroamerican sailors sprang up in port cities all over the world. The first electric tattooing implement was patented in the U.S. in 1891. The U.S. became a centre of influence in tattoo designs, especially with the spread of U.S. tattooers' pattern sheets.

The nautical, military, patriotic, romantic and religious motifs are now very similar in style and subject matter throughout the world; characteristic national styles of the early 20th century have disappeared.

In the 19th century released U.S. convicts and British army deserters were identified by tattoos, and later the inmates of Siberian prisons and Nazi concentration camps were similarly marked. Tattooing has long been especially characteristic of European and U.S. sailors. Before World War I tattooed symbols of other occupations were very common in continental Europe and sometimes occurred in England. Members of criminal gangs still frequently identify themselves with a simple tattooed device. During the late 19th century tattooing had a short vogue among both sexes in the English upper classes. In the 1960s, except for Euroamerican and Japanese types, and for special medical applications, tattooing is moribund or extinct in most parts of the world.

There are sometimes religious objections to the practice ("You shall not make any cuttings in your flesh on account of the dead or tattoo any marks upon you . . . ," Lev. xix, 28). Tattooing has been implicated in such disorders as skin cancer, and in 1961 the practice was sharply restricted by the New York city government because of the role of contaminated tattooing equipment in spreading hepatitis (*see* Gall Bladder, Biliary Tract and Liver, Diseases of).

Bibliography.—W. D. Hambly, *The History of Tattooing and Its Significance* (1927); H. Ebensten, *Pierced Hearts and True Love,* an Illustrated History of the Origin and Development of European Tattooing and a Survey of Its Present State (1953). (W. C. Sr.)

TATUM, EDWARD LAWRIE (1909–), U.S. biochemist and Nobel prize winner, was born in Boulder, Colo., on Dec. 14, 1909. He was educated at The University of Chicago and the University of Wisconsin (A.B., 1931; M.S., 1932; Ph.D., 1934). His Ph.D. thesis dealt with the nutrition and biochemistry of bacteria. Following a year of postdoctoral study in the Netherlands, Tatum went to Stanford university, where he collaborated with George Wells Beadle (*q.v.*) on research into the biochemical aspects of heredity. This work was carried out on the common fruitfly, *Drosophila,* and especially the red bread mold, *Neurospora,*

but the principles derived from it are of wide application in biology. One such application was to the study of the genetics of bacteria, which now became possible. Modern bacterial genetics was initiated by Tatum and Joshua Lederberg (*q.v.*). at Yale university in 1946. In 1957 Tatum went to the Rockefeller institute as a member and professor. In 1958 he shared the Nobel prize with Beadle and Lederberg. (N. H. H.)

TA-T'UNG, an industrial city of northern Shansi province, China, 165 mi. W. of Peking, at the junction of the Peking-Pao-t'ou and Ta-t'ung-P'u-chou railroads. Ta-t'ung is a major coal-mining centre, with cement plants, chemical factories and railroad repair shops.

The coal mines are situated in K'ou-ch'üan, a southwestern suburb. Coal is shipped primarily to the iron and steel mills of Pao-t'ou and T'ai-yüan (*qq.v.*).

The caves of Yün-kang, 10 mi. W. of Ta-t'ung, contain large Buddha figures dating from antiquity. Pop. (1953) 228,500. (T. Sd.)

TAUBER, RICHARD (1892–1948). British lyric tenor, was born at Linz, Aus., on May 16, 1892. He studied voice at Freiburg in Baden, conducting at Frankfurt am Main and made his operatic debut as Tamino in Mozart's *Zauberflöte* at the Neues Stadttheater in Chemnitz on March 2, 1913. He later sang for the state operas in Dresden, Vienna and Berlin, winning a reputation as a specialist in Mozart. After 1925 he appeared increasingly in light operas. He first appeared in England in 1931 and became a British subject in 1940. He toured the U.S., Australia and Africa as a singer and orchestra leader; he also appeared in and wrote music for motion pictures. His musical compositions include *Der Singende Traum* (1934) and *Old Chelsea* (1943). He died in London on Jan. 8, 1948.

TAUCHNITZ, the name of a family of German publishers and printers.

Karl Christoph Traugott Tauchnitz (1761–1836) established a printing business in Leipzig in 1796 and a publishing house in 1798. He published dictionaries, Bibles and stereotyped editions of Greek and Roman classics. The business was carried on by his son, Karl Christian Philipp (1798–1884), until it was sold in 1865.

More important, however, was the firm of Bernhard Tauchnitz, organized in Leipzig in 1837 by Karl Christoph's nephew Christian Bernhard (1816–1895; created Freiherr von Tauchnitz in 1860). The firm made its name chiefly through its "Collection of British and American Authors," begun in 1841. The "Tauchnitz edition" was popular in Germany, abroad and among English travelers, making available cheap paperbacked copies of books only otherwise accessible in expensive editions. More than 5,000 titles had been published by World War II.

At the end of the war the firm was re-established in Stuttgart and a new series of the "Authors" was begun in 1952. (M. Kl.)

TAULER, JOHANN (*c.* 1300–1361), German mystic, was born about the year 1300 in Strasbourg, and was educated at the Dominican convent in that city. From Strasbourg he went to the Dominican college of Cologne, and perhaps to St. James's college, Paris, ultimately returning to Strasbourg. In 1338–39 Tauler was in Basel, then the headquarters of the "Friends of God" (*see* Mysticism), and was brought into intimate relations with the members of that pious mystical fellowship. He died on June 16, 1361.

Tauler's sermons were printed first at Leipzig in 1498, and reprinted with additions from Meister Eckhart and others at Basel (1522) and at Cologne (1543). A modern edition was published by Julius Hamberger (Frankfurt, 1864), and A. W. Hutton edited Tauler's Sermons for Festivals under the title of *The Inner Way*. *See* H. Denifle, *Das Buch von geistlicher Armuth* (Strassburg, 1877); Carl Schmidt, *Johann Tauler von Strassburg* (Hamburg, 1841); S. Winkworth, *Tauler's Life and Sermons* (1857); R. A. Vaughan, *Hours with the Mystics*, 3rd ed., vol. i, pp. 214–307; Preger's *Gesch. der deutschen Mystik im Mittelalter*, vol. iii; W. R. Inge, *Christian Mysticism;* R. M. Jones, *Studies in Mystical Religion* (1909).

TAUNTON, a market town and municipal borough (1877) and the county town of Somerset, Eng., in the Taunton parliamentary division, 48 mi. S.W. of Bath by road. Pop. (1961) 33,178. Area 3.8 sq.mi. It lies on the river Tone in the Vale of Taunton Deane, sheltered on three sides by the Quantock, Blackdown and Brendon hills. The town was founded by the Saxon king, Ine, who built the first castle there about A.D. 710. The bishops of Winchester, who owned the manor, obtained their first charter from Edward the Elder in 904. Bishop Giffard of Winchester built the second castle early in the 12th century and founded a priory of Augustinian canons in 1127. During the Great Rebellion the castle was thrice besieged by royalists, before Sir Thomas Fairfax relieved the town in 1645, and at the Restoration Charles II deprived Taunton of its charter of incorporation. The castle was dismantled, apart from its gateway and the great hall where Judge Jeffrey's "Bloody Assize" took place after the failure of the Monmouth rebellion at the battle of Sedgemoor in 1685.

Taunton, in the diocese of Bath and Wells, is the seat of a suffragan bishopric. The churches of St. Mary Magdalene and St. James are Perpendicular in style. Huish's grammar school shares in the endowment of Robert Huish, who died in 1615; Taunton, King's college, Queen's college and Wellington are local public schools; in the town are a technical college and the Somerset College of Art. In Taunton is the Somerset County Cricket club ground. There are hunting and fishing in the district.

The Saturday market, dating from before the Conquest, has become one of the largest livestock and produce markets in the southwest. Shirts and collars and agricultural machinery are manufactured.

TAUNTON, a city of Massachusetts, U.S., and seat of Bristol county, is located 35 mi. S. of Boston at the head of navigation on the Taunton river, 14 mi. above Fall River. The centre of Taunton, the "Green," is about one mile from the Weir, the port of the city. Taunton was settled after purchase of land from the Indians and in 1639 became a town with representatives in the general court of Plymouth colony. The name was derived from the English home of many of the settlers. In the 18th century six towns were detached from the original grant.

Taunton was a base of operations during King Philip's War. In 1687 a town meeting declined to meet a tax levy by Sir Edmund Andros "without their own assent in an assembly." A Liberty pole with a flag inscribed "Liberty and Union" was raised on the "Green" in 1774 to express opposition to the Boston Port bill. During Shays's rebellion in 1786–87 the courthouse was twice besieged by insurgents who were dispersed each time by a judge, Gen. David Cobb.

Ironworks, established in 1656, were operated successfully until 1876; iron bars were sometimes used as a medium of exchange. Brickmaking and shipbuilding were other early industries. Silverware production, begun in 1824, remains important. Modern industries also include jewelry, plastics and electronics.

The nearby Dighton rock has inscriptions long attributed to the Norse, but believed now to be the work of Indians or of a Portuguese explorer. Taunton became a city in 1864. For comparative population figures *see* table in Massachusetts: *Population.* (Ca. M. C.)

TAUNUS, a wooded mountain range of western Germany in the *Land* (state) of Hesse. It lies between the Rhine and the Main on the south and the Lahn on the north, and is about 47 mi. long. Its southern edge stands 5 to 10 mi. back from the Main, but slopes steeply to the Rhine, and from Bingen downward forms precipitous crags, many of which are crowned with scenic ruins. It has an average elevation of about 2,000 ft., but the loftiest peaks occur in the east, where the Grosser Feldberg (2,887 ft.), Kleiner Feldberg (2,710 ft.) and Altkönig (2,618 ft.) dominate the Wetterau and the valley of the Main. Above the Rheingau (the slopes to the Rhine between Biebrich and Bingen), the altitude averages 1,500 to 1,700 ft. The geological core of the system consists of (?Archean) argillaceous schists, capped by quartzite (Devonian) and broken through in places by basalt. On the northern side, which sinks on the whole gently toward the Lahn, the higher greywacke formation (Devonian) attains a considerable development. The hills are generally well-wooded, with firs and beeches. The lower slopes are, where possible, planted with vineyards, orchards and chestnut and almond groves. The vineyards

of the Rheingau are specially famous, and yield many famous brands of wine. The Taunus is also famous for its mineral springs and health resorts.

Ruins survived in modern times of the castle of Kuno (above Falkenstein); of a fortress at Königstein; a Cistercian abbey at Eberbach; and two concentric lines of pre-Roman fortifications at Altkönig. The chief historical monument, however, is the Saalburg, a Roman fort serving as a centre of communications along the *limes* or fortified frontier line drawn from Rhine to Main by Domitian (*see* LIMES).

At Niederwald the gigantic "Germania" statue was erected in commemoration of the war of 1870. The steep crags of the western end of the Taunus, where they abut upon the Rhine, are rich in the romantic associations of the great river. There are the rock of the siren Lorelei; the old castles of Stahleck and Pfalz; and the quaint mediaeval towns of Kaub and St. Goarshausen.

TAURI, the earliest known inhabitants of the mountainous south coast of the Crimea (Herodotus iv, 103). Nothing is certain as to their affinities. They probably represent an old population perhaps connected with some Caucasus stock. They are not likely to be Celts. They were famous in the ancient world for their maiden goddess, identified by the Greeks with Artemis Tauropolos or Iphigenia (*q.v.*), whom the goddess was said to have brought to her shrine at the moment when she was to have been sacrificed at Aulis. Orestes sought his sister, and almost fell a victim to the Tauric custom of sacrificing to the maiden shipwrecked strangers, a real custom which was the ground of the whole myth. His adventures were the subject of plays by Euripides and Goethe. Towards the end of the 2nd century B.C. we find the Tauri dependent allies of the Scythian king Scilurus, who from their harbour of Symbolon Portus or Palacium (Balaclava) harassed Chersonese (*q.v.*). Their later history is unknown. (E. H. M.)

TAUROBOLIUM, properly the shooting of a bull, hence a rite of baptism by bull's blood, usually in connection with the worship of the ancient goddess Cybele, though not limited to it. Of oriental origin, its first known performance in Italy was in A.D. 134. Prudentius describes it in *Peristephanon:* the person undergoing the ceremony, clad in a toga and with golden crown and fillets on his head, takes his place in a trench covered by a grating on which a bull is slain.

The taurobolium was often performed as a measure for the welfare of the emperor, empire or community, its date frequently being March 24, the *dies sanguinis* of the annual festival of the Great Mother of the Gods and Attis (*qq.v.*). But its usual motive was the purification or regeneration of an individual, who was occasionally spoken of as "reborn for eternity" in consequence of the ceremony; more commonly the effect was considered to endure for 20 years.

TAURUS ("the Bull"), in astronomy, the second sign of the zodiac. It is a constellation of very great antiquity, the Pleiades and Hyades (*qq.v.*), two of its constituent star clusters, being possibly referred to in the Old Testament; Aldebaran, its brightest star, is mentioned by Hesiod and Homer. The Greeks fabled this constellation to be the bull which bore Europa across the seas to Crete, and was afterward raised to the heavens by Jupiter. Aldebaran is the principal object in the Hyades. Many of the Hyades (but not including Aldebaran) form an associated group of stars with a common motion. The measurement of the motions of these stars permits an accurate determination of their distance; they are situated at approximately 130 light years from the sun.

Aldebaran itself is a reddish giant star, located at a distance of about 80 light years from the sun. The constellation contains the "crab nebula," a gaseous mass which has been shown fairly conclusively to have originated in the outburst of the new star, or nova, of A.D. 1054. Photographs indicate that the nebula is expanding in size from year to year. The structure of the Milky Way in Taurus is complex, with dark, almost starless lanes and bright nebulosity.

TAUSEN, HANS (1494–1561), the protagonist of the Danish Reformation, was born at Birkende in Funen in 1494. The quick-witted peasant lad ran away from the plow at an early age, finally settling down as a friar in the Johannite cloister of Antvorskov near Slagelse. After teaching for a time at Rostock and Copenhagen, he was again sent abroad by his prior, visiting the Netherlands and making the acquaintance of the Dutch humanists. He was already a good linguist, understanding both Latin and Hebrew. Later he translated the books of Moses from the original. In May 1523 Tausen went to Wittenberg, where he studied for a year and a half, when he was recalled to Antvorskov. In consequence of his adherence to Lutheran doctrines he was first imprisoned in the dungeons of Antvorskov and thence transferred, in 1525, to the Grey Friars' cloister at Viborg in Jutland, where he preached from his prison to the people assembled outside, till his prior, whom he won over to his views, lent him the pulpit of the priory church.

Several young men in Viborg had studied at Wittenberg, and the burghers had already expelled their youthful bishop Jörgen Friis. Tausen no longer felt safe among the Franciscans; he discarded his monastic habit, and placed himself under the protection of the burgesses of Viborg. At first he preached in the parish church of St. John, then in the market-place from the church tower. When the Franciscans refused to allow him to preach in their large church, the mob broke in by force. A compromise was at last arranged, whereby the friars were to preach in the forenoon and Tausen in the afternoon. The bishop sent armed men to the church to arrest Tausen, but the burghers drove them back. In Oct. 1526 King Frederick I made Hans Tausen one of his chaplains, and charged him to continue "to preach the holy Gospel" to the citizens of Viborg, who were to be responsible for his safety. Tausen found a fellow worker in Jörgen Viberg, better known as Sadolin, whose sister, Dorothea, he married. He was the first Danish priest to marry. He was also the first to use Danish instead of Latin in the church services.

On the death of King Frederick, Tausen, at the instance of Rönne, was, at the *Herredag* of 1533, convicted of blasphemy and condemned to expulsion from the diocese of Sjaelland, whereupon the mob rose in arms against the bishop, who would have been murdered but for Tausen's intervention. Rönne then permitted Tausen to preach in all his churches on condition that he moderate his tone. On the final triumph of the Reformation Tausen was appointed bishop of Ribe (1542), an office which he held for 20 years.

TAUSSIG, FRANK WILLIAM (1859–1940), U.S. economist, whose outstanding contributions were in the field of international trade, was born at St. Louis, Mo., on Dec. 28, 1859. He was trained at Harvard university (A.B., 1879; Ph.D., 1883) and taught there for more than half a century. His influence was considerably extended by his editorship of the *Quarterly Journal of Economics* (1896–1936), by his widely translated *Principles of Economics* (2 vol., 1911) and by his public services in a variety of fields, from tariffs to industrial relations. Taussig sought to modernize the classical tradition of economics as formulated by John Stuart Mill. In his major theoretical study, *Wages and Capital* (1896), he substantially revised the old wages-fund doctrine by making it more flexible. In *International Trade* (1927) he developed most elaborately the traditional analysis to handle such phenomena as intergovernmental debts and the mechanism of international trade under nonmetallic monetary standards. In this area of international economics, Taussig rendered one of his most notable public services. He was appointed by Pres. Woodrow Wilson to the chairmanship of the new United States Tariff commission (1917–19), and he helped to prepare the tariff and commercial policy clauses of the treaty of Versailles (1919).

Taussig died in Cambridge, Mass., on Nov. 11, 1940.

See for bibliography, Jacob Viner *et al., Explorations in Economics: Notes and Essays Contributed in Honor of F. W. Taussig* (1936).
(J. DN.)

TAUTOLOGY. The notion of a tautology in the propositional calculus, explained in the article LOGIC, is due to C. S. Peirce (1885). The name tautology was introduced in 1921 by Ludwig Wittgenstein, whose philosophical use of the notion of tautology (*see* LOGICAL POSITIVISM) requires the extension of this notion, from the propositional calculus, to the full theory of types in-

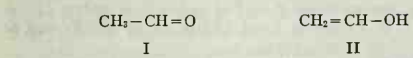
cluding the functional calculi of first and higher orders. This extended notion of tautology, explained somewhat cryptically by Wittgenstein and more satisfactorily by F. P. Ramsey in 1926, is in fact a less precise forerunner of what is now more usually known as *validity—see*, in the article LOGIC, the definition of *validity* in the functional calculus of first order, and, in SEMANTICS IN LOGIC, the definition of *satisfaction* on which this depends. Later logical positivists, especially Rudolf Carnap, amended Wittgenstein's doctrine in the light of the distinction that there is an effective test of tautology in the propositional calculus but no such test of validity even in first-order functional calculus. (Ao. C.)

TAUTOMERISM. A number of apparently pure substances have been discovered which behave as if they possess two or more distinct structures. In those instances in which the structures in question correspond to identical, or nearly identical, relative positions in space of all constituent atoms (*see* STEREOCHEMISTRY), this phenomenon is at present considered to find its explanation in the theory of resonance (*see* RESONANCE). In those other instances, however, in which the structures in question correspond to widely different relative positions of at least one atom, the phenomenon is explained in a different manner, and the term "tautomerism" has been coined for its description. The present article is concerned only with tautomerism in this sense, and no further mention of resonance will be made here.

It is commonly believed that any compound which exhibits tautomerism is an equilibrium mixture of two or more substances, and that these substances have the structures indicated by the behaviour of the mixture. A tautomeric mixture then must exhibit the reactions to be expected of all structures involved since all the corresponding substances are actually present. On this view, the reason why a tautomeric mixture appears to be a single pure substance is that its various components (or tautomerides, as they are often called) are capable of rapid interconversion; consequently, they are separable from one another only with considerable difficulty, if at all. Moreover, it usually happens that only one of the tautomerides is present to an appreciable extent in the equilibrium mixture.

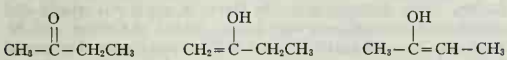
An alternative explanation of tautomerism, which may be correct in some instances, is that all the molecules present have the same structure but are able to react in two or more different ways. This further possibility need not, however, be discussed here.

A typical example of tautomerism is provided by acetaldehyde, which behaves in most respects as if it has the stable, so-called keto structure I, but in other respects as if it has instead the unstable, so-called enol structure II. This compound has never been separated into its two assumed components,



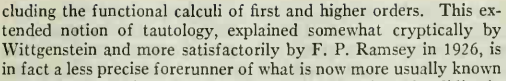
I II

and all attempts to prepare a substance (vinyl alcohol) with the structure II have led instead to ordinary acetaldehyde. The belief is universally held that acetaldehyde is essentially pure keto (I), and that only minute traces of the enol form (II) are present in equilibrium. For this reason, acetaldehyde is usually said to possess structure I. In methyl ethyl ketone, similarly, traces of the two enol forms, IV and V are supposed to be present in addition to the stable keto form III.



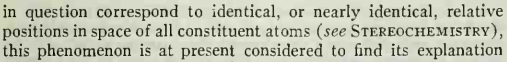
III IV V

The above interpretation of tautomerism would seem highly arbitrary and not at all convincing if no tautomeric compounds were known which can actually be separated into the tautomerides. Fortunately, a few such compounds have been discovered. Of these, the one which has received the most careful study is doubtless acetoacetic ester (ethyl acetoacetate). On the one hand, this substance appears to be a typical ketone, as in VI, since it forms an oxime with hydroxylamine, a hydrazone with hydrazine, and a cyanohydrin with hydrogen cyanide (*see* ALDEHYDES AND KETONES).



VI VII

Moreover, when treated with an excess of both sodium ethoxide and methyl iodide in alcoholic solution, it is transformed into a dimethyl derivative to which the structure VII can be definitely assigned. On the other hand, it appears to be a typical unsaturated alcohol, as in VIII, since it absorbs bromine very rapidly, gives a colour with ferric chloride, and has a pronounced acidic character. Moreover, when treated with diazomethane,

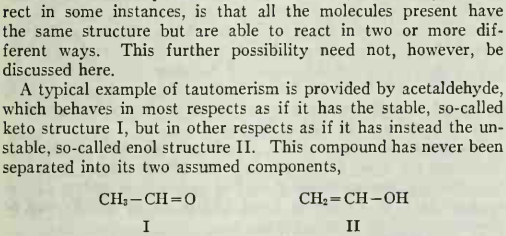


VIII IX

CH_2N_2, it is transformed into a methyl derivative to which the structure IX can be definitely assigned.

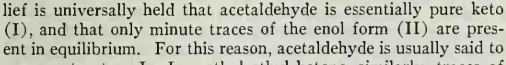
That this compound really is a mixture of two different substances with the structures under discussion was demonstrated in 1911 by L. Knorr, who found that the two components could be separated at −78° C., at which temperature their rate of interconversion is slow. At ordinary temperatures, the equilibrium mixture has been shown (K. H. Meyer, 1911) to consist of about 92.6% keto and 7.4% enol (in the pure liquid ester); these figures were obtained from a titration of the ester with bromine, with which the enol, but not the keto, form reacts rapidly.

A more complicated example of keto-enol tautomerism is presented by diacetylsuccinic ester, which has been obtained as the diketo form X, the keto-enol form XI, and the dienol form XII. (Actually, altogether five different forms of this ester are reported



X XI XII

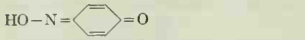
to have been isolated; two of these are apparently stereoisomers of other forms. *See* STEREOCHEMISTRY.)



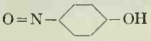
XIII XIV XV XVI

In addition to the above keto-enol tautomerism, a number of other types of tautomerism are also well known. For example, the amides of carboxylic acids (*see* AMIDES) exhibit the so-called lactam-lactim tautomerism. Thus, with acetamide, it is considered that the lactam form XIII predominates, but that the lactim form XIV may be present also in traces. Similarly, with a primary or secondary nitro compound (*see* NITRO COMPOUNDS), the nitro form, as in the structure XV of nitromethane, is considered to predominate greatly over the aci nitro form, as in XVI. A rather different type of tautomerism is illustrated by the fact that the same product results from the action of hydroxylamine upon benzoquinone as from the action of nitrous acid upon phenol. From its first method of preparation, this substance should be the monoxime (*see* OXIMES) of benzoquinone with the structure

XVII, whereas, from its second method of preparation, it should



XVII



XVIII

be *p*-nitrosophenol with the structure XVIII. The oxime form XVII is considered to be the more stable, although with a few analogous compounds both forms are reported to have been obtained.

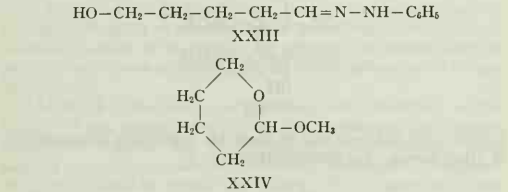
Prototropy and Anionotropy.—In the various types of tautomerism mentioned so far, the tautomeric structures differ in the position of a hydrogen atom (and also in the arrangement of the valence bonds). Most types of tautomerism, of which only a few have been described here, belong to this class; such tautomerism is sometimes designated by the special name of "prototropy" (from "proton," the nucleus of the hydrogen atom, plus the Greek *trepein*, to turn). A few examples are known, however, in which the structures differ in the position of some other atom or group of atoms (as well as in the arrangement of the valence bonds). Thus, the two unsaturated chlorides XIX and XX, and the corresponding alcohols XXI and XXII, are rather

XIX XX

XXI XXII

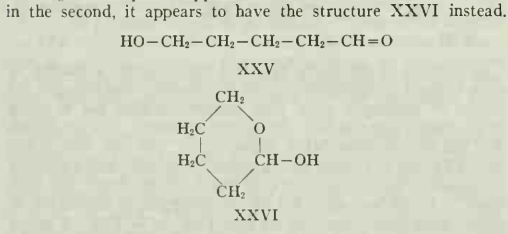
easily interconverted. The action of hydrogen chloride upon either alcohol gives a mixture of the two chlorides, and either chloride readily undergoes a partial spontaneous rearrangement into an equilibrium mixture of the two. Such tautomerism is sometimes referred to as "anionotropy" since the structures differ in the position of an atom or radical that is capable of forming a stable anion.

Ring-Chain Tautomerism.—A further important type of tautomerism can be illustrated with the example of 5-hydroxypentanal, which reacts with phenylhydrazine to give the phenylhydrazone XXIII, but with methyl alcoholic hydrogen chloride to give the cyclic product XXIV. In the first of these reactions,



XXIII XXIV

the original compound appears to have the structure XXV, but, in the second, it appears to have the structure XXVI instead.



XXV XXVI

This type of tautomerism occurs with those hydroxy aldehydes and ketones with which the cyclic form would have either a five or six-membered ring, and it is especially important in the sugar series. *See* CARBOHYDRATES. (G. W. WD.)

TAVERNER, JOHN (*c.* 1495–1548?), English composer, was born in Lincolnshire, possibly at Tattershall, and according to some accounts became organist of Boston church. In Nov. 1526 he was made master of the children at Cardinal college, Oxford (now Christ Church), and also acted as organist. During this period he wrote his church music. Taverner was then convicted of heresy but it appears that he was not sent to prison because his offense was slight. Taverner resigned his post at Cardinal college in April 1530 and some time during this year contributed three songs to Wynkyn de Worde's *Song Book* ("My Harte, my Minde," "Love wyll I" for three voices, and "The bella" for four voices), published in Oct. 1530. Having secured a lucrative post from Thomas Cromwell, he was concerned in the suppression of four friaries at Boston in Aug. 1538, later being given two of them. He died and was buried at Boston, probably in 1548.

A two-part song, "In Women no Season is Rest or Patience," is in the British museum, but, apart from this and the Wynkyn de Worde specimens, only Taverner's church music has been preserved. Of this there is a complete collection, with biographical notes, in the Carnegie edition of *Tudor Church Music*, i, iii. It includes 8 masses; 28 motets (some incomplete); 3 Magnificats and a Te Deum; a Kyrie known as the "Leroy"; and shorter pieces.

TAVERNIER, JEAN BAPTISTE (1605–1689), French traveller, was a pioneer of trade with India. He was born in 1605 at Paris, where his father Gabriel and uncle Melchior, Protestants from Antwerp, were geographers and engravers. He had already travelled much in Europe and was well acquainted with the principal European courts, when he started from Regensburg in 1630 for the Levant with two French fathers. In their company he reached Constantinople early in 1631, where he spent 11 months, and then proceeded by Tokat, Erzerum and Erivan to Persia. His farthest point in this first journey was Isfahan; he returned by Baghdad, Aleppo, Alexandretta, Malta and Italy, and was again in Paris in 1633. In Sept. 1638 he began a second journey (1638–43) by Aleppo to Persia, and thence to India as far as Agra and Golconda. His visit to the court of the Great Mogul and to the diamond mines was connected with the plans realized more fully in his later voyages, in which Tavernier traded in costly jewels and other precious wares, among the greatest princes of the east. The second journey was followed by four others. In his third (1643–49) he went as far as Java and returned by the Cape; in his last three journeys (1651–55, 1657–62, 1663–68) he did not go beyond India. In 1669 he received letters of nobility from Louis XIV and in 1670 purchased the barony of Aubonne, near Geneva.

Tavernier left Paris for Switzerland in 1687, and early in 1689 he passed through Copenhagen on his way to Persia through Muscovy. In that year he died at Moscow.

Tavernier's narratives of his travels are: *Nouvelle relation de l'intérieur du sérail du Grand Seigneur* (Paris, 1675), based on two visits to Constantinople in his first and sixth journeys; *Les six voyages de J. B. Tavernier*, 2 vol. (Paris, 1676) and *Recueil de plusieurs relations* (Paris, 1679).

TAVISTOCK, a market town and agricultural centre in the Tavistock parliamentary division of Devon, England, 15 mi. N. of Plymouth by road. Pop. of urban district (1951) 5,886. Area 3.3 sq.mi. It lies in a deep valley below the western edge of Dartmoor where one of the main routes into Cornwall crosses the Tavy. It was once the centre of an important mining district for tin and copper and is one of the four stannary towns of Devon. Its mediaeval importance owes much to the magnificent Benedictine abbey of St. Mary and St. Rumon founded in 961 by Orgar, earl of Devon. After destruction by the Danes in 997 it was rebuilt, and among its famous abbots were Lyfing, friend of Canute, and Aldred, who crowned Harold II and William and died archbishop of York. At the time of the Conquest it ranked as the wealthiest house in Devon, including the hundred and manor of Tavistock among its possessions. Most of the abbey was demolished at the dissolution and the lands given to the dukes of Bedford; in 1912 the council acquired some of this land from the duke. Remains of the abbey include the gateway which now houses the Tavistock library. The church of St. Eustachius dates from 1318 and possesses a lofty tower supported on four open arches; the "clothworkers' aisle" (1445) witnesses to the importance of the town as a wool centre. Kelly college, near the town, founded by Adm.

B. M. Kelly and opened in 1877, is now a public school for boys.

Henry I granted the abbots a market in 1105 and the famous Tavistock goosey fair in 1106. This takes place annually on the second Wednesday in October. In 1513 the abbots were granted a seat in the house of lords. Tavistock is a borough by prescription and returned two members to parliament from 1295 until 1867; in 1885 it was disfranchised.

TAVOY, a town and district in the Tenasserim division of Burma. The town is on the left bank of the river of the same name, 30 mi. from the sea. Pop. (1953) 40,312. A considerable coasting trade was developed with other ports of Burma and with the former Straits Settlements, which received the bulk of the tin ore for which the district became important. Silk weaving became a chief industry.

The district has an area of 5,390 sq.mi. It lies between Thailand and the Andaman sea of the Bay of Bengal, enclosed by mountains on three sides, viz., the main chain of the Bilauktaung on the east, rising in places to 5,000 ft., which, with its densely wooded spurs, forms an almost impassable barrier between Burmese and Siamese territory; the Nwahlabo in the centre, which takes its name from its loftiest peak (5,000 ft.); and a third range, under the name of Thinmaw, between the Nwahlabo and the sea coast. The Tavoy river is navigable for vessels of any burden. It is interspersed with many islands, and with its numerous smaller tributaries affords easy and rapid communication. The climate is on the whole pleasant. The annual rainfall averages more than 200 in. Pop. (1941) 211,729, showing an increase of 31,765 in the decade. The staple crop is rice. Forests cover an area of nearly 5,000 sq.mi., of which considerable areas are "reserved." Tin mining was developed comparatively recently. Iron ore is also mined.

See also TENASSERIM.

TAXACEAE (yew family), a small family of the Gymnospermae consisting entirely of evergreen shrubs and trees, containing only three small genera: *Taxus*, including the yew (*q.v.*), in Eurasia and various subspecies and varieties in North America; *Torreya* with five species in eastern Asia and North America; and *Austrotaxus*, restricted to New Caledonia. (E. D. ML.)

TAXATION. Taxes are compulsory levies on private units for general governmental purposes. The term taxes does not include governmental receipts from borrowing, gifts, reparations, or various direct charges for special services or benefits; *e.g.*, fees to cover the costs of licensing, public prices like postal rates, and special assessments levied to pay for sidewalks and similar specific improvements to real estate.

Clear-cut lines among the various types of public revenue are hard to draw, and each type shades into adjacent classes. For example, in a twilight zone between taxation and borrowing are the compulsory loans levied by Canada, Great Britain and the United States during World War II (in the form of refundable income taxes) and by Denmark and Israel in the 1950s. Also difficult to classify are social insurance contributions (not generally designated by law as taxes, except in the United States); they progressively take on the character of taxes as the linkage between the size of the individual's contributions and the scale of his benefits is weakened by legislative action. Though generally thought of as an instrument of revenue, taxes are recognized also as a major instrument of national economic policy. In war and in armed peace, nations call on tax increases to subdue inflation, and in depression they grant tax reductions to stimulate private markets and employment. Especially when government spending equals one-fourth to one-third of the national product, as in western Europe and North America, taxes play a crucial role not only as a source of revenue but as a balance wheel in the economy. Throughout the world, taxes are adjusted to promote economic growth.

History.—Only late in human history did governments come to rely on taxes, especially on taxes levied equitably and with the consent of the governed, for the bulk of their revenue. For centuries, the public domain was the chief source of public revenue, and taxation consisted largely of excises on domestic consumption and customs duties on foreign trade. It was the duty of the unprivileged classes—slaves, vassals, peasants, colonists and conquered peoples—to support the ruling classes, be they the free citizens of Athens and Rome, the lords and nobles of the feudal barony or the courts of the Louis. Taxes as a badge of freedom rather than a mark of bondage are a modern phenomenon.

In ancient Greece and Rome, income from mines, tribute from subjugated countries, gifts from wealthy citizens and taxes on trade and consumption filled virtually all revenue needs. Except for Rome's famed inheritance and capitation taxes (both foreshadowed by ancient Egyptian taxes) and certain emergency levies in Greece, direct taxes on income and wealth were unknown.

Taxes of any kind had little place in the agrarian feudal system of the middle ages. The king, the duke, the count, the baron —each in turn derived income from the land he held directly and levied rentlike dues on those who held land at his pleasure, *i.e.*, on those beneath him in the feudal hierarchy. From the 14th century on, as the feudal system was gradually broken up by dissipation of the public domain, by growth of industry and commerce and by centralization of government, these patrimonial revenues slowly gave way to taxes.

Land, as the primary source of wealth, also became the primary source of taxation during this period. Both local and central governments in England turned to land taxes, basing them at first on area (the old *hidage* and *carucage*) but later on annual rental value. The French counterpart was the arbitrary annual *taille* on estimated farm income. The early American colonies broadened the British type of land tax into a property tax whose base included land, houses, personal property and faculty (the rated earning capacity of one's trade or profession). With the great growth of foreign and domestic commerce between the 14th and 19th centuries, import and export duties and internal excises became important, especially in France, Spain and other commercial centres. Poll taxes and such presumptive levies as the hearth tax and window tax also flourished in this period.

Rebellion against arbitrary and oppressive taxation played a major role in modern history. Such milestones in British constitutional history as the Magna Carta (1215) and the Bill of Rights (1689) helped establish the principle of consent and representation in taxation. Failure to apply this principle to the colonies played no little part in the American Revolution. Much of the desperation that exploded in the French Revolution grew out of one of the most oppressive and inequitable tax systems of all time.

The establishment of representative democratic government, the advent of the industrial revolution and the development of modern ideals of social justice and equality left their imprint on both the principles and the forms of taxation. The 19th century saw expediency and oppression in taxation gradually make way for equity and uniformity. With the unfolding of new forms and differentiated sources of wealth, taxes on incomes and inheritances were adopted and refined, and property taxes were further broadened.

Great Britain adopted an income tax in 1798 and a death transfer tax in 1796, broadened both measures (after setting the income tax aside from 1816 to 1842) midway through the 19th century, and brought them to something like their modern form near the turn of the century. Although customs and excises continued to be an important source of revenue, the central government during the first half of the 20th century greatly intensified its use of direct taxes on income and estates. "Local rates" on real estate continued to be the mainstay of local tax systems.

Modern and effective use of income and death taxes was not made in the United States until about the time of World War I. Earlier, customs, excises on liquor and tobacco, and sales of land furnished the bulk of federal revenues. The peculiarly American general property tax provided most of the revenues needed at the state and local levels. But during World War I and in the interwar period, the federal government adopted an estate tax, developed its corporate and individual income taxes, expanded its excises and introduced social security pay roll levies. The states meanwhile tapped such new sources as gasoline taxes, retail sales taxes, motor vehicle levies and net income taxes, largely abandoning the property tax to the local units.

Under the pressure of World War II, central government taxes everywhere were brought to new peaks, both in their severity and

in their adaptation to such nonrevenue ends as curbing of inflation and prevention of war profiteering. After the war, most belligerents granted only modest tax reductions in the face of large debts, continued inflation and deteriorating relations between the Soviet and western worlds. Outbreak of the Korean War in 1950 touched off a new wave of tax increases, but within three or four years, such countries as Australia, New Zealand, the United States, Canada, Great Britain, the Netherlands and Germany were again reducing tax rates.

Modern Forms.—As modern economic society progressively exposed new facets of wealth, income and transactions to tax planners and legislators, they responded with a bewildering array of levies. The production and receipt of income, the ownership and transfer of wealth, the sale and use of commodities and services, the carrying on of a business and a host of other economic factors all became bases or objects of taxation.

The versatility of income as a tax base can be visualized by observing the gantlet of taxes it may have to run as it flows into, through and out of a corporation. At the point of entry, income may be subject to (1) taxes on gross receipts or gross income (often classed as consumption taxes because they are largely passed on to consumers in higher prices); (2) taxes on the net profits remaining after all expenses are deducted; (3) taxes on excessive profits, especially in wartime. Once inside the corporate walls, net income which is not distributed to stockholders may run athwart taxes on undistributed profits like the 1936–38 tax of 7%–28% in the United States or the British standard rate on retained earnings. At the point of exit, income paid out as dividends may encounter a distributed profits tax like the post-World War II British tax of 30%. As it reaches the stockholder, dividend income may be subject to (1) a progressive personal income tax applying a uniform schedule of rates to net income regardless of source (as in prevailing United States practice); (2) a schedular income tax applying different rates to different sources of income (as in Massachusetts); (3) a special tax or surtax on investment income (as in Canada); (4) special credits or allowances under the regular income tax (again, as in Canada). To this galaxy of income taxes should perhaps be added the low-rate levies on business pay rolls and individual wages so widely used to finance old-age and unemployment insurance programs.

Property remains the principal local tax base in the English-speaking world. Broadest of all property taxes is the U.S. general property tax, an impersonal levy on both real estate and personal property, measured by exchange value and levied against owners. Narrower and more personal is the British system of local rates, confined to real estate, measured by annual rental value, and in part levied as a tax on occupants. Even narrower are the rural land taxes of Asia, the middle east and much of Europe.

Most personal of the taxes on property ownership are the progressive taxes on personal net worth (assets minus liabilities) levied in the form of (1) annual net worth taxes, as in Scandinavia, Germany and the Netherlands; and (2) one-time capital levies, as in many European countries in the wake of both World Wars I and II. In the United States, constitutional barriers stand in the way of a national net worth tax.

Transfers of wealth by death are taxed in most western countries under both estate taxes (levied on the decedent's estate as a whole) and inheritance, succession or legacy taxes (levied on the heirs' distributive shares). Comparable taxes apply to transfers by gift.

Another major tax family are the consumption taxes, those which are generally passed on to consumers in higher prices even though initially levied on producers or distributors. The term hidden taxes ordinarily refers to this group of levies. In their more general forms, they are levied as turnover, gross income or value-added taxes at several stages of production; as sales or purchase taxes at the manufacturing, wholesale or retail stages; as use taxes on the consumer; and as tariffs or customs duties on imports and exports. In their more limited forms, they appear as selective excises on liquor, tobacco, gasoline, luxuries and many other commodities and services; or as profits from government monopolies of salt, liquor, matches, tobacco and the like.

Business taxes, in the broadest sense of the term, refer to all taxes arising out of the ownership and operation of businesses. More narrowly and usefully, it refers to a separate category of taxes applicable only to business entities as such and bearing no relation to the personal status of the owners or operators of the business. In this sense, it includes on one hand such general privilege taxes as incorporation fees, capital stock taxes, the occupational taxes popular in the southern United States, the German *Gewerbesteuer* and the French *patente*. On the other hand, it covers levies which single out certain types of business for special taxation, *e.g.*, in the United States the state taxes on insurance companies, usually measured by gross premiums, and on public utilities, measured by gross earnings or property. Corporate income and profits taxes are also often labelled business taxes.

Not readily classifiable as income, property, transfer, consumption or business taxes are such miscellaneous levies as poll taxes, prohibitory taxes like the United States levy on state bank notes, taxes on increments in wealth or land values and proposed graduated taxes on individual spending.

(For further details on the various types of taxes and their use, *see* TAXATION, LOCAL; CAPITAL LEVY; CORPORATION TAX; ESTATE AND INHERITANCE TAXES; EXCESS-PROFITS TAX; INCOME TAX; LAND TAX.)

General Classifications.—Tax literature and usage abound with general classifications. Apart from groupings by tax base or source, one of the most useful classifications is by type of effective rate structure, *i.e.*, progressive, proportional and regressive types. The distinction depends upon the ratio of tax liability to net income (or net worth). If the ratio rises as income rises—*i.e.*, if the tax takes a greater percentage of a person's income, the larger that income—the tax is progressive. If the ratio is constant, the tax is proportional. If it declines as income rises, the tax is regressive. The terms are applied both to particular taxes and to tax systems as a whole. For example, studies by Richard Musgrave for 1954 indicate that total tax burdens in the U.S. are roughly proportional in the income brackets below $7,500, where tax burdens average between 27% and 31% of income, and progressive above $7,500, rising to an average of 41% on incomes above $10,000. Among specific taxes, the personal net income, net worth, and death and gift taxes are almost universally progressive. Most property sales and excise taxes are levied at proportional rates but are regressive in effect; for example, a 5% sales tax takes a larger percentage of a small income than of a large income because the taxed items represent a larger percentage of the small-income than of the large-income budget.

Another widely used distinction is that between direct and indirect taxes. Commonly, the term direct denotes a tax which tends to be borne by the persons upon whom the government levies it. Indirect denotes a tax which tends to be shifted, *i.e.*, passed on to others in the form of higher prices received or lower prices paid by the taxpayer. Since it is difficult to determine precisely who bears the final burden of a given tax, the line between the two is not easy to draw. Customarily classed as direct are net income, net worth, death, gift and land taxes. Consumption taxes, many business taxes, pay roll taxes and taxes on realty other than land are ordinarily considered to be indirect. Relative emphasis on direct versus indirect taxes as thus classified varies greatly from country to country. Of total tax revenues (excluding social insurance levies) at all levels of government in 1950, direct taxes accounted for 56% in the United States, 52% in the United Kingdom, 32% in Germany and 26% in France; indirect taxes, for 29%, 40%, 53% and 67%, respectively; and property and wealth taxes, for 15%, 8%, 15% and 7%, respectively.

Another useful distinction, encountered particularly in discussions of custom duties, is that between specific and ad valorem taxes. A tax or duty is specific when it is based on some physical measurement or quantity, *e.g.*, so much per pound, per yard or per gallon. It is ad valorem when it is based on value and levied as a percentage of that value, *e.g.*, 15 mills per dollar (1.5%) of property value.

Objectives and Criteria.—The pervading objective of taxation, in concert with other government policies, is to maximize

the general welfare. Taxes contribute to this end by providing the financial foundation for the substantive functions of government and at the same time serving as an engine of social and economic betterment which nations can call upon to reduce excessive inequalities of wealth, to check inflation and war profiteering and to promote economic stability.

How can the broad welfare objectives of taxation be translated into guideposts for tax policy? In the search for an answer, the maxims propounded by Adam Smith in his *Wealth of Nations* in 1776 provide a traditional point of departure. He set forth the tests of equity, certainty, convenience and economy as follows:

I. The subjects of every state ought to contribute towards the support of the government, as nearly as possible, in proportion to their respective abilities; that is, in proportion to the revenue which they respectively enjoy under the protection of the state.
II. The tax which each individual is bound to pay ought to be certain, and not arbitrary. The time of payment, the manner of payment, the quantity to be paid, ought all to be clear and plain to the contributor, and to every other person.
III. Every tax ought to be levied at the time, or in the manner, in which it is most likely to be convenient for the contributor to pay it.
IV. Every tax ought to be so contrived as both to take out and to keep out of the pockets of the people as little as possible, over and above what it brings into the public treasury of the state.

Appropriate as they were in his day, both as standards and as reform slogans in taxation, Smith's canons depend for their present day validity largely on what is read into them. For example, Smith directed his maxim of certainty against the evil of lodging power over the amount and terms of taxes in the "tax-gatherer, who can either aggravate the tax upon any obnoxious contributor, or extort, by the terror of such aggravation, some present or perquisite to himself." In its modern context, the demand for certainty is usually a protest against too frequent changes in tax laws, regulations and court decisions. Moreover, the unfolding of the principle of progression in taxation has infused new meaning into Smith's maxim of equity. Even with modernized content, however, his canons furnish little more than three common-sense rules of tax administration and one basic principle, that of equity, for the distribution of tax burdens.

That this set of guides is no longer adequate is hardly surprising in view of the sweeping economic, social and political developments since Smith's day. Concepts of justice have changed profoundly. Methods of measuring economic strength have been refined. Democratic control of taxation has been won. Tax administration has vastly improved. Economic systems have become enormously more complex and interdependent. The sense of social responsibility and, with it, the role of government have steadily grown. In response to these changes, it was inevitable that the functions of taxation should be broadened and its standards recast. As adapted to modern circumstances, the canons may be re-examined under four major headings: (1) social justice; (2) consistency with economic goals; (3) ease of administration and compliance; and (4) revenue adequacy.

Social Equity.—That taxes should be just is a self-evident first principle of taxation. But in taxation, as in other fields, justice is a product not of scientific analysis but of a people's value judgments. As reflected in contemporary tax systems, the value decisions modern societies have made regarding social justice appear to require: (1) that taxes be impartial in their application—the "equity" test; and (2) that taxes, on balance, be designed to reduce economic inequalities—the "equality" test. Unfortunately the precise meaning and the practical dictates of these tests are not readily definable and have, in fact, been the subject of extended controversy.

It is, to be sure, widely agreed that equity in taxation requires impartial (not identical) treatment, *i.e.*, equal taxes on persons in like circumstances and reasonably differentiated taxes on persons in unlike circumstances. Thus, the equity criterion sets up the presumption that even when taxation is asked to serve such external ends as redistribution of income or recapture of excessive war profits, the internal structure of the taxes levied should adhere to the principle of reasonable classification and impartial treatment. But the nub of the problem is: What is reasonable classification in taxation—what characteristics of a person identify him with or differentiate him from other persons for taxing purposes? And once acceptable criteria have been developed for classifying people, what relative tax burdens should the different classes bear?

For a limited group of levies on the border line of taxation, it has been deemed equitable to classify people and tax them according to the quantity of benefits received from government. Where such benefits can be identified and measured, at least roughly, and the social interest does not demand a subsidy to the individual or group benefited, it is reasonable to apply the benefit principle. Gasoline taxes and social security employment levies are commonly justified on the ground that special benefits—the use of the highways and protection against insecurity—accrue to the taxpayer in more or less direct proportion to his in-payments.

But the difficulty of measuring most government benefits and the desirability of providing many services on the basis of need narrowly limit the application of the benefit principle. And the day when "taxation according to benefits received" was a battle cry in the struggle to shift the yoke of taxation from the overburdened masses to the exempt nobility has long since passed.

Accordingly, the principle of ability to pay is generally invoked as a basis for classifying and differentiating among people for tax purposes. Deceptively simple in appearance, this concept involves many difficulties of application. At the outset, one notes that ability to pay as such is subjective, an attribute of persons, not things. Yet its measure for tax purposes is necessarily objective, an external quantity such as income or wealth. An individual's ability may be gauged by his flow of income or spendings over a certain unit of time or his stock of wealth at any given time. Net income is widely recognized as the best index. But the amount of net income is only one element in the measurement of ability. Individual income tax laws also take into account (1) family obligations, recognized in marital and dependency exemptions, and (2) less universally, the source of the income, recognized in differential treatment favouring "earned" over "unearned" income. Even after this yardstick, properly calibrated, ranks individuals according to their income, the question remains as to how much more the person of high rank should pay than the one of lower rank.

Except for one or two fleeting references to progression, Adam Smith answered that taxes should be proportional to income. Nearly 75 years later, John McCulloch still held that "The moment you abandon . . . the cardinal principle [of proportion], you are at sea without rudder or compass, and there is no amount of injustice and folly you may not commit." Expressing the predominant modern view, F. W. Taussig replied that only through progression could democratic ideals of justice in taxation be achieved. As to the lack of precise standards for determining the degree of progression, E. R. A. Seligman added that it is better to approximate justice than to be clearly unjust.

While progression is, in practice, essentially a 20th-century phenomenon, it is sobering to reflect that the Indian sage Manu recognized the principle more than 3,000 years ago in these words:

"To make the burden of taxes equal . . . is not effected by a mere numerical proportion. The man who is taxed to the amount of one-tenth . . . of an income of 100 rupees per annum, is taxed far more severely than the man who is taxed an equal proportion of an income of 1,000 rupees, and to a prodigious bigness more severely than the man who is taxed an equal proportion of 10,000 rupees per annum."

In the development of ability-to-pay theories of progressive taxation, the "sacrifice" doctrines have played a major role. An offshoot of hedonistic or utilitarian economics, these doctrines find their most refined expression in A. C. Pigou's least-sacrifice theory. It holds that progression is necessary to minimize the aggregate direct burden of taxes on individuals. The theory stands or falls on the validity of its assumptions (1) that different people are about equally efficient as "pleasure machines," *i.e.*, that they have similar capacities for satisfaction, and (2) that the marginal utility of money diminishes as income increases, *i.e.*, that as an individual's income rises, the want-satisfying value of each additional unit of it falls. If these assumptions are valid, it follows from

(2) that the larger the income, the smaller the sacrifice in giving up a unit of it in taxes, and from (1) that the relation between income and sacrifice is similar for different people. Minimum sacrifice thus calls for progressive taxes which bear heavily on the less important units found at the margin of large incomes and lightly on the more important units found at the margin of small incomes.

Modern welfare economists question the stated assumptions, pointing out that they cannot be verified until a way is found not only to measure satisfactions but to compare them among different people. They are therefore disinclined to accept a defense of progression which is based on sacrifice doctrines.

Not satisfied with attentuated translations of "ability to pay" into "progression," many noted thinkers have rested their case for progressive taxation simply on a desire to reduce inequalities in wealth and income. Adolph Wagner argued in *Finanzwissenschaft* (1880) that defenses of progression on the basis of ability to pay become meaningful only in terms of greater economic equality. In his view, reducing inequality of incomes increases the general welfare by enabling more urgent needs to be satisfied at the expense of less urgent needs.

Taussig's later use of the equality criterion was based mainly on his interpretation of the democratic consensus that ". . . such great degrees of inequality as the modern world shows are regarded as not consonant with canons of justice." Henry Simons, writing in the 1930s, stated also that the case for sharp progression must rest on the ethical judgment that the prevailing degree (and/or kind) of inequality in wealth and income is "distinctly evil and unlovely." (Henry C. Simons, *Personal Income Taxation,* Chicago, 1938, p. 19.)

Consistency with Economic Goals.—The importance of harmonizing the economic overtones of taxation with the economic goals of the community—seen as efficient allocation and full employment of resources, rising living standards and economic stability—has long been recognized. But the concept of using taxes actively to promote a stable economy and full use of resources gained currency only toward mid-20th century, largely through Lord Keynes and his followers. Every tax became recognized as affecting the spending power and propensity of taxpayers as consumers and producers and their incentives to exert effort and initiative as workers, managers and investors. In terms of these economic effects, the ideal tax structure is one which (1) restricts private spending during inflation and expands it during depression and (2) interferes least with, or perhaps even stimulates, effort and risk taking.

The health and growth of a modern industrialized economy depend upon an adequate volume of demand for the goods and services it is capable of producing when fully employed. Whether that demand will be forthcoming depends on the aggregate spending of consumers, businesses and governments. When income is withdrawn from the spending stream by individuals who neither spend it nor invest it (or by businesses which neither distribute it nor spend it on inventories, plant and equipment), the net result is to dry up demand and employment. Alternatively, when individuals and businesses as a whole run deficits and augment their spending by borrowing money or drawing on assets, the result is to inflate demand and employment. Left to itself, the private economy will experience cyclical slumps and booms, aggravated perhaps by secular stagnation or inflation. In this setting, it becomes vital that central governments use taxes as part of a coordinated fiscal-monetary policy to stabilize the economy.

Tax policy in this context has to concern itself, first of all, with the cyclical stability of total spending. Flexible taxes serve as one means of resisting or correcting its instability, of restricting total spending in a boom period and expanding it in a slump. This approach calls for a tax system and a tax policy which increase tax collections relative to expenditures in booms and decrease them in depressions. It applauds progressive income taxes, whose revenues automatically expand and contract more than in proportion to the ups and downs of national income; current collection and withholding magnify this stabilizing impact by making tax revenues respond instantly to changes in income. Contracyclical tax policy also enlists short-term changes in tax rates to impound or free purchasing power. The British White Paper on Employment Policy (1944) stated that "the ideal to be aimed at is some corrective influence which comes into play automatically—on the analogy of a thermostatic control . . ." and suggested "variation of rates of taxation and the incorporation of some system of deferred credits (repayable in bad times) as a permanent feature of national taxation."

Unless contracyclical tax measures are already imbedded in the law, the necessary tax increases to combat inflation and tax decreases to counter deflation may be unduly delayed in the legislative and administrative process. This is especially likely in the United States, where tax proposals by the executive branch do not have the thrust provided by the British and Canadian parliamentary systems, under which the government in power stands or falls on the acceptance or rejection of its program.

Under the U.S. system, the executive proposes but the congress disposes—often slowly and only after successive modifications by the house of representatives through its ways and means committee, by the senate finance committee, by the senate itself and finally by the joint conference committee. Nonetheless, in severe economic fluctuations, it may be necessary to reinforce automatic variations in revenues with discretionary changes in tax rates to generate the required anti-inflationary surpluses or expansionary deficits.

In contrast with the wide area of agreement among economists on the broad principles of anticyclical tax policy, there is no consensus on whether tax systems should be biased toward stimulating or damping down total spending in the long run. For the classical economists of an earlier age, this was no problem. Full employment was largely taken for granted, and savings were seen as the key to progress, as an unmixed blessing to be disturbed as little as possible by taxes, especially progressive taxes. But the great depression of the 1930s raised the spectre of the mature or stagnant economy, beset with persistent unemployment, weak markets, redundant savings and lagging investment. It suggested to Keynes's followers that tax policy coupled with government expenditure should seek to activate dormant purchasing power and increase the flow of spending. This would call for taxes which bear heavily on the private incomes least likely to be spent and lightly on those most likely to be spent.

To the extent that high incomes are less likely, and low incomes more likely, to be spent by their recipients, progressive taxes would be welcomed under these circumstances.

Under the impact of World War II, the feared circumstances did not materialize. In fact, the economic record of most of the decade after the war suggested that, at least under conditions of armed peace, the problem of containing chronic inflation might replace the problem of sustaining aggregate demand. If this proved to be the case, advocates of progression would find little comfort in the economic argument. The more progressive the tax rates, the higher the level of tax revenues would have to be to achieve a given counterinflationary effect.

As taxes rose to new heights during and after the war, their disincentive effects became a source of increasing concern. Colin Clark developed the thesis that beyond levels of 25% to 30% of national income, taxes so weaken the will to work and invest and so stimulate wasteful business spending that further taxes become inflationary rather than deflationary in their effect. His thesis struck a popular chord in countries where the political will to tax had reached its limit. But it found little support among economists who, while acknowledging the adverse incentive effects of high taxes, pointed to (1) the important positive effects taxes also have on incentives, *e.g.,* decreasing a man's income by taxes may move him to work harder to restore his income to its pretax position, and (2) the existence in many countries of record levels of income, employment and investment side-by-side with the highest taxes in history, suggesting that the damaging effects of taxes were not yet evident.

Lack of firm statistical evidence bids fair to keep the incentive controversy alive for many decades.

Taxation is often called upon to serve economic ends other than

those concerned with aggregate spending and saving. Taxes are employed to subsidize the production of certain items, *e.g.*, protective tariffs, and to penalize the production of others, *e.g.*, the heavy taxes levied by dairying states on oleomargarine. In World War II, high excises were imposed on goods like automobiles and radios to divert resources from their production into war production. The elimination of unwarranted economic gain from war has been a primary objective of excess profits taxes. Taxation has thus come to be widely used as a conscious instrument of indirect economic control. (*See* TARIFFS.)

Ease of Administration and Compliance.—The exacting requirements of social justice and economic harmony in taxation are necessarily tempered by considerations of administrative ease and taxpayer convenience. Unless a tax can be administered at a reasonable cost to both government and taxpayers and without large-scale evasion by illegal means and avoidance through legal loopholes and subterfuges, it becomes unworkable no matter how laudable its economic and social objectives. The sense of equity is as surely violated by a tax which applies unequally to taxpayers in identical circumstances because of lax administration as it is by a tax which is basically unfair in its statutory provisions. From the standpoint of administration and compliance alone, taxes should be simple to compute and easy to understand (understanding being a matter, not of the wording of the law, but of the concepts and tax returns which confront the taxpayer), assessed on an obvious base easily arrived at, levied at moderate rates and payable in a convenient manner.

But the degree of simplicity required or of complexity permissible is not an absolute. Much depends on the nature of the tax; for example, the minimum complexity required to bring about tolerable equity in an excess profits tax is much greater than that required in a tobacco tax. Much also depends on the scope of the tax—how widely it applies and how deeply it cuts. The demands for ease and simplicity of the United States individual income tax were far less pressing at the end of the 1930s, when it applied at moderate rates to 4,000,000 persons, than during World War II or in the mid-1950s, when it applied at heavy rates to more than 50,000,000. Factors external to a given tax must also be taken into accounnt. The lack of co-ordination of federal, state and local tax systems, for example, imposes costs and irritations arising not out of a particular tax but out of the conflict and duplication among taxes and taxing jurisdictions. Even more important are the factors external to the tax system as a whole. The general level and standards of the civil service establish a boundary beyond which taxes cannot go in the direction of refinement and complexity. Administrative techniques and competence, in turn, depend in good part on the stage of development of society. A country that is sparsely settled and industrially undeveloped is likely to have very simple administrative machinery and therefore to rely heavily on tax with an easily discoverable base, *e.g.*, land taxes and indirect taxes on imports and exports. In more advanced countries, the machinery and the economic data on which it depends are better geared to direct, personal taxes like those on income and inheritances. The levels of public education, the attitudes toward government and the prevailing concepts of justice in taxation similarly condition tax forms and practices. They are vital in determining how complicated and refined a tax may be without jeopardizing its success in practice. As these limits of tolerance are broadened, the frontiers of taxation are automatically extended.

Revenue Adequacy.—The receipts of government must be sufficient to commandeer for the state that part of society's real resources which social and economic policy wishes to devote to collective use. Traditionally, it was demanded that taxes alone free the necessary resources for collective use each year, *i.e.*, balance the budget annually. This position was later modified to permit deficits in bad years, provided they were offset by tax surpluses in good years, *i.e.*, provided the budget were balanced cyclically. Many modern economists have gone further, abandoning the idea of a balanced central government budget as an end in itself. Their goal instead is a balanced economic budget for the nation, sufficient total spending to balance and sustain the income flowing from high-level production and employment. After World War II, the British, Canadian and to a lesser extent the United States government budgets were presented in the perspective of the national economic budget.

In this conception, taxes are not levied for revenue alone. The test of adequate yield as applied to the tax system is not that it cover a predetermined level of government expenditure but rather that, together with borrowing if necessary, it raise the required funds in the manner best calculated to further the end of full and stable employment. The quantitative role of taxation becomes interlocked with its qualitative role.

This reasoning applies primarily to central governments. Their fiscal policies may appropriately be designed to offset tendencies toward economic instability in the private sector of the economy. But for local governments, more closely analogous to private persons, taxation remains primarily a means of raising revenue. In the United States, for example, local governments differ from the central government in that they lack money and central banking powers, their debts are held externally, and they have neither the ability nor the responsibility to maintain high levels of employment. In this context, the traditional aims of stable tax yields and at least cyclically balanced budgets remain valid.

Shifting, Incidence and Effects.—To know only who bears the impact of a tax, *i.e.*, who turns the money over to the government or its agents, is to know very little indeed about the social and economic consequences of that tax. The tax may set loose forces which enable the original payer to obtain higher prices for the goods or services he sells or pay lower prices for those he buys, thus giving him relief through shifting. The tax may also affect both private and governmental consumption and investment and hence the national income. To pass final judgment on a tax, one needs to know how it affects the distribution of real income available for private use, called its incidence, and what repercussions it has on the level of national income, called its effects.

Unfortunately, the way in which a tax affects the distribution and level of income cannot be isolated for precise statistical study. Since conclusions must perforce be reached largely through deductive processes combining reasonable assumptions with analysis of economic forces, the theory of incidence and effects is heavily dependent on the economic theory of price and income determination. Though no definitive answers can be provided, a fairly general consensus on the incidence of the major taxes has been tentatively reached along the lines indicated above in the classification of taxes into direct and indirect.

Comparative Revenues and Burdens.—Interests in comparative tax statistics centre on (1) distributions of tax burdens among income groups; (2) relative reliance on different forms of taxes; and (3) international and intertemporal burden comparisons for total and particular taxes. Illustrative data on (1) and (2) are presented in *General Classifications* above. This section deals mainly with (3), but adds some information on (2). The accompanying table compares central government taxes in the United States, Britain and Canada for a pre-World War II, war and postwar year. First, it highlights the large absolute increase in total tax revenues from 1939 to 1944, an eightfold increase in the United States, sevenfold in Canada and nearly fourfold in Great Britain; from 1944 to 1957, an additional 80% in the United States and Canada, but only 50% in Great Britain. Second, the table makes clear the increased reliance on income taxes and the related decline in excise taxes in the United States and Canada in contrast with the comparatively stable tax structure of Great Britain.

Comparison of Central Government Tax Revenues and Sources in the United States, United Kingdom and Canada

Taxes	United States			United Kingdom			Canada		
	1939	1944	1957†	1939	1944	1957†	1939	1944	1957†
Total (billions)	$5.5	$42.1	$77.0	£1.0	£3.5	£5.3	$0.4	$2.7	$4.9
Individual income	18.8%	46.9%	46.0%	39.7%	38.0%	39.5%	13.1%	31.7%	29.4%
Corporate income, profits	23.3	36.2	27.4	2.2	15.1	4.1	19.5	29.4	27.9
Death and gift	6.6	1.2	1.8	7.7	3.0	3.3	..	.6	1.3
Excise	32.0	10.4	13.8	17.2	16.0*	19.9*	20.0	16.7	10.6
Sales	..	..	..	..	..	..	28.9	12.8	16.9
Customs	5.8	1.0	1.0	22.6	16.9	21.8	18.1	6.3	9.9
Employment	13.5	4.2	10.0	10.7	4.4	11.4†	..	2.3	3.5
Miscellaneous	..	..	..	..	6.6	..	.7	.3	0.5
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%

Sources: United States *Budget* Documents, United Kingdom *Financial Statements*, Canadian *Budget Speeches*.
Figures may not add to totals because of rounding.
*Includes purchase tax. †Estimated.

By itself, this table says little about the comparative burden of taxes in different years or in different countries. To give some indication of this, revenues are usually related to some measure of national income or output. In the three countries, total tax revenues (including those of all local governments, which comprised about 60% of total taxes in the United States and Canada in 1939 but less than 33⅓% in 1955, in contrast with about 20% in Britain in 1939 and only about 7% in 1955) as a proportion of national income moved from 21% in 1938 to 28% in 1944 and 30% in 1955 in the United States; from 25% to 43% to 42% in Great Britain; and from 21% to 32% to 26% in Canada.

The ratios of total tax collections to total output of goods and services (gross national product) in selected North American and European countries were as follows in 1955: Germany 35.3%; France 32.5%; United Kingdom 30.9%; Netherlands 30.3%; Austria 30.1%; Norway 26.5%; Italy 26.1%; Denmark 24.4%; United States 24.1%; Belgium-Luxembourg 23.6%; Canada 23.4%; Portugal 15.0%; and Spain 13.2%.

Valid international comparisons of tax burdens or tax rates are notoriously difficult to make. Comparisons of tax rates or per capita taxes are especially likely to be misleading because of marked differences in purchasing power of money (not reflected in exchange ratios), in standards of living, in distribution of income and the like. Even the ratios of taxes to national income or product—most widely accepted as a reliable index of comparative burdens—do not tell (1) whether taxes equalled, exceeded or fell short of government spending and (2) whether the respective governments spent the proceeds wisely or foolishly in terms of efficiency and honesty of administration and effectiveness in gauging and following the preferences of the body politic.

BIBLIOGRAPHY.—Harold M. Groves, *Financing Government,* 5th ed. (New York, 1958); Roy Blough, *The Federal Taxing Process* (New York, 1952); O. H. Brownlee and E. D. Allen, *Economics of Public Finance,* 2nd ed. (New York, 1954); Richard Goode, *The Corporation Income Tax* (New York, 1951); Ursula K. Hicks, *Public Finance* (London, 1947); J. M. Keynes, *How to Pay for the War* (New York, 1940); A. C. Pigou, *A Study in Public Finance,* 3rd rev. ed. (New York, 1949); E. R. A. Seligman, *Essays in Taxation,* 10th rev. ed. (New York, 1925); Henry C. Simons, *Personal Income Taxation* (Chicago, London, 1938); Josiah Stamp, *The Fundamental Principles of Taxation* (New York, 1936); William Vickrey, *Agenda for Progressive Taxation* (New York, 1947); Gerhard Colm, *Essays in Public Finance and Fiscal Policy* (Washington, D.C., 1955); Nicholas Kaldor, *An Expenditure Tax* (London, 1955; New York, 1956); J. H. Perry, *Taxation in Canada* (Toronto, London, 1951); Randolph E. Paul, *Taxation in the United States* (Boston, 1954). (W. W. HR.)

TAXATION, LOCAL. The proportion of government costs that falls on the local and central authorities, respectively, varies widely from one country to another. In general, the local share is a larger proportion of the total in the federal states than the unitary states. (For taxes levied by the central national authority, *see* TAXATION.)

All tax powers in unitary states rest with the central authority; local divisions have only delegated powers. In consequence, the local authorities are usually quite restricted in their choice of taxes. In federal states, on the contrary, the central government's powers of taxation have ordinarily been restricted, and the states or provinces have usually had wider independent powers of taxation than the central government itself in the beginning. Over long periods of time, however, the central governments' tax powers have been extended to meet their ever-increasing obligations, until they have seriously encroached on the tax powers of the underlying authorities. There is a tendency for federal taxes to overlap and ultimately to supersede state or provincial taxes.

This centralization of taxes is apparent in both federal and unitary states. It is the result of the growing centralization of most government activities that has come with improved facilities for transportation and communication. Few functions continue to be of strictly local concern, and few taxes can be successfully

TABLE I.—*Percentage of Government Revenues from Local Sources in Selected Countries Before World War I and After World War II*

Country	Year	Local per cent	Year	Local per cent
Unitary States				
United Kingdom	1913–14	43.7	1953–54	14.2
France	1913	20.1	1953	20.3
Federal States				
United States	1915	69.9	1955	33.2
Canada	1913	51.2	1951	30.0
Australia	1913–14	70.0	1955	26.6
Germany	1913–14	59.7	1954	44.7

administered by a small local jurisdiction. The same tendencies are at work for the different local levels of government, the larger jurisdiction growing continually at the expense of the smaller; so state and provincial taxes grow at the expense of municipal ones.

The percentage of government revenues from local sources for several countries is given in Table I. This shows both the greater centralization in the unitary states and the increasing centralization over a period of approximately 40 years. The two periods selected are periods in which the proportions were not distorted by the extraordinary demands of the financing of the two wars.

STATE AND LOCAL TAXATION IN THE UNITED STATES

State and local governments in the United States have a larger share of governmental functions than local governments in unitary states, and a proportionately larger share of governmental revenues and expenditures. There are at least three layers of government throughout the country—federal, state and local—each with its own tax powers. And except for a few combined city and county governments, there are at least two layers of local government. Typically, states are divided into counties and counties into townships or towns. These in turn are usually divided into school districts. Incorporated cities and villages sometimes replace these subdivisions and sometimes are in addition to them. Often there are special districts besides. All these have independent or delegated tax powers.

There is no sharp division of functions among the different levels of government. In general, schools are the responsibility of local governments, with some aid from the states. The highway system is a composite of state and local roads, often with state aid to the latter and federal aid to the former. Welfare activities originated as a purely local matter, but by degrees state and federal authorities assumed increasing responsibility for them. Police and fire protection were still primarily local duties, but by mid-20th century the same centralizing tendencies were beginning to be apparent here, also. The same forces that brought centralization of government activities in other countries were at work, likewise, in the United States.

There are as many state and local tax systems in the United States as there are states. The only restrictions on state tax powers are that states may not tax interstate commerce. In consequence, only customs duties are not available to state authorities. Every other important form of taxation employed by the federal government is at the disposal of the states and is used by some of them, although excises, with the exception of liquor, tobacco and gasoline taxes, are not widely employed. In consequence of this diversity, it is difficult to describe state and local tax systems.

The local subdivisions have only delegated tax powers. In practice they are usually restricted to the property tax for the major part of their income. Consequently, the wide diversity found among state tax systems does not extend to the tax systems of the smaller units of government.

The growing centralization of tax systems in the United States is shown in Table II. Federal government revenues exceeded those of state and local governments combined after World War II. With the outbreak of World War II there was no important increase in state and local taxes, whereas those of the federal government increased more than fivefold. Tables III and IV show the relative importance of different state and local taxes in 1902 and 1955 respectively. It is apparent that there was relatively little change in the nature of local taxes in this period, whereas the character of state tax systems was drastically revised.

General Property Tax.—The characteristic tax of the state and local tax systems of the United States is the general property tax. Property taxes began in colonial days as levies on specific forms of property, especially land and houses. The taxes were levied on capital value rather than rental value, but values were often determined by arbitrary measures. Land was often classified, for example, as woodland, pasture or arable land, and the value per acre was the same for all land of a given class. Houses were valued according to the number of doors, windows or chimneys. In the course of time, various classes of personal property were subjected to taxation. The items selected—*e.g.*, slaves, carriages, watches and clocks and money—were those that indicated

TABLE II.—*U.S. Federal, State and Local Tax Revenues*

Authority	1902	1925	1955
Federal	$526,000,000	$3,137,000,000	$63,291,000,000
State	155,000,000	1,282,000,000	12,735,000,000
Local	706,000,000	4,054,000,000	11,889,000,000
Total	$1,387,000,000	$8,473,000,000	$87,915,000,000
Federal	37.9%	37.0%	72.1%
State	11.2	15.2	14.4
Local	50.9	47.8	13.5
Total	100.0%	100.0%	100.0%

Source: Various publications of the Bureau of the Census.

TABLE III.—*U.S. State and Local Revenues From Specific Taxes, 1902*
(In millions of dollars)

Tax	State	Local	Total
Property	82	625	707
Alcoholic beverage	10	45	55
Corporation	44	10	54
Inheritance	7	—	7
Miscellaneous	12	26	38
Total	155	706	861

Source: Bureau of the Census, *Wealth, Debt, and Taxation, 1902* (1903).

TABLE IV.—*U.S. State and Local Revenues from Specific Taxes, 1955*
(In millions of dollars)

Source	State	Local	Total
Taxes			
Property	412	10,323	10,735
Personal income	1,004	143	1,237
Corporation income	737	7	744
Death and gift	249	8	257
General sales	2,667	453	3,120
Alcoholic beverage	550	20	570
Tobacco	459	49	508
Motor fuel	2,353	22	2,375
Motor vehicle	1,229	94	1,323
Pay roll	1,138	3	1,141
Miscellaneous	1,847	764	2,611
Total taxes	12,735	11,886	24,621
Nontax revenue (excluding grants)	3,943	5,921	9,864
Total revenue	16,678	17,807	34,485

Source: Bureau of the Census, *Reports on Government Finances.*

unusual taxpaying ability. When these levies became numerous, they were combined in a single general property tax.

The general property tax is a levy on all property: real and personal, tangible and intangible. Valuation for tax purposes is made by local assessors. The rate of the tax is determined annually by the different taxing authorities, usually by the process of dividing estimated expenditures that are not provided for through other revenue sources by the value of taxable property. In other words, the property tax is usually the elastic element in the tax system, varying with expenditures.

Up to the time of the American Civil War the tendency was, as indicated above, to combine various special levies into a general property tax. After the Civil War this tendency was reversed. One by one, various classes of property were withdrawn from the general property tax base, either to be taxed in other ways or to be exempted entirely. This tendency went so far in some states that only real estate remained subject to the tax.

The principal reason for this change was the increasing difficulty of reaching personal property, particularly intangibles. The difficulty arises partly from the ease with which such property can be concealed from the tax assessor, or even removed from any tax jurisdiction where taxes are heavy. Intangibles, in particular, so regularly evade the tax assessor that in many states the tax on this property is nominal only. The fact that one owner of stocks, bonds and mortgages may be subject to a heavy tax and another escape entirely makes the taxation of intangibles grossly unequal. Moreover, in so far as such property is reached there is double taxation. No taxpaying ability is created by placing a mortgage on a farm; yet the mortgage is added to the full value of the farm in the local tax base.

Tangible personalty, also, offers difficult problems of assessment. Household furniture and clothing, no longer new, have no real market value. Also, they are readily concealed from assessors who rarely cross the threshold. Inventories in shops and factories are more readily valued, but they can easily be reduced to abnormally low levels on assessment day. Consequently, personal property, as a result of evasion or legal exemption, has become a declining proportion of the property tax base.

Real estate, on the contrary, is both visible and immobile, and consequently bears the brunt of the property tax levies. Real estate assessments for the entire United States amounted to four-fifths of the total by the 1950s. Valuation of real estate is an expert's job, and local assessors have not always been chosen for their technical competence. This, together with political pressures, has resulted in serious inequalities in the distribution of the property tax burden.

In general, small properties are assessed at a higher percentage of true value than are large properties, principally because the assessor is better able to value a small house and lot than a large estate or department store. Also, there is a tendency to encourage new building through underassessment, and to compensate by leaving values unchanged on the older, deteriorating properties. (*See* also LAND TAX.)

Much real estate is exempt. The larger part of this is government property, and large state or federal holdings in small local taxing jurisdictions have often created serious financial problems. However, there has been some effort to compensate such local units through special state or federal assistance. Nonprofit organizations for educational, religious or philanthropic purposes have regularly been encouraged by tax exemption, and there has been a tendency to extend exemption to small home owners, through homestead exemptions, and to tenants, through the exemption of low-cost housing projects. These exemptions made serious inroads on the local tax base.

States endeavoured to assist their local subdivisions by obtaining their own revenues increasingly from other sources. Property taxes supplied 53% of state tax revenues in 1902 and only 3% in 1955. Even this assistance, however, left real estate taxes at high levels. Property owners have sought protection from excessive taxes through state legislation fixing the maximum rate that can be levied by the local authorities. These tax limits have been widely used but they have often proved arbitrary and have led to unbalanced local budgets or undesirable curtailment of local government services as well as to low tax rates.

Much of the hardship suffered by property taxpayers arises from inequalities in the tax. If real estate were assessed equitably, the tax would not be unduly burdensome in most areas. Equitable assessment would not, of course, make the tax a satisfactory measure of ability. However, approximately half of local expenditures are for street improvements, police and fire protection and other services bringing direct benefits to the property owners. Welfare activities, education and other functions which would not benefit property owners as such are supported in increasing measure by federal and state governments.

Consumption Taxes.—A variety of consumption taxes produce a large share of state tax revenues. The most important of this group of taxes are the motor fuel tax, taxes on alcoholic beverages and tobacco and the general sales taxes. The revenues from these different consumption taxes are given in Tables III and IV for 1902 and 1955, respectively; and the number of states employing such taxes is given in Table V.

Motor fuel taxes are levied at varying amounts—ranging from three to seven cents a gallon in 1956. They are quite regularly earmarked for highway construction or maintenance, and the states often share them with the local subdivisions for highway use. In isolated instances local governments have been authorized to levy independent taxes. The first motor fuel tax was levied in 1919. It spread to all other states in the course of the following decade, and by the late 1950s only the general sales tax produced more state revenue. In so far as the revenues of the motor fuel tax are used for highway purposes, it can be classified as a benefit tax, since the taxpayers contribute roughly in proportion to their use of the highways. A tendency developed to use a disproportionate share of the proceeds for rural highway systems and to give relatively little to the cities for their streets, but more and more states were recognizing the claims of urban highways.

TABLE V.—*Number of U.S. States (Including District of Columbia) Levying Various Types of Taxes, 1902–52*

Tax	1902	1912	1922	1932	1942	1952
General property	46	46	45	42	29	22
Personal income	..	2	11	19	34	33
Corporation income	..	2	11	20	33	33
Death and gift	27	38	45	47	48	48
General sales	..	..	1	6	22	32
Alcoholic beverage* (gallonage or monopoly)	..	..	..	..	49	49
Tobacco	1	3	11	20	39	41
Motor fuel	..	..	19	49	49	49
Motor vehicle licence	1	42	49	49	49	49
Pay roll	..	..	..	..	49	49

*Before prohibition liquor taxes were in the form of licences.
Source: Senate Doc. no. 69, 78th Cong. 1st sess. *Federal, State, and Local Government Fiscal Relations* (1943), and publicatioos of the Bureau of the Census.

Alcoholic beverage revenues from gallonage taxes and the profits of state liquor monopolies became an important source of state income in most states after the repeal of prohibition in 1933, and tobacco taxes spread rapidly after the beginning of the 20th century. These are comparatively easy taxes for the states to administer, and are usually defended as luxury taxes, although the cigarette taxes unquestionably fall heavily on the low-income groups.

General sales taxes also produce large revenues. Most of these taxes were introduced in the 1930s, when relief expenditures were growing, and existing revenue sources were shrinking. The nature of these taxes, and the dates of their introduction into state tax systems are given in Table VI. These taxes are regressive; *i.e.,* they are borne most heavily by persons of low income, particularly when food is included in the base, but they seem to be accepted with comparatively little protest. Consumption taxes as a whole produced nearly half of all state revenues in 1955 and about 4% of local revenues in this same year.

Personal Taxes.—State personal income taxes originated with a Wisconsin levy beginning in 1911. Other states followed until, by 1955, 32 were levying such a tax. This form of state taxation would probably have spread even more rapidly if the "equal and uniform" tax clauses in many state constitutions had not prevented the imposition of progressive rates. This tax became an important source of revenue—in some states the most important tax; but states were deterred from levying highly progressive taxes, partly through fear that wealthy residents would move to other states and partly because the federal government was well established in the field before most states had turned to it for new revenues.

TABLE VI.—*Sales, Gross Receipts, Gross Income and Use Taxes in Effect as of Jan. 1, 1957*

State	Basis of Tax	Effective Date Original Law		Rate of Tax
		Sales, etc.	Use	
Alabama	Retail sales of tangible personalty and amusement admissions	1936	1939	3%
Arizona	Retail sales of tangible personalty	1933	1956	¼%–2%
Arkansas	Retail sales of tangible personalty, room rentals, admissions and utility services	1935	1949	2%
California	Retail sales of tangible personalty	1933	1935	3%*
Colorado	Retail sales of tangible personalty and some utility services	1935	1936	2%
Connecticut	Retail sales of tangible personalty	1947	1947	3%
District of Columbia	Retail sale or rental of tangible personalty	1949	1949	2%
Florida	Retail sales of tangible personalty; rentals; admissions	1949	1949	3%
Georgia	Retail sales of tangible personalty	1951	1951	3%
Illinois	Retail sales of tangible personalty	1933	1955	2¼%
Indiana	Gross income of individuals; wholesale and retail sales	1933	None	½%–1¼%
Iowa	Retail sales of tangible personalty; admissions; some utility services	1934	1937	2½%
Kansas	Retail sales of tangible personalty; admission; some utility services	1937	1937	2%
Louisiana	Retail sales or rental of tangible personalty	1936	1938	2%
Maine	Retail sales of tangible personalty	1951	1951	2%
Maryland	Retail sales of tangible personalty and some utility services	1949	1950	2%
Michigan	Retail sales of tangible personalty and some utility services	1933	1937	3%
Mississippi	Gross income from various businesses	1934	1938	⅛%–3%
Missouri	Retail sales of tangible personalty; utility services; admissions	1935	None	2%
Nevada	Retail sales of tangible personalty	1955	1955	2%
New Mexico	Gross business income	1934	1939	⅛%–2%
New York	Optional local taxes on retail sales	1934	1934	†
North Carolina	Wholesale and retail sales	1933	1937	1/20%–3%
North Dakota	Retail sales of tangible personalty; admissions; some utility services	1935	1939	2%
Ohio	Retail sale or rental of tangible personalty	1935	1936	3%
Oklahoma	Retail sales of tangible personalty and certain services	1933	1937	2%
Pennsylvania	Selective retail sales	1956	1956	3%
Rhode Island	Retail sales of tangible personalty and some utility services	1947	1947	2%
South Carolina	Retail sales of tangible personalty	1951	1951	3%
South Dakota	Retail sales of tangible personalty; admissions; some utility services	1935	1937	2%
Tennessee	Retail sales of tangible personalty	1947	1955	3%
Utah	Retail sales of tangible personalty; admissions; some utility services	1933	1937	2%
Washington	Retail sales of tangible personalty	1935	1951	3½%
West Virginia	Retail sales of tangible personalty; gross income of business	1934	1951	2%‡
Wyoming	Retail sales and rentals of tangible personalty; admissions; some utility services	1935	1937	2%

*Additional local taxes at option of municipality.
†Taxes vary with locality. 3% in New York city.
‡Gross income taxes vary from .15% to 6%.

Local income taxes are found in a few large cities including, in 1955, Philadelphia, Pa.; St. Louis, Mo.; Washington, D.C.; Louisville, Ky.; and Toledo and Columbus, O. Also a number of smaller municipalities in the state of Pennsylvania levy such taxes. Except for St. Louis and Louisville, these taxes are in areas where there is no state tax. All have low rates. The revenues are substantial, and in Toledo in 1955 they exceeded property tax collections. However, few states have permitted such levies.

Inheritance, estate and gift taxes are long-established sources of state revenue, but wide variations in yield from one year to the next make them undependable, and they are rarely one of the larger revenue producers. Here, as in the case of personal income taxes, interstate competition for wealthy residents has tended to keep rates low, and the federal tax has become much heavier than those imposed by the states. These taxes have rarely been used by municipalities.

Business Taxes.—One of the earliest forms of taxation to be developed by the states was the corporation tax. The first taxes of this sort were mostly designed to replace or supplement unsuccessful efforts to tax intangibles under the general property tax. Since corporations obtain their charters from state authorities, the states have unquestioned right to levy privilege taxes in exchange for the right of the business in question to operate under a corporate charter. Moreover, the ease of assessment and collection of capital stock taxes, *e.g.,* on corporations which are already registered with state officials, simplified administrative problems.

Corporation taxes have taken many forms. Some are for the privilege of obtaining a charter, and others are for the annual privilege of doing business in the state. They sometimes apply to all corporations, and sometimes to a specific class of corporations, such as banks, insurance companies, railroads, power companies or business corporations. The base may be capital stock, the value of physical assets, gross receipts, net income or any other measure that the state can reach.

There is a tendency to replace earlier forms of corporation taxes with corporation net income taxes, but most of the states adopting corporation income taxes have continued to levy a variety of other corporation taxes. Local governments have had little opportunity to levy corporation taxes, but both counties and cities have commonly imposed a variety of business licences. These are a more important source of revenue in the south than elsewhere, but they are to be found throughout the United States. They are rarely important sources of revenue. State pay roll taxes are important revenue producers, but they are earmarked for unemployment benefits and make no contribution to general state expenditures.

Other Sources of Revenue.—The only other important tax is the motor-vehicle licence, employed by all states and some municipalities. This form of motor-vehicle taxation preceded the gasoline tax and has, for the most part, been retained at fairly high levels, although a few states have greatly reduced their licence taxes when they have introduced or increased gasoline taxes.

There are many fees charged for special services, and other income collected in the course of administering various state and local functions. Also, municipal public-service enterprises are for the most part self-supporting, although they incur deficits more frequently than do such enterprises in England. They rarely yield surpluses that can be contributed to the general revenue funds. Public-service enterprises are not responsible for either as large a share of municipal expenditures or revenues as in England. They are even less important in state finances.

Grants-in-Aid.—Grants-in-aid from federal to state and from state to local governments had become an important factor in state and local financing by the time of World War II. Federal grants were of little importance until the federal government adopted the social security system and began to participate in relief financing in the 1930s. These were, however, more than emergency assistance, and became a well-established method of financing state functions.

State aid to local governments has a much longer history; in fact it antedates the English grant system. Most of the earlier grants were for the development of the school systems. In 1902 nearly four-fifths of the total was for this purpose. Highways and welfare, however, received steadily growing assistance from state funds after the beginning of the 20th century, and many other local functions came to receive state assistance. Table VII gives the distribution of federal and state grants-in-aid according

TABLE VII.—*U.S. Federal and State Grants-in-Aid to State and Local Governments, 1955*
(In millions of dollars)

Function	To states	To local subdivisions
Education	299	3,154
Welfare	1,428	1,026
Highways	592	911
Other special functions	443	284
General purposes	—	591
Total aid	2,762	5,980
Total revenue for own functions	19,667	24,166
Percentage from aids	14.0	24.8

Source: Bureau of the Census, *State Government Finances in 1955.*

to functions. The total of state aid grew from $65,000,000 in 1902 to $6,000,000,000 in 1955, and from less than one-tenth to about one-fourth of all local funds in the same period.

The reasons for state aid are various. Most of the earlier grants were to stimulate local activity and improve local standards of performance. This was particularly true of the early school grants which were usually in the form of matching grants. During the 20th century, however, school grants were made more and more to assist the poorer districts to achieve minimum standards. The equalization grant grew at the expense of the matching grant. The same development became apparent in highway and welfare aid. Highway aid more often took the form of sharing the motor-fuel tax than grants from general revenues, but welfare grants were normally made from general revenue funds.

The effect of state aid was to improve local services greatly. Where equalization went so far, however, as to provide local districts with 80% or 90% of their revenues, difficult problems of control arose to prevent misuse of funds. For the most part, states did not extend controls over local governments to the extent that England did. Even with generous grants local governments were left comparatively free.

Evaluation of the State and Local Tax System.—The outstanding developments in state and local tax systems after 1900 were the rapid development of new sources of state revenue, and the tendency for states to assist local governments by leaving the property tax for their exclusive use, by sharing state taxes or increasing grants from general revenues, and by taking over formerly local activities for state administration and support. This is the same centralizing process apparent in other countries. In spite of this, local governments continued to struggle with unbalanced budgets and excessive property tax levies, while state governments usually found revenue sources adequate for their needs.

Many authorities suggested that the remedy probably lay, in some measure, in speeding up the process of state assistance, but there was danger that local authorities would lean too heavily on the state. The alternative was to improve administration of the tax on real estate, and reduce rather than continue to increase the properties exempted from this levy. Only thus could any substantial measure of local self-government be retained.

BIBLIOGRAPHY.—The best general description of state and local taxes and tax systems in the United States is to be found in H. M. Groves, *Financing Government*, 3rd ed. (New York, 1950). The most comprehensive work on the general property tax is J. P. Jensen, *Property Taxation in the United States* (Chicago, 1931). Accounts of the tax systems of specific states are available for many states in various histories of state tax systems.

Current developments and discussion of special problems will be found in the publications of the National Tax association, both the annual *Proceedings* and the *National Tax Journal*. The publications of the Tax institute, including *Tax Policy* and an annual *Symposium* on some specific current tax problem, are also invaluable for this purpose.

Statistics of state and local tax systems are to be found in a series of publications of the census bureau of the department of commerce. The series includes decennial reports on *Wealth, Debt, and Taxation*, and annual reports on *Financial Statistics of States and Financial Statistics of Cities*. Occasional studies have been published by the National Industrial Conference board on the *Cost of Government in the United States*, and by the Tax institute on *Tax Yields*. Individual states also publish financial reports with considerable regularity.

Summaries of state and local systems are published decennially by the bureau of the census, and almost annually, since 1930, by the Commerce Clearing House, Inc. The 13th edition of this was published in 1952 and is entitled *Tax Systems*.

LOCAL TAXATION IN ENGLAND AND WALES

The sphere of local government in England and Wales, measured in terms of revenues and expenditures, is far more limited than is that of the national government. It is more limited, also, than the sphere of local government in federal states, and in many other unitary states. In 1931–32 less than one-third of government expenditures in England and Wales was incurred by local authorities, and an even smaller proportion of revenues came from independent local sources.

There are a variety of local authorities, and most areas are subject to overlapping local jurisdictions. The large cities are organized as county boroughs and, except for the parish, normally comprise a single local government. The area outside these county boroughs is divided into administrative counties, and these in turn are subdivided into municipal boroughs, urban districts or rural districts. The smallest unit of local government, the parish, is found throughout England and Wales, but by the 20th century it no longer exercised important local powers.

There was constant pressure to transfer functions from the smaller to the larger jurisdiction. Until the middle of the 19th century the parish was the most important area of local government. In the early years of the 20th century it had become purely nominal in urban areas, and even in rural areas its activities had become few and unimportant. Parish expenditures declined to a fraction of 1% of all local expenditures. The total number of local units, excluding the parishes, declined from about 1,900 in 1929 to approximately 1,500 in 1956.

This centralizing tendency resulted in the transfer of functions from parish to district, from district to county and from county to central government. Poor relief, for example, was originally the responsibility of the smallest unit of government, the parish. In 1934 it became primarily a national function and in 1948 the entire responsibility for public assistance was transferred to the national government. In 1946 health, also, became the responsibility of the national government.

The county is principally responsible for education, police and highways. Housing, minor roads and trading services are largely the responsibility of the districts and boroughs.

The local revenue system is very simple. All local units are supported largely from two sources—rates and grants. The first is the only tax the local authorities are permitted to levy, except for a few small county licences that produce only a fraction of 1% of local revenues. Grants come from the general revenues of the national government. The development of local rates and grants, and their relation to national tax revenue, is shown in Table VIII. It is apparent that in the century preceding World War II, there was a clear trend toward sharing an increasing proportion of national tax revenues with local governments through the system of grants. This trend was reversed with World War II, even though the absolute amount allotted for grants continued to increase.

World War II increased national obligations, rather than local. The increased tax revenues were necessitated by war expenditures; but greater central government participation was required also for civilian functions such as housing. In 1953–54 local rates were less than 10% of combined national and local tax revenues and grants had risen to more than half of all local revenues.

Rates.—*History and Development.*—Local governments derived their authority to levy rates, the one important local tax,

TABLE VIII.—*British National and Local Revenues, 1842–43 to 1953–54*

Year	United Kingdom			England and Wales				
	Millions of Pounds		Per cent of tax revenue to grants	Millions of pounds			Per cent from	
	National tax revenue	Grants		Local rates and grants	Rates	Grants	Rates	Grants
1842–43	50.2	0.6	1.3	9.1	8.9	0.2	97.3	2.7
1872–73	64.9	2.4	3.7	19.7	18.6	1.1	94.4	5.6
1913–14	163.0	27.1	16.6	94.4	71.3	23.1	75.5	24.5
1928–29	685.3	108.3	15.8	259.8	166.0	93.8	64.1	35.9
1939–40	1,017.0	217.2	21.4	383.2	201.3	181.9	52.1	47.9
1953–54	4,606.0	485.0	10.5	790.9	376.7	414.2	47.7	52.3

Source: *Statistical Abstract of the United Kingdom.*

from the Act for the Relief of the Poor of 1601. It had been common practice, however, long before the law of 1601, to levy comparable rates. The act of 1601 provided that every inhabitant of the parish should be taxed for the support of the poor "according to the ability of the same parish." No measure of ability was specified, and the tax developed into a personal tax levied in proportion to the rental value of the property occupied. Personal property was sometimes included in the assessment in the early history of the rates but had been definitely exempted by the middle of the 19th century. Thus rates became levies on those persons occupying land or buildings, designated as "rateable hereditaments," within the jurisdiction of the local rating authority.

Vacant property is exempt from the rates in England and Wales. In Scotland, half of the tax is levied on the owner and half on the occupier. The owner's share is levied on his property whether it is occupied or not.

The occupier is assessed on rental value, which is usually determined by the actual rent paid for the specific property or similar properties.

According to the Rating and Valuation act of 1925, sec. 68, "gross value means the rent at which a hereditament might reasonably be expected to let from year to year if the tenant undertook to pay all usual tenant's rates and taxes, and if the landlord undertook to bear the cost of the repairs and insurance, and other expenses, if any, necessary to maintain the hereditament in a state to command that rent."

Specialized owner-occupied properties must receive special treatment. These are mostly business properties and the valuation is related to profits. Machinery is exempted. Provision was made for valuations every five years under the Rating and Valuation act of 1925, but actual changes in value became comparatively infrequent. This act did not change methods of valuation materially, but it attempted to secure uniform valuations throughout the country.

The base of the local rates was reduced from time to time by successive "deratings." The Agricultural Rates act of 1896 exempted one-half of the rental value of agricultural land from local rates. In 1923 this exemption was increased to three-fourths, and in 1929 the total rental value of agricultural land was exempted. Agricultural buildings were exempted in 1925.

The Local Government act of 1929, which derated all agricultural land, also derated industrial properties and railroads, reducing the "net annual value" to which the rates apply to 25% of full value. As early as 1848, the land used for the actual railroad line, but not including the property used for stations, had been derated by this same percentage. Complete exemption was provided for the property of educational, religious and other nonprofit organizations. This left as the most important ratepayers the occupiers of houses and commercial buildings. The proportion of the total valuation attributable to these by mid-20th century was approximately 95%.

Administration of Rates and Distribution of the Burden.—The valuation areas after 1925 were the boroughs and districts. The elected councils of the boroughs and districts were made responsible for the valuation. To ensure some degree of uniformity among valuation districts these were grouped into larger assessment areas, each with an assessment committee appointed by the rating authorities and the county council. The function of these committees is to supervise valuations, hear complaints and equalize.

Most counties also have full-time county valuation officials, to ensure uniformity within the county. A Central Valuation committee for all of England and Wales serves in an advisory capacity. Railroad property, formerly valued piecemeal by local officials, was valued as a unit after 1930 by a national Railway Assessment authority. The valuation is based on revenues.

Each local tax authority fixes its own rates, but all the rates levied in a given area are levied as a consolidated rate. Collections are also combined. In case of delinquency, the local authorities have recourse only to the occupier's estate, which may be only personal property. If the delinquency is deliberate the taxpayer may be imprisoned. The administration is largely the responsibility of boroughs and districts.

According to the early legislation authorizing rates, they were to be apportioned according to taxpaying ability. Later, however, this conception of taxpaying ability was modified in recognition of the fact that rentals are not closely related to incomes and tend to take a larger proportion of small rather than large incomes. This makes the levy regressive. Moreover, rates vary considerably from one district to another, whether such variations are measured in terms of the taxable valuations or per capita. This is the result of wide differences in per capita wealth in different areas and the tendency for local needs to be in inverse proportion to wealth.

In so far as the rates are used for services, bringing direct benefits to the ratepayers, a case can be made for regarding rates as benefit taxes. If the occupiers receive benefits from local government roughly in proportion to rates paid, the test of ability is irrelevant. The original poor rates obviously did not bring any such direct benefits to ratepayers. But many of the costs of local government met from the rates are of direct benefit to the ratepayers.

Consequently there was a tendency to classify the rates as benefit taxes. Some authorities even took the position that rates are not strictly taxes at all, but charges for special services. However, the national government did not subscribe to this theory. Since the first half of the 19th century it had been assisting local authorities through an expanding system of grants, which was directed increasingly to the relief of ratepayers, and especially to the equalization of rates in different areas.

Grants-in-Aid.—The national government introduced grants-in-aid to local authorities early in the 19th century. The earliest of these went to individual local governments for specific and unusual expenditures. The first grant for an entire class of local authorities was introduced in 1835, with the reimbursement of counties by the central government for half the costs of criminal prosecutions and transportation of prisoners. This and many of the succeeding grants were made specifically for the relief of local ratepayers. The functions aided were those regarded as of national importance. Another important reason for the growth of grants was the desire of the central government to improve the standards of local services.

By the end of the 19th century the national government was assisting local governments by grants for police, education and various welfare and health activities. Highways, for which grants had been provided intermittently during the 19th century, were given regular aid beginning in 1908. Housing grants began in 1919.

Practically all the earlier grants were for specific functions and were usually fixed at a given percentage of approved local expenditures. The percentage met from national funds was substantial, usually amounting to from 25% to 75% of costs. Thus they tended to stimulate expenditure for the activities aided, and to raise the level of local administration through the setting of national standards for these services.

The guiding motive in the development of the grant system prior to 1929 was the desire to stimulate local authorities to expand services for certain functions and to improve standards of local administration. In some instances relief to local ratepayers was a consideration, but only for school grants was any effort made to equalize resources.

Some indirect equalization was achieved through the enlargement of local districts and through national assistance with the welfare burden. Unemployment insurance and old-age benefits, particularly, gave local ratepayers relief. The total rate burden was probably not excessive in 1929, but inequalities in rates were very great. Moreover, the heaviest rates tended to fall on depressed areas. Rates for poor relief tripled between 1919 and 1927, and the increases were greatest in the areas with the greatest unemployment. This was the principal reason for the 1929 changes.

The Local Government act of 1929 was part of the government's recovery program. The aim was to reduce local rates, particularly for agriculture and industry. There was no thought of stimulating new and improved local government services. The block grant provided by this act was partly compensation for derating, partly

a substitute for specific grants abolished by the act and partly additional funds for general rate reduction.

Greater equalization was attained under this act, partly through the transfer of certain local functions to larger local districts, and partly through the block grant. This grant was distributed according to an elaborate formula which weighted population in accordance with the proportion of children under five years of age in the population and with the proportion of the unemployed. It was also weighted to favour those districts with a low per capita rate base and relatively few inhabitants per mile of highway.

The formula was never completely satisfactory. At first, however, the larger sums available afforded real relief. The total of rates declined from £166,000,000 in 1928 to £146,000,000 in 1932. Rates increased after that and exceeded the 1928 level in 1936. In 1948 the block grant was replaced by "exchequer equalization grants" which increased the total distributed and were designed to achieve greater equalization. Changes in the amount of grant revenue as a result of these acts are given in Table IX.

The tendency for the national government to increase its assistance to local governments continued after the Local Government act of 1929. New housing subsidies were introduced in 1930 and 1931. In 1935 the national government took over the entire administration of able-bodied poor relief, instead of increasing grants for this purpose. It did not assume the whole cost, requiring local authorities to contribute 60% of their 1932–33 expenditures for this purpose. This ensured local government authorities of no increase in these costs but also prevented decreases.

TABLE IX.—*Local Government Grants, England and Wales, for Selected Periods*

(In millions of pounds)

Purpose	1928–29	1930–31	1936–37	1953–54
Education	41.7	43.9	44.8	226.4
Highway	14.3	19.2	9.7	20.0
Housing	10.7	11.9	14.0	37.2
Police	10.3	10.4	11.8	32.5
Other specific	10.8	8.3	14.1	34.9
Equalization	—	45.0*	45.3	63.2
Total	93.8	138.7	139.7	414.2

*For derating losses 22.3, for loss from discontinued grants 16.3, additional revenue 5.0, supplementary grants for special hardship 1.4.
Source: Ministry of Health, *Annual Report.*

In 1938–39, £33,000,000 out of £244,000,000 spent for relief was contributed by local authorities. The 1940 Old Age and Widows Pensions act transferred a large number of people from local public assistance to unemployment aid, and an act of 1948 transferred all important public assistance costs to the national government. During World War II the national government financed a variety of emergency costs through grants or direct expenditure; and in 1946 it assumed full responsibility for the health function.

Trading Services and Miscellaneous Sources of Taxation.—The third important source of British local revenue, after rates and grants-in-aid, is the earnings of trading services. These municipal industries comprise water supply, gas and electricity, tramways, buses, docks, harbours, canals and private street improvements. There is no tendency to use the trading services as a source of net income or to subsidize them from rates. For the most part charges cover the costs of the services in question. Sometimes, however, there are deficits to be met from rate funds, or surpluses that may be applied to rate reduction.

Other sources of net income are of very minor importance. There are a few county licence taxes, including dog licences, that bring in small revenues. Minor revenues are derived from various administrative fees. Special assessments are rare.

Local authorities can obtain money for capital improvements through loans. Local debts more than doubled in the decade from 1920 to 1930, largely because of the rapid development of municipal housing projects. They increased relatively little in the decade from 1930 to 1940. In 1940, however, the net debt of local authorities exceeded £1,500,000,000. About one-third of this sum was for trading services, an even larger share for housing and the remainder for miscellaneous local activities. For some years after World War II local authorities could borrow only from the exchequer. This greatly restricted their capital financing. The right to borrow in the open market was restored, however, in 1953.

Evaluation of the Local Revenue System.—The local revenue system of England and Wales is like that of most countries in that it depends primarily on local real estate for its tax base. Only with such an immovable base can heavy local taxes be administered successfully. Also, such a base is appropriate in so far as local services benefit the owners or occupiers of local properties. The English system differs from most local tax systems in its simplicity. Increasing dependence on the central government for aid is characteristic of local government financing everywhere. However, grants-in-aid were introduced in England somewhat earlier than in most other countries and became a more important source of local revenues than elsewhere. This raised important problems concerning the relation of the central and local governments.

The need for extensive equalization seems not to have been apparent before World War I, although some effort was made from time to time to use grants at least partly for this purpose. In the years following that war, the decline of some of England's basic industries led to acute need in certain areas, together with comparative affluence in other areas where industry was expanding.

The national government attempted to meet this situation by assuming a larger share of the cost of support for the unemployed, by widening the area of charge and by providing larger grants distributed on equalization bases.

These measures proved of genuine assistance, but serious inequalities continued. Complete equalization would result in the national government's contributing three-fourths, and in some areas even nine-tenths, of the revenues for a large number of local authorities. This increased the danger—always present when the spending authorities are not directly responsible to the taxpayers—of wasteful use of funds. One of the merits claimed for the block grant was the removal of central government controls. But, while many minor restrictions were withdrawn, important controls remained.

For the most part, the national government extended central supervision and control along with the grants-in-aid, but if the controls were adequate for a system in which the larger share of the income took the form of grants, the freedom of local self-government, which is the principal virtue of independent local authorities, would be greatly curtailed. Moreover, there was some danger of maintaining the *status quo* in depressed areas, where emigration of workers might better be encouraged.

The gains of derating were also questioned. In so far as local rates were benefit taxes it was difficult to justify the withdrawal of industrial and agricultural values from the tax base. And the decrease in the local tax base decreased local independence somewhat in proportion. These, however, were tendencies to be found in local revenue systems in all industrialized countries. The centralizing trend is apparently irresistible.

BIBLIOGRAPHY.—An excellent account of the local revenue system is to be found in Herman Finer, *English Local Government*, part v (London, 1933, New York, 1934); Edwin Cannan, *History of Local Rates in England* (1912) is the classic study in this field. The best account of the history of grants-in-aid is to be found in Sidney Webb, *Grants in Aid*, rev. ed. (Boston, 1920). A later, comprehensive study of the grant system is Mabel Newcomer, *Central and Local Finance in Germany and England* (New York, 1937). J. R. and U. K. W. Hicks, *Standards of Local Expenditure* (Toronto and London, 1943) is an excellent analysis of the distribution of the burden of local expenditures among ratepayers and others.

Official statistics of local revenues appear in various government documents. An annual summary is to be found in the *Statistical Abstract for the United Kingdom*. Detailed annual figures for rates and rateable values of each local authority are published in the annual report of the ministry of health on *Rates and Rateable Values: England and Wales*. Also, the *Annual Report* of the ministry of health devotes a chapter to developments in the field of local finance. Detailed figures of all revenues for each local authority appear in the ministry of health, *Annual Local Taxation Returns: England and Wales*.

There are many special reports of government committees and royal commissions, dealing with local revenues. Among the most useful of these are the Royal Commission on Local Taxation, *Final Report*, Cd. 638 (1901); Department Committee on Local Taxation, *Final Report*, Cd. 7315 (1914); Committee on Local Expenditure for England and Wales, *Report*, Cd. 4200 (1932).

LOCAL TAXATION IN OTHER COUNTRIES

The characteristics of the diverse local tax systems found throughout the world cannot be summarized briefly. Instead, some of the more important deviations from the English and U.S. patterns have been described below.

France.—The French system has many characteristics that are typical of European countries. It differs from that of England or the United States in its high degree of centralization. France, before World War II, was a unitary state that did not delegate to local governments the degree of independence that English local authorities acquired. It is divided into departments, which are in turn subdivided into communes. Administration is highly centralized, with commune authorities, for the most part, responsible to department authorities, and these responsible to the national authorities. In consequence, local revenues and expenditures are a comparatively small part of the total.

The second important difference between the French system on the one hand and the English and American on the other is that a large proportion of local revenues come from *centimes additionnels*. These are local levies on national tax bases, a form of taxation unknown in England and the United States. Before the introduction of the income tax in France, in 1917, the national government had four direct taxes to which departments and communes could add *centimes*. The taxes in question were on rental value of real estate, doors and windows, presumptive income (determined mostly by rentals) and business. After the income tax was introduced the national government gave up all these except for a modified real estate tax. The old bases were retained, however, for local use, except for the door and window tax, and the commune taxes were continued as additional levies on department tax bases.

Local governments did not share in the national income tax, but the national government made substantial grants to communes for various purposes. These became an increasing proportion of local revenues, exceeding one-fourth of the total in 1950. In addition, communes and departments have substantial independent tax sources. These include a variety of consumption taxes, amusement taxes and direct taxes for roads and other purposes. The amount of revenue from grants and local sources is given in Table X.

TABLE X.—*Local Revenues in France, 1934 and 1953*

Source	In 000,000,000 francs		Per cent	
	1934	1953	1934	1953
Direct taxes	5.9	172.0	39.3	21.0
National grants	2.0	178.5	13.3	21.7
All other revenue	7.1	470.2	47.4	57.3
Total revenue	15.0	820.7	100.0	100.0

Source: *Annuaire Statistique* (1938 and 1955).

Germany.—Before Hitler, Germany was a federal state and as such developed independent federal and state tax systems; or rather, the federal system was superimposed on existing state systems with relatively few changes in the latter. The state systems presented much the same diversity as did those of the United States.

Prior to World War I, the personal income tax was the central tax of most state and local systems. This tax was commonly supplemented by a trio of taxes on land, buildings and trade. The land and building taxes were usually levied on owners, and the trade tax on occupiers. The base of these taxes was determined by the state; and local taxes took the form of additions to state taxes.

Local authorities were permitted to determine, within limits, the rates they wished to levy on the state tax base for income, land, building and trade taxes. These rates were added to state levies. In this respect the German system resembled the French system closely. It differed in the variety of taxes from one state to another and in the greater independence of *Kreis* and commune authorities. In some states the local authorities were permitted to levy independent taxes on income or property, and commonly did so in preference to using the state tax base. In others the local governments depended almost entirely on additions to state taxes, having few independent tax sources at their disposal.

These direct taxes produced more than three-fourths of all state and local revenues, the income tax alone producing more than half of all state and local revenues in 1913–14. There were, however, a variety of stamp, beverage and entertainment taxes that produced substantial income.

Under the Weimar constitution, after World War I, the national government took over personal and corporation income taxes and shared these and other central government taxes with the states. The states in turn redistributed part of the proceeds to local governments. The result was to circumscribe state and local tax powers materially and to unify tax systems for the entire *Reich*. By 1934, more than half of state tax revenues and more than one-fourth of local tax revenues came from *Reich* taxes. This use of shared taxes was found in other countries, but Germany developed it much further than any other important nation. Nevertheless, direct taxes on land and buildings and trade and a variety of consumption taxes remained at the disposal of state and local governments.

The larger share of state and local costs fell on local governments in 1932–33. Their expenditures were almost double those of the states. State expenditures exceeded local for police and education, but other important functions were primarily the obligation of local authorities.

Centralization increased with the Nazi government. The federal form of government was abandoned, and the states became mere subdivisions of the central government. Local independent tax powers remained only by sufferance. Thus the German system became more highly centralized than the French.

Following World War II, state taxes were restored, and in western Germany the state and local share of the total was 38% in 1953–54.

Other European Countries.—Other European countries offer many deviations from the French and German patterns. The limited local additions to central taxes are, however, a very common form of local taxation. Shared taxes are much less common. They usually appear when the central government takes over a formerly local tax source.

Local consumption taxes and other indirect levies are found in varying degrees. In some countries they play a more important part than in either France or Germany. There is a tendency, however, for national governments to obtain a larger share of their taxes from indirect levies than for local governments. Income taxes tend to be primarily national levies, but real estate taxes are almost universally a local source of revenue.

The local authorities of continental Europe have a much greater variety of taxes at their disposal than either the English or U.S. local governments; but they have less freedom in determining the amount of the levy.

BRITISH COMMONWEALTH

Canada.—The Canadian tax system resembles that of the United States in many ways. Both are adapted to the federal form of government, and in both the different levels of government have wide powers of taxation with substantial overlapping in the taxes of different jurisdictions.

In both, municipal governments lean heavily on real estate taxes and state or provincial governments get large revenues from liquor, gasoline and motor vehicle licence taxes. Also, there is wide variation in the tax systems of the different provinces.

There are, however, important differences. The Canadian municipal property tax differs both from the general property tax of the United States and from the English rates. The tax is levied on capital value as in the United States, but it is usually limited to real estate and often has exempted improvements entirely or has assessed them at a lower value than land.

The rapid rise in land values, especially during the opening years of the 20th century and the desire to encourage settlement led to the attempt, in the western provinces especially, to place the entire burden of local taxation on land alone.

This proved to be a very unstable tax base. Land booms were followed by collapse. In some cities the tax base was cut in half in

a short space of time. The result was frequent crises in municipal financing, and after World War I there was an effort to broaden the tax base. Improvements were generally included in assessments for local taxes, although often at a fraction of full value. Personal property was exempted but the real estate tax was often supplemented by a tax on business rentals.

The provinces enjoyed a much wider variety of tax sources, but

TABLE XI.—*Canadian Provincial and Municipal Revenues, 1951*
(In millions of dollars)

Revenue source	Provinces	Municipalities	Total
Real and personal property tax	6	469	475
Liquor taxes and profits	144	—	144
Gasoline tax	178	—	178
Corporation taxes	188	—	188
Other taxes and fees	189	91	280
Total taxes, profits and fees	705	560	1,265
Dominion grants	127	20	147
Other revenue	240	91	331
Total revenue	1,072	671	1,743

Source: *Canada Year Book, 1955.*

beginning with World War I, the Canadian government tended to encroach on provincial sources, even going so far as to conscript the personal income tax at the time of World War II. The federal government, however, offered the provinces generous compensation in the form of substantial grants for specific functions and general purposes. In 1940 these amounted to more than one-fourth of provincial revenues from taxes and grants. The largest were for old-age pensions and unemployment benefits, but there were also substantial sums that the provinces were free to spend as they chose.

By agreement with all the provinces excepting Quebec, these arrangements were continued, with relatively minor changes, after the war. The more important local tax revenues are given in Table XI.

Australia.—Australia, like Canada, has a federal government. The states have independent and widely varying tax systems; and municipalities here, as elsewhere, obtain their principal revenues from real estate taxes.

The local rates, as they are called, differ from the English rates and vary somewhat in the different states. In most states there are two separate levies—one on the unimproved capital value of land and one on the improved value.

In some states the latter is on rental value and thus resembles the English rates. In other states this, also, is on capital value. Real estate is assessed for these levies by a state official. The amount of the levy is, however, determined by local officials and varies with needs from one municipality to another and from one year to the next.

Personal property is not subject to these taxes.

Municipalities have no other important independent taxes. They obtain substantial grants from state governments and obtain some income from public-service enterprises and administrative charges. The revenues from all sources for 1946–47 are shown in Table XII.

TABLE XII.—*Australian State and Municipal Revenues, 1954–55*
(In millions of pounds)

Source	State	Municipal	Total
Rates	—	51.7	51.7
Inheritance tax	21.6	—	21.6
Motor vehicle tax	27.8	—	27.8
Other taxes	40.1	.8	40.9
Total taxes	89.5	52.5	142.0
Grants	169.8	14.4	169.8*
Business undertakings†	197.3	.3	197.6
Other revenue	36.3	18.0	54.3
Total revenue	492.9	85.2	563.7*

*Municipal grants have not been included in these totals to avoid double counting since these are paid from state revenues. †All receipts reported in state accounts, and only profits in municipal accounts. State business expenditures exceeded income.
Source: Commonwealth Bureau of Census and Statistics, Australia, "Public and Private Finance," *Bulletin No. 47.*

Because of the sparse population in much of Australia the states undertook many functions ordinarily left to local governments, including education, health, welfare and much of the cost of highways. Municipal expenditures are largely for public-service enterprises and public improvements. States obtain revenues from a large variety of taxes, as shown in Table XII. They no longer have at their disposal, however, the income taxes that provided them with more than half of their revenues in the period prior to World War II. Their largest single source of revenue is from the inheritance tax.

The states also obtain exceptionally large revenues from land sales and from public-service enterprises.

The Australian commonwealth, like the Canadian government, assisted the underlying governments with an extensive series of grants. Some of these were in compensation for tax sources transferred to the federal government, and some were for aid to specific functions. Also, some were special grants to particularly needy states. Beginning with World War I, the commonwealth began to encroach seriously on state sources of revenue.

With the outbreak of World War II, the commonwealth appropriated the income tax in spite of vigorous protests from the states. State revenues from this source were replaced by a commonwealth grant of the same amount. This arrangement was continued after the war. Thus the centralizing tendencies apparent in other countries appeared in Australia also.

See also Index references under "Taxation" in the Index volume.

BIBLIOGRAPHY.—The best summaries of the tax systems of the different countries of the world are to be found in *Tax Systems.* Thirteen editions of this have been published up to and including the 1952 edition. An excellent account of the Canadian system is to be found in the series of reports of the Royal Commission on Dominion-Provincial Relations, published in 1939 and 1940; J. Harvey Perry, *Taxation in Canada* (1951). For Australia, one of the better studies is the *Second Report of the Royal Commission on Taxation, 1934* (1934). A good account in English of the French system is to be found in R. M. Haig, *The Public Finances of Post-War France* (1929). An account of the German system is to be found in M. Newcomer, *Central and Local Finance in Germany and England* (1937).

Detailed statistics for all countries are to be found in their official government reports. (M. NR.)

TAX FARMING, a method of tax collecting in which the collecting, instead of being done directly through officials of the state, is done indirectly through contractors. In ancient Rome and the empire companies or syndicates of tax farmers, known as *publicani,* were formed to farm the indirect tax sources of the state and, in addition, such sources of the state revenues as mines, fisheries, etc. (*See* PUBLICANI).

In monarchical Europe revenues were nearly all farmed, but the system was gradually narrowed to only indirect taxes. France, from the 16th through the 18th centuries, was entirely in the hands of financiers because of the hopeless condition of its revenues. From 1681 to the French Revolution the indirect taxes were farmed collectively to a single company of 40 capitalists known as farmers-general. Appointed by the king for six years they paid an annual fixed sum in advance. Not only was the system wasteful, but the harsh methods of collection led to hatred of the tax farmers by the people. At the time of the Revolution, 32 of the farmers-general were executed and the entire system was abolished

TAXICAB: *see* MOTOR TRANSPORTATION; TAXIMETER.

TAXIDERMY, the art of lifelike representation of animals (and especially mammals and birds) by means of their prepared skins, has, by a natural association of ideas in museum practice come to include the making of habitat groups, in which animals are exhibited with their naturally associated plants, with or without a painted landscape background.

Zoological collecting is related to taxidermy in origin and in the techniques of preparation of bird and mammal specimens. (*See* ZOOLOGICAL COLLECTING.)

History.—Taxidermy, like museum collecting of animals in general, evidently has its origins in the preservation of skins, horns and skulls as hunting trophies or as travel souvenirs. The early perfection of tanning methods made possible the use of skins as clothing, rugs and bedcovers. The introduction, in the 18th century, of chemical poisons for protection of skins, hair and feathers from destructive insects made possible the accumulation of great private and royal collections of mounted birds and mammals that

preceded the development of modern museums. Taxidermy as an art is now carried on almost entirely under the auspices of larger museums. As a more or less skilled trade, the mounting of game birds, fishes and especially the heads of horned mammals continues in the hands of commercial establishments.

The origins of modern taxidermy may be seen in the report of skins of the European bison or aurochs mounted over a framework of boards by the Austrian Baron Siegmund von Herberstein, in the great hall of his residence some time before the year 1550. Preservation of exotic or rare creatures with an educational purpose in view rather than primarily as trophies represents the beginnings of museum preparations of exhibits; indeed, such collection and preservation represent the beginnings of museums of natural history.

Early crude attempts to fill out skins of animals to lifelike size consisted in merely stuffing the sewed-up skin with hay or straw, or, at a later date, with excelsior. Such specimens might have shown the hair or plumage well, but could only crudely represent lifelike form. There seem to have been critical attempts to improve methods of taxidermy, and even to supply vegetation and painted backgrounds, early in the 19th century. Thus an account of the larger African mammals in *Bullock's Museum,* published in 1815, states that:

> Various animals, as the lofty Giraffe, the Lion, the Elephant, the Rhinocerous, etc., are exhibited as ranging in their native wilds and forests; whilst exact Models, both in figure and colour, of the rarest and most luxuriant plants from every clime, give all the appearance of reality; the whole being assisted with a panoramic effect of distance and appropriate scenery, affording a beautiful illustration of the luxuriance of a torrid clime.

The effort to mount large mammals in natural positions and even in positions representing great activity appears in the work of the Maison Verreaux, a commercial establishment in Paris under the direction of Jules Verreaux. Verreaux combined the roles of naturalist, explorer, taxidermist and businessman, and the Maison Verreaux formed the prototype of Ward's Natural Science establishment in the U.S. For the International exposition in Paris in 1867, the Maison Verreaux created a group showing an Arab courier on a camel attacked by two lions; this group was purchased for the American Museum of Natural History in New York city in 1869.

The subsequent history of taxidermy is mainly American, and centres to an extraordinary degree in the group of active young men who gravitated to Ward's Natural Science establishment at Rochester, N.Y. This unique business, founded by Henry A. Ward, engaged in the supply of museum materials—from minerals to mastodons and from birds of paradise to orangutans—to schools, colleges and universities during the period from 1860 to about 1890. These decades were likewise the period of active growth of museums in the U.S., which were an additional market for Ward.

William T. Hornaday and his fellow taxidermists at the Rochester establishment, William Critchley, Frederick S. Webster, George B. Turner, Henry E. Denslow and later Carl E. Akeley, seem to have been in extraordinary agreement in the hope of raising taxidermy to the status of an art. The accompanying cartoons from a set in the *Ward's Natural Science Bulletin* (1883) illustrate the feeling of the Ward group against the crude effects obtained by their colleagues who "stuffed" animals. The realization that natural positioning of animals depends on knowledge of their habits in life and of their anatomy, and in turn on a sculptural process to produce the manikin on which the skin is to be applied, was by no means new; but the Ward taxidermists experimented more actively with clay and plaster of Paris and with various techniques for making manikins than had any other group.

Among the young men who came to Rochester to practise taxidermy by working at the Ward institution was Carl E. Akeley (*q.v.*). Akeley, benefiting especially by the advice of Critchley, the bird taxidermist, and possessing great native mechanical and artistic ability, found the process that represents the greatest single advance in the taxidermy of large mammals.

This process, briefly, consists in a method of transfer of the form of a clay model to a light, strong and permanent manikin on which to mount the skin. The animal is modeled in the attitude desired for exhibition, in clay. The clay model is sculptured to represent the muscular surface of the animal, with attention to the forms of the muscles involved in the particular posture chosen. The clay form is then converted into a plaster of Paris mold in such a way as to form a three-piece waste mold. By the application of burlap, composition (originally papier-mâché) and, in large animals, a layer of wire cloth to the inside of the mold, there results (after removal of the mold) a light, strong, durable form, the manikin, with a surface well adapted to hold the skin of the animal it is to represent, while it is at the same time a satisfactory copy of the clay model. So successfully does this process meet the requirements of scientific and artistic taxidermy, that with only minor changes it still is widely used by all larger museums.

(K. P. S.; X.)

General Taxidermy Methods.—With the decreasing use of stuffed animals as parlor ornaments and trophies of the chase, and their increasing use as visual aids to education in both museums and schools, taxidermy must include preparation of mounts for displaying not only birds and mammals but the whole range of animal life—from mammals to worms and sponges. In addition to dealing with skins, the scope of taxidermy can be extended to include the making of life-sized replicas by various methods and the making of accessories to simulate a natural setting for the animals.

Mammal and bird taxidermy involves arranging the treated skin over a prepared manikin or model and inserting glass eyes. A supplementary process involving the replacement of the skin by Celluloid, adapted from a process developed for reptiles, is mentioned below. Bird and small mammal mounts usually have only a partial manikin, and require flexibility in the limbs by the use of wires, to secure the desired effect in final manipulation.

Reptiles and amphibians when prepared as stuffed skins or as skins drawn over manikins do not look lifelike to a discriminating eye, and age does not treat them kindly. Models made from plaster molds of the fresh animals and cast in plaster, or better in wax, and then painted can be really convincing if the taxidermist is sufficiently skillful with the brush, and they are relatively permanent.

However, Leon L. Walters, at the Chicago Natural History museum, developed a more elaborate process of casting with a Celluloidlike material that has some special merits. The material used is cellulose nitrate or cellulose acetate dissolved in a suitable solvent. The proper oil colours for each part of the pattern of the animal to be cast are added to bottles of this thick liquid. This material is painted on the inside of the mold, each colour in its proper place, until the taxidermist decides the right intensity of colour and translucency of subsurface colour are achieved. This thin layer of Celluloid material is then reinforced by a suitable backing.

When the animal used in making the mold has been well arranged and the application of the colours has been successful, a wonderfully lifelike replica is revealed when the plaster mold is removed. However, since the patterns are painted from the inside or the back, the final result cannot be judged until the replica is finished and dry; consequently some surface painting or touching up is often necessary.

This "Celluloid skin" process can be used for hairless or scantily haired mammals, as has been demonstrated with replicas of a hippopotamus and of narwhales, and can be used for the naked soft parts of birds; *e.g.,* the Celluloid head of a vulture may be incorporated with the feathered skin of the rest of the bird to give a complete mount. Hair also can be transferred to a "Celluloid skin." This is done by putting the raw wet skin over a clay model; making a plaster mold over it, the hairs being embedded in the plaster; leaving the skin to macerate for a few days until the hairs will pull out of the skin by the roots; then removing the skin, leaving the hairs embedded in the plaster mold. The next step is to paint the inside of the mold with the appropriately coloured Celluloid which forms the artificial skin and grips the roots of the hairs and holds them in exactly the original position. This artificial skin is then reinforced, the plaster mold soaked

and broken off, and the remaining particles combed out of the hair. A combination of methods can be used in mounting the same animal.

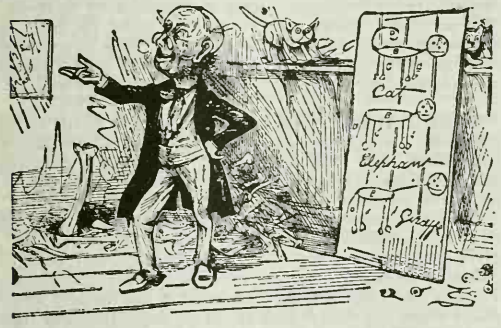
Fish are often mounted by making a mold of the side of the fish to be exhibited and then skinning the fish and fitting the skin into the mold. There it dries, is reinforced, and then the mold is removed and the natural colours painted onto the skin to restore

The cartoons below, which date from 1883, illustrate the turning point in the history of taxidermy from "stuffing" to "modeling," and emphasize also the influence of the Ward group in the development of modern methods.

THE PRINCIPLES LAID DOWN

The triumphs of the art of Taxidermy, displayed in Bird-stuffers' windows and in the museums of many of our institutions of learning, have ever had a fascination for our funny man, S————. He always had a feeling that some deep, changeless though *esoteric* principle must govern these remarkable, weird efforts.

At the Taxidermists' Exhibition, at New York, S. met an old Professional, who, by his scathing criticisms of the work of all others and his lofty praise of his own genius, quite convinced him that this Professional who had, for fifty years past, blessed New York with his interesting productions, might give him some clue to the above mentioned mystery, S. therefore followed him down to Broadway to his sanctuary, saw his work and heard his discourse.



"All animal forms," said the Professional, "were intended by Nature to be alike, though a series of blunders and accidents have irreparably marred the design; it therefore rests with the Taxidermist to restore the universal similarity. All animals are composed of a *head* (A), a *body* (B), *legs* (C), and a *tail* (D); a fact incontrovertibly proving our theory. Bones are an abuse which have crept in, and should be suppressed"

"Touching the similarity theory, take the dog: add antlers, remove the collar, and screw in a shorter tail, and he becomes a *deer*. Again, take the dog shown before: remove the legs, ears and tail, affix flippers—and you have a seal!"

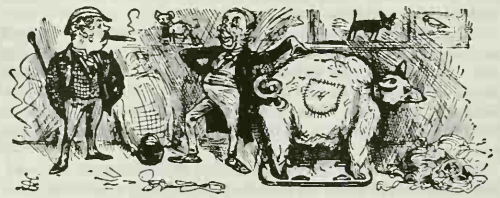
"The sole difference between animals, admissible by the Taxidermist, is the difference of *habits*—expressed by posture. The wild animal, when at home, is always engaged in *creeping down a rock;* while the tame animal has two poses—*standing up and smiling, and lying on a footstool and smiling*"

AN UNAPPRECIATED TRIUMPH

"The skins of animals," continued S—'s instructor, "were originally intended by Nature to be filled out to their fullest extent, though the failures in carrying out this principle are many and glaring. Among such the greyhound is a painful example."

"A gentleman possessed one of these animals, which died, and its skin was brought to us for manipulation"

"The result was a great success; yet the owner was dissatisfied! 'My dear, sir,' said I, 'we *can't* stuff it any tighter—human strength cannot do it.' However, to please the gentleman, we promised to try"



"Well, sir, I invented a powerful air-pump, and blew that skin out until the result was a marvel of perfection in canine form. Would you believe it, sir? the owner grumbled more than before! Well, sir, we joyfully took the specimen off his hands, and exhibited it as an example of what our art could attain to, and we daily expect an offer of a fabulous sum for it from the National Museum"

the lifelike appearance. Such skins have a strong tendency to crack; hence often a plaster cast of the fish is made and painted to resemble the fresh skin. Fins must be added. Carved plastic fins have been used, but some unfortunate experiences with cracking of plastic fins have caused opinion to veer back toward the use of natural, dried fins or fins made of wax.

So little of the soft parts of sea urchins and starfish (echinoderms) are exposed that the dried animals themselves usually are exhibited. Their soft-bodied relatives, the sea cucumbers, however, must be modeled and carved in wax.

Arthropods with stiff exoskeletons, such as insects, are usually simply killed, legs, wings and bodies posed in the proper position and dried. However, in preparing large-bodied insects, such as some moths, grasshoppers and beetles, the abdomen should be opened and the viscera removed.

For crabs and lobsters, not only must the viscera and some of the flesh be removed, but also the shell must be painted to restore the fresh colours.

Caterpillars, which are soft-bodied, need special treatment. The contents of the body are squeezed out through an opening cut in the anal end of the animal. The empty skin is then inflated, and when dried can be coated inside with a thin plastic, injected through the opening with a hypodermic needle. In another method the empty skin is inflated with hot wax. Some caterpillars also can be cast in wax.

Mollusks usually are displayed as dried shells. In cases where it is advisable to show the soft parts protruding from the shell, this can be done with wax, modeled and carved; such functions as a snail carrying its shell and crawling along on its broad flat foot may be shown thus.

Large octopuses can be modeled in clay in parts and cast in papier-mâché or plaster, and the casts assembled and painted. Small specimens can be cast in wax.

Worms are perhaps best preserved by fixing in a preserving solution and then embedding in a clear plastic, such as a sirup of polyester resin, to which a catalyst is added to cause it to jell; the material then is cured with heat. Such embedding is used by commercial biological supply houses that provide material for schools and also sell kits for embedding biological material. Not only worms but butterflies, small shrimps, frogs and bats have been embedded. These mounts have the advantage of permanency and resistance to the hazards of handling, important in a school museum.

Sometimes dry, brittle material such as crinoids can be embedded to advantage. These objects, fixed in the centre of a block of clear plastic, can be seen from all angles and can be very attractive.

Jellyfish, sea anemones and their relatives the corals (coelenterates), whose soft, intricately patterned bodies contain a large amount of water, present special problems. Corals usually are displayed as dried specimens, without the soft, minute animals that built the rocklike coral. Jellyfish are sometimes exhibited arranged in jars of preserving fluid, which leaves much to be desired.

Very beautiful glass replicas of jellyfish and sea anemones are to be seen in some museums. They are exquisite examples of the glass blower's art, but it seems impractical to have more of them made and they could be considered *objets d'art*.

As an alternative to realistic models, jellyfish may be presented as reverse carvings. This is done by engraving, with a grinding tool, the shape of the jellyfish on the back of a sheet of clear plastic. Viewed from the front, the effect is that of a jellyfish embedded in plastic. This method also can be used effectively for some other animals, such as small aquatic translucent shrimps and arrowworms.

Sponges are usually simply dried, with their soft parts removed, but embedding of small sponges, or parts of large ones, is also advisable to show what the living animal looks like.

Protozoans, of necessity, must be shown mounted on microscope slides. Their structure can be shown by enlarged diagrams or models or reverse carvings.

When the prepared animal is displayed in a setting to show its natural surroundings (*i.e.*, a habitat group), the complicated process of reproducing all the accessories becomes the responsibility of the taxidermist. The soil may be represented by earth mixed with shellac; rockwork by coloured plaster or papier-mâché; water by a sheet of plastic; natural grass and moss may be dried and coloured; leaves may be made of wax and cotton; stems of wax and wire; natural branches and tree trunks may be used, or constructed of hardware cloth and papier-mâché and coloured; and a landscape painted on the curved background. The foreground and background should be well blended so that the effect is that of seeing the animal in the out-of-doors.

The mixture of real parts of animals with replicas in plaster, wax and plastic may seem disconcerting at first to the person who expects to see the real animal displayed. But considering how little of the real animal is present in a "stuffed" deer, for instance, the step from that to a completely artificial model is small. Since the object of museum exhibits is to inform about the appearance of animals rather than to preserve trophies, the mixture of models and real skins seems justified.

BIBLIOGRAPHY.—William T. Hornaday, *Taxidermy and Zoological Collecting* (1891); R. W. Shufeldt, "Scientific Taxidermy for Museums," *Rep. U.S. nat. Mus., 1892* (1894); Oliver Davie, *Methods in the Art of Taxidermy* (1894); Montagu Browne, *Artistic and Scientific Taxidermy and Modelling* (1896); John Rowley, *Taxidermy and Museum Exhibition* (1925); Carl E. Akeley, *In Brightest Africa* (1923); Leon L. Walters, "New Uses of Celluloid and Similar Material in Taxidermy," Field Museum of Natural History *Museum Technique Series*, no. 2 (1925); Leon L. Pray, *Taxidermy* (1943); A. A. Wood, "Preparing Insect Displays," Canada Department of Agriculture *Science Service*, no. 1032 (1958). (A. L. RD.)

TAXIMETER, an apparatus used on taxicabs to compute the fare for riding and waiting time. For the taxicab owner's information the meter also totals the fares recorded and the mileage driven.

The fare-indicating device in the meter is operated by a flexible shaft to record distance traveled and a clock to record waiting time, but only one of these means can operate on the fare drums at one time. The meter is put into operation by the movement of a flag that has three positions, which are: vacant, hired, not recording.

The not-recording position means that waiting time is not recording; it is used when the passenger is paying his fare or when a tire is being changed, etc. A fourth position known as double tariff is sometimes used in Europe and South America. In this position the meter will record at a higher rate of fare, such as a night rate.

The meter is driven by a flexible shaft running from the speedometer outlet at the transmission, and a reduction box is provided that reduces the flexible-shaft speed to approximately 60 revolutions per mile. Change gears are provided under the meter to compensate for differences in tire diameters. Some meters issue a receipt to the passenger for the fare paid and print a record of the fare and mileage on each trip for the owner's information. (B. C. P.)

TAXODIACEAE, a small but very important family of the Coniferales (*see* CONIFERS) which include some of the largest trees known. The California redwood and the other American and oriental genera are important timber trees. Many species are planted for ornament.

The genera customarily recognized are: *Sciadopitys* (umbrella pine), *Sequoiadendron* (big tree), *Sequoia* (coast redwood), *Metasequoia* (dawn redwood), *Taxodium* (bald cypress), *Glyptostrobus, Cryptomeria* (Japanese cedar or redwood), *Athrotaxis, Taiwania* and *Cunninghamia*. All occur in the northern hemisphere in America, Asia or both, except the genus *Athrotaxis* from Tasmania. The trees are monoecious and have scalelike or needlelike leaves, often in frondose branchlets.

BY COURTESY OF BERNICE H. KERR

BALD CYPRESS (TAXODIUM DISTICHUM) AT CYPRESS GARDENS, FLA.

The staminate, or pollen-bearing, cones are solitary (terminal or axillary) or clustered. The spirally arranged scales bear three to nine sporangia. The ovulate, or seed-bearing, cones are solitary and terminal on a branch and are composed of many spirally arranged scales fused more or less completely with the bract which becomes thick and woody. The two to nine seeds are erect in some genera, and inverted in others. Embryogeny in the family ranges from the extreme type of cleavage polyembryony in *Sciadopitys*

to the rather advanced form in *Sequoia*. The number of cotyledons varies from two to eight. (R. W. H.)

TAXONOMY. The science of classification in a broad sense, usually restricted to biological classification and specifically to the classification of plants and animals. Taxonomy is a word derived from the Greek *taxis* ("arrangement") and *nomos* ("law"). Taxonomy is thus the methodology and principles of systematic botany and zoology, and sets up arrangements of the kinds of plants and animals in the form of hierarchies of superior and subordinate groups. The major taxonomic categories are the plant and animal kingdoms; the subordinate groups are successively the phylum (the phyla are the main subdivisions of the kingdom), class, order, family, genus and species, each such category including one or more of the groups next subordinate to it. Further subdivisions within this hierarchy are employed only when a comprehensive classification of a group of animals or plants expresses finer degrees of relationship of the component groups (*see* BOTANY: *Taxonomy;* ZOOLOGY: *Classification*).

The purpose of taxonomy is to develop a convenient and precise method of classifying knowledge that conforms to fundamental principles of biology. Need for naming the systematic categories (many of which are recognized in common speech) produces an intimate relation between taxonomy and zoological and botanical nomenclature (*see* BOTANY; ZOOLOGY; ZOOLOGICAL NOMENCLATURE). The taxonomic name provides an index symbol to the biological literature containing all recorded information about the species or group.

Taxonomy depends on descriptive biology for most of its materials, but descriptions of plants and animals, and most alterations of the system of classification, are now made within the framework set up by the taxonomists of the past. This framework is the slow growth of two centuries of proposal and critique. Some taxonomic problems are subject to experiment, and taxonomy does not rest upon descriptive morphology alone. Structure, however, is an indication of subtle genetic, physiologic and growth processes, so much progress depends upon correlations of structural differences and similarities. In the pre-Darwinian era, the categories of the classificatory system (especially the species and genus) were thought of as reflecting some mystical relation to the ideas of the Creator. In the light of the evolutionary theory (*see* EVOLUTION, ORGANIC), the classification of organisms tends toward a more and more exact reflection of their natural relations; *i.e.*, of their phylogeny (*see* ANIMALS, PHYLOGENY OF).

Modern taxonomy depends upon the reproduction of basic genetic homology (*see* HOMOLOGY) through evolutionary time together with modification by means of genetic mutation, recombination, reproductive isolation and natural selection. All life, so far as is known, is interrelated and genetically continuous in the past, and all life has undergone adaptive evolution and branching of taxonomic groups in geologic time. A few living species have not changed sufficiently to require taxonomic reassignment since Mesozoic times (over 70,000,000 years), and several living genera are recognized from fossils in Triassic deposits (over 150,000,000 years old). One living genus of brachiopods (*Lingula*) is known from fossil shells of Ordovician age (about 350,000,000 years old).

The basic method of taxonomy is to compare and weigh the characteristics of the structure of plants and animals by the methods of comparative anatomy, and to interpret the major subdivisions in the light of comparative genetics, biochemistry, physiology, embryology, behaviour, ecology and geography. Systems of classification based upon a single set of characters, or drawn up from a single viewpoint, must be reinterpreted by the study of all available knowledge of the forms concerned.

Classification may begin purely as a matter of convenience, as was the case with early botanical classifications, which developed in association with the use of herbs and plant products in medicine. Such classification based on the more superficial distinctions rather than on the more fundamental similarities and differences is termed artificial. (*See* BIOLOGY: *Early Attempts to Classify Living Things.*) The distinction between an artificial and a natural classification was necessarily obscure in the century before Darwin.

The basic units of the classificatory system are readily recognizable kinds or species (*see* SPECIES: *Biological Species*) of plants and animals. Some of these are distinguished in common knowledge, like the horse and the ass, or white oak and red oak. These instances also illustrate the first step in the process of classification, the grouping together of the various single-hoofed animals related to the horse (horse, ass and zebra) as the horse genus, and similarly the grouping of acorn-bearing trees into the oak genus. By the simple device of combining the name of the genus with that of the species, the resulting names indicate the relationships and allow for additional species to be included when discovered. The technical scheme of generic and specific names in Latin, established by Linnaeus in the mid-18th century, has proved satisfactory and has in fact greatly stimulated the exploration of the world for the distinguishable kinds of plants and animals.

The merit of the Linnaean system lies in its suitability for continued correction, as well as expansion, in the light of increasing knowledge, without loss of its essential character. A large proportion of our knowledge of plants and animals comes to be correlated with the system of classification. The amplification and correction of the classification is a continuing active function of taxonomic research.

The definition of species (*cf.* A. E. Emerson) as "an evolved (and probably evolving), genetically distinctive, reproductively isolated, natural population" is analytical, practical and understandable in the light of biological theory. Different species populations, whether sexual or asexual, have different genetic characters that may be detected by quantitative incidence of genes, qualitatively different genes, polygenic characters, or chromosomal changes in arrangement and numbers (*see* GENETICS), all of which are reflected in the structure or the living activity of the component individuals.

It is estimated that roughly 330,000 plant species and over 1,000,000 animal species have been named. Descriptions of newly discovered species add between 15,000 and 20,000 each year. Over half of the known species of plants and animals are insects (class Insecta), and about one-third of all known species of animals are beetles (order Coleoptera). There is probably no single species of living or extinct plant or animal that is of no importance to man, either economically or for his basic understanding of organic nature.

As the knowledge about the variability of species increased, and when species began to be studied as populations, it was realized that the collecting of specimens constitutes only a method of sampling. A subordinate category, the subspecies, was established for series of organisms only partially reproductively isolated and often connected by transitional (or intergrading) populations.

Geographic gradients of variation, either uniform or separated by "steps," as topographic, climatic or other barriers reduce genetic exchange, form an important aspect of variation within species. Such gradients of variation are termed clines (Julian Huxley), and those in which there are interruptions or irregularities are referred to as step-clines. To distinguish as named subspecies the segments of a uniform variation gradient (particularly of a single character) appears to be fallacious. The segments of a sloping step-cline, however, may correspond exactly to a series of subspecies defined on other grounds.

The subspecies of zoological systematics is defined most often on a geographic and ecologic basis, and is derived from the Linnaean concept of variety subordinate to the species. Variety, as a term, has been used to cover so great a diversity of variation, equally including mutation, phenotypic variation and geographic form, as to be useless in a precise taxonomy. Transition to the subspecies concept in botanical nomenclature has lagged behind that in zoology. A great variety of infraspecies terms has been proposed, and is useful in the analysis of variation and evolutionary process. (*See* works listed in *Bibliography:* G. C. Robson and O. W. Richards; E. Mayr.)

Turning again to the species level, a most important clarification of the concept lies in the recognition that purely morphological definition must be subordinated to the concept that the species is composed of populations in which variability is inherent. The "closet naturalist" (without experience with living populations in

their natural habitat) was exposed to the fallacy of describing variant individuals as species. Much of modern taxonomic investigation is occupied in the re-evaluation of the work of earlier taxonomists by means of more adequate collections, anatomical study, and concepts from the fields of genetics, embryology, physiology, ecology, psychology and geography.

Re-examination on the basis of larger collections (*i.e.*, a representative sample) may unite a whole series of supposedly distinct forms into one, or may separate distinct forms that had long been included in a single variable species. A few species have now been subjected to comprehensive genetic analysis, especially among the fruit flies of the genus *Drosophila*, and such analysis throws much light on taxonomic problems.

The morphological characteristics distinguishing genera are ordinarily (but not necessarily) more sharply defined than those distinguishing species, and the distinctions between genera are regarded as having an origin more remote in time than those between their included species. Thus the concept of the genus comes to have especial importance in paleontology (*see* GEOLOGY: *Historical Geology*). The genetic and biochemical reality of the genus or other more inclusive taxonomic category is attested by instances in which one species of parasite, symbiote or predator is associated exclusively with several species within a single genus, or several genera within a single family.

So many genera are familiar in ordinary thought and language, like the oaks (*Quercus*), the maples (*Acer*), the cats (*Felis*), some of the squirrels (*Sciurus*), etc., that there should be no difficulty in the wider application of the concept of the genus as a natural group of related species. Definition of genera by means of characters thought to be of generic value may lead to an arbitrary and unnatural assortment of species. The synthetic assortment of known species into their evident groups, as if they were building blocks, thus becomes an extremely important check on generic classification. It has been discovered that characteristics suitable for the definition of genera in one group may be useful only in the distinction of species in another. When the number of species of a genus is large, these may fall into subordinate groups, which are referred to as subgenera. When a single species has a number of sharply distinctive characters separating it from a related group of species within a genus, it may be placed in a monotypic genus by itself.

Study of fossil remains of both plants and animals obviously introduces the dimension of time into variation, and this affects all of the categories, and would complicate the distinction of the groups of lower rank even more were it not for the fact that the interruptions in the paleontological record supply convenient gaps between species and genera. As paleontological data accumulate, important modifications of the classification of the higher groups become necessary (*see* below).

Geographic series of subspecies, connected by intergradation where they meet, or similar series with overlapping variation (as on chains of islands), are matched by series of geographically representative but fully distinct species. Ecological differences in the habitats of the component members of such series has led to local adaptation. When former geographically separated species have been brought together by a breakdown of the barriers between them, the two species may form a hybrid population, one species may eliminate its competing relative, or each may occupy different ecological niches in the same region. Ernst Mayr clarifies this problem by the distinction of species as sympatric, if their areas of distribution overlap, and allopatric, if their geographic ranges are distinct.

THE HIGHER GROUPS

Just as well-known sharply defined species and readily recognizable genera serve to introduce the words "species" and "genus" in the more precise sense of taxonomy, common knowledge is reflected in the language for some of the larger divisions at the other end of the system of classification. The primary division, into the plant and animal kingdoms, appears in various languages. A few groups like the sponges, molluscs, insects, birds and ferns, correspond rather precisely with the more exact zoological and botanical categories.

The taxonomy of the higher groups may be pursued analytically, beginning at the top and searching for the successive subdivisions of the two kingdoms, or it may be studied by the synthetic method, beginning with the more factually defined groups nearer the bottom of the scale, and especially with the genera. The arrangement of the groups above the species level has been improved by additional information from many sources, particularly from paleontology, comparative anatomy, embryology and ecology. Related taxonomic groups of whatever rank may be presumed to share varying degrees of genetic identity by means of descent from common ancestry, and the ranking of taxonomic categories is at least roughly correlated with the comparative number and complexity of shared homologous characters.

The difference between an artificial and a natural classification of the higher groups rests primarily on the distinction between analogy and homology (*see* HOMOLOGY). The discernment of both homology and analogy in the structures of animals closely or remotely related, forms the contribution of comparative anatomy to taxonomy. Because homologues and analogues may be associated in the same structure, the practical distinction may be far from simple. As wings, the wings of a bird and of a bat are analogous and not homologous, and must be presumed to have arisen independently and not one from the other. As fore-limbs, the wings of a bird and those of a bat are plainly homologous, as they are also with the arm of man or the foreleg of an elephant, and support other evidence as to the relation of these several forms within the subphylum Vertebrata. In a broad sense, homology may be of quite a different type, like the correspondence between the alternating generations of ferns and flowering plants so brilliantly demonstrated by W. F. Hofmeister.

Homology of structures predicates the common inheritance of the characters under consideration, while independently derived analogous structures that resemble each other are said to be convergent (*i.e.*, the wings of birds and insects). Independently derived functional adaptive evolution of a basically homologous character (*i.e.*, the stiff tail feathers of woodpeckers and creepers), or independently evolved vestiges of a formerly adapted homologous structure (*i.e.*, the reduced eyes of cave fishes and cave salamanders), are often referred to as examples of parallel evolution.

The successive subordination of the series of groups, with steadily increasing number of the lower categories in most types (not in all) is reflected in the familiar comparison with a branching tree, with its trunk, main branches, smaller branches and finally twigs. Actually when the major groups are traced backward in the light of paleontology, the origins prove to be remote in time and apparently close together, and one is tempted to write of a phylogenetic "selaginella" or "fernclump" rather than of a "tree." Thus all of the main groups of living mammals trace back to the Early Tertiary or Late Cretaceous, and most of the principal groups of animals, the phyla, appear in the earliest known fossil-bearing strata of Cambrian age approximately 550,000,000 years old.

The exact relations of branch to branch or of branch to trunk are often unknown, being for the most part lost in the imperfections of the paleontological record. There are accordingly numerous groups at every level in the classification that are regarded as equivalent, and much effort has been expended to make them reasonably so. It may be pointed out that much difference of opinion over taxonomic groups in general applies to their rank in the hierarchy rather than to their validity as definable groups.

Many of the intermediate categories, especially the orders and families, are essentially the resultant of evolutionary specialization in a single direction (like the orders for the ichthyosaurs or whales, or the family of sea snakes). A different type of order or family is represented by a group that retains (or reacquires) the capacity for divergence into several major types of environment (*i.e.*, terrestrial, subterranean, arboreal and aquatic), and whose members thus compete with a wide variety of other animal types for specific ecological niches. Examples are supplied by the order Rodentia, the Saurischian dinosaurs, or the weasel family, the Mustelidae. Thus the plastic is often sharply distinguished

from the specialized or end group. This involves a generalization as to the "irreversibility of evolution." Complex identical genetic patterns (*see* GENETICS) never originate twice nor, if lost, ever appear again in evolutionary history. Parallel mutations of genes and chromosomes may be repeated, but almost all taxonomic characters are genetically far more complex. Still another type of nonequivalent categories is evident in the difference between the grades "order" and "family" in an ancestral group like the Reptilia, and a derived group such as the Mammalia.

Another problem is introduced by paleontology in groups in which the evidence of ancestral approximation (at the point of divergence of the branches of the phylogenetic tree) has been discovered. The various genera close to the point of origin may either be associated with their later derived and more distinctive descendants, in a vertical classification, or with the primitive forms of the other branches to which they may be more closely similar in actual morphology, in a horizontal classification. This antithesis of phylogenetic and morphological definition of groups is notably exemplified by the Early Tertiary ancestry of horses, rhinoceroses and tapirs. In both types of classification, the more inclusive taxonomic categories are monophyletic and are presumed to have originated from a single ancestral species.

The tendency to give equal importance to all comparative characters of organisms often leads to bias and to errors of judgment. Because the order of taxonomic classification is based upon homology through descent, principles of evolution are both substantiated by taxonomic data, and in turn, may be used for taxonomic interpretation. For example, it is possible to distinguish homology from analogy and both from fortuitous resemblance, strongly adaptive from weakly adaptive or nonadaptive characters, primitive from derivative characters, rapidly evolving from slowly evolving characters, persistent conservative characters from labile changing characters, qualitative from quantitative characters, progressively functional characters from regressed vestiges, relatively complex from relatively simple characters, and individually acquired characters from hereditary characters.

Since some classification is essential for the establishment of scientific botany and zoology, taxonomy in its origins is one of the most elementary of disciplines. The record-keeping function of taxonomy affords an essential continuing service to all of the biological sciences. As these have grown to maturity, they have returned to draw problems and insights from the parent science, and the parent taxonomy has been infused with new interest from the light thrown upon its material and its categories by comparative genetics, embryology, physiology, ecology, biogeography and behaviour. The synthetic science of ecology, growing out of the older less critical natural history, depends implicitly on sound development of systematic botany and zoology.

Future taxonomic progress may be expected by the discovery of new species, both living and fossil; relating taxonomic data to investigations in other biological sciences; further understanding of evolutionary processes; development of new analytical techniques in mathematics, morphology and basic life processes; and also by verifying the facts and interpretations of past taxonomists as well as weeding out past errors of fact and theory. Basic research in taxonomy continues to have important effects in the applied sciences of medicine and agriculture.

BIBLIOGRAPHY.—Alfred E. Emerson, "Taxonomy and Ecology," *Ecology*, 22:213 (1941); Julian Huxley, *Evolution, the Modern Synthesis* (1942), (ed.), *The New Systematics* (1940); Ernst Mayr, *Systematics and the Origin of Species* (1942); E. Mayr, E. G. Linsley and R. L. Usinger, *Methods and Principles of Systematic Zoology* (1953); G. C. Robson and O. W. Richards, *The Variation of Animals in Nature* (1936); George Gaylord Simpson, "The Principles of Classification and a Classification of Mammals," *Bull. Amer. Mus. Nat. Hist.*, vol. 85 (1945); E. T. Schenk *et al., Procedure in Taxonomy* (1956).

(K. P. S.; A. E. E.)

TAY, the longest river in Scotland, has a catchment area of 2,400 sq.mi. (6,216 sq.km.) covering nearly all Perth and parts of Angus and Argyll. Its discharge is among the highest in the United Kingdom. The headwaters are at about 2,500 ft. (762 m.) on the north side of Ben Luí, east of Dalmally. Several small streams gather to form the Fillan Water, which, after passing through Lochs Dochart and Iubhair, is known as the River Dochart. The river then assumes a northeasterly direction for many miles among Highland hills of 2,000–4,000 ft.; it then enters the head of Loch Tay along with the River Lochay. On leaving Loch Tay it receives the River Lyon, and about 5 mi. (8 km.) downstream of Aberfeldy it curves eastward to Logierait, where there enters, as a left-bank tributary, the flow from the great Garry-Tummel System, a main stream receiving the Rannoch-Tummel. The joint stream flows south-southeast to Dunkeld, receives the Highland River Bran on the right bank, and turns east-southeast through the gap between Birnam Hill and Newtyle Hill into fertile lowlands. It turns southwest at the great elbow at the Isla junction near Cargill and southeast once more near the junction with the Almond. After flowing through a gap in the low hills continuing the line of the Sidlaws, on which the city of Perth stands, the river turns east to enter the Firth of Tay at the confluence with the Earn. From Newburgh in Fife, the firth runs northeast past Dundee to open out into the North Sea off Buddon Ness in Angus. The firth broadens gently to 3 mi. upstream of Dundee, narrows to 1 mi. between Broughty Ferry and Tayport, and finally broadens to 3 mi. near its mouth, 120 mi. (193 km.) from the headwaters of Ben Lui.

Dundee is accessible to shipping at all stages of the tide. Small vessels can reach Newburgh and even Perth, but navigation is impeded by shifting sandbanks. (A. T. A. L.)

TAY, LOCH, the largest lake in Perth, Scot., is situated in Breadalbane, approximately in the middle of the county. It is a glacially excavated ribbon lake about 355 ft. (108 m.) above mean sea level and generally from 90 to 600 ft. in depth. It is 14½ mi. (23 km.) long from Killin—where the Rivers Lochay and Dochart converge at the head—to Kenmore, where the River Tay takes egress at the foot; the lake runs northeast in the central stretch south of Ben Lawers (3,984 ft. [1,214 m.]) and east-northeast at either end. The breadth is from just under half a mile to just over 1 mi. The spring salmon fishing is well known. Loch Tay lies among fine heathy hills between 2,000 ft. and almost 4,000 ft., with some woodland on lower slopes; the views to Ben Lawers from the minor road south of the loch are excellent, while the main road on the north side rises higher and gives wider, if less varied, panoramas. Near Kenmore a small wooded islet contains the tumbled ruins of a priory founded in 1121 by Alexander I in memory of his queen, Sibylla, daughter of Henry I of England, who was buried there. (A. T. A. L.)

TAYLOR, ANN (1782–1866), afterward Mrs. Joseph Gilbert and **TAYLOR, JANE** (1783–1824), English writers for children, were born in London Jan. 30, 1782, and Sept. 23, 1783, respectively. Jane was a lively and entertaining child, and composed plays and poems at a very early age. Their father and mother held advanced views on education, and under their guidance the girls were instructed not only in their father's art of engraving, but in the principles of fortification. Ann introduced herself to the publishers Darton and Harvey by a rhymed answer to a puzzle in the *Minor's Pocket Book* for 1799, and Jane made her first appearance in print in the same periodical with "The Beggar Boy." The publishers then wrote to Isaac Taylor asking for more verses for children from his family, and the result was *Original Poems for Infant Minds* (2 vol., 1804–05), by "several young persons," of whom Ann and Jane were the largest contributors. The book had an immediate and lasting success. They followed up this success with *Rhymes for the Nursery* (1806); *Hymns for Infant Minds* (1808, 2nd ed. 1810); *Country Scenes* (1806); *City Scenes* (1806); a less-known collection, *Signor Topsy Turvy's Wonderful Magic Lantern; or, The World Turned Upside Down* (1810); and *Original Hymns for Sunday School* (1812). In 1813 Ann married a Congregational minister, Joseph Gilbert, and Jane went to live at Ilfracombe with her brother Isaac. In 1816 Jane returned to Ongar, where the family had been settled for some years, and died there on April 13, 1824. Ann Gilbert died at Nottingham on Dec. 20, 1866.

TAYLOR, (JAMES) BAYARD (1825–1878), U.S. author known primarily for his lively travel narratives and for his translation into English of Goethe's *Faust*, was born in Chester county, Pa., Jan. 11, 1825. He was a restless student at Bolmar's academy

in West Chester, Pa., and the academy at Unionville, Pa., before he was apprenticed to a printer at the age of 17. In 1844 his first book, a volume of verse, *Ximena*, was published. He then made an arrangement with the *Saturday Evening Post* and the *United States Gazette* to finance a trip abroad in return for publication rights to letters describing his travels. His experiences in Europe resulted in the extremely popular *Views Afoot* (1846). In 1847 he settled in New York and began a lucrative career in journalism. *Eldorado* (1850) recounted his trials as a *Tribune* correspondent in the 1849 California gold rush. He continued his trips to remote parts of the world—to the orient, to Africa, to Russia—and became renowned as something of a modern Marco Polo. In 1862 he became secretary of the U.S. legation at St. Petersburg, and in 1878, shortly before his death, became U.S. minister to Germany. He died on Dec. 19, 1878.

Some works of this later period were *The Poet's Journal* (1862); the novels *Hannah Thurston* (1863), *John Godfrey's Fortunes* (1864) and *The Story of Kennett* (1866); and the translation of *Faust* (1870–71) which remains his best-known work. Although greatly lauded in his day, Taylor was undistinguished as a novelist and poet, being at his best in his travel narratives and in an occasional richly textured poem, such as those in *Poems of the Orient* (1855).

Collected editions of Taylor's poetical and dramatic works appeared in 1880.

Bibliography.—M. H. Taylor and H. E. Scudder, *Life and Letters of Bayard Taylor* (1884); A. H. Smyth, *Bayard Taylor* (1896); Richard Cary, *Genteel Circle; Bayard Taylor and His New York Friends* (1952).

TAYLOR, BROOK (1685–1731), English mathematician, noted for his contributions to the development of the calculus, was born at Edmonton in Middlesex on Aug. 18, 1685. He was educated at St. John's college, Cambridge, and studied mathematics under John Machin and John Keill. He obtained in 1708 a remarkable solution of the problem of the "centre of oscillation," which, however, remained unpublished until May 1714 (*Phil. Trans.*, vol. xxviii, p. 11), when his claim to priority was disputed by John Bernoulli.

Taylor's *Methodus Incrementorum Directa et Inversa* (1715) added a new branch to the higher mathematics, now designated the "calculus of finite differences." Among other ingenious applications, he used it to determine the form of movement of a vibrating string, by him first successfully reduced to mechanical principles.

The same work contained the celebrated formula known as "Taylor's theorem" (*see* Calculus, Differential and Integral), the importance of which remained unrecognized until 1772, when J. L. Lagrange realized its powers and termed it the basic principle of the differential calculus.

In his essay on *Linear Perspective* (1715; rev. ed., 1719) Taylor set forth the true principles of the art in an original and more general form than any of his predecessors; but the work suffered from the brevity and obscurity which affected most of his writings.

Taylor was elected a fellow of the Royal society early in 1712, sat in the same year on the committee for adjudicating the claims of Newton and Leibniz, and acted as secretary to the society from 1714 to 1718. From 1715 his studies took a philosophical and religious bent. Taylor died on Dec. 29, 1731, at Somerset house, and was buried at St. Ann's, Soho. *See* also Taylor's Theorem.

TAYLOR, DAVID WATSON (1864–1940), U.S. marine architect who built the first ship-model testing establishment in the United States, at the Washington (D.C.) navy yard, and who formulated certain basic principles of ship design. Born in Louisa county, Va., March 4, 1864, he graduated from the U.S. Naval academy in Annapolis, Md., and the Royal Naval college, Greenwich, Eng., with the highest grades made until then at those institutions. His talent for writing in concise, readable, scientific language was displayed in his first book *Resistance of Ships and Screw Propulsion* (1893).

Beginning in 1899, Taylor undertook the first studies to find what characteristics of a ship hull governed its water resistance. By a methodical series of extensive proportions, internationally known since 1910 as the Taylor Standard series, he determined the actual effect of changing those characteristics and thus made it possible to estimate in advance the resistance of a ship of given proportions. His *The Speed and Power of Ships* (1910), which sets forth this knowledge, is still an informative publication in its field.

From 1915 until 1921, Rear Admiral Taylor was responsible for the design and construction of surface and submarine vessels and of aircraft for the U.S. navy, including the first plane to fly the Atlantic. He made many other contributions to aeronautics in 15 years of service on the National Advisory Committee for Aeronautics.

Among many distinguished awards, he received the John Fritz gold medal in 1931 for outstanding achievement in marine architecture, and the first David W. Taylor medal in 1936 from the Society of Naval Architects and Marine Engineers. In 1937 the new U.S. naval model basin at Washington was named the David W. Taylor model basin in his honour. He died in Washington, July 28, 1940. (H. E. Ss.)

TAYLOR, EDWARD (*c.* 1645–1729), American clergyman and poet whose poems were not published until 1939; he is now generally considered the foremost writer of verse in colonial America. He was born in Leicestershire, Eng., in or near Coventry, probably of a dissenting family. Unwilling to subscribe to the required oath of conformity because of his staunch adherence to Congregational principles, he gave up schoolteaching, emigrated to New England and was immediately admitted by Pres. Increase Mather as a sophomore at Harvard college. After his graduation in 1671, he was settled as minister in the frontier village of Westfield, Mass., where he remained until his death. Small in stature, he was remembered by his contemporaries for his intensity of nature and his personal charm. He married twice and became the father of 13 children, most of whom he outlived. His grandson Ezra Stiles (*q.v.*), president of Yale college during the American Revolution, inherited most of his manuscripts and the greater part of his 200-vol. library.

By his express desire, Taylor's 400-page quarto manuscript "Poetical Works" was not published by his heirs. It came into the possession of Yale in 1883 by the gift of a descendant, and the best of Taylor's verse was published in 1939. The important poems fall into two broad divisions. "God's Determinations Touching His Elect" is an extended verse sequence thematically setting forth the grace and majesty of God as a drama of sin and redemption. The "Sacramental Meditations," about 200 in number, were written over a period of 44 years, from 1682 to 1725. Taylor described them as "Preparatory Meditations Before My Approach to the Lord's Supper."

The achievement of Taylor is his adroitness in fusing an ecstatic religious vision with earthly realities by means of homely diction and imagery drawn from the daily round of Puritan life. Central to all his poems is the typical metaphysical mode: the extravagant trope and the association of image and idea which is intended by its tension to draw poetic sparks.

The Poetical Works of Edward Taylor (1939), edited by T. H. Johnson, is a selection of poems, together with a biographical sketch, critical introduction and notes. It is supplemented by gleanings in the *New England Quarterly* (1943), *Publications* of the Colonial Society of Massachusetts (1943) and the Yale university *Library Gazette* (1954). (T. H. Jn.)

TAYLOR, FREDERICK WINSLOW (1856–1915), U.S. engineer, was born March 20, 1856, at Germantown, Pa. In 1875 he was apprenticed as patternmaker and machinist in a small Philadelphia shop. As early as 1880 or 1881, while gang boss in the employ of the Midvale Steel company, controversy with workers led him to realize the necessity of accurate measurement of what constituted a day's work on any operation. This in turn led to a series of careful experiments resulting in important achievements in two fields; in mechanical engineering, the discovery with Maunsel White of a new method of tempering tool steel permitting metal-cutting operations at high speed (1898); and in the field of management, perfection of that system of shop management which later came to be known as scientific management. The remainder of his life was spent in forwarding the prin-

ciples of this system, by practising the new profession of consulting engineer in management (1893–1901), lecturing and writing. He died March 21, 1915.

TAYLOR, JEREMY (1613–1667), English clergyman and writer who represents the best of the religion and culture of the Anglican royalists during the period of the Great Rebellion, and whose ornate prose is among the glories of the English language. He was baptized on Aug. 15, 1613, at Holy Trinity church, Cambridge, the son of Nathaniel Taylor, a barber. He was educated at the Perse school, made a sizar of Gonville and Caius college in 1626, given a junior fellowship in 1633, and ordained in the same year. When he was deputizing for his friend Thomas Risden as lecturer at St. Paul's in 1633, his eloquence and good looks made a sensation, and brought him to the notice of Archbishop Laud, who was greatly impressed, but considered him too young for immediate preferment. Two years later, by a stretch of authority, he presented Taylor to a fellowship at All Souls college, Oxford. A friendship with Francis a Santa Clara (Christopher Davenport), a learned Franciscan, led to an often repeated and always vigorously repudiated accusation that Taylor had a leaning toward the Church of Rome. In March 1638, William Juxon, bishop of London, presented him to the living of Uppingham, where he married Phoebe Langsdale on May 27. In the following November Taylor returned to Oxford to preach the annual Gunpowder plot sermon, the first of his published works. It is, in the main, a conventional attack on the recusants but some passages give promise of future greatness. In Nov. 1642 Taylor had left Uppingham and was with the king at Oxford, where he published his first important work, *Of the Sacred Order and Offices of Episcopacy*. In it he assembles with persuasive skill most of the arguments then in circulation in favour of episcopacy, but adds little new. In 1643 he was made D.D. by royal decree and given the small living of Overstone, which he never visited. He next appears in Wales where he was taken prisoner with the royalist forces at Cardigan castle in 1645. On his release he joined the grammarian William Nicholson in setting up a school at Newton hall in Carmarthenshire, near Lord Carbery's seat at Golden Grove where he served as chaplain.

The next ten years were the happiest, and the most fruitful, of Taylor's life. His literary genius was at its height. *The Liberty of Prophesying* (1647) is a noble plea for religious liberty to be granted to all whose principles would not lead to the overthrow of the state or destroy the foundations of the Christian faith. After the Restoration, the breadth of toleration pleaded for in the book caused Taylor some embarrassment and he is said to have made an effort to suppress it, but there is little doubt that it expressed his real beliefs though he was not always able to practise them. *The Great Exemplar* (1649) consists of discourses, meditations and prayers on "the life and death of the ever blessed Jesus Christ." It is the earliest of his devotional works and must have shown him the direction in which his gifts could best be exercised. *The Rule and Exercises of Holy Living* (1650) and *The Rule and Exercises of Holy Dying* (1651) so clearly supplement each other that they are usually regarded as one book. They were written to help members of the Church of England who were deprived of a regular ministry by the disturbances of the Great Rebellion, but their wisdom and beauty have made them popular with all denominations. *Holy Living* contains encouragement to the life of prayer, general directions for godly living and prayers and meditations suited to both these subjects. With illustrations drawn from the classical moralists and a wide range of other reading, as well as from Christian teaching, Taylor inculcates moderation, cheerfulness, contentment and reliance upon the divine goodness. By 1651 when he wrote *Holy Dying*, he had lost both his wife and his patroness, Lady Frances Carbery. He had been her spiritual director and close friend, and the oration which he delivered at her funeral is one of the finest in the English language. The first chapter of *Holy Dying* exhibits the full glory of Taylor's prose. The beauty of his language, the richness and variety of his fancy, match the nobility of his theme. Death is presented as an inevitable good for those who live and die as Christians. John Wesley owed much to these two books.

Twenty-eight Sermons (1651) and *Twenty-five Sermons* (1653) *Preached at Golden Grove* make up *Eniautos, a Course of Sermons for all the Sundaies of the Year* (1653). They are full of beautiful cadences and delicately wrought images. *The Real Presence and Spirituall of Christ in the Blessed Sacrament* (1654) attacked the doctrine of transubstantiation. *Unum Necessarium* (1655), a study of repentance intended as prolegomena to a work of casuistry which he had long been meditating, contained a repudiation of Calvinistic doctrines of original sin and predestination which plunged him into controversy with high churchmen and Presbyterians alike. *The Golden Grove* (1655), a manual of instruction and prayers, contained also *Festival Hymns*, a collection of not very successful poems in irregular verse.

Taylor had by this time become the spiritual counselor of the diarist, John Evelyn, who helped him with gifts of money. After two short imprisonments for his religious principles (1655), he left Golden Grove and lived at Mandinam, an estate in Wales belonging to Joanna Bridges whom he had recently married. Later he moved to London and ministered to scattered congregations of Anglican royalists, frequently visiting great houses in the country. *A Discourse of Friendship* (1657), written at the request of Mrs. Katherine Philips ("the Matchless Orinda"), is his only secular work. *A Collection of Offices* (1658), intended as a substitute for the proscribed Book of Common Prayer, shows his wide acquaintance with eastern liturgies. Taylor had become intimate with Edward, 3rd Viscount Conway, who offered him a chaplaincy on his large estates in Ulster. He went there in June 1658 and for the next two years was mainly occupied with *Ductor Dubitantium—a Great Instrument for the Determination of Cases of Conscience* (1660), dedicated to Charles II. It is vast both in bulk and learning, a maze of illustration and digression, but remains of value for its profound scholarship and as the first complete work of casuistry in English.

Taylor's reputation for unorthodoxy militated against an English appointment, which he would have much preferred, and he was made bishop of Down and Connor (1660), to which was added, in 1661, the administration of the diocese of Dromore. He was also made a member of the Irish privy council and vice-chancellor of Trinity college, Dublin. There he was instrumental in reconstituting the university after the disorders of the Great Rebellion. In his diocese he worked equally hard, but was not happy. The Scots who had colonized Ulster were followed over by Presbyterian ministers who were resolutely hostile to episcopacy. Thirty-six of them were in possession of church livings, and these Taylor deprived after they had refused all his advances. He filled their places with men from England or of English extraction. At his first visitation (1661) he preached on *The Minister's Duty in Life and Doctrine*, the best of his post-Restoration sermons, in a style less elaborate than his earlier work. At the request of his brother bishops he wrote a *Dissuasive From Popery* (part i, 1664; part ii, 1667). Both are marred by overstatement and inaccurate quotation. *Chrisis Teleiotike* ("The Unction Which Perfecteth"; 1664) is the best study of confirmation to appear before the 19th century. Taylor died on Aug. 13, 1667, in Lisburn, of a fever caught while visiting a sick man. He was buried on Sept. 3 in Dromore cathedral, which he had himself rebuilt.

The Whole Works of Jeremy Taylor were edited, with a life, by Reginald Heber, 15 vol. (1822); there was a revised edition by C. Page Eden, 10 vol. (1847–54). *The Golden Grove, Selected Passages From the Sermons and Writings*, was edited by Logan Pearsall Smith, with a bibliography by R. Gathorne-Hardy (1930).

Humane, generous, impulsive, handsome, of great personal charm, Taylor was among the most lovable of men. His prose, filled with a joyous prodigality of beauty, is one of the chief glories of the English language. In it he gave his genius full rein so that the sentences run on and on, gathering new beauty as they go, with complete indifference to orderliness or structure. In a learned age Taylor was one of the most learned and he poured all the wealth of his wide reading into his books, drawing perhaps most often upon the moralists of Greece and Rome. He was a close observer of nature and could describe with an inspired rapture the sheen on the breast of a dove, the groping tendrils of a vine or the petals

of a newly opened rose. In spite of the vicissitudes of his life his religious faith was of singular serenity and it is this which constitutes much of the charm of his devotional writings. He never threatens, but draws his reader to holiness by his own love of it. He has been accused of borrowing from other writers in his devotional books, but they are in fact entirely original. In controversy he was never at his best and in much of his other work he was apt to spoil his effects by overaccumulation of detail and superfluous argument.

As a theologian Taylor shared the Laudian love of beauty and order and respect for antiquity, but he was also influenced considerably by his contemporaries William Chillingworth and Henry More.

BIBLIOGRAPHY.—The earliest authority on Taylor's life is *A Funeral Sermon Preached at the Obsequies of . . . Jeremy Lord Bishop of Down by George Rust*, which can be relied on. Anthony à Wood's *Athenae Oxonienses* (1691) and *State Papers, Charles II, Domestic and Irish*, add details of interest. F. R. Bolton, *The Caroline Tradition of the Church of Ireland* (1958), supplies some details of Taylor's Irish period. E. Gosse, *Jeremy Taylor* (1903), has useful comments on Taylor as a writer but must be used with caution as to fact. Coleridge admired Taylor greatly and references to him are scattered throughout his prose works, especially in *Notes on the English Divines*, ed. by Derwent Coleridge (1853), and *Literary Remains*, vol. 3 (1838). *See* also C. J. Stranks, *The Life and Writings of Jeremy Taylor* (1952).
(C. J. Ss.)

TAYLOR, JOHN (1578–1653), English pamphleteer and journalist, commonly called "the Water-Poet," was born at Gloucester, Aug. 24, 1578. He was sent to a grammar school but became as he said, "mired in Latin accidence" and was apprenticed to a Thames waterman. He served in the navy and saw action at Cadiz (1596) and Flores (1597). Returning to London, he worked as a waterman, held a semiofficial post at the Tower for some years and was an active member of the Watermen's company. He increased his income and his fame by dedicating verses to noble patrons and by a series of extraordinary journeys, which he described entertainingly in oddly-titled pamphlets. With a companion as light-hearted as himself, he went from London to Queenborough, Kent, in a paper boat with two stockfish tied to canes for oars. He amused the court and the public, especially in his verbal battles with another eccentric traveler Thomas Coryate. *The Pennyles Pilgrimage . . . of John Taylor . . . on foot from London to Edenborough* (1618) describes a journey perhaps suggested by Ben Jonson's similar undertaking, though Taylor denied any suggestion of burlesque. He reached Braemar and, returning through Leith, met Jonson, who good-naturedly gave him 22 shillings to drink his health in England. He undertook a journey to Prague—where he was entertained by the queen of Bohemia (1620)—for a wager, as he did the travels described in *A very Merry, Wherry-Ferry-Voyage, or Yorke for my money* (1622) and *A New Discovery by Sea, With a Wherry From London to Salisbury* (1623). When the Civil War began, Taylor, who took a public house in Oxford (where he stayed, in Oriel college, during the plague of 1625) wrote royalist pamphlets, but when the city surrendered (1645) he returned to London and kept "The Crown" (later "The Poet's Head") off Long Acre until his death in Dec. 1653.

Editions include: *All the Workes . . . beeing 63 in Number* (1630), reprinted, 3 pt. (1868–69); *Works . . . Not Included in the Folio Volume of 1630*, 5 pt. (1870–78); C. Hindley (ed.), *Works*, a selection (1872) and *Miscellanea Antiqua Anglicana*, vol. ii and iii (1873).

See R. B. Dow, "John Taylor" in *Harvard Summaries of Theses* (1931).

TAYLOR, JOSEPH (*c.* 1586–1652), English actor, is mentioned in the folio Shakespeare of 1623 as one of the 26 who took principal parts in all of those plays. He was with the duke of York's company in 1610, but involved himself in a lawsuit with John Heming, by leaving them to join the Lady Elizabeth's in the next year. Taylor appeared in the actor list of the latter company in *The Honest Man's Fortune* (1613) and *The Coxcomb* (1613).

From 1616 he was one of the prince's company, but left it to join the King's men in 1619. He remained with them until he retired, playing several of Burbage's parts, including Ferdinand in *The Duchess of Malfi* and Hamlet. There is a legend that he was trained by Shakespeare to play Hamlet. In many of Beaumont and Fletcher's plays he had a leading role, and he is one of the ten actors who signed the dedication of the first folio of those dramatists (1647).

Taylor led the King's men, with J. Lowin, after Heming and H. Condell, and in 1630, after Heming's death, obtained two shares in the Globe, and one in the Blackfriars theatres. In 1639 he was appointed yeoman of the revels.

Taylor died at Richmond, where he was buried on Nov. 4, 1652.

TAYLOR, MAXWELL DAVENPORT (1901–), U.S. army officer, a pioneer in air-borne warfare, was born at Keytesville, Mo., Aug. 26, 1901, and graduated from the United States Military academy in 1922. Taylor assisted in the organization of the U.S. army's first air-borne division, the 82nd, early in World War II. At great personal risk, he passed through enemy lines 24 hours before the Allied invasion of Italy in 1943 to confer with Italian leaders in Rome concerning possible air-borne seizure of Roman airfields. Later he led the 101st air-borne division in the assaults in Normandy and the Netherlands, and his division gained international fame for its defense of Bastogne during the "battle of the Bulge" in 1944.

As commanding general of the 8th army in 1953 Taylor directed UN forces in Korea during the closing phases of the Korean War. He then served as army chief of staff (1955–59) and three years later was appointed chairman of the joint chiefs of staff. In 1964 he became U.S. ambassador to the Republic of Vietnam, which was being given increasing military support by the U.S. He resigned this post in July 1965. (W. Si.)

TAYLOR, TOM (1817–1880) prolific English dramatist and editor of *Punch*, was born at Bishop Wearmouth on Oct. 19, 1817. He studied at Glasgow university and at Trinity college, Cambridge (becoming a fellow in 1842), and was professor of English at University college, London (1845–47). Meanwhile he read for the bar, was called in 1846 and practised as a barrister until in 1850 he became assistant secretary, and in 1854 secretary, of the board of health. He was later transferred to the home office. He retired in 1871 and died at Wandsworth on July 12, 1880.

Taylor first made his name as a dramatist, getting four burlesques produced at the Lyceum in 1844 and making his first hit with *To Parents and Guardians* (Lyceum, 1846). He wrote over 100 plays, among the best-known being *Our American Cousin* (1858), produced by Laura Keene in New York city, where E. A. Sothern created the part of Lord Dundreary, *Still Waters Run Deep* (1855) and *Ticket of Leave Man* (1863). In addition Taylor was a busy man of letters. After working on the staff of *Punch* for many years he succeeded Shirley Brooks as editor in 1874, and remained editor until his death.

Amateur actor and artist, occasional art critic—benevolent and industrious if somewhat ponderous in manner—he was the universal go-between in mid-Victorian theatrical and artistic society, and introduced Lewis Carroll to his illustrator John Tenniel and Ellen Terry to her first husband G. F. Watts. His completion of C. R. Leslie's *Life and Times of Sir Joshua Reynolds*, 2 vol. (1865) has had lasting value.

See W. Tolles, *Tom Taylor and the Victorian Drama* (1940).
(D. Hn.)

TAYLOR, WILLIAM (1821–1902), U.S. Methodist missionary, was born in Rockbridge county, Va., on May 2, 1821, of Scotch-Irish and English parentage. On March 15, 1843, he was admitted on trial in the Baltimore conference of the Methodist Episcopal Church, and in 1847 was ordained elder. Taylor was sent to California in 1849 by the Missionary society, arriving in San Francisco, via Cape Horn, when it was still a city of tents. Standing on a box or a whiskey barrel in the plaza, he made his powerful voice heard by as many as 20,000 people and made many converts.

Taylor's evangelistic tours, taken more or less independently of the church authorities, took him throughout the United States and Canada (1856–61), and to England (1862), Australia (1863), the West Indies (1864), India (1870–75), South America (1877–78, 1880, 1883–84), South Africa (1866, 1884). On his first South American visit (1877–78) he organized a system of self-supporting

schools on the west coast. Elected missionary bishop for Africa in 1884, he established a chain of missionary stations in the Congo and self-supporting missions in Liberia. His steadfast belief that "God had taken William Taylor into a peculiar partnership" filled him with self-confidence and courage and also made him impatient with official direction and supervision. For years he was a thorn in the flesh to bishops and board secretaries. In 1896 he was retired by the general conference, and he died on May 18, 1902.

Taylor wrote 14 or more books, including *Seven Years' Street Preaching in San Francisco* (1856); *Christian Adventures in South Africa* (1867); and *Story of My Life* (1896).

See Wade Crawford Barclay, *History of Methodist Missions*, vol. iii, pp. 179 ff. (1949–57); Oscar Macmillan Buck, "William Taylor" in *Dictionary of American Biography*, vol. xviii, pp. 345 ff. (W. C. B.)

TAYLOR, ZACHARY (1784–1850), 12th president of the U.S., was born in Orange county, Virginia, Nov. 24, 1784. Within the year his father, Richard Taylor, an officer in the American Revolution, made a new home for his family in Kentucky where Washington appointed him collector of the port of Louisville, then on the border where Spaniards and Indians were still a menace. In this environment Zachary Taylor spent his youth. He gained what little schooling he was to enjoy from Elisha Ayres, a New England tutor, who was employed by the elder Taylor. In 1806 young Taylor volunteered during the difficulties caused by Aaron Burr's southwestern schemes and saw brief military service. Two years later, through the instrumentality of his relative, James Madison, and others, Taylor was given a commission by President Jefferson as first lieutenant in the 7th infantry.

As captain and major he served through Harrison's Indian campaign and the War of 1812 in the northwest territory and at the close of the war temporarily left the army because reduction to peacetime strength had lowered his rank. In 1816 President Madison reappointed him to his old rank and there ensued 20 years of garrison life at various posts, varied in 1832 by an expedition against Black Hawk. At the outbreak of the Seminole War (1835) he was an infantry colonel in the northwest; soon orders came for him to take his command to the Everglades. In this bewildering warfare he spent three years and was promoted to the rank of brigadier-general by brevet for his service in the battle of Lake Okeechobee. He was able to make little headway against the Indians and in 1840 asked to be relieved. He was assigned to duty in Louisiana and while there established a home in Baton Rouge and acquired a large plantation in Mississippi. In 1841 he was given command of the second department of the western division of the army with headquarters at Ft. Smith, Ark.

When Texas was acquired in 1845 Taylor was ordered by the war department to occupy the borders of Texas; by the spring of 1846 he had moved down to the mouth of the Rio Grande with a force of over 3,000 men and established Ft. Brown. A strong Mexican force was stationed at Matamoras and soon there were border skirmishes that resulted in war. Even before war was declared on May 13, Taylor had defeated the Mexicans at Palo Alto (*q.v.*) and Resaca de la Palma on May 8 and 9. The news of these victories and of the occupation of Matamoras on May 18 was received with great appreciation throughout the country and various persons, including Whig politicians and editors, began to look upon Taylor as a possible presidential candidate. After months of waiting for wagons and boats, Taylor moved to the head of navigation on the Rio Grande, established a depot at Camargo and set out for Monterrey, which he captured on Sept. 21 after a hard-fought three-day battle. When President Polk disapproved the eight weeks' armistice granted by Taylor upon the surrender of Monterrey, and secretary of war Marcy began to make suggestions as to the movements of his subordinates, Taylor became convinced that the administration was trying to discredit him for political purposes. His letter of Nov. 5 to Gen. E. P. Gaines criticizing Polk and Marcy found its way into the press, called forth a rebuke from the secretary of war and increased the friction between Taylor and Washington authorities. When his superior, Gen. Winfield Scott, preparing an expedition against Vera Cruz and Mexico City, ordered most of Taylor's seasoned troops to join him, Taylor was further incensed. Although instructed to concentrate his remaining troops at Monterrey and remain on the defensive, he crossed the mountains to Saltillo and on Feb. 22, 1847, engaged the army of Santa Anna in battle at Buena Vista (*q.v.*). Although outnumbered 4 to 1, the Americans, inspired by the leadership of "Old Rough and Ready" and aided by skillful artillery fire, won a brilliant victory and saved the north of Mexico. At home the feeling was general that Taylor had been stripped of his troops by a jealous administration and left to perform the impossible. This ended his military activity but his political availability increased.

Shrewd Whig politicians had become increasingly convinced of Taylor's potentialities as a candidate. Very early in the war, Thurlow Weed of New York began pushing him and popular enthusiasm made his task easier. His correspondence was skillfully handled; he was reluctant, he maintained, and with truth, to be a seeker for the office; he had never voted and though a Whig in sympathy was not an avowed partisan. As a slaveholder also he was certain to gain votes in the south where there was a good deal of suspicion that northern influences dominated the Whig managers. These advantages were sufficient to ensure success and in the Whig convention on the 4th ballot, Taylor won the nomination from Clay, Webster and Scott. An opportune split in the Democratic party ensured his victory and on March 5, 1849, he was inaugurated.

Few presidents have entered office with less knowledge of what was expected of them. The new executive expected to be a nonpartisan president, leaving to congress all legislative matters and confining himself to executing the laws. With Colonel Bliss, his son-in-law, to phrase his thoughts, he, at first, was more or less influenced in his policies by Vice-President Fillmore, but it was not long before the superior astuteness of Senator Seward, Fillmore's rival in New York state, accomplished results. After the new president became convinced that it was his duty to build up the Whig party by means of clever patronage, deserving Whigs began to come into their own.

With these preliminary lessons learned, Taylor found his most perplexing problem to be the status of the newly acquired territory. As a soldier he was eager to see stable government established. In order to hasten this he authorized Congressman King of Georgia to go to California and urge an application for statehood. When congress met, Dec. 1849, the president had the pleasure of reporting that California was ready to become a state and that New Mexico would be ready soon. He recommended the admission of California and hoped that the slave issue would not be injected into the situation. When southern representatives objected because California had prohibited slavery, he resented their protests. He declared himself willing to accept any law providing for admission, and warned southern leaders that he would take the field if necessary to carry out such a law and hang any who resisted. His attitude was the natural reaction of a military executive to those opposed to the regular process of law and order. Consequently he was not in favour of conciliating those opposed and referred contemptuously to the Compromise Measures of 1850 (*q.v.*) as the "Omnibus Bill." In other respects his administration was not proving a sinecure; charges were made against his secretary of war, Crawford, to the effect that he had used his official position to pursue a claim successfully before the secretary of the treasury by which he benefited financially. This charge involving the secretaries of war and the treasury as well as the attorney general, deeply mortified Taylor, who prided himself on scrupulous honesty, and he determined to reorganize his cabinet. Then came the fatal July 4, 1850, when exposure to heat at the Washington monument ceremonies and injudicious eating combined with worry brought on a serious intestinal disturbance which resulted in Taylor's death five days later. The doctors diagnosed his illness as cholera, but it seems to have been acute gastroenteritis.

A combination of honesty, simplicity, determination and common sense had brought Taylor successfully through a variety of difficult situations and by a strange turn of the wheel carried him to the White House. In 1810 Taylor had married Margaret Smith (1788–1852), who bore him six children, three of whom survived him. One daughter, Sarah Knox Taylor, married a young sub-

altern, Jefferson Davis, in 1835, surviving the marriage but three months. His only son, Richard Taylor (1826–79) graduated from Yale in 1845 and after a short service in his father's camp, became a sugar planter and politician in Louisiana. At the outbreak of the Civil War he entered the Confederate army and served in various capacities, finally commanding an army which surrendered in May 1865.

BIBLIOGRAPHY.—Brainerd Dyer, *Zachary Taylor* (1946); Holman Hamilton, *Zachary Taylor, Soldier of the Republic* (1941) and *Zachary Taylor, Soldier in the White House* (1951); William K. Bixby, *Letters from the Battlefields of the Mexican War* (1908); Justin H. Smith, *War With Mexico*, 2 vol. (1919). (R. F. N.; B. Dy.)

TAYLOR'S THEOREM, the statement of a formula discovered by Brook Taylor (*q.v.*) in 1712 and published three years later in his *Methodus incrementorum directa et inversa* (1715). It may be written in the form of an infinite series:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \cdots \qquad (1)$$

and under suitable conditions it makes possible the computation of the value of $f(x+h)$ to any desired accuracy when the value of $f(x)$ and its derivatives $f'(x)$, $f''(x)$, . . . , for a particular fixed value x are known; *e.g.*, with

$$x = 1, \log_e (1+h) = 0 + h - \frac{h^2}{2} + \frac{h^3}{3} - \frac{h^4}{4} + \cdots$$

Taylor's formula with remainder may be written as a finite sum:

$$f(x+h) = f(x) + hf'(x) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x) + R_n \qquad (2)$$

where

$$R_n = \frac{h^n}{n!}f^{(n)}(x+\Theta h)$$

and Θ is a fraction between 0 and 1. Several other formulas have been devised for R_n. Sometimes one of these may be used to check on the accuracy obtained by using a given number of terms of (1). Moreover, formula (2) is valid in some cases when formula (1) is not.

See W. A. Granville, P. F. Smith and W. R. Longley, *Elements of the Differential and Integral Calculus.* (L. M. G.)

TCHAIKOVSKY, PETR ILICH (1840–1893), Russian composer, born at Votkinsk, in the province of Viatka, on May 7, 1840, was the son of a mining engineer, who shortly after the boy's birth removed to St. Petersburg (Leningrad) to become director of the Technological institute there. While studying in the school of jurisprudence, and later, while holding office in the ministry of justice, Tchaikovsky picked up a smattering of musical knowledge, but he was never suspected at first of possessing any special musical talent, of which there had been no previous traces in his family. Nevertheless, the seriousness of his musical aspirations revealed itself in due course and led him to enter the newly founded conservatory at St. Petersburg under Zaremba, where he was induced by Anton Rubinstein, its principal, to take up music as a profession.

With the former he fared better than the latter who, though worshiped by Tchaikovsky, seems never to have appreciated the gifts and promise of his young pupil who was destined none the less completely to eclipse his own fame. Very different was the attitude of Nicholas Rubinstein (Anton's brother) who showed his belief in the young musician in the most practical manner by inviting him to become, in 1866, practically the first chief of the recently founded Moscow conservatory, since Serov, whom he succeeded, never took up his appointment.

In Moscow, while engaged at the conservatory, Tchaikovsky wrote among other things his first opera *The Vojevode,* which was, however, a failure when produced in 1869. In the meantime he had met in St. Petersburg Balakirev, Stassov, Rimsky-Korsakov and other representatives of the then "advanced" school in Russian music, who exercised a stimulating influence on him, although neither then nor later was he ever wholly in sympathy with their ideals.

Also at this period he fell desperately in love for a time with the opera singer Désirée Artôt who, however, shattered his hopes by marrying someone else. He found consolation in his art, and his third opera *The Oprischnik*, was one of many works composed at this time, others including the pianoforte Concerto in B flat minor, the Third Symphony and another opera *Vakoula the Smith*. *The Oprischnik* was produced at St. Petersburg in 1874 and *Vakoula the Smith* in 1876, but neither had any success.

A happier fate attended the piano Concerto in B flat minor, which though severely condemned in the first instance by Nicholas Rubinstein (to whom it was originally dedicated) as clumsy and unplayable, as a mere duel between the piano and orchestra, and so on, was afterward taken up with enthusiasm by Hans von Bülow (to whom the composer transferred the dedication) and performed by him everywhere with the greatest success. At the same time it is only fair to add that Tchaikovsky, though much hurt by Nicholas Rubinstein's strictures, considerably modified the work on the strength of his criticisms and that Rubinstein in turn ultimately recognized the merits of the concerto and played it constantly. To the period of the late 1870s belong the Quartet in E flat, the ballet *Swan Lake* and the *Francesca da Rimini* fantasia among other compositions, while during this period also, namely in 1877, Tchaikovsky first began to work on the opera of *Eugen Onegin*, destined to become the most popular of all his stage works and first heard at the Moscow conservatory in March 1879.

Meanwhile the more personal side of the composer's career had been given a romantic touch by his acquaintance with his lifelong benefactress Nadezhda Filaretovna von Meck (1831–94) and his deplorable fiasco of a marriage. In 1876 he had aroused the interest of Mme. von Meck, the widow of a wealthy railway engineer and contractor. She had a large fortune and she began by helping the composer financially with commissions for work, but in 1877 this took the more substantial shape of an annual allowance. The strangest feature of their association consisted in the fact that, by agreement, they never met, though they corresponded with one another continually. In 1890 Mme. von Meck, imagining herself—apparently a pure delusion—to be ruined, discontinued the allowance; and though Tchaikovsky was then no longer really in need of it, he was deeply wounded by the manner in which she had made the announcement, and to the end of his days never entirely recovered from the shock which his abnormally sensitive nature had received from the termination of their relations.

Not less distressing was the episode of his marriage. Tchaikovsky married Antonina Ivanovna Milyukova on July 6, 1877, but the union proved an impossible one through no fault of hers but simply through his own abnormality of temperament; and it resulted in separation in October.

Tchaikovsky, who was deeply affected, went abroad with his brother to recover and was restored in due course to a more satisfactory state, but such happiness as he was to enjoy throughout the remainder of his career was qualified in an ever increasing degree by the tendency to melancholy and morbid introspection which, from the first, had been a conspicuous characteristic of his disposition.

Nor can it be said that he did not have many causes for discomfort and disillusionment in the ill luck and indifferent appreciation which attended in so many instances the performances of his works. But he continued to produce none the less with unchecked ardour and the Fourth Symphony may be named among the other works which he composed during his period of recuperation abroad.

Later, after his return to Russia, came the *Italian Capriccio* and the *1812 Overture* among other things. In 1879 he wrote his *Maid of Orleans* (produced in 1880) and his Suite No. 1 for orchestra. In 1881 died Nicholas Rubinstein, to whose memory he dedicated the pianoforte Trio in A minor. To the next five years belong the *Manfred* symphonic poem and the *Hamlet* overture fantasia, the operas *Mazeppa* and *Charodaïka,* the *Mozartiana Suite* and the big Fifth Symphony, of which even the last-named, though now reckoned among his finest achievements, failed initially to elicit any great enthusiasm. Nor were the opera *The*

Queen of Spades (which later vied with *Eugen Onegin* in popularity) or the delightful ballets *Swan Lake, The Sleeping Beauty* and *The Nutcracker* any more favourably received when first produced, so far at least as the critics were concerned. Better fortune, however, attended *The Nutcracker* music in suite form which soon found unlimited favour and quickly established itself as one of the most popular pieces of its class ever written. (Incidentally, it may be noted that Tchaikovsky himself had the poorest opinion of this music when he was actually writing it.)

But though, after the time-honoured experience of genius, Tchaikovsky gained full appreciation so tardily in his own country, in foreign countries his claims were much more readily recognized. In 1887 the first of several concert tours took him to Leipzig, Berlin, Prague, Hamburg, Paris and London, and won many warm friends for his music in every instance, not least so in London where he conducted the *Serenade* for strings and the Variations from Suite No. 3 at a Philharmonic concert.

Later, in 1891, he paid a visit to the United States, where he conducted six concerts—four in New York city, one in Baltimore and one in Philadelphia—and again he was everywhere received with unbounded enthusiasm, the works performed being the *Coronation March*, Suite No. 3, two sacred choruses, the piano Concerto in B flat minor and the *Serenade* for strings.

In 1893 Tchaikovsky sketched his Sixth Symphony, becoming more and more absorbed in his task as the work proceeded and more and more firmly convinced that it was the greatest thing he had ever done, a judgment subsequently confirmed in full measure by the world at large; for this was the symphony afterward known as the *Pathétique*, and under that title destined to enjoy such world-wide fame and popularity as perhaps no similar work had ever known before. Yet even this symphony had a cold and indifferent reception on the occasion of its actual first performance (Oct. 28, 1893) at St. Petersburg, though in this instance the verdict was speedily reversed, both in Russia and elsewhere.

But the composer was not fated to witness and enjoy its triumphs. For within ten days of its production the world was shocked by the news of his wholly unexpected death—from an attack of cholera, caused by the drinking of unfiltered water—on Nov. 6, 1893.

As to the precise value of Tchaikovsky's music and the place which it is likely ultimately to take critical opinion has been sharply divided, but with a prevailing tendency to accord far less importance to his work than some consider its due. Almost inevitably the enormous popularity enjoyed by so many of his works at the hands of the general public has strengthened this tendency and the impartial student may well wonder if, even when the worst has been allowed, the process of critical belittlement which has been witnessed during recent years in the case of Tchaikovsky, has not gone far beyond the requirements of the case.

While his work as a whole may be unequal, there is in the finest of his orchestral compositions, which represent him at his best, a breadth and scope, a wealth of inspiration and imagination, and a command of his materials which may well secure for them a far greater measure of favour at the hands of posterity than is anticipated by some of his critics. How, indeed, the title of a great master can be denied to the composer of such things as the Fifth and Sixth symphonies, the *Francesca da Rimini* fantasia, the piano Concerto in B flat minor, and the pianoforte *Trio* it is indeed hard to understand. Tchaikovsky's operas are admittedly on a lower plane than the best of his orchestral writing, though containing many delightful pages, and much the same applies to the songs, albeit these also include many examples—he wrote over 100 in all—of rare beauty and charm.

BIBLIOGRAPHY.—Modeste Tchaikovsky, *The Life and Letters of Peter Ilitch Tschaikovsky*, 3 vol. (1900–02), Eng. trans. by Rosa Newmarch, 1 vol. (1906); Rosa Newmarch, *Tschaikovsky, His Life and Works* (1900); E. Markham Lee, *Tschaikovsky*, vol. ii in the "Music of the Masters Series" (1904). (H. A. Sc.)

TCHEREPNIN, ALEXANDER (1899–), Russian composer, the son of Nicolai Tcherepnin (*q.v.*), was born in St. Petersburg on Jan. 21, 1899. He studied music under his father, Anatol Liadov and others. In 1921 he went to Paris and, after writing somewhat in the style of Prokofiev at first, came under various French influences. His works include two operas, five ballets, incidental music for plays, orchestral music including three piano concertos, chamber music, many other piano works and some songs.

TCHEREPNIN, NICOLAI (1873–1945), Russian composer, was born at St. Petersburg on May 14, 1873. He first studied law, but in 1895 turned to music and entered the conservatory as a composition pupil of Rimsky-Korsakov. In 1901 he became conductor of the Belayev Symphony concerts and soon afterward of the Maryinsky theatre.

During 1908–14 Tcherepnin conducted for Diaghilev's Russian ballet in western Europe. In his ballet *Narcissus*, written for that company, French influences began to show, but he reverted to tendencies which in his work as a whole are seen to range from Glazunov to early Stravinsky. In 1918–21 he was director of the Tiflis conservatory, but in 1921 he settled in Paris. He died at Issy-les-Moulineaux on June 26, 1945.

He wrote five other ballets, two masses, symphonic poems, a string quartet, etc.; but apart from numerous piano pieces and songs his output was not large.

TCZEW (German DIRSCHAU), a town of Gdansk province, Poland; pop. (1950) 24,600; on the left bank of the Vistula, 20 mi. S. from Gdansk (Danzig), at the junction of the main railway lines Danzig-Bydgoszcz and Berlin-Kaliningrad (Königsberg).

The river is there crossed by two fine iron bridges. The chief industries are railway workshops and the manufacture of sugar, cement and agricultural implements. It was at Tczew that some of the first battles of World War II were fought.

The Poles blew up bridges across the Vistula in an effort to halt the Germans advancing from East Prussia and Danzig, but the town was occupied in the early stage of the campaign.

TEA. The name given to the leaves or leaf buds of the tea tree or tea shrub; the manufactured product prepared from the green leaves by fermentation and firing; the drink brewed from the manufactured product; the afternoon social function at which the beverage is served.

The word "tea" comes from the Chinese local Amoy dialect word *t'e*, pronounced "tay." In Cantonese it becomes *ch'a*, pronounced "chah." The name traveled in this form to Japan, India, Persia and Russia. The Dutch brought the form *t'e* to Europe via Java. The English word "tea," originally pronounced "tay" and later "tee," was derived through the Dutch. The word is not to be found in the Bible, the works of Shakespeare, or any other publication in English, previous to the latter half of the 17th century.

Early History.—There is a Chinese legend which places the introduction of tea drinking in the reign of the mythological Emperor Shên Nung, about 2737 B.C., but the earliest credible mention is A.D. 350, when tea was described by Kuo P'o in the *Erh Ya* or ancient Chinese dictionary. The original area of production was in that portion of southeast Asia which includes bordering provinces of southwest China, northeast India, Burma, Thailand (Siam) and Indochina. Cultivation and the use of the beverage spread throughout China and Japan under the patronage of Buddhist priests, who sought a means of combating intemperance. Tea had its first handbook in the *Ch'a Ching*, written by the Chinese scholar, Lu Yu, about A.D. 780. In it the product appears as a kind of cake made from leaves that had been steamed, crushed and molded. To prepare it the cake was fired or toasted and shredded, after which it was steeped in salted boiling water. Tea was taxed under the T'ang dynasty in China A.D. 780.

The earliest notice of tea in Japanese literature dates from A.D. 593, and its cultivation from A.D. 805. Japan also had its book of tea, but not until A.D. 1200.

The favourite legend of the origin of tea concerns Bodhidharma or Daruma, a Buddhist saint, who, having fallen asleep over his devotions, cut off his eyelids on awakening and threw them on the ground, where they took root and grew up as a bush, the leaves of which, when dried and infused in hot water, produced a beverage that would banish sleep.

From China the cultivation of tea spread to Japan from whence

it was introduced into Java in 1684 by Andreas Cleyer, a German naturalist and doctor of medicine. Later, in 1827, J. I. L. L. Jacobson introduced the first tea plants from China into Java and became known as the founder of Indonesia's tea industry. Tea growing in Formosa dates from the early 19th century.

Indigenous tea was discovered in northern India (upper Assam) by Major Robert Bruce in 1823, but the founding of the Indian industry dates from the appointment by Gov. Gen. Lord William Charles Cavendish Bentinck, in 1834, of a committee to formulate a plan for the introduction of cultivation.

In Ceylon, tea dramatically replaced coffee, destroyed by the coffee rust (*Hemileia vastatrix*), in the late 1870s. Russia began growing tea in 1847; Uganda, Nyasaland and Kenya 1900. The U.S. experimented with its cultivation under the direction of Dr. Charles U. Shepard at Summerville, S.C., from 1890 to 1915, when it was given up as impracticable from an economic standpoint. The climate was suitable, but labour was too high, and there was no tariff protection after the Spanish-American War tax of ten cents per pound was repealed (1903).

The first account of tea reached the Arabs A.D. 850; the Venetians in 1559; the English in 1598; the Portuguese in 1600. The Dutch brought the first tea to Europe about 1610; it reached Russia in 1618; Paris, 1648; and England and America about 1650.

Tea was first mentioned in European literature in 1559 as *chai catai,* "tea of China," in *Voyages and Travels* by Giambattista Ramusio, a noted Venetian author. The first printed reference to tea in English appeared as *chaa* in a 1598 translation of the *Travels of Jan Hugo van Lin-Schooten,* a Dutch navigator.

R. L. Wickham, in charge of the English East India company's agency at Firando, Japan, achieved the distinction of supplying the first reference to tea by an Englishman. In a letter, dated June 27, 1615, to the company's agent at Macao, Wickham said: "I pray you buy for me a pot of the best *chaw.*" This is probably the earliest pidgin English for the Chinese *ch'a.*

The earliest mention of the use of milk in tea appeared in a book of travel by the Dutch author, Jean Nieuhoff, who observed the practice in Canton, China, in 1655.

The first newspaper advertisement for tea appeared in *Mercurius Politicus,* London, in 1658. It announced: "That Excellent and by all Physitians approved China drink, called by the Chineans *Tcha,* by other nations, *Tay,* alias *Tee,* is sold at the Sultaness Head Cophee House in Sweetings Rents, by The Royal Exchange, London." About this time metal or leather trade coins, used in lieu of small change, and known as tea and coffee tokens, were much in vogue in London's 2,000 coffee houses.

Samuel Pepys, in his famous diary for 1660, records that: "I did send for a cup of *tee* (a China drink) of which I never had drank before." In 1662 Edmund Waller wrote the first eulogy of tea in English verse and at the same time eulogized England's first tea drinking queen, Catherine of Braganza. The poem begins:

> Venus her Myrtle, Phoebus has his bays;
> Tea both excels, which she vouchsafes to praise.

It is to Anna, duchess of Bedford, that we owe the custom of afternoon tea. She had tea and cakes served at 5 o'clock because, to quote her, she "had a sinking feeling."

After the coffee houses, tea was sold by chemists' shops, glass sellers, milliners, silk mercers and chinaware dealers. There is even a reference to a "short-hand author" selling tea—his only other trade. As late as 1805, grocers in the Strand were weighing tea in apothecaries' scales.

The world's greatest tea monopoly, the English East India company, or, as it was popularly known, John company, flourished from 1600 to 1858, when it was dissolved. Meanwhile, it figured in many adventures, including the introduction of China and India teas into England and America and the enactment of the Tea act of 1773, which caused the Boston Tea Party and helped precipitate the American Revolution. The Tea act attempted to perpetuate John company's monopoly of the tea trade, a thing distasteful alike to British and American merchants. There followed, not one, but six American tea parties.

Here it is interesting to note that, just as England had started out to be a nation of coffee drinkers before John company began its propaganda for tea, America started as a nation of tea drinkers; it was the Boston Tea Party that cast the die for coffee. By mid-20th century the tea and coffee habits of the two countries were reversed; on a per capita, pound basis, England drank about 5 times as much tea as coffee, the U.S. drank about 25 times as much coffee as tea.

One of the most picturesque eras in maritime history was the day of the clipper ships, which transported tea from China to England and America. Its most celebrated chapter was perhaps that of the Great Tea race of 1866, which started at Foochow and ended 99 days later at London after covering 16,000 mi. Out of 11 clipper ships, "Taeping" was the winner at the London dock; "Ariel," the winner at the Downs, ten minutes ahead of her rival.

Three noted American pioneers made their fortunes in the China tea trade. They were John Jacob Astor, Thomas Handasyd Perkins and Stephen Girard.

The advent of the London pleasure gardens in the 18th century brought tea into the open for the first time in England, where it was drunk in public by both sexes. Among the historic London tea gardens, Vauxhall, Ranelagh and Marylebone became famous.

Tea in Literature and the Arts.—Tea and tea drinking have been celebrated in many ways in literature and the fine arts more than 12 centuries. A notable contribution to tea by the Japanese was its glorification in the art and practice of *cha-no-yu,* tea ceremony, literally, "hot water tea." The origin of the teapot is the Chinese wine jug. Dr. Nicholas Brady, chaplain to the court of William and Mary, called tea "the sovereign drink of pleasure and of health," in his poem, "The Tea Table." William Cowper wrote of "the cups that cheer but not inebriate," borrowing the phrase from Bishop Berkeley. Paul Revere, the American patriot, was famous too, for his antitea cartoons and his artistic silver tea service. "Waltzing Matilda" is an Australian billycan tea song. Many beautiful paintings have had tea as their inspiration. They are to be found in the world's leading art galleries and museums along with fine examples of silver and china teapots.

Production.—The world's production of tea in the 1950s totalled about 1,300,000,000 lbs. annually from an area of approximately 2,000,000 ac. These figures do not include production or acreage of China or the U.S.S.R., both of which have no bearing on the world supply as exports outside of the two areas are negligible. Authoritative sources indicated 1951 production in China of 360,000,000 lb. and in the U.S.S.R. of 40,000,000 lb. For many years prior to World War II China was the largest tea producing country in the world, but India assumed that position in 1941, accounting for more than half of the known world total.

While the total yield at best is an approximation, the amount actually placed on the world markets is obtainable from official sources, and according to official reports the quantity exported in 1952 was 918,200,000 lb. Table I shows tea exports of principal countries in 1948–52.

About 1930, increasing surplus supplies of tea on world markets led to unremunerative prices to producers. Negotiations for the restrictions of exports from India, Ceylon and what was then the Netherlands East Indies led to the International Tea agreement which came into operation on April 1, 1933, and which established a quota system of exports on a percentage basis of fixed standard export figures. The agreement was renewed in practically the same form from time to time until March 30, 1950, when a new agreement was reached by India, Ceylon, Pakistan and Indonesia which continued the main features of the previous agreement as regards

TABLE I.—*Tea Exports of Principal Producing Countries*
(In 1,000 lb.)

Country	1948	1949	1950	1951	1952
India	348,652	487,262	401,688	448,941	410,092
Ceylon	295,838	297,573	298,099	305,171	314,495
Indonesia	19,760	48,350	62,850	88,387	70,268
British East Africa	24,372	23,038	28,835	30,058	29,885
Pakistan	32,686	27,245	16,246	47,171	23,971
Formosa	10,485	27,904	16,463	24,963	20,583
Japan	8,884	15,947	15,708	19,179	20,783
Others*	34,177	26,466	35,201	35,885	28,153
Total*†	774,854	953,785	875,180	999,755	918,200

*Estimated.
†Includes among others Indochina, Malaya, Mozambique, Brazil, Southern Rhodesia.

regulation of exports and control of the extension of tea cultivation. After 1940 the export quotas were placed as high as 130% of the standard figure in order to assure unrestricted production of tea while none was available from China, Japan and Indonesia, and in the years following World War II when demand was ahead of the supply. Administration of the International Tea agreement was entrusted to the international tea committee, London.

The bulk of the world's tea is grown in Asiatic countries; Ceylon was second to India in 1952, with 24% of the total world production exclusive of China and the U.S.S.R. Next were Japan 9%, Indonesia 6%, Pakistan 4%, British East Africa 3%, Formosa 2% and other sources 2%.

On an exporting basis in 1952 India also led with 44½% of the world total; Ceylon followed with 34%; Indonesia 7½%; British East Africa 3%; Pakistan 2½%; Japan and Formosa 2% each; and all others about 4½%. Tea is grown commercially in 19 countries.

Consumption.—About two-thirds of the world's known tea production is shipped each year to importing countries and is consumed there barring stocks held in the principal markets, notably London. The largest tea consuming countries are the United Kingdom, which consumes half of the world's annual exports, the United States, Australia and Canada. The British Commonwealth accounts for the bulk of the world's tea trade—more than two-thirds of all imports and about 90% of recorded exports in 1952.

TABLE II.—*World's Consumption of Tea*
(Net imports in 1,000 lb.; per capita in pounds)

Country		1946	1948	1950	1952
United Kingdom	Total	347,840	404,394	357,808	472,901
	Per Capita	8.8	8.16	8.30	9.0
United States	Total	91,667	89,259	113,563	92,749
	Per Capita	.53	.57	.61	.66
Australia	Total	50,623	47,985	60,303	52,956
	Per Capita	6.85	6.5	6.7	6.5
Canada	Total	27,016	34,738	54,876	45,308
	Per Capita	3.2	2.6	3.4	3.4
Egypt	Total	10,778	29,745	35,186	35,760
	Per Capita	.62	1.5	1.75	1.78
French Morocco	Total	8,753	16,129	23,565	27,389
	Per Capita	1.03	1.87	2.74	3.22
Union of South Africa	Total	11,861	17,633	19,734	23,529
	Per Capita	1.58	1.61	1.7	1.9
Iraq	Total	3,860	13,532	15,335	19,000*
	Per Capita	.8	2.82	3.1	3.9
Ireland	Total	16,173	25,804	24,099	17,241
	Per Capita	4.9	8.07	7.9	5.8
Netherlands	Total	11,913	13,115	18,849	16,186
	Per Capita	1.1	1.3	1.8	1.58
West Germany	Total	‡	1,600	3,698	6,058
	Per Capita	‡	.03	1.2	1.5
Others*†	Total	89,416	130,866	136,484	150,117
Total World*		669,900	824,800	863,300	959,200

*Estimated.
†Includes tea imported into producing countries for domestic consumption.
‡Negligible.

The per capita consumption in the principal consuming countries (1952 estimated) was: United Kingdom, 9 lb.; Australia, 6.5 lb.; Ireland, 5.8 lb.; New Zealand, 5.1 lb.; Iraq, 3.9 lb.; Canada 3.4 lb.; French Morocco, 3.2. lb.; South Africa, 1.9 lb.; Egypt, 1.78 lb.; the Netherlands, 1.58 lb.; United States, .66 lb. (W. H. U.)

Botany.—While the tea plant was originally designated by Linnaeus under the binomial system in 1753 as *Thea sinensis*, it is now placed in *Camellia* as *C. sinensis* (family Theaceae). Later Linnaeus recognized two species as previously described by John Hill, viz., *T. viridis* and *T. bohea*. It was erroneously assumed that the latter was a source of black tea, while *T. viridis* was held to produce green tea. In 1843, however, Robert Fortune found that black and green tea were produced from the leaves of the same kind of plant by varying the process of manufacture. There are many synonyms including *Thea bohea, T. viridis, T. cantonensis, T. assamica, Camellia thea, C. theifera* and others. At the same time various authors proposed numerous varieties, but many of these have been shown to be untenable, and there is no agreement among botanists or horticulturists as to varietal limits. A part of the difficulties was due to the facts that early authors had wholly inadequate collections of reference material and that the tea plant in cultivation, like all cultivated plants, varies in characters, dependent on local conditions and on the special qualities for which various strains were selected and propagated.

While the accepted binomial stems from Linnaeus in 1753, the species was well known to earlier European botanical authors, including Gaspard Bauhin (1623), Dr. Jacob Bontius (1658), Leonard Plukenet (1696), Paolo Boccone (1697) and others, for Linnaeus gives no less than ten references to European pre-Linnaean botanical literature. Linnaeus originally proposed two genera, *Thea* and *Camellia*, but it has been clearly shown that two groups cannot be recognized here, because of intergrading characters. The generic name *Camellia* was retained for this group by action of the International Botanical congress. (*See* THEACEAE.) (X.)

The tea plant is an evergreen shrub, which in its natural state grows to a height of 15 to 30 ft., but the tea planter keeps it pruned down to a height of 3 to 5 ft. In its general appearance and the form of its leaf, it resembles the myrtle; its blossoms are white, with yellow anthers, and suggest the wild rose. These are succeeded by the tea fruit, containing three seeds. The plant grows best in the monsoon climate of the tropics, from sea level up to 6,000 ft.; in more temperate zones, where there is danger of frost, it must be kept to low elevations. It does best in the higher altitudes of the tropics. The commercial tea belt of the world is largely confined to a ring of mountains around or near the equator, within 42° N. and 33° S.

Tea usually is grown from seed in nursery beds. On large tea estates the young plants are transplanted to the prepared tea fields when they are 6 to 8 in. high, usually in six months' time. They are set out in rows, 3 to 6 ft. apart. In two years they reach a height of 4 to 6 ft. Then they are cut down to something less than a foot. By the end of the third year they are ready for plucking. Chinese, Japanese and Formosan tea growers (mostly small family gardens) do not set out nurseries but propagate by planting the seeds directly where they are to be grown or by layering, *i.e.*, transplanting. Weeding, cultivation between the rows and pruning go on regularly at different periods for 25 to 50 years, the average life of a tea bush. When in full bearing—about the 10th year—a bush yields as much as a quarter of a pound of leaf in its several flushes during the season.

The quality of the tea is at its best at different seasons in the various tea countries. Teas are plucked all the year around in southern India. Ceylon and Indonesia. Seasonal teas are those coming from northern India, the China greens, congous, Japan teas and Formosa oolongs. The plucking is done by native men and women who strip the leaves directly into bamboo baskets or duck bags at their sides, or who toss them deftly over their shoulders into baskets hung upon their backs by thongs passing over their foreheads. In northern India, during the first flush, only the tips of the new green shoots are plucked. There remain three or four leaves of which two may be plucked with the tip in the second flush and the remainder in the succeeding flushes. In Ceylon, a tea plucker, using both hands, plucks about 30,000 shoots a day. Approximately 3,200 shoots are needed to make one pound of manufactured tea. It takes three to five years to bring a tea garden into commercial production.

Kinds of Tea.—Generally speaking, teas may be divided into three classes: (1) fermented or black; (2) unfermented or green; and (3) the semifermented or oolong teas. Tea comes from practically the same plant in all countries, the differences in the various classes being due to methods of manufacture and local climatic, soil and cultivating conditions. There are about 1,500 different growths and about 2,000 possible blends.

The fermented, or black, teas include the China congou, or "English breakfast," teas, subdivided into north China congou (black leaf) from Hankow and the south China congou (red leaf) from Foochow; India, Ceylon, Java, African and. since 1929, Japan and Formosa blacks. The unfermented, or green teas, include two main varieties—Chinas and Japans; also, such green teas as may be manufactured in India, Ceylon and Indonesia. The semifermented (oolong) teas are obtained from Formosa and Foochow in China.

Manufacture.—Black and green teas result from different manufacturing processes applied to the same kind of leaf. After plucking, the leaf is withered by being spread on bamboo trays in the sun, or on withering tats within doors. The process takes 18 to 24

hours. Next it is rolled by hand or by machines. The object of rolling is to break the leaf cells and liberate the juices and enzymes sealed within. The roll may last as long as three hours. Then it is taken to the roll breaker and green leaf sifting machine and after that fermented in baskets, on glass shelves or on cool cement floors under damp cloths for ½ to 4½ hours. The firing process (drying) follows, in pans or baskets or in firing machines. It takes 30 to 40 min. The difference between black tea and green tea is the result of manipulation. Green tea is manufactured by steaming without fermentation in a perforated cylinder or boiler, thus retaining some of the green colour. Black tea is allowed to ferment after being rolled and before firing. In the case of back tea the process of fermentation, or oxidation, reduces the astringency of the leaf and, it is claimed, develops the colour and aroma of the liquor. In making green tea, the fermentation process is arrested by steaming the leaf while it is green and by light rolling before drying.

Semifermented or oolong teas have some of the characteristics of both black and green teas. They are allowed to wither in the process of manufacture and partially ferment.

After firing, teas are graded by being put through cutting, sifting and sorting machines, before being packed in wood or metal tea chests. The principal grades are broken orange pekoe, consisting of terminal buds with portions of the finest leaf that have broken off in the manufacture; orange pekoe, consisting of some terminal buds and the finest leaves; pekoe and pekoe Souchong, consisting of coarser leaves; pekoe dust, dust or fine broken tea.

Trade Characteristics.—The teas that grow and are plucked all the year 'round are those of southern India, Ceylon and Indonesia. The best qualities in southern India are the fermented teas produced in December and January. In Ceylon the finest quality fermented teas are produced in the months of February, March, April, July, August and September. Ceylon teas are known chiefly by their garden marks. Fermented teas of Indonesia are best in July, August and September. These teas are also known as Javas or Sumatras or by their garden marks.

Northern India's fermented teas are best when made from the second flushes of July, August and September. In Assam and the Duars, however, the finest teas are made in November and December. Indian teas are named after certain districts, the best known being Darjeeling. The first pickings from April to October of the north and south China blacks are the best. The north China congous are the burgundies of China teas; the south Chinas are the clarets. The best China greens are made in June and July although they are obtainable from June to December. In Japan the season is from May to October. There are several crops, but the first and second (May and June) are best. Japans are the white wines of teas. In Formosa there are five recognized crops, viz., spring, summer (two), autumn and winter. The season extends from April to November. First and second summer crops (June and July) are the finest. They have been called the champagnes of teas.

Chemical Aspects and Tea Brewing.—The principal chemical constituents of tea are caffeine, tannin and essential oil. The first supplies the stimulating quality, which, because of its purity, is without harmful reaction; the second, the strength of body; and the third, the flavour and aroma, chief factors in determining cup value in a trade sense.

Arguments against the healthfulness of tea are frequently founded on the experience of individuals or on the assumption that tea is taken in excess; however, tea is the beverage of people in many parts of the world, living in every kind of climate, and generalizations are dangerous.

In 1935 Dr. C. R. Harler, a chemist of the Indian Tea association, said: "A cup of tea contains, on the average, a little under a grain of caffeine and about two grains of tannin. The medical dose of caffeine suggested in the British Pharmaceutical codex is 1 to 5 grains and of tannin 5 to 10 grains. It will be realized, therefore, that in the average cup of tea these two most important constituents of tea are present in very small amounts, especially when it is remembered that the caffeine is injected gradually and the tannin is fixed by proteins during its journey through the alimentary tract. The infusion is faintly acid, almost neutral. Gastric juice is a thousand times as acid.

"When milk is added to tea, the tannin is fixed by the casein in the milk. Sugar added to the tea merely sweetens it and adds to the value of the drink as a food. The addition of milk to tea robs it of practically all its astringency.

"When the infusion is drunk, it passes first into the stomach, where the sugars are absorbed as ordinary foods, and the ingestion of caffeine begins. The comforting effect of the warmth of the drink is at once felt, but the stimulus due to the caffeine comes about a quarter of an hour later."

TABLE III.—*Chemical Analysis of Leading Teas**

	Caffeine per cent	Tannin per cent	Soluble solids per cent
Java (Indonesia) black teas	2.7–4.4	6–20	16–26
Japan green teas	2.0–3.3	4–12	16–26
China black teas	2.0–3.7	5–10	16–22
Formosa oolong teas	3.1–3.7	12–23	23–25
India black teas	2.0–3.0	6–10	22–25

*Five-minute infusion.

Source: J. J. B. Deuss, *L'Agronomie Coloniale* (Paris, 1924), No. 80 and C. R. Harler, *The Culture and Marketing of Tea* (New York, 1933).

In the British Isles, Australasia, North America and the Netherlands, where most of the black teas of India, Ceylon and Indonesia are consumed, the beverage is drunk primarily for its stimulating effect. The alkaloid caffeine, responsible for the stimulus, has a slightly bitter taste when taken in medicinal quantities, but the small amount present in the usual cup of tea is practically tasteless. Usually, sugar, milk or cream is added.

In Russia a pot of strong China tea is brewed and put in the samovar. A little of the extract is put in a glass, hot water added; also a slice of lemon, sugar or jam. Americans have made iced tea a favourite summer drink.

The ideal preparation of tea is one which extracts a maximum of caffeine and a not excessive amount of tannin. Such a preparation also conserves the aroma and flavour—evanescent qualities easily lost by careless preparation. From the chemist's point of view the two main essentials in tea brewing are fresh water, freshly boiled, and a three- to five-minute infusion. The tea used should be suitable to the water of the locality in which one lives. Tea infuses more readily in soft water than in hard. Small, thick, full teas are required for soft water; brisk, full, flavoury teas where the water is hard. Liverpool water, which comes from the Welsh mountains near by, is said to be the softest water in England and the best for making tea. Water fresh from the tap is the first essential. The second is to bring it to a bubbling boil and pour it at once upon the tea leaves. To delay means that the water will go flat, and flat water will spoil the finest tea. Hence the discouraging concoctions on steamships and railways.

Generally speaking, a five-minute infusion is best for fully fermented teas like those of India, Ceylon or Indonesia, but George F. Mitchell, a U.S. supervising tea examiner, reached a slightly different conclusion in a series of tests in 1904–05 at the U.S. department of agriculture.

These investigations demonstrated that a perfect cup of tea from a chemical standpoint was, unfortunately, not a perfect cup of tea from the consumer's standpoint. From a chemical standpoint the length of time in which the maximum amount of caffeine and total soluble matter was extracted, with the minimum amount of tannin, averaged three minutes after boiling water had been poured on the leaves. After that time more tannin was extracted, with only a small amount of caffeine. Usually, however, this makes a very "skinny" cup of tea, lacking body and a certain amount of pungency that is desired by all tea drinkers; and, of course, if cream or milk is added to this beverage, it takes away from the small amount of pungency, so that the cup appears even less desirable.

For this reason Mitchell concluded it best to brew tea, that is to be drunk without cream or milk, from three to four minutes, and that to be drunk with cream or milk from four to five and even six minutes, as in some teas the real flavour does not come out until after a six-minute infusion.

Stated in general terms for all loose teas, this, then, is the best practice: (1) buy the highest grade of the kind of tea suited to your taste and the locality where you intend to use it; (2) use

freshly drawn, slightly soft or slightly hard cold water from the tap or faucet; (3) bring it to a bubbling boil; (4) allow one rounded, standard teaspoonful of tea for each cup of tea required; (5) pour freshly boiling water over the tea leaves in a heated earthenware, porcelain, or glass pot and let them steep for three to five minutes, depending upon the kind of tea used. Stir while infusing; (6) pour off the liquor into another heated china pot; never use the leaves a second time.

BIBLIOGRAPHY.—C. P. Cohen-Stuart, "A Basis for Tea Selection," *Bull. Jard. Bot. Buitenzorg*, iii, 1:193–320, pl. 21–31, bibliography pp. 306–319 (1919); Kakuzo Okakura, *The Book of Tea*, 14th ed. (New York, 1928); F. W. F. Staveacre, *Tea and Tea Dealing* (Toronto, 1929); H. J. Moppet, *Tea Manufacture: Its Theory and Practice in Ceylon*, 2nd ed. (Colombo, 1931); E. C. Elliott and F. J. Whitehead, *Tea Planting in Ceylon* (London, 1931); H. K. Rutherford, *Planters' Note Book*, 8th ed. (London, 1931); Agnes Repplier, *To Think of Tea!* (Boston, 1932, London, 1933); C. R. Harler, *The Culture and Marketing of Tea* (New York, 1933); William H. Ukers, *All About Tea* (New York, 1935), *The Romance of Tea* (New York, 1936); J. J. Fox, "Caffeine, Tea and Coffee," *Allen's Commercial Organic Analysis*, vol. vii (1930); E. A. Houghton Roberts, "Chemistry of Tea Fermentation," *Advances in Enzymology*, vol. ii, p. 113 (1942); V. D. Wickizer, *Coffee, Tea and Cocoa* (Stanford, 1951), *Tea Under International Regulation* (Stanford, 1944); Commonwealth Economic Committee, *Plantation Crops* (London, 1950); Ernest Hainsworth, *Tea Pests and Diseases and Their Control* (Cambridge, 1952); R. D. Morrison (International Tea Committee), *Tea: Memorandum Relating to the Tea Industry and Tea Trade of the World* (London, 1945); E. A. Curtler, *The Cultivation and Manufacture of Tea in Ceylon and India* (1932); International Tea Committee, *Bulletin of Statistics* (London, 1953); Morris B. Jacobs, ed., *The Chemistry and Technology of Food and Food Products*, vol. ii, pp. 1683–1705, 2nd ed. (New York and London, 1951). (W. H. U.)

TEA-CADDY, a box, jar, canister or other receptacle for tea. The word is believed to be derived from *catty*, the Chinese pound, equal to about a pound and a third avoirdupois. The earliest examples that came to Europe were of Chinese porcelain, and in shape resembled the ginger-jar. They had lids or stoppers likewise of china, and were most frequently blue and white. The English kilns at first imitated them, but speedily devised forms and ornament of their own.

But tea-caddies were not for long confined to porcelain or faïence. They were presently made in a variety of materials, and in an equal variety of shapes.

TEA CEREMONY, a time-honoured institution in Japan, is rooted in the principle of Zen, a sect of Buddhism, and is founded upon the adoration of the beautiful in the daily routine of life, such as preparing a meal and tea, cleaning the house and garden, etc. The Japanese word for the ceremony, *cha-no-yu* (literally "hot water tea"), means an aesthetic way of entertaining guests, who, after they have been served a meal, are served thick, weak tea made of pulverized tea leaf stirred in hot water. Everything is done according to the established rules of etiquette. The entertaining takes place in the tea room (*chaseki*), usually a small building constructed to suggest refined poverty, with great forethought given to the choice of material and the construction. The room is usually about nine feet square, or smaller. It contains an alcove for ornaments and a small sunken fireplace, which is used for heating the kettle during winter; in summer a brazier with a charcoal fire is used.

Each guest enters the room in humility by crawling through a small opening less than three feet square.

The founders of the tea ceremony emphasized the following four qualities: (1) harmony among guests and utensils; (2) respect, not only among the participants but also for the utensils; (3) cleanliness, derived from Shinto, the ancestral religion of Japan, which requires each participant, before entering the *chaseki*, to wash his hands and rinse his mouth as symbolic gestures of cleansing himself externally and internally; and finally, (4) tranquillity, which suggests the indescribable quality of mellowness which is given to the article by long and fond use.

BIBLIOGRAPHY.—E. S. Morse, *Japanese Homes and Their Surroundings* (1885); F. F. Brinkley, *Japan—Its History, Arts and Literature* (1901); Kakuzo Okakura, *The Book of Tea*, new ed. (1925). (J. HAR.)

TEACH [THATCH or THACH], **EDWARD** (d. 1718), English pirate, popularly known as Blackbeard, is believed to have been born at Bristol. He is said to have gone out to the West Indies during the war of the Spanish Succession, to have engaged in privateering, and after the declaration of peace (1713) to have turned pirate, but he is not actually heard of in this capacity till the end of 1716. The following year he captured a large French merchantman, rechristened her "Queen Anne's Revenge," and converted her into a warship of forty guns. His outrages in the Spanish main, the West Indies, and on the coasts of Carolina and Virginia, quickly earned him an infamous notoriety. He made his winter quarters in an inlet in North Carolina, the governor of which colony was not above sharing in the proceeds of his crimes, but the governor of Virginia at last dispatched two sloops, manned from the British warships on the station, to cut him out.

On Nov. 22, 1718, Lieut. Maynard boarded Teach's sloop, after a sharp fight, and himself shot the pirate dead.

TEACHERS, TRAINING OF. Teacher training is generally concerned with four matters: (1) preparation in subject matter; (2) professional training, including psychology, methods and practice teaching; (3) experience and training in the social functions which a nation considers significant for teachers, or in which it wants all citizens to become experienced; and (4) further training in new methods and ideas. In this article the focus will be on teacher training for public elementary and secondary schools.

HISTORY

The teacher training program historically has received attention as a result of problems in primary or secondary education, and has therefore been developed in terms of needs for improvement of the total school system.

France and Germany were beginning teacher education as early as the 17th century and have been most influential in shaping patterns of training in other countries. With the secularization of schools in Germany, work in teacher education was greatly extended through the influence of Wilhelm von Humboldt, who was named minister of education in 1809. He helped establish a planned educational program, set up regular standards and examinations for teachers and encouraged the development of pedagogical seminars. In 1826 a year of trial teaching was established as a requirement of all candidates for teaching positions. In 1831 examinations in pedagogy as well as in the main subjects of the higher school course were designated as requirements for secondary school teachers.

Teacher education in the United States gained acceptance largely between 1820 and 1865. It was in the northeast that the normal school (the original name for teacher-training schools) took root. The common school with equal opportunity for people of all classes led inevitably to support of a normal school to train teachers; the widening range of ability of students attending school brought an emphasis on teaching methods. Thus because of widespread establishment of common schools, Samuel R. Hall opened his private normal school at Concord, Vt., in 1823. In 1839 the first of the public normal schools was established by the state of Massachusetts. Traditionally, however, the training of teachers had been one of a purely liberal education, and the effort to professionalize teacher education started a continuing controversy. The Yale catalogue of 1829–30 required completion of a liberal education appropriate for all before the beginning of any professional training. However, Thomas Jefferson at the University of Virginia and George Ticknor at Harvard began experimenting in elective programs with subjects of utilitarian value for preprofessional work, and this influence helped the new normal schools to establish themselves.

The training of teachers dates, in England, from the beginning of the 19th century, when the National Society for Promoting the Education of the Poor in the Principles of the Church of England and its Nonconformist rival, the British and Foreign School society, first attacked the problem of elementary education on a national scale. Both these societies adopted the "monitorial system," the principle of which was that picked older scholars should

learn their lessons from the adult teacher in charge of the school and hand them on to the younger children. Some training of the instruments of this system was found to be necessary and training schools were accordingly set up to give it. One of these developed into the Borough Road training college, which is still in existence. About 1840 James Kay (afterward Sir James Kay-Shuttleworth), a far-seeing official to whom, more than to any other single man, the English educational system is due, began the movement which replaced the monitors by "pupil-teachers"; *i.e.,* boys and girls who were regularly apprenticed for a period of five years from the age of 13 and both learned the art of teaching and continued their education under the head teacher of an elementary school. Kay's scheme also involved training colleges where the professional and academic education of at least the best pupil-teachers could be carried to a higher stage after their apprenticeship was completed. He himself founded the college which, as St. John's college, Battersea, London, was taken over by the National society and furnished a model widely imitated.

These early patterns of teacher education had their beginnings at a time when educational responsibility rested primarily with the church. When governments began to assume civil responsibility for establishing and maintaining schools, they also began to determine the preparation and qualification of teachers. As the percentage of children enrolled in school in the United States increased, and higher numbers continued into the secondary schools, standards of certification were raised and collegiate study was increasingly expected of elementary as well as secondary school teachers. The period from 1920 to 1930 has been described as that in which normal schools became teachers colleges in answer to this expectation. In England the first major development toward increasing educational opportunity came in 1918, but it was not until 1944 that it assumed significant proportions.

IMPACT OF UNIVERSAL EDUCATION

In the 20th century, as the privileges of education were extended to more and more people throughout the world, certain universalities in teacher preparation began to appear. These were recognized through the activities of the International Bureau of Education and later of the United Nations Educational, Scientific and Cultural organization (UNESCO) in promoting international exchange of information and the systematic gathering of educational data.

The predominant institution for the education of elementary teachers was a normal school following completion of the elementary school. Generally these institutions gave courses in both general and professional education. A few countries offered education for elementary teachers at the level of higher education, either through the university or through higher schools of education. Still other countries offered a combination of work beginning at the secondary level and extending into the higher education level.

The preparation of secondary school teachers was both academic and professional. Academic preparation was provided for in several ways: (1) through work at the regular universities; (2) through work at special institutions providing both academic and professional training; (3) by bringing in persons with special proficiency (engineers, for example, might teach mathematics, or pharmacists, chemistry); (4) in special schools for music, physical education or art; (5) in countries with no higher educational institutions, teachers were sent abroad for training. Professional training was provided in a variety of ways.

Just as the beginnings of teacher education took place almost simultaneously in the various countries of western Europe, so in the 1940s there was a nearly universal review of the adequacy of education in an effort to extend opportunity to a larger number of the population in the various countries (*see* ILLITERACY; SECONDARY EDUCATION). The British Education act of 1944 delineated the responsibility of the local education authorities to provide for primary, secondary and further education for all pupils suitable to their "ages, abilities and aptitudes" and including appropriate practical instruction and training. Similar plans for the extension and modification of public education were expressed in the report (1946) of the Commission Ministerielle d'Etude, La Reforme de l'Enseignement in France, which furnished guidelines for modification of the French educational program.

It was at this time that article 26 of the United Nations Universal Declaration of Human Rights, which states that everyone has a right to education, made its influence felt in the recognition throughout the world of the need to extend compulsory education. The effect of these developments on teacher education was direct. Extension of teacher education was needed because of the extension of compulsory education, the increases in population of school age, and the extension of the years of schooling. In Costa Rica, for example, there were 3,500 students in secondary schools in 1948; 17,500 in 1956. In Afghanistan in 1954–55 there was a 17.3% increase in enrollment in village schools and a 31% increase in enrollment in secondary schools over the previous year. Enrollment in primary schools of Iraq was 90,000 in 1944; 258,000 in 1954. In Syria the 1944 enrollment was 77,000; in 1954, 245,000. The result of this increase in school enrollments on teacher training institutions is shown in figures of the 1953 International Conference on Public Education. Nine of the 66 countries reporting opened new primary teacher training institutions, and seven had extended the period of training for primary teachers during that one year alone.

The International Bureau of Education has not only brought together information such as that given above, but has, since 1931, conducted a yearly International Conference on Public Education, after 1945 in co-operation with UNESCO. As early as 1935 pressures were felt on teacher education programs, and the 1935 conference was devoted to problems of the training of primary teachers. In 1953 the problem of primary teacher education was again the subject of the conference, while in 1954 problems in the training of secondary school teachers were considered. The recommendations of these conferences reflected practice and also served as guideposts to ministries of education in formulating procedures in teacher education. These recommendations revealed common directions in systems of training and curricula which were generally considered to be desirable:

1. There appeared to be a definite trend in many countries to move away from the wide variety of teacher training institutions and toward the university or its equivalent for training both primary and secondary school teachers.

2. There was an attempt to extend the training of each group toward a longer period of general education for primary school teachers and toward more professional training for secondary school teachers.

3. There was recognition of the desirability of increasing contacts between primary school and secondary school trainees to facilitate continuity and transfer between schools.

4. There was general agreement that the qualifications for teaching should include social and psychological characteristics beyond those of intellectual ability alone.

5. There was recognition of the need for extensive psychological training for all teachers.

6. There was agreement that teacher training should include a strong practical aspect covering experience in various types of schools, extended probation under supervision and experience with extracurricular and community activities.

7. There was emphasis on training teachers to select and use a variety of methods appropriately.

Courses and Requirements.—The report on *Secondary Teacher Training* (UNESCO, 1954), giving responses of 58 countries to a questionnaire, listed the professional courses for secondary teacher training at that time under three headings: education, psychology and practice teaching. Prevalent among education courses were those which dealt with teaching methods and a general orientation to education. The psychology courses provided for general psychological orientation and child or adolescent psychology. Opportunities for experimental and practical research were numerous. Practice teaching was uniformly provided but differed in the length of time and type of work done by the student. In Austria a year of probationary service under the guidance of experienced teachers was provided. In the second half of the

year probationers were wholly responsible for a given subject in a class. They took part in all school activities throughout the year. In the Federal Republic of Germany practice teaching included two years of preparatory service following academic studies with two four-week periods of practical work in a school during vacations. Training included all aspects of school life.

The *Training of Primary Teacher Training Staffs* was prepared for the International Conference on Public Education in 1957. It presented descriptions of the subjects that were taught in primary teacher training establishments in 63 countries. Examination of the reports indicated the following as persistent requirements in the area of general education: mother tongue, a second language, history, mathematics, sciences and geography. Professional subjects closely approximated the course offerings indicated in the requirements for preparation of secondary teachers. More specialized provision was made for the study of child psychology, together with more emphasis on methods. Persistent offerings in special subjects were physical culture, drawing, penmanship, music, handwork, domestic economy, agriculture and rural studies, health and arts and crafts. These subjects reflected a trend to fit teachers with skills that would enable them to contribute to national needs in agriculture, health and domestic patterns. Furthermore, they indicated that primary education universally dealt with a wider variety of cultural emphases than reading and writing alone, and that these subjects were of sufficient significance for recognition in the curricula of teacher education.

The four-year course in pedagogical schools for the training of primary school teachers in the U.S.S.R. in 1955–56 revealed many similarities when compared with the four-year course of the State Teachers college at Oneonta, N.Y. (now State University of New York College at Oneonta). In the U.S.S.R. the teachers took Russian language and methods of teaching, literature, mathematics and methods of teaching, physics, chemistry and mineralogy, natural science and methods of teaching, geography and methods of teaching, history and methods of teaching, history of the Communist party, psychology, pedagogy, history of pedagogy, logic, school hygiene, penmanship and methods of teaching, drawing and methods of teaching, singing and methods of teaching, physical education and methods of teaching, practical work in shop and agriculture, observation and practice teaching, and electives. In Oneonta the course included art, child development, written composition and speech, mathematics, music, physical education, introductory science, biology, history, the child and his curriculum, health education, industrial and practical arts, literature, geography, U.S. history and government, practice in elementary school teaching, seminar in elementary education, sociology, and electives.

The fundamental difference in the two programs was in the emphasis which the Soviet program gave to teaching methods in direct relation to subjects, while the Oneonta course gave separate consideration to the child and his curriculum, which was a two-year integrated methods course.

Standards.—The quality of teacher training has been of much concern. In England the area training organizations (A.T.O.) have been given responsibility for all types of students in training. The A.T.O. "involves the pooling of resources and facilities by the universities, the training colleges, and the local education authorities into a teacher-training center usually based administratively in a university and comprising a system of schools, institutes, faculties, and teachers in service working together." (Herbert Schueler, "Education in the United Kingdom of Great Britain," p. 142, *Comparative Education,* ed. by A. H. Moehlman and J. Roucek, Dryden Press, 1951.)

The National Council for the Accreditation of Teacher Education in the United States has worked on such problems as setting up criteria for selection of candidates in teacher education, the balance between general and professional courses, adequate supervision of practical teaching, selection of laboratory experiences and the encouragement of research to determine the effectiveness of applied criteria.

A threat to the quality of training has been the shortage of teachers. At the same time that attention was being given to improving quality and extending time for programs of teacher training in a majority of countries, those same countries were forced to set up emergency training measures. Many countries provided increased living allowances for students in training, in order to encourage more people to enter the profession. Others temporarily shortened the training periods. Appointments were made with required part-time study to bring the emergency teacher up to standards. In the United States the National Science foundation equipped traveling science laboratories which served the double purpose of giving enriched experience to children and training in contemporary science to teachers. The continental classroom television program was another effort in the United States to bring current information in specialized fields to teachers already in service.

Supplementary Education and Training.—As in other professions, newer methods and new developments in the fields of knowledge necessitate further education for teachers with experience. In many countries seminars or institutes are conducted by the government. The content reflects the major concern of the governmental agency. For example, Czechoslovakia, Belorussia and the U.S.S.R. required a summer training course for teachers to prepare for greater emphasis on practical and polytechnic training in the secondary curriculum. Materials centres either in a central place or in a bookmobile have been used to bring new ideas to teachers on the job. In the United States local school systems have frequently sponsored courses for their own teachers. (*See* also SUMMER SCHOOL.)

No article on teacher training would be complete without attention to the significant role of teachers associations in improving the quality of teaching and of teacher training. Even before education became a civil responsibility, members of the profession met to share ways of improving their work. They have cooperated with government and institutions of higher learning in modern efforts. They work with the A.T.O. in England, and they have developed significant and influential plans for educational improvement in Ireland, Scotland, Australia and Canada. In Ontario, for example, they have conducted summer workshops for teachers and developed a permanent training centre. The National Education Association of the United States with its various affiliates has conducted continuous research for the improvement of the profession. (*See* also EDUCATION, SOCIETIES OF.)

Other significant developments in further education include research centres. In Brazil, for example, the Brazilian Centre of Educational Research was established in 1955. "The purposes of these centres are: (1) to study cultural and educational conditions and trends in the development of each region and of Brazilian society as a whole, with a view to formulating, by degrees, a national education policy; (2) to draw up plans and make recommendations and suggestions for the reform of Brazilian education in every region and at all levels; (3) to prepare reference works, textbooks and educational materials and to conduct special studies on school administration, curricula, educational psychology and philosophy, the training of teachers and any other matters connected with the improvement of teacher training; (4) to arrange training and refresher courses for school administrators, those responsible for educational guidance, educationists and primary teachers in teacher training and primary schools." ("Education in Brazil," *Education Abstracts,* vol. x, no. 9, p. 6, UNESCO, Paris, Nov. 1958.)

Newer communication media have also influenced teacher training. The use of video tape made it possible for a teacher to teach a group of students, then view his own work immediately, and thus perhaps improve his teaching. Tape recorders have been used the same way in addition to their use in speech and foreign languages.

The 20th century has seen the wide use of international exchange in the training of teachers. The technical assistance programs of UNESCO, the International Cooperation administration and the Fulbright scholarship programs in the United States, the British council, the Ford and Rockefeller foundations, the American Friends Service committee and the Southeast Asian Board of Scholarships and Exchange of Professors have made international

exchange significant in the training of teachers throughout the world. These exchanges helped to establish certain common directions in teacher education, but they also emphasized the fact that teacher education must be rooted in the culture and expectation of the particular country in which it functions.

See also Elementary Education; Pre-Elementary Education; Secondary Education; Teaching, Methods of.

Bibliography.—*General:* U.S. Department of Health, Education and Welfare, Division of International Education, *Education in the USSR*, Bulletin No. 14 (1957); I. L. Kandel, *New Era in Education* (1955); UNESCO, *Studies in Compulsory Education*, vol. i-xvi dealing with this topic in separate nations or geographic areas (1951–56).

Teacher Education: Merle D. Borrowman, *The Liberal and Technical in Teacher Education* (1956); C. A. Richardson, Helene Brule and Harold E. Snyder, *Education of Teachers in England, France, and the U.S.A.* (1953); UNESCO, International Bureau of Education, No. 117 *Primary Teacher Training* (1950); No. 155 *Secondary Teacher Training* (1954); No. 151 *XVIth International Conference on Public Education*, Proceedings (1953); No. 159 *XVIIth International Conference on Public Education*, Proceedings (1954); No. 188 *XXth International Conference on Public Education*, Proceedings (1957); No. 182 *Training of Primary Teacher Training Staffs* (1957). (T. D. R.)

TEACHERS' PENSIONS: *see* Pensions: *Teachers' Pensions.*

TEACHING, METHODS OF. Teaching may be defined as the stimulation of learning. Learning is an active process carried on through doing, reacting, undergoing; the learner participates in his own education. The outcomes of learning are facts, generalizations, attitudes, appreciation, skills and abilities. (For the psychology of learning *see* Learning.)

Methods of teaching include those things that are done to stimulate learning. These methods must of necessity take into account the nature of learning and of the learner, the objectives or desired outcomes of learning, and must differ with learners of differing levels of maturity and with the differing objectives of general and special education.

Method, Technique and Device.—Method is any organized, systematic procedure through which any activities that go beyond the routine are carried on. Techniques and devices are parts within a method: they are subsidiary procedures designed for a specific purpose, item or moment. Systematic method is of particular importance in the fields of knowing, thinking, learning and teaching.

Considerable confusion exists in the educational world because of misuse of the word "method." There are repeated references to:

the recitation method	the lecture method
the study method	the socialized method
the textbook method	the drill method
the individual method	the group method
the project method	the visual method

None of these is a method; they are subsidiary techniques, or devices.

Early Methods.—Learning in primitive times was through imitative play and direct participation. The learner's interest was spontaneous; his method, trial and error. There was not, in any modern sense, a method of teaching. Advances in the complexity of living, with the rise of an intellectual tradition of tribal history and religion, brought about the development of primitive teaching methods. Demonstrations of skills to be performed, plus enforcement of drill or practice, appeared. Tribal history and religion was lectured, chanted in sagas and dramatized in the form of priestly accounts, songs and pageants. Severe tests of courage, some with torture and pain, were thought to teach courage and devotion to the values of the tribe.

Eventually, as civilizations progressed, textbooks developed to set forth the facts to be learned; learning was by memory and by imitation. Teaching methods consisted of setting tasks, enforcing practice or memorization, and testing for results.

Aristotle (384–322 B.C.) set forth in his *Organon* the system of deductive thinking and learning (*see* Aristotle; *The History of Aristotle's Writings and Philosophy*). Deductive logic arrived at many erroneous beliefs because the original premise was wrong. The deductive method is, none the less, a sound teaching method, particularly in specialized education and with advanced students. The basic principle, generalization or law is set down first, together with illustrations. The student is lead to analyze the principle and examples, and then to find new ones in the field of study.

Francis Bacon (1561–1626) in his *Novum Organum* (1620) presented the method of inductive verification of hypotheses. Knowing and learning started not with acceptance of a stated definition or law but with observation of the world around the student. Scrutiny of the facts and of problems related thereto called for the development of hypotheses to guide the search for further facts. Hypotheses were developed both by reasoning and by experimental tryout. Definitions, laws and theories were derived, not accepted on authority. Modern scientific knowledge resulted from this reform in logic. A new approach to knowledge, a new method of knowing had emerged. The inductive process now stood alongside the deductive. The method of teaching ultimately based on this new development directed students to observation and analysis of phenomena in the world, with comparisons over a series of specific instances leading to generalization. (*See* also Scientific Method.)

Francis Bacon is the man chiefly associated with the new method but many others were involved: Roger Bacon (*c.* 1220–*c.* 1292), John Locke (1632–1704), Thomas Hobbes (1588–1679). René Descartes (1596–1650) wrote *Discourse on Method* (1637), still of value today. The early writers, to be sure, including Francis Bacon, made errors. Certain things about induction were overlooked to be corrected by later writers. All the shortcomings affected early methods of teaching. Two centuries later another unfortunate error arose. Thomas H. Huxley (1825–1895) urged his contemporaries to "sit down and learn from the facts." The significant word "from" was overlooked. Teaching method concentrated, and still does in some schools, on pounding in facts to be memorized and given back in recitation. Facts by themselves are inert, dead, meaningless; to have meaning they must be related to other facts, to hypotheses, to theories. Facts must be functional and learned in functional situations. Modern methods have gone a long way toward correcting this error.

Space prohibits inclusion of the interesting account of the relation between these basic methods and the derived methods of teaching. Part of the story is apparent in this abbreviated account.

Development of Modern Methods.—Methods of teaching consisted largely of setting individual tasks, materials to be learned (memorized) and recited individually to the teacher. Copies were set to be imitated. Repetitive drill was a prominent method.

The beginnings of change came with the Institute of the Christian Brothers (*q.v.*) in France (1684) and with Andrew Bell (1753–1832) and Joseph Lancaster (1778–1838) in England. The Christian Brothers introduced manuals for teaching setting forth methods to be used, and developed methods for simultaneous class instruction. Lancaster and Bell similarly developed methods for teaching classes instead of individuals. Despite some advance, the methods of teaching were still largely devoted to hearing recitation of material committed to memory and to directing repetitive drill.

Johann Comenius (1592–1670) presented in theory many elements of modern methods though he had little effect in his own time. He related teaching method to the maturity and development of the learners and stressed preparing the learners for the lesson to come, in contrast to assignment without explanation; and indicated the values of gentleness, persuasion and pleasantness in contrast to harsh authoritarian methods. He advocated class instruction and he was one of the first to warn against fatigue in learners, which defeats the best methods.

Jean Jacques Rousseau (1712–78) is usually cited as the originator of the modern period even though he was not a teacher, nor did he himself develop systematic methods. His stress on the periods of development in children, on their interest, curiosity and capacities, on the necessary gradation of materials, while discussed as theory, influenced many new developments. J. H. Pestalozzi (1746–1827) also advocated adaptation to the nature of children, using methods based on love and kindness. He also advocated experimentation to discover new methods of teaching. He introduced study objects with induction to be based on them. Primarily he urged and practised group instruction in place of

individual recitation. The teacher was a part of the group and not a taskmaster, or leader apart.

J. F. Herbart (1776–1841) developed the "five formal steps" of preparation, presentation, association, generalization and application, or the inductive-deductive procedure, as an organized method of teaching, widely used in Europe and the United States for several decades. He also stressed the interest and intellectual background of the learner as factors in determining methods to be used.

Friederich Froebel (1782–1852), chiefly associated with the kindergarten and the teaching of very young children, stressed the nature of childhood, kindness and social participation as factors in developing methods.

In the United States, William James (1842–1910) emphasized the biological point of view, that of the learner as a behaving organism. Clues to method were to be derived from observation of the activities of the learner. The philosophic analysis of ends and values by John Dewey (1859–1952), B. H. Bode (1873–1953) and W. H. Kilpatrick (1871–) contributed strongly to developing methods. E. L. Thorndike (1871–1933) and C. H. Judd (1873–1946) corrected overemphasis upon motor activity, recalled attention to the importance of central mental activity and stressed the scientific approach. The latter gave basis for adapting methods to individual differences and to problem solving and thinking as methods superior to drill and memorization.

Argument between two schools of thought, teacher-centred or child-centred education, made profound contributions to teaching method along many lines already indicated by earlier theorists. The modern view is that teaching methods are based on original, inventive and creative thinking by teachers, and these methods take due account of both the nature of the learner and the demands of society.

See also ELEMENTARY EDUCATION; EDUCATION, HISTORY OF.

General Method of Teaching.—The two dominant general methods of teaching are (1) the assign-study-recite-test procedure; and (2) the unit. Each of these is useful for differing purposes, with differing levels of maturity. Each has values, and each is subject to grave abuse; each is dynamic and improvements appear regularly.

In the historical development of the assign-study-recite-test method, movement has been steadily away from a routine, rule-dominated method toward an increasingly functional one. The assignment was first the arbitrary assignment of pages, exercises, topics or chapters. The Herbartians added intellectual preparation through recall of related information to orient learners. Differentiated assignments and two- and three-track systems appeared. Problems and projects with varied assignments were developed to contribute to a total learning experience. The study period was at first unsupervised study (memorization) of assigned text material. Supervised study of assigned materials appeared, followed by supervised study of supplementary references. Home-room guidance of study is widely used, sometimes with study coaches for individuals and small groups, and formal teaching of study procedures is used.

Recitation at first consisted of individual answers to questions, later individual and group reports, and still later the socialized recitation with some varied activities. The term "work period" is used instead of recitation by many schools. Testing, traditionally through arbitrarily selected essay questions, gave way to standardized tests, which in turn gave way to informal homemade tests and to greatly improved essay questions calling for understanding.

The modern unit method is an effort to align school experience with life experience. Many developments hinted at by earlier writers affect this. Units are initiated, preferably through co-operative selection and definition of a purpose, by pupils and teacher. An organization for study, work or problem solving, and for the distribution of work and responsibility, is also arranged co-operatively. The working period of a unit may include any type of learning activity, depending upon the objective. Individuals and small committees carry on various activities; they may consult printed sources, interview persons, listen to lectures, make excursions, perform experiments, hold group discussions, hear committee reports and make analyses thereof, make original and creative contributions, gather exhibits of real materials, construct apparatus and prepare illustrations. The demonstration of learning is through balanced and tested essay questions calling for application or explanation in functional situations of what has been learned. Inventories, interviews and anecdotal records of behaviour are made and problem situation tests are devised.

The assign-study-recite-test method is centred in the subject matter, which is proper with advanced students and in specialized areas. Mastery of the material for use is the objective. The unit is centred upon the individual. The objective is the personal-social-moral development of the individual, as well as mastery of knowledge.

The abbreviated account here emphasizes high lights and differences, but there are large areas of similarity between the two methods. Each uses necessary subject matter and learning activities, and each contributes to the development of the learner. Emphases differ with accompanying differences in procedure.

See also EDUCATION, PHILOSOPHY OF.

BIBLIOGRAPHY.—Boyd H. Bode, *How We Learn* (1940); W. H. Burton, *The Guidance of Learning Activities*, 2nd ed. (1952); John Dewey —any of his monographs, articles or books; Ernest R. Hilgard, *Theories of Learning*, rev. ed. (1956); Wm. H. Kilpatrick, *A Reconstructed Theory of the Educative Process* (1935); Henry C. Morrison, *The Practice of Teaching in the Secondary School*, rev. ed. (1931); S. C. Parker, *The History of Modern Elementary Education* (1912). The last reference is still the best on methods in elementary schools, historically analyzed. (W. H. BU.)

TEAK (*Tectona grandis*) is one of the most valuable world timbers. Known from ancient times—the Sanskrit name is *saka*—teak has been widely used in India for more than 2,000 years. The generic name *Tectona* is from the Malayan word *tekku*. Teak is a large deciduous tree of the verbena family (Verbenaceae), with a straight but often buttressed stem, a spreading crown and four-sided branchlets with large quadrangular pith.

The leaves are opposite, or sometimes whorled in very young specimens, 1 to 2 ft. long and 6 to 12 in. wide. On sucker shoots and seedlings the leaves are much larger, often from two to three feet long. In shape they somewhat resemble those of the tobacco plant, but their substance is hard and the surface rough. Small white flowers, very numerous, in large, erect, cross-branched panicles, terminate the branches. The fruit is a drupe, two-thirds of an inch in diameter. The bark of the stem is about one-half an inch thick, gray or brownish-gray, the sapwood white; the unseasoned heartwood has a pleasant and strong aromatic fragrance and a beautiful golden-yellow colour, which on seasoning soon darkens into brown, mottled with darker streaks. The timber retains its aromatic fragrance to a great age.

Teak thrives best in districts with a mean annual rainfall of more than 50 in. and a mean annual temperature between 75° and 81° F. Native to India, Burma and Thailand, its natural northern limit of growth in these areas is about the 25th parallel, although plantations have been established as far north as the 32nd parallel in the Punjab. The tree is not found near the coast; the most valuable forests are on low hills up to 3,000 ft. Stands are also found in the Philippines, in Java and elsewhere in the Malay archipelago, although there is no record of its being native in those places. Teak grows on a great variety of soils, but there is one indispensable condition—perfect drainage or a dry subsoil. On level ground, with deep alluvial soil, teak does not always form regularly shaped stems, probably because the subsoil drainage is imperfect.

During the dry season the tree is leafless; in hot localities the leaves fall in January, but in moist places the tree remains green until March. At the end of the dry season, when the first monsoon rains fall, the new foliage emerges.

The principal value of teak timber for use in warm countries is its extraordinary durability. In India and in Burma beams of the wood in good preservation are often found in buildings several centuries old, and instances are known of teak beams having lasted more than 1,000 years. In one of the oldest buildings among the ruins of the old city of Vijayanagar, on the banks of the Tunga-

bhadra in southern India, the superstructure is supported by planks of teakwood 1½ in. thick. These planks, examined in 1881, had been in the building for 500 years and were still in a good state of preservation and showed the peculiar structure of teak timber in a very marked manner. In the wall of a palace of the Persian kings near Baghdad, which was pillaged in the 7th century, two archaeologists in 1811 found pieces of Indian teak which were perfectly sound. In the old cave temples of Salsette and elsewhere in western India, pieces of teak which must have been more than 2,000 years old have been found in good preservation. The timber is practically imperishable under cover.

Teak is used for shipbuilding, fine furniture, door and window frames, wharves, bridges, cooling tower louvres, flooring, paneling, railway cars and venetian blinds. An important property of teak which affects utilization is its extremely good dimensional stability. It is of medium weight (45 lb. per cubic foot), strong and of average hardness. Termites eat the sapwood but rarely attack the heartwood; it is not, however, completely resistant to marine borers. Although the tree flowers freely, few seeds are produced as many of the flowers are sterile. Spread by self-sown seed is impeded by the forest fires of the dry season, which in India generally occur in March and April, after the seeds have ripened and have partly fallen. Although germination is slow, in its youth the tree grows with extreme rapidity. Two-year-old seedlings on good soil are five to ten feet high, and instances of more rapid growth are not uncommon. In Burmese plantations teak trees on good soil have attained an average height of 60 ft. in 15 years, with a girth, breast high, of 19 in. In these plantations it is estimated that the tree under favourable circumstances will attain a diameter of 24 in. (girth 6 ft.) at the age of 80. Normal exploitable girth limits are 7 ft. 6 in. for good forests, and 6 ft. 6 in. for poor teak forests. In the natural forests of Burma and India, teak timber with a diameter of 24 in. is never less than 100 and often more than 200 years old. Mature trees are generally not more than 100 to 150 ft. high.

Although by far the largest volume of teakwood produced is used in India and Burma, Ceylon and the east, much of the best quality is exported to other parts of the world. In the early 1960s Burma produced 75% of the world's teak supply. India, Thailand, Java and Ceylon ranked next in production. Teak is also produced commercially in Vietnam.

Burma teak comes mainly from the Pegu Yomas and Chindwin drainage areas, and is floated down the Sittang and Irrawaddy rivers to Rangoon. The most important teak forests in India are found in Madhya Pradesh, Mysore, Madras, Coorg, Bombay and Travancore-Cochin. Practically all Indian teak is utilized within the country, and the supply falls short of actual requirements. Thailand teak is shipped from Bangkok mainly to European markets. Teak has been cropped in Central America and elsewhere in the new world, but it has never been commercially successful.

See United Nations Food and Agricultural Organization, *Country Reports on Teak* (1956). (W. L. S.)

TEAL was originally the name for a small duck of Eurasia, *Anas* (*Nettion*) *crecca*, about 15 in. long with a cinnamon and green head, and its American counterpart, the green-winged teal. The word teal in the 19th century came to be used for a number of species of small fresh-water ducks (*q.v.*) related to the mallard (genus *Anas*). Among these are the blue-winged and cinnamon teal of America, *A.* (*Querquedula*) *discors* and *A. cyanoptera*, and the hottentot teal, *A. punctata*, of Africa. (A. L. Rd.)

TEALL, SIR JETHRO JUSTINIAN HARRIS (1849–1924), British petrologist, was born at Northleach, Eng., on Jan. 5, 1849. While at school at Cheltenham, he became interested in fossils. In 1869 he entered St. John's college, Cambridge, to read mathematics, but he was won for geology through listening to his college tutor Thomas G. Bonney. He attended Adam Sedgwick's final course of lectures and, after graduating in 1872, won the first Sedgwick prize in 1874 and was elected fellow of his college in 1875.

For several years he traveled as a university extension lecturer, attracting large classes in many parts of Britain. This brought him into contact with amateurs such as Samuel Allport of Birmingham. Encouraged by Bonney and H. Rosenbusch, Teall turned especially to petrology; and when he retired from his lectureship he concentrated on research and the preparation of his chief work, *British Petrography* (1886).

Teall's first important petrological paper (1883) dealt with Devonian andesites of the Cheviots. It led him, along with others, to recognition of hypersthene as a common constituent of andesites throughout the world. From 1884 to 1887 Teall was responsible for 12 papers extending geographically from Sutherland to the Lizard. His favourite, *The Metamorphism of Dolerite Into Hornblende-Schist*, appeared in 1885 and was based on evidence furnished by the northwest Highlands of Scotland. This led Sir Archibald Geikie in 1888 to persuade Teall to join the geological survey as petrographer. Teall continued as petrographer until 1901, when he was appointed director. He had been made a fellow of the Royal society in 1890. He retired in 1914 and was knighted in 1916. He died at Dulwich, Eng., on July 2, 1924.

He made valuable contributions also to the knowledge of quartz-dolerites in England and Scotland, metamorphism at the Lizard, magma-invaded vesicles, Devonian plutonics at Garabal hill and in the southern uplands of Scotland, association of radiolarian chert, pillow lavas and serpentine (the Steinmann trinity), borolanite, banded gabbros, contact-metamorphism including studies of cordierite, corundum and dedolomitization, albite-schist and phosphatic deposits.

See *Geological Magazine* (1909) for portrait and bibliography; E. B. Bailey, *Geological Survey of Great Britain* (1952). (E. B. B.)

TEASDALE, SARA (1884–1933), U.S. poet and Pulitzer prize winner, whose sonnets and short personal lyrics were considered unusually outspoken for women poets of her day, was born in St. Louis, Mo., on Aug. 8, 1884. A voracious reader, she was tutored at home and attended a private school, where she wrote verse. At an early age she toured Europe and the near east, and she contributed articles to many magazines of the period.

In 1907 her first book, *Sonnets to Duse and Other Poems*, was published (when Miss Teasdale wrote the sonnets, she had not met the Italian actress Eleonora Duse). Her works include *Helen of Troy and Other Poems* (1911); *Rivers to the Sea* (1915), which assured her fame; *Love Songs* (1917); *Flame and Shadow* (1920); *Dark of the Moon* (1926); *Stars Tonight* (1930); *Strange Victory* (1933); and *Collected Poems* (1937). She used short words and plain images. In 1917 she received the Pulitzer prize for poetry, and edited *The Answering Voice: One Hundred Love Lyrics by Women* (2nd ed., 1928); she also edited an anthology for children, *Rainbow Gold* (1922).

Sara Teasdale's life was tragic. A sickly, pallid girl from her youth, her health caused her much concern, and she appears to have lived within herself with few, if any, close companions. She died in New York city on Jan. 29, 1933.

TEASEL (*Dipsacus sylvestris*) is a common plant of the English countryside, with a tall (up to 6 ft.), stout, rigid, prickly stem, bearing large, spreading, opposite leaves with prickly midribs. The lavender flowers in conspicuous terminal, oblong heads, up to four inches long, are subtended by long, spiny, upward-curving bracts. Fuller's teasel, *D. fullonum*, in which the bracts are hooked, may be only a form of *D. sylvestris;* its dry heads are used to comb up the nap on cloth. Both plants are native to Europe and Asia; the common teasel is extensively naturalized in North America, but fuller's teasel is only sparingly adventive about wool mills in the eastern and middle states. (N. Tr.)

TEATE MARRUCINORUM: *see* Chieti.

TEBESSA (the Roman Theveste), a town of Algeria, Africa, lies on a plateau 146 mi. S. of Bône and 12 mi. W. of the Tunisian frontier. Pop. (1954) 20,431. The hub of the town is Kasbah square and there is a modern church built from the ancient ruins.

The central feature of the triangular-shaped town is the walled Byzantine citadel, a square with 12 towers and 4 gateways. Solomon's gate, on the east, was named after the Byzantine general who built the ramparts in 530; and on the north is the splendid Roman quadrifrontal arch of Caracalla (214), each side of which is flanked by a pair of monolithic columns. The temple of Minerva, built in 211, is now a museum containing remarkable mo-

saics and sculptured oil lamps. However, the most beautiful historic building is the Christian basilica, 1 mi. N. from the town centre and restored in the 1930s. It has traces of three successive constructions: the subterranean church and its catacombs (1st century); the first basilica (315–350), one of the strongholds of the Donatist heresy; and the upper basilica (after 350). Together with gardens and the remains of a 6th-century Byzantine military establishment, it makes one of the loveliest sights of Algeria. Outside Solomon's gate are the ruins of a Roman amphitheatre, and to the south of the town hot baths were uncovered.

A prehistoric site, Tebessa was an outpost of Carthage in the 7th century B.C. and in 146 B.C., a Roman garrison. At its height in A.D. 200, it declined in the 5th century under Vandal dominion and in the 6th century was occupied by Byzantines and later by Berbers. With the Arab invasion of the 7th century it disappeared from history. The Turks stationed a small garrison of janizaries there in the 16th century, but after French rule began in 1840 Tebessa was developed as the easternmost of the Algerian gateways to the south.

Tebessa is linked by rail to Bône and is also a road junction, and its airport ($\frac{1}{2}$ mi. away) is served by French airways. In addition to working the important phosphate mines of Kuif (northeast of Tebessa and linked by rail), the people live by trade, especially in sheep, esparto and grain, and there is a thriving handicraft production of carpets. (R. Gu.)

TEBU ("men of Tu"; *i.e.*, "of the rocks"), a nomad Negro-Berber race of the eastern Sahara. Their westernmost settlements are the oases of Agram, Kawar and Jebado, their northernmost the district of Gatron (Qatrun) within the Fezzan frontier, while south and southeast they merge in the Negroid populations of Kanem, Bornu (Chad basin), Wadai and northwest Darfur. But the bulk of the nation is concentrated in Tibesti or Tu, hence their name. Two main divisions are the northern Teda or less Negroid Tebu, and the southern Daza or more Negroid Tebu. Less closely connected are the Baele of the eastern and southeastern oases and the Zoghawa (Zaghwa) of Darfur.

The Tebu are usually identified with the Garamantes of Herodotus, whose capital was Garama (Idrisi's Germa) in Phazania (Fezzan), and of whom Ptolemy spoke doubtfully as Ethiopians (Negroes?). But Leo Africanus transfers them to the Berber connection, whose fifth great division he deals with under the names of Gumeri (Garamantes?) and Bardaei or Bardoa, that is, the Teda of the Bardai oasis, Tibesti.

TECHNETIUM, element 43, was the first example of the creation of a hitherto unknown element by artificial means. The name from the Greek word meaning artificial, commemorates this fact and has been accepted by the International Union of Chemistry. It is the homologue of manganese (25) and rhenium (75) in the periodic system. Formerly known as masurium, it follows molybdenum (42) and precedes ruthenium (44). The chemical symbol is Tc.

This element has been extensively sought in natural ores by several authors, notably I. and W. Noddack, who in 1925 claimed to have detected its X-ray spectrum in the same series of experiments that led to the discovery of rhenium. However, the findings of the Noddacks were not confirmed by other experimenters. It is now known that its natural existence is most improbable because it would contradict well-established rules of nuclear stability. However the possibility of the existence on the earth of extremely small amounts of some radioactive isotope of technetium, possibly Tc^{98}, is not ruled out.

In 1937 E. Segrè and C. Perrier found several radioactive isotopes of element 43 in molybdenum targets that had been bombarded with deuterons in the cyclotron. Using the methods of radiochemistry they gave positive proof that some of the radioactivity was caused by element 43. At the same time they established several chemical properties of the new element.

Technetium can be produced by several types of nuclear bombardments, such as neutrons on molybdenum or deuterons on molybdenum, and is also a product of the fission of uranium (O. Hahn and F. Strassmann, E. Segrè and C. S. Wu). The latter nuclear reaction promises to supply the largest amounts of the element. If it existed in nature, its geochemistry would probably be similar to that of rhenium.

The element is metallic in its properties; it is isomorphous with rhenium; and the isotope with mass number 99 has a density of 11.49 gr./cm.^{3} Metallic technetium becomes a superconductor at 11.2° K. In the course of a chemical analysis technetium would separate in the second group (sulfides insoluble in acid solution) and would usually appear with rhenium, although it can be separated from this element by a distillation at 180° C. in a current of hydrogen chloride. Its separation from molybdenum, which is often of practical importance, can be obtained by precipitating the molybdenum from an acetic acid solution with 8-hydroxyquinoline. Technetium stays in solution.

Among the technetium compounds are the oxides TcO_2, Tc_2O_7, the sulfide Tc_2S_7 and the important ion TcO_4 which gives pertechnetates similar to permanganates and perrhenates.

Not less than 13 isotopes of technetium are known ranging in mass number from 92 to 105. They are all radioactive, some emitting electrons, some decaying by orbital electron capture and some undergoing isomeric transitions. Following these isomeric transitions the characteristic X-ray radiation of technetium is emitted and this provided an opportunity for the spectroscopic observation of the K X-ray lines of this element before macroscopic amounts of it were available.

Beginning in 1944, substantial progress in the study of technetium was made in the laboratories of the atomic energy agencies of several countries. Taking advantage of the fact that the technetium isotope of mass 99 has a beta decay period of about 200,000 years, it was possible to isolate many grams of it. This isotope is obtained either as a fission product, or by neutron capture of molybdenum followed by a beta decay.

If the technetium formed in nuclear reactors in existence in the late 1950s were extracted there would be an annual production of several hundreds of kilograms, quite comparable to the annual production of some of the rarer natural elements. Thus the possibility of practical applications for this element is not excluded. In this connection it is interesting to note that it has been reported that technetium compounds and alloys are very effective in the prevention of corrosion.

The availability of relatively large amounts of technetium permitted study of it through the ordinary methods of chemistry and physics, without taking advantage of its radioactive properties. Investigations were conducted on its X-ray spectrum by external excitation, the crystal structures of the element and of its compounds, and the optical spectral lines. Also several chemical compounds were prepared in pure form and studied thermodynamically and otherwise, confirming the general analogy with rhenium.

In 1952 it was reported by Paul W. Merrill that spectra of stars of the S type show strong technetium lines. The astrophysical reason for this remarkable fact is not known.

See C. Perrier and E. Segrè, "Some Chemical Properties of Element 43," *J. Chem. Phys.*, 5:715 (1937); G. Friedlander and J. W. Kennedy, *Nuclear and Radiochemistry* (1955); E. Anders, "Technitium and Astatine Chemistry," *Am. Rev. Nucl. Sci.*, vol. 9 (1959). (Eo. S.)

TECHNICAL AID, assistance, usually in education, public health and agriculture and taking the form of training in and demonstration of advanced techniques, provided to the nationals of economically underdeveloped countries on an intergovernmental or multilateral basis by the more economically developed nations. In the United States, technical aid is sometimes called the Point Four program, after the fourth point of proposed U.S. policy enumerated in Pres. Harry S. Truman's inaugural address of Jan. 20, 1949, outlining "a bold new program for making the benefits of our scientific advances and industrial progress available for the improvement and growth of underdeveloped areas." Technical assistance is provided through the U.S. foreign aid programs and such agencies as the Organization of American States, the Colombo plan and the various United Nations specialized agencies.

See Foreign Aid Programs.

TECHNICAL EDUCATION. Technical education (Greek *techne*, art or craft) may be defined as instruction in the sciences

and skills required for the practice of trades or professions, especially those involving the use of machinery or scientific equipment. In Great Britain a distinction is increasingly being drawn between technological education, a term used to describe the higher levels at which advanced knowledge of theory is important, and technical education, directed primarily at the acquirement of skill in techniques. The term vocational education embraces all studies related to any form of employment. (H. C. D.)

THE UNITED STATES

The first technical instruction in the United States appears to have been given in connection with the U.S. Military academy, organized in 1802 at West Point, N.Y. Naturally the instruction was concerned with military problems. In 1824 Rensselaer Polytechnic institute was founded at Troy, N.Y., and started the work of training "civil," so-called in distinction to "military," engineers. Rensselaer conferred its first engineering degrees in 1835. Harvard opened the Lawrence scientific school and Yale established the Sheffield scientific school in 1847. The University of Michigan began instruction in engineering in 1853. The Polytechnic Institute of Brooklyn was founded in 1854, Cooper Union in 1859 and the Massachusetts Institute of Technology in 1861.

In 1862 congress passed the Morrill Land Grant act and thus laid the foundation of the land-grant colleges from which have grown many of the state universities. The act in setting aside public lands for the support of these colleges specifically prescribed that they shall teach "such branches of learning as are related to agriculture and the mechanic arts." Under the impetus of this act and the economic pressure resulting from the needs of a new country engineering grew apace.

Thus, Cornell university was founded in 1865, the University of Illinois in 1867, etc.

Technical education of university grade grew therefore under two auspices, namely, private ownership and state ownership. In the first group were to be found technical schools of high grade, independent of general educational functions as illustrated by the Massachusetts Institute of Technology, Rensselaer Polytechnic institute and California Institute of Technology. Again many large privately endowed universities such as Harvard, Yale, Cornell, etc., have engineering colleges which are integral parts of a group of colleges that comprise the university. Others such as Stanford and Princeton have engineering courses integrated with general education. In some of the state universities the colleges of engineering and agriculture are separated from the general courses not only educationally but geographically as well. Thus, the University of Michigan proper is at Ann Arbor while the Michigan State College of Agriculture and Applied Science is at East Lansing some distance away. The increase in the number of technical schools of collegiate rank is noteworthy. In 1866 there were but six engineering colleges of established reputation and only 300 men had been graduated in the previous 31 years. By mid-20th century there were more than 150 degree-granting colleges with a total enrolment of about 150,000.

While there is considerable variation in the entrance requirements of this large group, in general a high school education is required.

In general, the entrance requirements of the privately endowed colleges are somewhat higher than those of the state institutions, the latter from the nature of state educational systems being closely articulated with the high schools of the state.

There is also a considerable degree of uniformity in the curricula of the better type of technical college. This is necessarily so in a measure since engineering is an application of the basic sciences of mathematics, physics and chemistry and their derivatives. Analytic mechanics, which is a key course, is usually completed by the middle of the third year, leaving at least one and one-half years of the four-year course for engineering fundamentals and applications. There is considerable variance, however, in the matter of instruction for specific technical callings. Practically all colleges recognize four principal fields of engineering, namely, civil, mechanical, electrical and chemical. For the most part also the higher technical schools recognize the necessity of offering a limited amount of instruction in specialized fields such as heat-power, radio, industrial, hydraulic engineering, etc., in the senior year, usually as optional studies. There is always, however, a strong pressure from the industrial field in the direction of highly specialized courses with their beginnings much earlier and not a few colleges have yielded to this pressure. Thus, there are many colleges offering engineering courses of good quality and accredited courses in aeronautical, agricultural, architectural, ceramic, chemical, civil, electrical, electrochemical, fuel technology, general, industrial, mechanical, metallurgical, mining, naval architecture and marine, petroleum, and sanitary engineering. For the most part, however, educators have become convinced that the student's time is best spent during the first three years, at least, in acquiring fundamentals. Specialization should be deferred as long as possible and any extended special study should be carried to a fifth-year study. All technical colleges of repute offer graduate study.

Until comparatively recent years the technical colleges were the principal centres of research work in engineering and the chief sources of engineering data. The growth of commercial research laboratories has changed this situation but there is still an immense amount of research conducted in college laboratories. There is a growing tendency for industrial enterprises to cooperate with college laboratories in the prosecution of special problems. If properly conducted such joint efforts are most fruitful since the commercial concern can supply the practical data for the problem and the funds for its prosecution while the college can supply the necessary apparatus and knowledge.

The growing tendency of the engineer to take a leading place in industrial administration gave rise to a need for a broadening of the curriculum by the introduction of economics, accounting and subjects of a more liberalizing character such as history. While a certain amount of such topics were introduced into the standard four-year courses any great addition necessitated a lengthening of the curriculum and five-year courses came into use.

No outline of U.S. technical education would be complete without mention of the co-operative methods in many colleges whereby the student spends part of his time in practical work and part in study, thus obtaining his practical training and theoretical work at the same time. Students in such courses may "earn as they learn" and contact with industry stimulates interest in theory. There is also an important group of polytechnical institutes offering instruction in technology somewhat below the level of the higher schools but of great importance. Pratt institute in Brooklyn, N.Y., may be cited as an excellent illustration. The high schools during this period also tended more and more to prepare boys and girls specifically for life and to lay less accent on college preparation. Many state educational departments continued to offer technical training in the high school grades. (D. S. K.)

UNITED KINGDOM

England and Wales.—Technical education in the modern sense did not develop in England on any substantial scale until the last decade of the 19th century. The mechanics' institutes, founded in great numbers from 1823 to about 1860, were intended to instruct workingmen "in the scientific principles of arts and manufactures," but for the most part they quickly became diverted to social and literary rather than industrial education. The Great exhibition of 1851, which confirmed fears that Britain's industrial primacy was being threatened, led the government to establish a science and art department, but the grants that it distributed did little beyond developing an arid examination system.

The 1867 international exhibition at Paris shook British complacency. It led to a royal commission which, reporting in 1875, moved the livery companies of the City of London to found in 1880 the City and Guilds of London institute to promote the teaching of applied science. In 1883 the institute opened the Finsbury Technical college for the training of artisans and in 1884 the Central Training college at South Kensington to train technical teachers and the higher ranks of industry. Meanwhile, quite independently, Quintin Hogg opened in 1882 the Polytechnic in

Regent street, London.

By that time another royal commission on technical instruction (1881–84) was sitting. Its recommendations brought about the Technical Instruction act, 1889, which empowered the newly created county councils and the county borough councils to give limited aid to technical education out of public funds. In 1891 the Local Taxation (Customs and Excise) act unexpectedly made further large sums of money (popularly called "whiskey money") available for this purpose. Consequently, the decade 1889–99 saw great activity in the provision of buildings for the teaching of applied science and technology. In addition to public effort, funds provided by the City of London Parochial Charities act, 1883, aided the building in London of more polytechnics.

In 1899 the Board of Education act created a single national department charged with the supervision of public education throughout England and Wales. The Education act, 1902, similarly made the county and county borough councils general purpose education authorities, responsible for supplying or aiding the supply of secondary and technical as well as elementary education. Unfortunately, the preoccupation of the board of education and the local education authorities with the development of the newly established secondary system considerably slowed down provision for technical education for more than ten years.

Nevertheless, three important developments took place during that period. About 1905 trade schools for boys and girls leaving elementary schools began to be established in London, Middlesex and elsewhere. In 1913 these schools were made eligible for grant as "junior technical schools." In 1907 the Imperial College of Science and Technology was founded by royal charter which merged the Royal College of Science, the Royal School of Mines and the City and Guilds of London Institute Engineering college into a single institution (though functioning largely as three bodies) devoted to the highest levels of education and research in science and technology, especially in their application to industry. In 1908 the college became a school of the University of London. The third development was the most important of all. This was the foundation, between 1900 and 1909, of six universities: Birmingham, Liverpool, Manchester, Leeds, Sheffield and Bristol. The importance to technical education lay in the fact that three of these universities—Birmingham, Leeds and Sheffield—had grown out of colleges mainly devoted to applied science and were already strongly biased thereto, while the others were much more disposed to grant it a prominent place in their program than the older universities. This remained the case.

The period between World Wars I and II was one of slow and fitful but continuous expansion and improvement of facilities for technical education, made almost exclusively by the local education authorities. The years of greatest progress were 1925–29, when Lord Eustace Percy was president of the board of education. He not only encouraged building but set on foot a number of investigations, which resulted, among other things, in the board of education *Report on Education for Industry and Commerce* (1928), the Clerk *Report on Education for the Engineering Group of Industries* (1929) and the Goodenough *Report on Education for Salesmanship* (1931). A scheme for a big expansion of facilities was prepared during the late 1930s, but World War II prevented its being implemented.

A previous development calls for special note. Shortly after World War I the board of education, in co-operation with professional institutions related to industry, established an entirely new system of examination and certification for technicians, leading to the award of national certificates and diplomas. The certificates, available to part-time students only, were awarded on the cumulative results of concurrent industrial experience and technical education. They were of two grades: ordinary, to be obtained normally after a three-year course, and higher, following a further two-year course. The first scheme begun was for mechanical engineering, in 1921. In 1958, 15 technologies had such schemes, and the national certificate had long been established as a nongraduate technician's recognized qualification.

World War II brought to England full recognition of the importance of technical education in an industrial society. At its outbreak the country found itself dangerously short of technologists, technicians and skilled craftsmen and of means to train them. By ingenuity and improvisation—the latter including the introduction of two- or three-shift training programs into many technical colleges—the shortages were temporarily overcome. Technical colleges alone gave special courses of training to nearly 300,000 men and women.

Meanwhile the nation looked to the future. In 1943 the ministry of labour asked industry to consider the recruitment and training of juvenile employees after the war; as a result, by 1956 more than 100 sections of industry had drawn up national apprenticeship schemes including systematic education and training, and the number of young employees receiving such training in working hours had increased ninefold. In 1945 a committee on higher technological education recommended the upgrading of technical college courses to train technologists and the development of joint consultation between education and industry for planning of vocational education. The latter recommendation was carried into effect in 1947–48 by the creation of the National Advisory Council on Education for Industry and Commerce and ten regional advisory councils for further education.

The Education act of 1944 made it the statutory duty of the local education authorities to secure sufficient facilities for all types of education (except university education) in their areas. But until 1947 hardly any new building was possible, so that with a greatly increased demand for vocational education many technical institutions became grossly overcrowded. The demand was partly a result of the government's Further Education and Training scheme, under which men and women who had served in the armed services or on work of national importance during World War II could receive grants for higher education and training. It was also a result of a progressively increasing demand from industry for vocational training, especially during working hours, of all kinds for employees of all ranks.

In 1946 the ministry of education urged technical colleges to undertake more research and invited their co-operation in establishing national schools devoted to advanced work in particular branches of technology. By 1951 six new national colleges had been established for foundry work, rubber, leather and food technology, horology and instrument technology and heating, ventilating, refrigeration and fan engineering. In 1953 the government announced its intention to double the number of students at the Imperial College of Science and Technology (from 1,600 to 3,000) by 1962, and offered technical colleges increased grants for approved high-level courses. In 1954 it decided to expand largely technological faculties in selected universities. In 1956 it announced a £100,000,000 five-year plan for developing technical colleges in Great Britain which included their regrouping in four categories, a 50% increase in their production of technologists and a doubling of daytime craftsman training. A diploma in technology ranking with an honours degree was created and nine colleges were made colleges of advanced technology.

Scotland.—Modern technical education in Scotland began with the founding by George Birkbeck in 1800 of a class for instructing Glasgow operatives in "mechanical and chemical philosophy." In 1823 this class broke away from the Andersonian institute to which it was attached to form the first Mechanics institution. Two years previously there was opened in Edinburgh the School of Arts, claimed to be the first institution in Great Britain founded expressly to give scientific instruction to workingmen.

Until 1897 the English science and art department covered Scotland as well, and consequently statutory development of technical education was broadly parallel with England, except that under the Educational Endowments (Scotland) act, 1882, schemes were made and money was provided for establishing and maintaining technical schools. Among other changes, the Andersonian institute and the Glasgow Mechanics institution were merged in what became the Royal Technical college, Glasgow, and the School of Arts in the Heriot-Watt college.

Codes of regulations made in 1899 and 1901 organized elementary technical instruction in "continuation classes." To encourage

advanced work the 1901 code provided that outstandingly good and well-equipped technical colleges could be exempted from its regulations and grant-aided on special terms. The Education (Scotland) act, 1908, recognized 11 of these "central institutions," as they became known.

In 1956 there were 17, including 3 agricultural and 2 veterinary colleges. Among them the Royal Technical college, Glasgow, and the Heriot-Watt college, Edinburgh, were affiliated to the universities in these cities. Other technical education was given in "continuation class centres."

In the 1950s the pattern of technical education in Scotland was (nomenclature apart) closely similar to that in England, except that the latter had no designated central institutions. Five regional advisory councils had been established.

Northern Ireland.—Before 1922 education throughout Ireland was regulated from England. In the 19th century, despite the nonindustrial character of the country, technical education developed remarkably. It received an added stimulus in 1899 when a department of agriculture and technical instruction was created. In 1922, 39 technical schools managed by local authorities were handed over to the Northern Ireland government. They included the Belfast College of Technology, one of the largest and best in the British Isles.

The Education (Northern Ireland) act, 1923, further stimulated technical education by abolishing the restriction on expenditure to a 2*d.* rate and merging it in the general expenditure of the local authorities. Junior technical and commercial schools expanded, especially in the provision of day classes.

The Belfast College of Technology prepares students for degrees in engineering and related subjects offered by the Queen's university. Among its students are also trades scholars, who are apprentices granted scholarships by the ministry of education for extended courses.

REPUBLIC OF IRELAND

The department of education set up in 1924 took over the functions of the technical instruction branch of the department of agriculture and technical instruction established by the Agricultural Technical Instruction (Ireland) act, 1899. Following the report of the Commission on Technical Education, issued in 1927, which noted especially the lack of technical instruction in rural areas, a system of rural continuation schools was developed offering full- and part-time classes to meet the needs of their localities. Technical colleges and schools maintained by local authorities were established in all cities and large towns.

The National University of Ireland established faculties of engineering at two of its three constituent colleges, Cork and Dublin, and the University of Dublin (Trinity college) also established an engineering faculty. Both universities offer degrees in agriculture.

BRITISH COMMONWEALTH NATIONS

Australia.—With increasing industrialization, much attention was paid to technical education, especially after 1940. In all states there are technical secondary schools or courses. In South Australia all apprentices in "proclaimed" trades and areas are required to attend trade schools for one half-day a week and one evening. Most apprentices in New South Wales attend technical colleges on a similar basis.

Many new technical colleges and schools have been built in Australia. In 1947 the New South Wales University of Technology was founded; it provides degree courses in civil, electrical, chemical and mechanical engineering, mining, applied mathematics, physics and chemistry. The Gordon Institute of Technology at Geelong in Victoria includes a textile college which meets the requirements of the whole commonwealth for highly trained executives for the wool industry. The School of Mines of Western Australia at Kalgoorlie similarly covers the mining industry. The University of Queensland runs correspondence courses in technical education up to degree standard.

New Zealand.—At the secondary level there are technical high schools in the large towns; in the smaller ones consolidated schools offer technical courses. The Auckland and Canterbury colleges of the university have engineering schools and Otago has a school of mines that grants its own diplomas. The Lincoln and Massey agricultural colleges are affiliated to the university, of which they form the school of agriculture.

Canada.—Recent years have seen great educational developments, especially in the prairie provinces. In both English and French educational systems there are technical high schools and trade schools at the secondary level. Technical institutes covering the senior high school years and beyond were increasing in the 1950s. The Nova Scotia Technical college, and the agricultural, pharmacy and veterinary colleges in Ontario are of university status, as is Macdonald college, the agricultural faculty of McGill university, Montreal, Que.

The western universities and New Brunswick have schools of engineering and agriculture.

Union of South Africa.—Though each province had its own department of education, all forms of university and vocational education were financed by the union education department, which also conducted the examinations. Technical education was provided in universities, technical colleges and schools and agricultural schools. The universities of Cape Town, Natal and Stellenbosch had schools of engineering and agriculture, those of South Africa and the Witwatersrand engineering schools.

India.—Great efforts were made during the 1950s to improve the country's facilities for technical education. In many elementary and some secondary schools instruction in handcrafts was made compulsory. Beyond that, technical education was provided in universities and technical institutions at four levels: institutions which could be entered from the seventh standard of a high school and offering courses leading to a certificate; institutions requiring matriculation for entry and offering diplomas; institutions offering degrees or equivalent courses; and institutions doing mainly postgraduate and research work.

The responsibility for education was assumed by the states, but the central government, through the All-India Council for Technical Education, co-ordinated and improved facilities. It maintained the Delhi polytechnic, established similar national establishments such as the Institute of Technology at Kharagpur in West Bengal (1950) and promoted advanced studies and research.

Pakistan.—At the time of partition in 1947 Pakistan had three engineering colleges and some industrial institutes. The Council of Technical Education prepared a national development plan. In 1951 the first of six technical secondary schools was opened at Karachi and the establishment of the Karachi polytechnic (assisted by the Ford foundation) begun. By 1953 the establishment of new engineering colleges in the Punjab and the North-West Frontier Province was in hand and Dacca engineering college was re-equipped. Technicians were being sent overseas for practical training. The universities had faculties of engineering and agriculture and there were agricultural colleges at Lyallpur and Sakrand.

EUROPEAN COUNTRIES

France.—Technical education was brought under state control in 1892 and codified by the *Loi Astier* of 1919. Provisions were made for three main levels: apprenticeship, intermediate and higher. The first and second were usually entered from the elementary school; the third was open only to students possessing the *baccalauréat* awarded at the end of a secondary course.

At mid-20th century elementary school children could get the rudiments of technical education in the *collèges modernes* (formerly *écoles primaires supérieures*) and the *cours complémentaires,* both available from the age of 12. For elementary school leavers there was, after 1945, a great development of *centres d'apprentissage*, which offered to young people between the ages of 14 and 18 a three-year course of full-time trade training along with general education, leading to the *certificat d'aptitude professionelle,* awarded for a wide variety of crafts. At a higher level there was the development of *collèges techniques* and of regional schools of agriculture, the latter run by the ministry of agriculture.

At the highest level are the national schools such as the École Polytechnique, which trains engineers for state service, both military and civilian, the École Centrale des Arts et Manufactures, whose aim is to produce future industrial leaders, and the Écoles Nationales d'Agriculture at Grignon, Montpellier and Rennes. There are similar national schools for horticulture, veterinary science, aviation and mining.

German Federal Republic.—During the 19th and early 20th centuries Germany built up a comprehensive national system of vocational training for workers of all grades. The Hitler regime perverted this, and the collapse of Germany in 1945 temporarily laid it in ruins. It was later built up again on similar lines, at varying rates in the different *Länder*.

In the 1950s there were four main levels: *Berufschule, Fachschule, Höhere Fachschule* and *Technische Hochschule*. All boys and girls leaving elementary schools at 14 were due to enter *Berufschulen*, and by 1951 compulsion was said to be generally operative throughout the republic. They had to attend for three years, eight hours of instruction a week being required. Alternatively, in some *Länder* they could attend a full-time *Berufsfachschule* for one year or more. *Fachschulen* served students who completed their basic vocational education and wished to be more highly qualified in their trades; these schools were usually full time for one year. *Höhere Fachschulen* are equivalent to English technical colleges, though usually possessing fewer departments, and recruit mainly from the *gymnasien* (grammar schools).

Technische Hochschulen, of which there were 13 in the Federal Republic in the 1950s, had university status. The famous Charlottenburg institution in Berlin was made a full university in 1946 to compensate for the fact that the Friedrich Wilhelm university was in the Russian sector of the city.

Netherlands.—A system of continued education with a vocational bias for children leaving the elementary school is followed by junior technical schools (*ambachts scholen*) giving general and vocational education, trade schools (*vakscholen*) giving more intensive training in one trade (or group of allied trades), and technical schools, training for supervisory posts. At university level are the Technical High school at Delft, offering degrees up to the doctorate in engineering, and the Agricultural institute at Wageningen, offering doctorates in agricultural science to students qualified as agricultural engineers.

Belgium.—Technical education is largely in the hands of private bodies though it was placed under the supervision of the ministry of education in 1945. At Verviers a municipal École Supérieure de Textiles training textile engineers was established; by study there and at Liège university a degree can be obtained. The Faculté Polytechnique de Mons, devoted principally to engineering, has university status.

Denmark.—There are many technical schools and three institutions for advanced instruction, of which two, the Royal Veterinary and Agricultural college and the Technical University of Denmark (both in Copenhagen) grant degrees. Some of the newer folk high schools have vocational biases, notably those preparing women for nursing.

Norway.—Some continuation schools provide vocational courses for young people leaving the elementary schools. Youth schools cater for rather older students (aged 17–20) who wish to prepare for entry to more advanced vocational training. Trade schools for crafts, industrial occupations, seamanship and domestic science are numerous. At university level there are the Norwegian Institute of Technology at Trondheim and the Agricultural College of Norway at Vollebekk.

Sweden.—Practical lower secondary schools and upper secondary schools emphasize vocational work; the former offer a certificate for it. Full-time technical instruction is given in trade schools. There are two national technical colleges, the Royal Institute of Technology at Stockholm and the Chalmers Institute of Technology at Göteborg, both under the same governing board; they grant degrees up to the doctorate. Many of the folk high schools include vocational and quasi-vocational work in their programs, especially in gardening, farming, forestry, domestic science and gymnastics.

Finland.—There are numerous trade, technical and agricultural schools. The National Institute of Technology, at Helsinki, grants bachelors' and doctors' degrees. The people's high schools teach handicrafts, agriculture for men and domestic subjects for women.

Italy.—After the elementary school, which gives a five-year course from the age of 6 to 11, pupils wishing to learn a trade can (on passing an entrance examination) enter the *scuola di avviamento al lavoro*, which offers three-year vocational courses and a diploma ranking in status with the secondary school certificate.

Alternatively, elementary school children may (again on passing the necessary entrance examination) go to the *scuola media* (lower secondary school) between 11 and 14, which also offers prevocational courses. After the *scuola media* students can (by examination) enter an *istituto tecnico* for a four-year course leading to the *abilitazione* (professional diploma).

General postschool technical education is administered by provincial associations under the control of the ministry of public education. Courses granting vocational qualifications are conducted by professional bodies, such as the National Institute for Training and Improvement of Industrial Workers (I.N.A.P.I.), which work under the control of the ministry of labour. At the highest level the government maintains polytechnics in Milan and Turin, and schools of engineering in Rome, Naples, Bologna, Padua, Palermo and Pisa.

Switzerland.—There are many excellent technical schools administered by the cantons, to which the federal government delegates responsibility for education. The École Polytechnique at Lausanne constitutes a faculty of the cantonal university. The famous Swiss Federal Institute of Technology (Eidgenössische Technische Hochschule) at Zürich is maintained and administered by the federal government, ranks as a university and has an international reputation. It was organized in 11 schools of architecture, civil, electrical and mechanical engineering, agriculture, agricultural engineering and surveying, forestry, chemistry, mathematics, physics, natural science, military science and pharmacy. It also established a department of optional studies teaching the humanities, politics and economics.

Turkey.—Steady progress toward a comprehensive system was made after the act of 1942, which outlined a ten-year plan. Mobile courses teach smithing and carpentry in villages. Primary school leavers can enter three- or five-year trade schools and go on from there to two- or four-year technical schools or a school for master builders. There are also evening trade schools for men and women. The Technical University of Istanbul, created in 1944 from the former Imperial Engineering school, admits graduates of *lycées* and technical schools to a five-year course. In the 1950s preparations were made for technical universities at Trabzon and Ankara. There are training colleges for technical teachers (men and women) at Ankara.

BIBLIOGRAPHY.—Annual reports of the Ministry of Education, England and Wales, the Scottish Education Department and the Ministry of Education, Northern Ireland; *Further Education* (H.M.S.O., London, 1947); *Technical Education*, report of the Advisory Council on Education in Scotland (H.M.S.O., London, 1946); publications of the Association of Technical Institutions of the United Kingdom (London).

Australian federal and state government publications; Australian Council for Educational Research publications; Indian government, *Facilities for Technical Education in India* (1948), *Review of Education in India, 1950–51;* Canadian dominion and provincial government publications.

P. F. R. Venables, *Technical Education* (London, 1955); H. C. Dent (ed.), *Year Book of Technical Education and Careers in Industry* (London, annual from 1957); P. J. Idenburg, *Education in the Netherlands* (The Hague, 1951); W. F. Holtrop, *Vocational Education in the Netherlands* (1951); Italian government, *Education in Italy* (Rome, 1952), *Guida D. Annuario della Scuola e della Cultura* (Centro Didattico Nazionale, Florence and Rome, 1951); Ragnar Lund (ed.), *Scandinavian Adult Education* (Copenhagen, 1949); Nils Hjelmtveit, *Education in Norway* (Norwegian government, 1946); Bertil Östergren, *Higher Education in Sweden* (Stockholm, 1952); Turkish government, *Education in Turkey* (Ankara, 1955).

United Nations Educational, Scientific and Cultural Organization, *Education in a Technological Society* (Paris, 1952); American Council on Education, *Universities of the World Outside U.S.A.*, ed. by M. M. Chambers (1950). (H. C. D.)

TECHNOCRACY. The movement known as technocracy originated in the United States in 1932. It sought to give expression to a group of principles which for many years had found place in the writings of a few engineers, economists and students of science. Its basic principles were: (1) Social phenomena are measurable and the laws of social control may be derived from these measurements. (2) The application of the machine to the production of goods and services has made it impossible to measure the value of the goods produced in terms of any single commodity (gold, for example) and thus the engineer had destroyed the price system. (3) The expansion of credit, sometimes termed the creation of debt, under the capitalistic system has disturbed the relative claims of capital and labour to the goods produced to such an extent that the capitalistic system has already collapsed. (4) The economics of the social order are too complicated to be understood and controlled by politicians. Control should be placed in the hands of engineers and scientists.

TECHNOLOGY: *see* AUTOMATION; INVENTIONS AND DISCOVERIES; INDUSTRIAL REVOLUTION.

TECK, a ruined castle in Württemberg, at the foot of the Swabian Alps, about 20 mi. S.E. of Stuttgart.

The duchy of Teck was first formed out of family possessions by Adalbert, younger son of Conrad, duke of Zähringen, after his father's death in 1152; but after the territory had been acquired by the counts of Württemberg (1381), the ducal title disappeared on the extinction of Adalbert's line (1439). The German king Maximilian I, however, revived the title for the dukes of Württemberg in 1495, but the elector Frederick renounced it on becoming king of Württemberg in 1806. King William I of Württemberg in 1863 gave the title of prince of Teck to Francis, count of Hohenstein, who was created duke of Teck by King Charles I of Württemberg in 1871. Francis' daughter Mary married George, duke of York (later king as George V), in 1893. His eldest son, Adolphus Charles (1868–1927) in 1917 renounced all his German titles for himself and for his issue and was instead created marquess of Cambridge.

TECUMSEH (TECUMTHE or TECUMTHA) (*c.* 1768–1813), American Shawnee chief, was born probably in the old Shawnee village of Piqua, near the site of Springfield, O., between 1768 and 1780. While still a youth he took part in attacks on settlers passing down the Ohio river and in widely extended hunting expeditions or predatory forays to the west and south; and he served in the Indian wars preceding the treaty of Greenville in 1795. His eloquence and his self-control made him a leader in conferences between the Indians and whites as early as 1800. In 1804 and 1805, when the Indians of the northwest became aroused by a series of treaties calling for new cessions of their territory and by the prospect of war between Great Britain and the United States, the opportunity was presented to Tecumseh and his brother, popularly called "the Propbet," to put into operation a scheme which followed the ambitious dream of Pontiac. With some scattered Shawnee clans as a nucleus, the brothers proceeded to organize.

Their actual plan was to combine all the Indians from Canada to Florida in a great democratic confederacy to resist the encroachment of the whites. The twofold influence that was to dominate this league was the eloquence and political ingenuity of Tecumseh and the superstitious reverence aroused by the Prophet. This program alarmed the whites along the northwestern border. In the course of the next three years Gov. William Henry Harrison of Indiana held interviews with each of the brothers, and during one of these, at Vincennes in 1810, the respective leaders narrowly avoided a hostile encounter. Nevertheless, the Prophet and Tecumseh reiterated their determination to remain at peace with the United States if the recent cessions were given up and if no purchases were made thereafter without the consent of the tribes. Harrison refused to consider this arrangement, and, during Tecumseh's absence in the south, made a hostile move against "the Prophet's town." The latter ventured to meet him, but was defeated Nov. 7, 1811, in the famous battle of Tippecanoe, which broke the personal influence of the Prophet and largely destroyed the confederacy built up by Tecumseh.

Tecumseh still professed to be friendly toward the United States, probably because his British advisers were not ready to open hostilities, but a series of border outrages indicated that the fatal moment could not long be postponed. With the outbreak of the War of 1812 he was commissioned a brigadier general in the British army and participated in the skirmishes which preceded Gen. William Hull's surrender at Detroit. In the battle of the Thames on Oct. 5, 1813, Tecumseh was killed. The Prophet died in 1834.

BIBLIOGRAPHY.—Benjamin Drake, *The Life of Tecumseh and of His Brother the Prophet* (1841); Ethel T. Raymond, *Tecumseh,* "Chronicles of Canada Series," vol. xvii; H. J. Webster, *Harrison's Administration of Indian Territory* (1907); Don Seitz, *Uncommon Americans* (1925).

TEDDER, ARTHUR WILLIAM TEDDER, 1ST BARON (1890–), British air marshal, who had a large share in the success of the Allied landings in Normandy in World War II, was born at Glenguin, Scot., in 1890 and educated at Whitgift school and Magdalene college, Cambridge. He was commissioned in the army in 1913 and served in France in World War I, being seconded to the royal flying corps in 1916. He was granted a permanent commission in the Royal Air Force in 1919.

In 1934, after service in a number of training establishments, he was appointed director of training. From 1936 to 1938 he was air officer commanding the far east command. He then returned to the air ministry as director of research and development. Early in World War II he became deputy air member for development and production at the ministry of supply. In 1941 he was appointed air officer commanding in chief in the middle east, and subsequently he took command of all the Allied air forces in that area. During his command the air force played a notable part in expelling the enemy from north Africa. In 1944 he was deputy commander of the Allied expeditionary force under Gen. Dwight D. Eisenhower and played a large part in planning and controlling the Allied landings in Europe. Much of his success was due to the efficient manner in which he co-operated with other Allied forces. Between 1946 and 1950 Lord Tedder was the first peacetime chief of the air staff and senior member of the air council. He was created knight grand cross of the Order of the Bath in 1942 and elevated to the peerage in 1946. (E. B. BN.)

TEDDINGTON, a part of Twickenham (*q.v.*), Eng., mainly residential, lies on the Thames close to Bushy park and Hampton Court palace. Teddington lock is the biggest on the Thames. The National Physical laboratory, the National Chemical laboratory, the Admiralty Research laboratory and the Paint Research station are scientific establishments at Teddington. Eminent residents have included Peg Woffington, the actress, who died there in 1760 (*see* WOFFINGTON, MARGARET); Robert Dudley, earl of Leicester; William Penn, the Quaker; Richard Doddridge Blackmore, the author of *Lorna Doone,* who also died there; and Noel Coward, the playwright and actor, who was born there. (W. H. Jo.)

TE DEUM, opening words of the ancient Latin hymn *Te Deum laudamus* ("We praise Thee O God"), at one time believed to have been written by St. Ambrose of Milan and now generally ascribed to Nicetas, bishop of Remesiana in Dacia. Its noble character speedily procured its incorporation in the liturgy of the Latin church while it has been set separately by countless composers as a hymn of thanksgiving for special occasions.

TEEMANT, JAAN (1872–?), Estonian statesman, was educated in law at St. Petersburg university and began to practise in Reval (Tallinn). In 1905, however, he associated himself with the revolutionary movement for which he was condemned to death by the Russian government. He escaped to Switzerland but after a number of years he returned to Russia and served a sentence of imprisonment. After the 1917 revolution he took an active part in paving the way for an Estonian government, was a member of the constituent assembly, 1919–20, and after 1920 of the Estonian diet. He rose to power as a leader of the Agrarian party and was prime minister from Dec. 1925 to 1927 and again in 1932 but failed to be re-elected to parliament in 1938. He was arrested and deported in 1940 by the Soviet occupation forces, after which his fate was unknown.

TEES, a river of England, rising on the east side of Cross Fell (*see* PENNINES), 8 mi. N. of Appleby in Westmorland and flowing 70 mi. S.E. and E. to the North sea. Part of its earlier course forms the boundary between the counties of Westmorland and Durham. From its source to Middleton-in-Teesdale it flows through bleak moorland with hills rising 2,500 ft. and more above the sea. Two famous falls above Middleton-in-Teesdale where the brown, peaty river traverses the hard, black dolerite of the Whin Sill, are known as Caldron Snout and High Force. At Caldron Snout the water dashes over huge steps of rock and at High Force it falls from a vertical rock 70 ft. high into a swirling pool below. From a point immediately below Caldron Snout to its mouth the river forms the boundary between Durham and Yorkshire almost

without a break. Below Caldron Snout trees appear, contrasting with the broken rocks, and as the river flows through the wide vale past Middleton-in-Teesdale the scenery becomes gentler. The town of Barnard Castle, Eggleston abbey and Rokeby hall—well-known through Sir Walter Scott's poems—are passed, and then the valley begins to open out, and the river crosses the rich plain south and east of Darlington. The Tees now becomes an important commercial highway, having on its banks such ports as Stockton-on-Tees and Middlesbrough and forming an outlet for the chemical plant at Billingham and the ironworking district of Cleveland in the North Riding of Yorkshire. The drainage area is 708 sq.mi.

TEETH are hard white structures found in the mouth in man and many animals and used variously for mastication of food, seizing and holding objects, combat, cutting, chiseling and other incidental purposes. Each tooth consists of a crown and one or more roots. The crown is the functional part that is visible above the gum. The root is the unseen portion that supports and fastens the tooth in the jawbone. The shapes of the crowns and the root characteristics vary in the different parts of the mouth and from one animal to another.

The different modern dentitions have been attained under the pressures of biological laws of evolution. Some dentitions are very restrictive and adapted to a particular diet. Carnivores cannot prepare the food that makes up the diet of a horse, nor can sheep handle a meat diet. Other dentitions, like that of man, are very general (omnivorous), and can prepare both flesh and vegetable foods. Modern man's dentition, though variable, is showing signs of general reduction in size and form, as compared with the dentition of fossil man.

Occasionally the statement is made that man gradually is losing all his teeth. This is based on the observation that certain teeth fail to develop in some individuals and are of a reduced form in others. To conclude that the trend must continue is a fallacy. The future survival and selection pressures of evolutionary laws will determine the fate of man's dentition, not the fact that the present shift is in a given direction.

Human Dental Anatomy.—Man has two sets of teeth during his lifetime. The first set is acquired gradually between the ages of six months and two years. As the jaws grow and expand, these teeth are replaced one by one by the teeth of the second set. The first set is known as the milk, deciduous or primary dentition. The second set is called the permanent or secondary dentition.

The right side of a dentition is, except for minor variations, a mirror image of the left. The upper teeth differ from the lower and are complementary to them; the relationship is similar to that of a mortar and pestle. There are 5 deciduous teeth and 8 permanent teeth in each quarter of the mouth, making a total of 32 permanent teeth to succeed the 20 deciduous ones. Each quadrant has four specific types of teeth, which have different shapes and uses. These are the incisors, the canines (or cuspids), the bicuspids (or premolars) and the molars.

Incisors.—The permanent incisors, of which there are two in each quarter of the mouth (total number eight), are used principally for biting and cutting. They have a delicate tactile sense that enables them to be used for identifying objects in the mouth by nibbling. The biting portion is wide and thin, making a sort of chisel shape or cutting edge. The upper central incisor is the largest. The lateral incisor, next in line, is smaller and sometimes irregular in form. The lower ones are much smaller and are very similar to each other. They bite against the upper incisors, either edge to edge or slightly inward on the tongue side. An edge-to-edge biting position is acquired more and more as the teeth wear down, particularly among peoples whose main foods are coarse or gritty.

Cuspids.—Next to the incisors is the canine, or cuspid (total number four). It is at the corner of the mouth, the third tooth from the mid-line. It frequently is pointed and rather peglike in shape. The cuspid is absent in many mammals that have developed protective horns; in some other animals it is a great defensive weapon by virtue of its size. In man the cuspid tooth has the function of cutting and tearing. Its length and position are such that it does not normally protrude beyond the level of the other teeth, and this permits a side-to-side chewing motion that is not possible in the other primates, which have long canines.

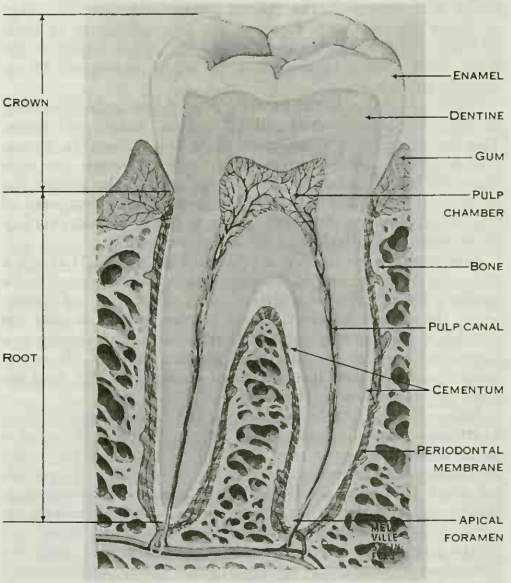
Bicuspids.—The two bicuspids, or premolars (total number eight), are next in line. They replace the baby molars but have a somewhat different function. Each has two cusps, although the lower second bicuspid frequently has a small third one. The first lower bicuspid often has a compressed appearance, as though it had been squeezed on the sides. This gives a prominence or sharpness to the forward sloping edge that is reminiscent of the scissors action against the long upper cuspid of man's remote ancestors. The roots of the bicuspids are reduced forms compared with those of fossil and modern anthropoids. The upper first bicuspid generally has two roots, sometimes one; the second generally has one root, sometimes two; the lower bicuspids have one root each. Occasionally they are found divided in two.

Molars.—The three molars (in each quadrant; total number 12) are the teeth farthest back in the mouth. They are used exclusively for grinding. The upper molars have four (sometimes three) cusps, whereas the lowers have five (sometimes four). The number and position of the cusps and the character of the grooves between them differ among individuals. Generally the upper molars have three roots and the lowers two.

In persons of European origin the first molar usually is large, the second smaller and the third the smallest. In some other groups the third molar is larger than the second, though smaller than the first.

Deciduous and Permanent.—In the milk or deciduous dentition there are only 20 teeth, 10 in each jaw, 5 in each right and left segment. They include, in each quadrant, 2 incisors, 1 cuspid and 2 molars. The molars are replaced by the permanent premolars or bicuspids. The incisors and cuspids are replaced by the permanent incisors and cuspids. In the deciduous dentition there are no predecessors to the three permanent molars. The deciduous dentition differs from the permanent in its smaller size, whiter colour, greater constriction of the necks of the teeth, and in the fact that the roots of the molars are widely spread to give a wider base and an accommodation for the developing permanent tooth germs.

Eruption.—Emergence or eruption of teeth from the gums is

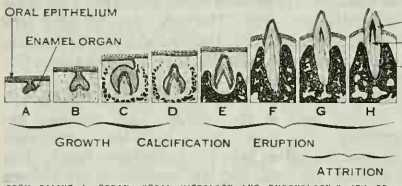


BY COURTESY OF AMERICAN DENTAL ASSOCIATION

FIG. 1.—CROSS SECTION OF A HUMAN MOLAR

variable among individuals both as to timing and as to order. In the deciduous dentition the lower first or central incisor comes in first, between the sixth and ninth month. This is followed by the upper central and lateral incisors. The other teeth erupt at four- to six-month intervals, the dentition being completed by two years of age. The sequence of permanent tooth emergence follows different patterns in different individuals. In some, the lower first molar is the first permanent tooth to appear; in others it is the lower central incisor. This occurs about the sixth year.

From this time on there is a succession of loss of deciduous



FROM BALINT J. ORBAN, "ORAL HISTOLOGY AND EMBRYOLOGY," 4TH ED. (1957), THE C. V. MOSBY COMPANY; MODIFIED FROM I. SCHOUR AND M. MASSLER, "STUDIES IN TOOTH DEVELOPMENT: THE GROWTH PATTERN OF HUMAN TEETH," "J. A. D. A.," VOL. 27 (NOV. AND DEC. 1940)

FIG. 2.—DIAGRAMMATIC ILLUSTRATION OF THE LIFE CYCLE OF THE TOOTH

(A) Initiation (bud stage); (B) proliferation (cap stage); (C) histodifferentiation (bell stage) and active morphodifferentiation; (D) apposition; (E) intraosseous; (F) eruption into oral cavity; (G,H) attrition

teeth and emergence of permanent teeth in the presence of jaw-bone growth, expansion and change. This process continues intermittently with rest periods until about age 12, when the second molars arrive. The third molars, or wisdom teeth, usually do not come in until about age 18 or 20. Occasionally a deciduous tooth is retained into adulthood when its successional tooth fails to develop or when the permanent tooth is in an abnormal position. This follows a definite pattern genetically, along with other phenomena of the dentition.

Histology.—Teeth are composed of several tissues (fig. 1).

Dentine.—The main bulk or core of each tooth is dentine, a calcified tissue made up of approximately 30% organic matter and water and 70% inorganic matter, principally calcium and phosphorus. Dentine is a highly sensitive tissue with a circulation of lymph in its tubules. The tubules extend from the inner pulp cavity to the periphery, more or less perpendicularly to these surfaces. The dentine extends almost the entire length of the tooth, being covered by enamel on the crown portion and by cementum on the roots.

Enamel.—Enamel is the hardest tissue in the body, consisting of 96% inorganic material and 4% organic matter and water. This tissue receives its nourishment, slight as it is, from the dentine that it covers. Some circulation of fluids in enamel has been shown by use of radioisotope tracer materials. The enamel has no nerve supply. Microscopically, it is seen to be composed of fine, hexagonal prisms arranged at right angles to the outer surface of the tooth and extending to the underlying dentine.

Cementum.—Cementum is another calcified tissue. It affords a thin covering to the root of the tooth and has the appearance of a modified type of bone. In a few instances, as in horses and rabbits, where the enamel is folded, the cementum extends beyond the confines of the root surface and fills these folds, thus serving as a substantial part of the functional chewing surface. It serves normally as a medium for attachment of the fibres that hold the tooth to the surrounding tissues.

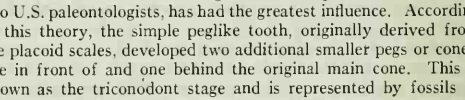
Pulp.—Nerves, lymph, blood vessels and fibrous tissue make up a structure generally known as the pulp or "nerve," occupying a cavity located in the centre of the tooth. The pulp canal is long and narrow with an enlargement in the coronal end. It extends for about four-fifths the length of the tooth, with communication to the general nutritional and nervous systems through the apical foramina at the end of the roots. Lining the entire wall of the pulp cavity are the resting odontoblasts that originally functioned to produce the tubules of the dentine. Upon stimulation from decay or other external contact to the dentine proper, the odontoblasts may again resume activity and produce another layer of so-called secondary dentine. This differs somewhat from the first dentine; it is less well organized and in certain respects resembles bone. This new addition of tissue (secondary dentine) is a protective mechanism, designed to counter the losses from the advancing carious or wear processes.

Embryology.—Teeth are produced by specialized tissues induced and organized from simple elements of the ectodermic and mesodermic layers of the embryo. Ten distinct and separate thickenings of certain cell groups occur in the epithelial covering of each jaw. Each of these stimulates some of the underlying mesodermic cell elements to transform their character for the specialized function of producing a protoplasmic substance in the form of tubules, which later calcify and form the dentine.

At the same time the cells of the thickened epithelial layer also differentiate into enamel rod producers. The sum total is a tooth germ in which the process of enamel and dentine production goes on simultaneously from a common starting line-up of cells, the enamel cells (ameloblasts) proceeding outward and the dentine cells (odontoblasts) inward (fig. 2). Thus, the enamel and dentine are developed in direct contact with each other. The tissues soon are calcified by deposition of calcium salts in the matrices of processes and tubules. The ameloblasts remain on the surface after enamel formation is completed. After eruption of the tooth they frequently are a source of annoyance by their affinity for stains and surface discoloration. At the termination of dentine production, the odontoblasts remain as a lining of the pulp chamber, ready to resume function by formation of the secondary dentine upon the proper stimuli (*see* above).

Evolution and Comparative Anatomy.—Teeth and bone have figured conspicuously in evolutionary studies and theory; they constitute major evidence in prehistoric studies. The teeth, particularly, persist through time by virtue of their calcified structure and their resistance to erosion and the various conditions of burial. With a few exceptions, the history of the evolution of tooth form and structure is complete. Comparison of the many modern dentitions with their predecessors reveals their origins and transition. Modern man's dentition differs from that of his fossil forebears in the size of teeth and in the proportionate development or modification of structural parts. Similar points of difference are found between man and the other anthropoids. The differences between tooth forms and combinations in the animal classifications are striking.

Various theories have been proposed to explain the evolution in form from the simple conelike tooth of the early reptiles to the complicated patterns of human molars. In early discussions of the subject disputes over some sequences and stages were common. The fossil record, however, now indicates clearly the pathway of these changes. The theory of trituberculy, developed largely on the basis of research and writings of E. D. Cope and H. F. Osborn, two U.S. paleontologists, has had the greatest influence. According to this theory, the simple peglike tooth, originally derived from the placoid scales, developed two additional smaller pegs or cones, one in front of and one behind the original main cone. This is known as the triconodont stage and is represented by fossils of

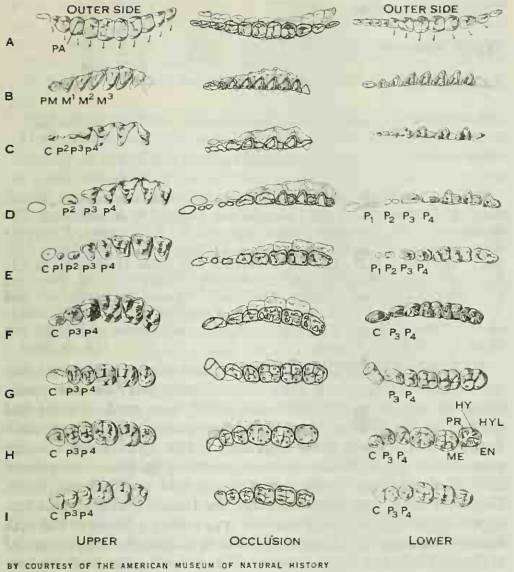


BY COURTESY OF THE AMERICAN MUSEUM OF NATURAL HISTORY

FIG. 3.—THREE MAJOR STAGES IN THE EVOLUTION OF TEETH

(A) Homodont stage in which all the teeth are the same form; (B) pre- and post-canine stage in which the dentition has been divided into anterior and posterior parts by the elongation of the canines; (C) heterodont stage in which the teeth have taken on specialized forms and functions, differing in the various locations of the dental arcade

extinct mammals. The two added structures were supposed then to have moved externally in the upper teeth and internally in the lower, to form a triangle. Specimens are known that show that these cusps evolved in this triangular position and did not migrate. This tritubercled tooth is considered the basic ancestral form from which all modern mammalian dentitions were derived, by addition and subtraction as well as by modification of cusps. The names



BY COURTESY OF THE AMERICAN MUSEUM OF NATURAL HISTORY

FIG. 4.—NINE STRUCTURAL STAGES IN THE EVOLUTION OF TEETH FROM THE SIMPLE HOMODONT STAGE (A) TO THE COMPLEX HOMINOID HETERODONT STAGE (I)

Key: PA = paracone; P = premolar; M = molar; C = canine or cuspid; PR = protoconid; HY = hypoconid; HYL = hypoconulid; ME = metaconid; EN = entoconid

of the cusps as developed in the Cope-Osborn system still are used in the literature, although later evidence shows a few discrepancies in the plan. As an example, the internal anterior cusp of modern upper primate molars bears the name protocone, indicating that it is homologous with the original reptilian single-coned tooth. It is in reality a later added cusp; the paracone, outer anterior cusp, is the representative of the first cusp. Nevertheless, the organization of thinking and the impetus given this field of study by Cope and Osborn's theories were extremely significant.

The first teeth were not teeth as they are now known. They were not calcified, nor were they firmly attached to the underlying supporting structures and tissues; neither were they confined to the jaws. Even today this is true of some lower forms of animals.

Fish display the earliest trends in the evolution of teeth. In some, the scales at the edge of the mouth are used as teeth and are replaced by the ones next in line as they are needed. Some fish have calcified structures as in the higher animal forms, generally confined, however, to a single cone shape. These are constantly formed in the jaw and erupt from time to time to succeed their predecessors as they are worn out or lost.

A brief description of the teeth of the members of the vertebrate classes follows.

Cyclostomata.—In the lampreys and hags the teeth are horny cones.

Elasmobranchii.—These include the cartilaginous fish, such as sharks and rays, in which the teeth are arranged in several rows. The hind ones move forward as those in the front row fall out. In some of these fish the teeth are set in positions like flat pavement blocks; in the rays they are like flat plates that crush against each other. This flatness gives them the appearance of the molars of man. Except for the saw of the sawfish (*Pristis*), they are not embedded in sockets. The body of the tooth, the dentine, is not like dentine of the higher vertebrates; it resembles bone, has no pulp cavity and is called osteodentine.

Teleostomi.—The teleostean and ganoid fish show great variability in teeth. In the sturgeon there are no teeth at all, while in other fish each of the several bones bounding the mouth, including the branchial arches, bears teeth. The pike's mouth and pharynx are completely armoured with teeth. The pike and hake have hinged teeth that bend to allow food and prey to pass down the throat, but become erect and bar exit of anything that has gone by. The hinge is a soft tissue attachment with elastic ligaments. Most of the Teleostomi have bonelike osteodentine, but the hake and a few others have vasodentine, which has many blood vessels running through it.

Amphibia.—These include the frog, toad and salamander, among others. Although the teeth are less numerous than in the first three classes, they do occur regularly on the vomer as well as on bones that make up the lower and upper jaws proper. The toad has no teeth, and the frog has none in the lower jaw. The extinct tailed amphibians, the Stegocephalia, are of particular interest because of their complex convoluted type of dentine, known as labyrinthian. They are called labyrinthodonts.

Reptilia.—These have considerable variety in the dentition. The first separation of anterior from posterior teeth by the specialized elongation of the canine tooth occurred in this class in a primitive thereapsid reptile, such as Ictidopsis.

In the Chelonia (turtles) there are no teeth, although the ectodermal ingrowth (dental band) from which they are developed in other animals is present in the turtle embryo. The place of the teeth in these reptiles is taken by horny jaw cases.

Among the Ophidia, the nonpoisonous snakes have two rows of teeth in the upper jaw, one on the maxilla and another on the palatine and pterygoid bones, while in the lower jaw there is only one row. These teeth are sharp pegs anchylosed to the bones and so strongly recurved that no prey that has passed the teeth can escape. The poisonous snakes have special modified teeth, the commonly known poison fangs in the maxilla of each side; the fang has a deep groove or canal running down it that conducts the poison from the poison gland. In the colubrine snakes, such as the cobra, the poison fang is always erect, but in the viperine snakes, such as the adder and rattlesnake, there is a mechanism by which the specialized tooth is erected only when the jaws are opened for striking. At other times the poison fang lies flat in the roof of the mouth.

In the lizards, or Lacertilia, the teeth usually consist of a series of pegs in the upper and lower jaw, each resembling the one in front of it. In some lizards, as the chameleon, the teeth are anchylosed by their bases to the bone; in others, as the iguana, they are fused by their sides to a wall-like ridge on the lateral surface of the bone. In the former case the dentition is called acrodont, in the latter pleurodont.

The teeth of the Crocodilia are fitted into definite sockets as in mammals and are not anchylosed with the jaws. This arrangement is called thecodont. In this group the 1st, 4th and 11th teeth of the lower jaw are larger than any of the others. In the upper jaw, the 3rd and 9th teeth from the mid-line on each side are enlarged. This regularity suggests the pattern of division of the dentition into the various tooth groups that are seen in the mammals generally.

Aves.—The modern birds are distinctive in having no teeth, though the fossil record shows that at one time they did. There are three known forms: (1) The archaeopteryx, about the size of the modern pigeon, had sharp pointed teeth. The jaws were long and slender, and the conical teeth arose from the margins. These teeth were distributed in a single row and resembled very closely the teeth of certain modern reptiles. (2) Ichthyornis had about 20 sharp pointed teeth arranged much like those of archaeopteryx. The tips of the teeth of this bird were slightly curved. They were implanted in sockets and had wide-open roots and a heavy cone of dentine covered with a cap of enamel. (3) Hesperonis had teeth that were not set in sockets but were placed in a groove with small bony partitions between the individual roots. The fossil birds all had teeth composed of osteodentine capped with enamel.

In these first six classes of vertebrates the dentitions are quite simple in form, being in the main of the cone type, except for particular specializations. Also in most of these classes there is a continuous succession of teeth, so that replacements for lost teeth

are always at hand. This arrangement for continual succession of teeth is called polyphyodont.

Mammalia.—Among the mammals may be found a very wide variety of dentitions and specializations. This class has the heterodont type (many-shaped) as opposed to the homodont (single-shaped) dentition that is found among the other classes. In the mammals the polyphyodont dentition, or continuous succession of teeth, is reduced to a diphyodont dentition; *i.e.*, there is only one replacement of the first set. In the marsupials the reduction is carried still further, for only one premolar in each segment of the jaw is replaced, while in the toothed whales there is no succession at all. When one set has to do duty throughout life the dentition is called monophyodont.

Some of the specializations among the mammals are extreme. The rodents, for example, have incisors that continuously grow throughout life. The horses, members of the Ungulata, have large teeth that are folded and have very long roots. These long, heavily ridged teeth are practically as efficient as the continuously growing variety. Cats, dogs and other carnivores likewise have teeth that evolved from the tritubercled form as well as the other orders; however, the dentition of these animals developed in a direction that would serve in cutting, grasping and shearing rather than grinding. Some mammals lost parts of their dentition rather than added to them. Some, though only a few, of the so-called edentates have lost all their teeth, but even the toothless members have vestiges of teeth in the early developmental stages.

All the primates have similar dentitions. The main differences between them are in the emphasis of expression of parts, in the relative sizes of parts, in the over-all size of the dentition and in the actual number of the types of teeth. The Lemuroidea exhibit a certain coarseness on the surface of the teeth, in contrast with human teeth. They also tend to vary in the number of premolars, from two to four in each of the four quadrants. The aye-aye has even lost all the canines; it retains only the upper and lower central incisor, one upper premolar and none of its lower premolars. The anthropoids follow a rather consistent formula as to number of teeth; the main exception is that the Platyrrhina, South American monkeys, have three premolars in each of the four quadrants, whereas the old world forms (Catarrhina) have two premolars in each quadrant. Some Platyrrhina also have two rather than the usual three molars.

Although following a basic pattern throughout, modern man's dentition does show considerable variation from one population to another. The main differentiating characteristics are to be found in the characteristics of the back or tongue side of the upper incisor teeth (absence or presence of shovel shape), in the expression of the areas near the gum margin of the posterior teeth (ridges, cusps and grooves), in the general wrinkling or smoothness of the occluding surfaces and in the over-all size of the dentition.

Failure of certain teeth to erupt has been reported among populations. Absence of one or more third molars, or wisdom teeth, has been reported in 19% of the white populations of central Europe; absence of upper lateral incisors is reported in from 2% to 3% in most groups. Reductions in the form or pattern of certain teeth and modification or absence of certain cusps are also frequently noted. These variations follow a regular pattern, with the third molars, lateral incisors and second premolars being the teeth most commonly affected. Differences in the frequencies of these occurrences have been used in anthropological studies to describe the dentition of various population groups. One population may be distinguished from another by comparing the different frequencies of these features. The frequencies reflect the characteristics of the genetic composition of populations.

Dental Formulas.—It is convenient to indicate the number of each kind of tooth in one side of the upper and lower jaw in the various mammalian orders by the use of dental formulas. The order and number are indicated from left to right, as in the case of man's formula: incisors $\frac{2}{2}$, cuspids $\frac{1}{1}$, premolars $\frac{2}{2}$, molars $\frac{3}{3}$. This is condensed into $\frac{2 \cdot 1 \cdot 2 \cdot 3}{2 \cdot 1 \cdot 2 \cdot 3}$. Some typical formulas are:

Primitive mammalian ..... $\frac{3 \cdot 1 \cdot 4 \cdot 3}{3 \cdot 1 \cdot 4 \cdot 3}$

Typical marsupial ..... $\frac{5 \cdot 1 \cdot 3 \cdot 4}{4 \cdot 1 \cdot 3 \cdot 4}$

Anthropoidea:

Platyrrhina ..... $\frac{2 \cdot 1 \cdot (2\text{-}3) \cdot (2\text{-}3)}{2 \cdot 1 \cdot (2\text{-}3) \cdot (2\text{-}3)}$

Catarrhina ..... $\frac{2 \cdot 1 \cdot 2 \cdot 3}{2 \cdot 1 \cdot 2 \cdot 3}$

Bibliography.—P. M. Butler, "Studies of the Mammalian Dentition," *Proc. Zool. Soc. Lond.*, series B 109 (1939); A. A. Dahlberg, *The Dentition of the American Indian* (1949); W. K. Gregory, *Evolution Emerging* (1951), "A Half Century of Trituberculy," *Proc. Amer. Phil. Soc.*, vol. lxxiii (1934); C. F. A. Moorrees, *The Aleut Dentition* (1957); B. Orban, *Oral Histology and Embryology* (1944); P. O. Pedersen, *The East Greenland Eskimo Dentition, Numerical Variations and Anatomy* (1949); J. T. Robinson, "The Dentition of the Australopithecinae," *Transv. Mus. Mem.*, 9 (1956); G. G. Simpson, *Tempo and Mode in Evolution* (1944). (A. A. Dg.)

TEETH, ARTIFICIAL, artificial substitutes replacing missing natural teeth. Such replacement dates from ancient times, when human and animal teeth and wood were used. In the last quarter of the 18th century one-piece restorations of porcelain began to be used, and at the beginning of the 19th century Giuseppangelo Fonzi, an Italian dentist practising in Paris, introduced individual porcelain teeth mounted on gold or platinum bases. Porcelain teeth were introduced into the United States about 1817, reputedly by Anthony A. Plantou (1774–1837). Early in the 19th century, through the efforts of J. Leon Williams, the appearance of artificial teeth was much improved by designing the teeth to harmonize with the shape of the face. Teeth were made in three typical facial forms, square, tapering and ovoid. At the same time the effectiveness of the chewing surfaces of the back teeth was improved by the anatomical designs of Alfred Gysi of Switzerland.

In modern times porcelain, plastics and glass teeth are made in such a great variety of shapes, sizes and colours that almost any natural teeth can be closely matched. Modern artificial teeth are usually fastened to plastic bases that resemble the natural gums. This combination of artificial teeth fastened onto a base is an artificial denture and is commonly referred to as a set of false teeth or a dental plate. Artificial teeth are also used in dental bridges, which are cemented to natural teeth.

Artificial teeth are shipped from the factory inserted in wax on plastic cards or blocks. A card usually contains a set of six front teeth (two cuspids and four incisors of either the upper or lower jaw) or a set of eight back teeth (four bicuspids and four molars of either jaw). However, the dentist or dental laboratory may purchase singly or in any combination desired. The normal number of natural teeth is 32, but the four third molars (wisdom teeth) are not supplied in artificial sets because there is usually not room for them in an artificial denture.

Porcelain teeth are prepared by fusing a mixture of feldspar (orthoclase), flint (silica) and clay (kaolin) at about 2,400° F. The high-melting flint and clay are fused together in a vitreous mass formed by the relatively low-melting feldspar. Suitable opacities and colours are obtained by small amounts of metallic oxides such as those of titanium, chromium, nickel, cobalt, iron and uranium. The uranium oxide causes the artificial teeth to glow (fluoresce) in ultraviolet light in much the same manner as natural teeth.

A common method of making porcelain teeth consists of carving the desired shapes and sizes out of plaster of paris or other suitable material. These models are made one-fifth larger than the finished product, to allow for shrinkage during fusion. Split metal molds are made from these carved models. Starch, powdered porcelain and water are mixed to a thick paste and packed into the molds. The powdered porcelain used in these pastes may be varied in composition in order to obtain the desired opacity and colour. The packing of the mold is done in stages. The porcelain that resembles the translucent enamel covering of natural teeth is packed into the appropriate position in the mold. A more opaque porcelain, which resembles the dentine core of natural

teeth, is placed in another part of the mold. The molds are closed and heated to bake the porcelain to a hard state so that the teeth can be removed for a final fusing. Other methods employ flexible rubber molds into which the porcelain paste is vibrated until it is condensed sufficiently to be removed from the mold for final fusing.

Approximately one-fifth of all modern artificial teeth are made of plastics, consisting principally of polymethyl methacrylate. Relative to porcelain teeth, plastic teeth are less brittle, easier to grind and polish, make a firmer union with the plastic base and make less noise when being used. However, plastic teeth have poorer chewing efficiency and do not wear so well as porcelain teeth. It is for these reasons that plastic teeth have not supplanted porcelain teeth. *See* also PROSTHODONTICS.

See E. W. Skinner, *The Science of Dental Materials,* 4th ed. (1954); Stanley D. Tylman, *Theory and Practice of Crown and Bridge Prosthesis,* 3rd ed. (1954). (G. B. D.; G. C. PE.)

TEGEA, an ancient Greek city of Arcadia, on a plateau enclosed by Mt. Parthenium and Mt. Maenalus on east and west, and by transverse ranges which separate it from the plateau of Orchomenus and from the Eurotas valley. Tegean territory occupied the southern part of this plain; the northern half belonged to Mantineia (*q.v.*). Its geographical position accounts for the conflicts which arose with Mantineia and with Sparta.

Tegea was one of the most ancient cities of Peloponnesus; tradition ascribed its concentration (*synoecism*) out of eight or nine primitive cantons to a mythical king Aleus. As several Cretan townships passed for colonies of Tegea, oversea connections may be inferred in prehistoric days. The prominence which legend assigns to its king Echemus in opposing the Heraclid invasion shows that it was one of the chief Peloponnesian communities in pre-Dorian days. For several centuries Tegea screened Arcadia against expanding Sparta; ultimately subdued about 550 B.C. it was allowed to retain independence and Arcadian nationality. During the Persian invasion the Tegeans displayed a readiness unusual among Peloponnesian cities; in the battle of Plataea they were the first to enter the enemy's camp. A few years later they headed an Arcadian and Argive league against Sparta, but after losing two pitched battles at Tegea and Dipaea they resumed their former loyalty about 468–467. In 423 there was open war with the Mantineians, and when the latter rebelled against Sparta and allied themselves with Argos and Athens, the Tegeans stood firmly by Sparta's side: in the decisive battle of Mantineia (418) their troops had large share in the overthrow of the coalition; and during the early 4th century Tegea continued to support Sparta against the Mantineians and other malcontents. But after the battle of Leuctra the philo-Laconian party was expelled with Mantineian help.

Tegea henceforth took an active part in the revival of the Arcadian league and in alliance with Thebes against Sparta (371–362), and the defection of Mantineia confirmed its federalist tendencies. The foundation of the new federal capital Megalopolis threw Tegea somewhat into the shade. Hostile to the Macedonians, in 266 it joined the Chremonidean league against Antigonus Gonatas. To the incorporation of Mantineia into the Achaean league (233) Tegea replied by allying itself with the Aetolians, who in turn made it over to Cleomenes III of Sparta (228). From the latter it was transferred by Antigonus Doson to the Achaean league (222); in 218 it was again occupied by the Spartans but reconquered in 207 by the Achaean general Philopoemen. In Augustus' time Tegea was the only important town of Arcadia, but its history throughout the Roman and Byzantine periods is obscure; it ceased to exist as a Greek city after the Gothic invasion of 395. The site is now occupied by the small village of Piali.

The temple of Athena Alea at Tegea is described by Pausanias as excelling all others in Peloponnese. The original temple built by Aleus, the founder of the city, was superseded by a larger one which was destroyed by fire in 395 B.C. Rebuilding was entrusted to Scopas, the great sculptor; and probably he also provided the pediment sculptures, representing at the front, the hunt of the Calydonian boar, and, at the back, the battle of Achilles and Telephus. *See* GREEK ART.

TEGERNSEE, a lake in upper Bavaria, situated in beautiful mountain country, 2,380 ft. above the sea, 34 mi. south from Munich by rail to Gmund, a village on the north shore. The lake is 4 mi. long, 1 mi. broad, and about 237 ft. deep. Its waters discharge through the Mangfall into the Inn. It is one of the most frequented summer resorts in the vicinity of Munich. The village of Tegernsee, on the east bank, has a 15th-century parish church and a ducal castle (formerly a Benedictine monastery).

TEGNÉR, ESAIAS (1782–1846), the greatest Swedish poet of his age. Born at Kyrkerud in Värmland, on Nov. 13, 1782, he was thrown on his own resources by his father's death in 1792. He entered Lund university in 1799, graduated in 1802 and in 1812 was appointed professor of Greek. He continued to lecture at Lund until 1824, when he became bishop of Växjö.

Tegnér first won literary fame with his patriotic poem *Svea,* which was awarded the highest prize of the Swedish Academy in 1811. In 1812 he joined the *Götiska förbundet,* a romantic, nationalist league founded by Erik Gustaf Geijer and others, and between 1812 and 1813 published some of his finest poems in its periodical, *Iduna.* The solemn "Skaldens morgonpsalm" and "Sång till solen" were particularly notable. His topical poems—*e.g.,* "Hjälten," a tribute to Napoleon—are also remarkable. He was originally associated with the romantic movement, but later reacted against its emotional and mystical aspect, *fosforismen,* so-called from its periodical *Fosforos,* and represented by the work of Atterbom. "Strength and clarity" were what he demanded of poetry in the famous *Epilog vid magisterpromotionen i Lund 1820.* He had entered the Swedish Academy in the previous year. His greatest poetic achievement was *Frithiofs saga,* a cycle based on an Old Icelandic saga; it became a national poem and was translated into almost every European language. Other outstanding poems of this period were *Nattvardsbarnen* (1820) and the verse narrative *Axel* (1822; Eng. trans. by R. G. Latham, 1837), modeled on Byron.

The year of the publication of *Frithiofs saga* (1825) marked not only a peak in Tegnér's development as a poet, but also a turning point in his life. From then on he was a sick man, and an unhappy love affair added to his suffering. He describes his state of mind in the mid-1820s in a moving poem, *Mjältsjukan* ("Melancholia"). After 1826 his duties as bishop left him little time for poetry and yet some of his famous poems such as "Sång den 5 april 1836" (commemorating the 50th anniversary of the founding of the Swedish Academy) and "Den döde," a love poem dedicated to a young married woman in Växjö, date from this period. He devoted special attention to education in his diocese, and his addresses at the end of the school year are of great interest. As bishop he was automatically a member of the *riksdag* and had to spend long irksome periods in Stockholm. Originally a liberal, he became increasingly conservative and entered into controversies with the successful liberal spokesmen of the time in parliament and press. These disputes and his ill-health led to embitterment. Following a stroke in 1840 he showed signs of mental disorder. During a period of convalescence he wrote "Kronbruden," an idyll in hexameters which is his last important poem. He died at Östrabo, near Växjö, on Nov. 2, 1846.

Tegnér's poetic reputation has been confirmed by posterity. He is also regarded as Sweden's most brilliant orator and wittiest letter writer. His letters did not become widely known until his posthumous works were published by his grandson in 1873–74, at a time when the public was becoming weary of lyrical romanticism and ready to appreciate Tegnér's realistic prose style, distinguished by sparkling wit and pregnancy of expression.

An edition of Tegnér's works, by E. Wrangel and F. Böök, was published in 10 vol. (1918–25). A complete edition of his letters was begun by the Tegnér society (1953).

BIBLIOGRAPHY.—G. Brandes, *Esaias Tegnér* (1878); F. Böök, *Esaias Tegnér,* 2 vol. (1946); E. Wrangel, *Tegnér i Lund,* 2 vol. (1932); A. Werin, *Esaias Tegnér. Från Det eviga till Mjältsjukan* (1934). (AT. WN.)

TEGUCIGALPA (tā-gōō-thē-gahl'pą), the capital of Honduras and of the department of Francisco Morazán; 3,200 ft. above

sea level, on the Choluteca, at the head of a highway to the port of San Lorenzo on Fonseca bay. Tegucigalpa is the only capital city of the American republics without railway connections. Pop. (1961) 134,075, including the twin city of Comayagüela. Tegucigalpa became capital of Honduras, a status it had previously shared with Comayagua, in 1880. During the 18th century the neighbourhood was famous for its gold, silver and marble, but in modern times the mines and quarries have greatly declined in value and farming is the chief local industry.

TEHERAN (Tehran), the capital of Iran and the chief city of Teheran *ostan* (province). The city is situated on the northern fringe of the central plateau, 10 mi. to the south of the Elburz (Alborz) mountains which skirt the southern shore of the Caspian sea. The city, which has a mean altitude of 3,800 ft., stands on ground rising gently toward the north. Pop. (1956) 1,512,082; (1964 est.) 2,200,000.

In the summer, during weekends and on public holidays, many of the citizens escape from the heat by going to the high country to the north and northeast, where there are cool resorts surrounded by magnificent mountain scenery. In winter there is good skiing on the hills above Lashkarak, 10 mi. to the northeast. The snow-capped cone of Mount Demavend, an extinct volcano (18,580 ft.), 44 mi. E.N.E., is visible in clear weather.

Teheran possesses few buildings of historical interest and none of any real antiquity. In the centrally situated Maidan-i-Sipah (army square) is the Shahrdari (town hall) and nearby are various ministry buildings, the police headquarters (a modern building in Achaemenian style), and the archaeological museum which houses a most interesting collection of antiquities from Susa and other ancient sites. To the south of the Maidan-i-Sipah is the Gulistan palace (early 19th century), which is now a museum containing many objects of historical interest, including the so-called Peacock throne (*see* Delhi: *Old Buildings and Monuments*). Nearby, in the bazaar quarter, is the Masjid-i-Shah mosque, also dating from the early 19th century. Some distance to the east-northeast of the Maidan-i-Sipah is the Baharistan palace, in which the *majlis* (parliament) meets; it also contains a large library. Just south of the Baharistan palace is the Sepahsalar mosque, now a theological college and library. In the western part of the city is the Kakh-i-Marmar (marble palace), one of the residences of the shah; it has a tiled dome modeled on that of the beautiful Shaikh Lutfullah mosque in Isfahan. In the northwest, at Jalaliyeh, is the University of Teheran (founded 1935), the buildings of which are in a large and well laid-out garden. In the southern part of the city, at the foot of the Khiaban-i-Pahlavi (Pahlavi avenue), is the large and imposing railway station. There are two public gardens, the Bagh-i-Milli, to the south of the Maidan-i-Sipah, and the Bagh-i-Shah, near the western outskirts.

Communications and Industries.—Teheran is the focal point of the Iranian railway system, with lines radiating to Kazvin (Qazvin) and Tabriz, to Ahvaz and thence to Khorramshahr (on the Karun river where it enters the Shatt al Arab) and to Bandar-e Shahpur on an inlet of the Persian gulf (the Trans-Iranian line), to Meshed in the east, to Bandar-e Shah, a port on the Caspian coast, and southeast to Kashan. The city is also an important road centre, as it is linked with Rasht and Bandar-e Pahlavi (formerly Enzeli) to the northwest, with Tabriz (and thence to the Soviet and Turkish frontiers), with Hamadan, Kermanshah and the Iraq frontier, with Qom, Isfahan, Shiraz and Bushire to the south, with Yazd, Kerman and Zahedan to the southeast, with Shahrud and Meshed to the east, and with the Caspian coast at Chalus to the north. The main airport is at Mehrabad, 5 mi. W. of the city, served by international airlines.

Teheran has factories for the manufacture of textiles, cement, cigarettes, sugar, glass, leather, small arms and ammunition.

History.—Compared with cities such as Isfahan, Teheran is of comparatively recent foundation. It is believed to have come into existence some time in the 9th century A.D., but, as it was only 5 mi. from the then large and flourishing city of Ray (*q.v.*), it remained a small and insignificant town until after the almost total destruction of that city by the Mongols in 1220. During the Safavid period (1501–1736) Teheran slowly developed into a town of medium size. In the second half of the 18th century, Karim Khan Zand was so impressed with the situation of Teheran that he thought for a time of making it his capital but eventually chose Shiraz instead. In 1785 Agha Mohammed Khan, the founder of the Qajar (Kajar) dynasty, besieged and captured Teheran and thereupon made it the seat of his government, since when it has been the capital of the country. Fath Ali Shah (1797–1834), Agha Mohammed's successor, embellished the town, completing the Gulistan palace and erecting the Masjid-i-Shah and a number of other buildings. During the long reign of Nasir ad-Din Shah (1848–96) Teheran was largely rebuilt and grew appreciably in size. After the accession of Reza Shah Pahlavi in 1925 the process of modernization, which had already begun, was greatly speeded up. Many broad, asphalted thoroughfares were cut through congested areas, thus clearing slum districts and facilitating the flow of traffic. The development of the city is particularly noticeable in the north and northwest. Many modern buildings have been erected and those which use the Achaemenian and Sassanid styles externally have modern interiors. Between 1925 and 1955 the population increased more than sixfold. (M. A. Ma.; L. Lo.)

Teheran Ostan (or Central Province) comprises a large part of the southern slopes of the Elburz mountains (*q.v.*), a number of districts in the interior highlands, the northwestern fringe of the great central plain and part of the Daryacheh-ye-Namak desert. The province is divided into the *shahrestans* of Teheran, Karaj, Qazvin, Damavand, Mahallat, Saveh, Qom, Tafresh, Rey, Shemiranat, Garmsar and Khomeyn. Pop. (1956) 2,874,511; area 23,460 sq.mi. (60,761 sq.km.).

Teheran *ostan* is the focal point of Iran's chief roads and railways. About a third of the country's industries are located there, the chief being the manufacture of consumer goods, sugar refining, textiles, coal mining, iron founding and the production of munitions and cement. The province is connected by oil pipeline with Abadan refinery and with the new oil field of Alborz near Qom. The Karaj and Latiyan dams provide a water supply for Teheran as well as electric power and water for irrigation. (H. Bo.)

Bibliography.—Sir Herbert Thomas, *Travels in Persia, 1627–1629* (1928); Guillaume-Antoine Olivier, *Voyage dans l'Empire Othoman, l'Egypte et la Perse* (1807); L. Lockhart, *Persian Cities* (1960).

TEHERAN CONFERENCE: *see* World War II Conferences, Allied.

TEHRAN: *see* Teheran.

TEHRI-GARHWAL and UTTAR KASHI, two districts in the extreme northwest of Uttar Pradesh, India. Formerly a princely state (Tehri) set up under a Rajput dynasty by the British after the Nepalese war in 1815, the combined area was administered with the United Provinces till 1938, and then controlled through the Punjab residency till partition. It was merged with Uttar Pradesh on Aug. 1, 1949, and formed into a separate district for purposes of administration, known as Tehri-Garhwal. Since 1960 it has been divided into two, the northern part being Uttar Kashi (in Uttarakhand division), the southern Tehri-Garhwal (in Kumaun division). The populations in 1961 were Tehri-Garhwal (347,736), Uttar Kashi (122,836). Areas were respectively 1,745 sq.mi. and 3,018 sq.mi. Lying entirely in the Himalayas, the area rises in places to over 20,000 ft. The sources of the Ganges and the Jumna rise in this area, and associated with them are the Hindu shrines Gangotri (*q.v.*) and Jamnotri, both of which lie in Uttar Kashi. The Bhagirathi river originates in a glacier known as *Gaumukh* and joins the Alaknanda at Devaprayag to form the Ganges. The Jumna originates west of the peak of Bandarpunch (20,720 ft.). Because of rugged topography and forests, only about 4% of the total area is cultivated; rice, wheat and barley are grown. Narendranagar, the headquarters of Tehri-Garhwal, is 16 mi. N.E. of Hardwar. The headquarters of Uttar Kashi is the town of Uttar Kashi. (S. S. Bh.)

TEHUANTEPEC, ISTHMUS OF (Istmo de Tehuantepec), lies between the gulfs of Campeche and Tehuantepec, with the Mexican states of Tabasco and Chiapas on the east and Veracruz and Oaxaca on the west. It includes that part of Mexico lying between the 94th and 96th meridians of longitude west. The isthmus is 137 mi. across at its narrowest part from gulf to

gulf, or 120 mi. to the head of Laguna Superior on the Pacific coast. The Sierra Madre breaks down at this point into a broad, plateaulike ridge, whose elevation, at the highest point reached by the Tehuantepec railway (Chivela pass), is 781 ft. The northern side of the isthmus is swampy and densely covered with jungle. The whole region is hot and malarial except on the open elevations where Pacific winds render it comparatively cool and healthful. The annual rainfall on the Atlantic or northern slope is 48 in. and the maximum temperature about 95° F. in the shade. The Pacific slope has a light rainfall and drier climate.

Since the days of Hernán Cortés, the Tehuantepec isthmus has been considered a favourable route, first for an interoceanic canal and later for an interoceanic railway. Its proximity to the axis of international trade gave it some advantage over the Panamá route, which was counterbalanced by the narrower width of Panamá. When the great cost of a canal across the isthmus compelled engineers and capitalists to give it up as impracticable, numerous projects for the construction of a railway were initiated. The railway was finally opened to traffic in 1907; it is 192 mi. long, extending from Puerto México on the gulf of Campeche to Salina Cruz on the gulf of Tehuantepec. There are rail connections to Veracruz, Mexico City and the Guatemalan border. Salina Cruz is a free port with large dry docks. The town of Tehuantepec is the centre of the distinctive Tehuanas, whose social organization is matriarchic. The Tehuana women are noted for their beauty and for their colourful fiesta dress.

(J. A. Cw.)

TEHUELCHE, an almost extinct Indian tribe once occupying the plains of Patagonia from the Río Negro to the Strait of Magellan. Their great stature and physical vigour have created a giant tradition in European literature. As the arid, wind-swept nature of Patagonia unfits it for agriculture, the Tehuelche necessarily were nomadic hunters. Before the 16th century they lived much as did their kinsmen, the Ona (*q.v.*), in Tierra del Fuego, but the advent of the horse resulted in greatly enlarging their sphere of action and in various cultural changes.

TEIGNMOUTH, JOHN SHORE, 1st Baron (1751–1834), governor-general of Bengal (1793–98), was born in London on Oct. 8, 1751. He was educated at Harrow and went out to India as a writer in the Bengal civil service in 1769.

Shore became an acknowledged authority on revenue matters; he sat on the governor-general's supreme council (1787–89) and had much influence with Lord Cornwallis (then governor-general). The zamindari revenue system was adopted on his recommendation though he did not approve of a permanent revenue settlement. As Sir John Shore he became governor-general (1793) and continued the approved policy of noninterference in the affairs of the Indian states. He aroused criticism by failing to help his ally, the nizam of Hyderabad, against the Marathas in 1795, but he acted with vigour in replacing Wazir Ali by Saadat Ali as ruler of Oudh on very favourable terms. He was recalled (1798) on account of his concessions to a combination of officers discontented with Cornwallis' economies. He was created Baron Teignmouth in the Irish peerage for his handling of the Oudh affair. (*See* also India-Pakistan, Subcontinent of: *History.*)

From 1807 to 1828, as Lord Teignmouth, he was a member of the board of control (of Indian affairs), but as a member of the evangelical "Clapham sect" he became more interested in religion than politics. From 1804 until his death on Feb. 14, 1834, he was the first president of the British and Foreign Bible society.

See Charles Shore, 2nd Baron Teignmouth, *Memoir of the Life and Correspondence of John, Lord Teignmouth,* 2 vol. (1843); H. Furber (ed.), *Private Record of an Indian Governor-Generalship* (1933).

(T. G. P. S.)

TEIGNMOUTH, a seaside resort and residential town in the Tiverton parliamentary division of Devon, Eng., 16 mi. S.S.E. of Exeter by road. It lies on the north side of the Teign, and the urban district includes the parish of Shaldon on the south side of the river. At Shaldon the Ness headland and estate are a public park. Pop. of urban district (1961) 11,576. The town spreads along the valley and over the wooded uplands to Little Haldon (800 ft.) where 43 ac. of heathland belong to the National trust.

Teignmouth possessed a church of St. Michael as early as 1044, when what is now East Teignmouth was granted by Edward the Confessor to Leofric, bishop of Exeter. In the middle ages Teignmouth was a flourishing port, able to supply 7 ships and 120 mariners to the Calais expedition of 1347, and depending chiefly on the fishing and salt industries. Teignmouth was raided by the French in 1340 and again in 1690. It was a 19th century resort. The harbour is still used, and there is sea and river fishing.

TEILHARD DE CHARDIN, PIERRE (1881–1955), French Jesuit paleontologist and thinker whose originality invites comparison with that of such devout pioneers as Pico della Mirandola, Ramon Llull or even Origen, was born on May 1, 1881, at Sarcenat, near Orcines, Puy-de-Dôme. Having acquired from his father a habit of keen observation and an interest in geology, he was sent in 1892 to the Jesuit college of Mongré, at Villefranche-sur-Saône. In 1899 he joined the Jesuit noviceship at Aix-en-Provence. Because of the expulsion of the Jesuits from France, his subsequent studies were carried out in Jersey (1901–05) and finally, after an interval of teaching in Cairo (1905–08; during which time he carried out his first explorations in the Fayum), at Hastings in England, where he was ordained priest in 1911. He then studied paleontology under Marcelin Boule at Paris (1912–14). He served during World War I as a regimental stretcher-bearer (battles of Ypres, Verdun, Chemin des Dames and the Marne) with the rank of corporal, though he was offered the post of chaplain with the rank of captain. He taught at the Institut Catholique in Paris for some years but soon began his visits to China, where in Dec. 1929 he was concerned in the discovery of an exemplary specimen of *Sinanthropus.* Further travels took him to the Gobi desert, Sinkiang, Kashmir, Java and Burma. He spent 1939–46 at Peking in near-captivity. After World War II he returned to Paris. In 1951, however, making a detour by way of South Africa, he went to the United States to spend the rest of his life with research teams there (apart from another visit to South Africa in 1953). He died in New York on April 10, 1955.

Most of Teilhard's writings were scientific. His philosophical books were the product of long meditation: *Le Milieu divin* appeared in 1926–27, *Le Phénomène humain* in 1938–40. *L'Apparition de l'homme* (1956) and *La Vision du passé* (1957) consist of opuscula put together for the collected edition of his works. He aimed at a metaphysic of evolution, holding (against Bergson) that it was a process converging toward a final unity. As a pupil of Léonce de Grandmaison in theology his thought was Christocentric, and he revitalized with his talk of cosmic redemption the ideas of Irenaeus and Duns Scotus about the recapitulation of all things in Christ. Natural optimism did not prevent his preoccupation with the place of sin in the evolutionary process. His broad syntheses often met with incomprehension or suspicion.

Bibliography.—C. Cuénot, *Pierre Teilhard de Chardin . . .* (1958); N. Corte, *Pierre Teilhard de Chardin,* Eng. trans. (1960); C. E. Raven, *Teilhard de Chardin: Scientist and Seer* (1962). Books and articles about Teilhard are listed annually in *Archivum historicum Societatis Iesu.*

(J. H. Cn.)

TEIRESIAS, in Greek legend, a famous Theban seer, son of Eueres and Chariclo, a descendant of the *Spartoi* (*see* Cadmus). He was blinded early in life either because he saw Athena (or Artemis) bathing naked, or because he told Zeus and Hera that women had far more pleasure of love than men, thus bitterly offending the latter. At all events, he was comforted by the gift of seercraft and granted a very long life, finally being allowed, even in Hades, to retain his wits and not be a mere phantom. As a result of killing coupling snakes he turned into a woman; years later he repeated the action, and became a man again. He died just after the capture of Thebes by the Epigoni. His grave was at the Tilphusian spring; but there was a cenotaph of him at Thebes, and also in later times his "observatory," or place for watching for omens from birds, was pointed out. He had an oracle at Orchomenus, but during a plague it became silent and remained so in Plutarch's time.

TEISSERENC DE BORT, LÉON PHILIPPE (1855–1913), French meteorologist, noted for his discovery of the stratosphere, was born in Paris on Nov. 5, 1855, the son of an engineer. He began his scientific career in 1880, when he entered

the meteorological department of the Central Bureau of Meteorology in Paris under E. E. W. Mascart. In 1883, 1885 and 1887 he made journeys to north Africa to study geology and terrestrial magnetism, and during that period published some important charts of the distribution of pressure at a height of 4,000 m. In 1892 he became chief meteorologist to the bureau but resigned in 1896 and founded a private meteorological observatory at Trappes, near Versailles, where he carried out investigations on clouds and the problems of the upper air. In 1898 he published an important paper in *Comptes Rendus* detailing his researches by means of balloons into the constitution of the atmosphere. His discovery of the so-called isothermal layer, or stratosphere, as it is now generally called, will always stand out as one of the most important events in the study of the upper atmosphere. He also carried out investigations in Sweden and over the Zuider Zee, the Mediterranean and the tropical region of the Atlantic, and fitted out a special vessel in order to study the currents above the trade winds. He collaborated with Hugo Hildebrandsson in *Les Bases de la météorologie dynamique* (1900–05). In 1908 he was awarded the Symons gold medal by the Royal Meteorological society. He died at Cannes on Jan. 2, 1913.

TEKAKWITHA (Tegakwitha, Tegakouita), **KATERI** (1656–1680), the first North American Indian ever presented to Rome for sainthood, was born probably at Ossernenon (Auriesville), N.Y., in 1656. Her father, a pagan Mohawk chief, her mother, a Christian Algonkian, and her brother died of smallpox when she was four. The disease left her with pitted skin, poor eyesight and damaged health. Staying with her anti-Christian uncle, she was deeply impressed at the age of 11 by the lives and words of three passing Jesuits, the first Christians she had spoken to. For years she refused marriage proposals and virtuously performed drudgeries common to squaws. At 20 she was instructed in religion and baptized Katherine by the Jesuit Jacques de Lamberville, who had been astonished at her knowledge and practice of Christian ideals. In her home village she was jeered, stoned, calumniated and threatened with torture. She narrowly escaped death by fleeing 200 mi. through forests to the Christian Indian mission of St. Francis Xavier at Sault St. Louis near Montreal. There in summer she toiled in the fields and during the winter hunts she carried heavy loads. Practising her faith heroically in suffering, prayer and kindness, she merited the title "Lily of the Mohawks." She died at the age of 24, on April 17, 1680. Her practice of heroic virtue was recognized by the Sacred Congregation of Rites in 1943.

See *Positio of the Historical Section of the Sacred Congregation of Rites on the Introduction of the Cause for Beatification and Canonization and on the Virtues of the Servant of God, Katherine Tekakwitha, the Lily of the Mohawks* (1940). (J. V. J.)

TEKIRDAG (Rodosto), a city of European Turkey, on the Sea of Marmara, 78 mi. W. of Istanbul. A Greek settlement founded probably in the 7th century B.C., it was originally called Rhaedestus or Bisanthe, later Rodosto. Since 1923 it has been officially known under its Turkish name of Tekirdag ("prince's hill"). Rodosto was for centuries the port of the Adrianople (Edirne) area, but its trade suffered when Dédéagatch (Alexandroupolis), on the Aegean sea, became the terminus of the railway up the Maritsa river. The city was conquered by the Turks in 1360; it was occupied by the Russians in 1877–78, by the Bulgars in 1912 and by the Greeks in 1920–22. The Greek church of Panagia Rheumatocratissa contains the graves, with Latin inscriptions, of Hungarians banished from their country early in the 18th century by the German captors of Buda. The population of Tekirdag in 1914 was estimated at 25,000, half of it Greek. In 1927, after the Greek-Turkish exchange of populations, Tekirdag had only 14,569 inhabitants; by 1960 it had 23,905. It is the chief town of the *vilayet* (pop. [1960] 274,806) of the same name.

TEKTITE, the name given by F. E. Suess (from the Greek *tektos,* "melted"), for the natural glass objects found in various parts of the world and thought to be either meteorites or the result of impact on earth by meteorites, asteroids or comets. They range in weight from less than a gram to about eight kilograms and are of varied colour, shape and surface sculpture. Most are black, some are green, a few are yellow. Some are lustrous and some have a delicate sheen from minute alternating ridges and furrows which swirl over the entire surface, following the contorted internal structure. The younger, less-corroded tektites include spherical, elliptical, lenticular, teardrop, dumbbell, disk and button forms. They have been used as charms or amulets; and in folklore were regarded as an index of the richness of alluvial gold and tin deposits in which they are often found.

Tektites came to rest in groups at widely different times in the earth's history, with none arriving in between. They range from a few thousand years old (Australia) to about 50,000,000 years old (Texas). They are known on every continent except Antarctica and South America and are usually given distinctive names according to where they are found. Among these names are moldavites (Bohemia), australites, billitonites and javanites (Indonesia), indochinites, philippinites, bediasites (Texas), Darwin glass (Tasmania) and Libyan desert glass (Sand sea, Egypt). They are known from the Ivory Coast and Georgia. The americanites from Colombia and Peru, originally believed to be tektites, are now thought to be volcanic glass. The button form of australite, showing two periods of melting, furnishes strong evidence that tektites came from outside the atmosphere. Such a form can be explained only if a sphere of glass plunged into our atmosphere. It would melt on the forward side, the area of greatest friction, but as the fragment slowed, some glass would flow toward the lee and curl to form a flange. There is equally strong evidence that most tektites formed from sedimentary rocks like those on earth. Sand-size glass particles in tektites, identified by Virgil E. Barnes as lechatelierite (fused quartz), have a consistent earthly counterpart only in the quartz of sandy and shaly sedimentary rocks (*see* Silica: *The Silica Minerals*). Chemically, tektites are more closely related to sandy and shaly sedimentary rocks than to any other rock group. Because of this G. Linck believed tektites came from another planet blanketed by sedimentary rocks.

L. J. Spencer thought tektites might be formed from meteorite impacts on earth. Harold C. Urey in 1955 suggested that the impacting bodies might be asteroids, but in 1957 proposed that shock waves from comets might be a more logical means for their formation. If the button form was produced by impact or shock waves, then molten glass must have been ejected beyond the atmosphere to solidify, before plunging back into the atmosphere.

See also Meteorites: *Glass Meteorites;* Geochemistry: *Meteorites and Geochemistry.*

Bibliography.—F. E. Suess in *Jb. Geol. Reichsanst.,* 50:103–382 (1900); G. Linck in *Neues Jb. für Miner.,* 57:223–236 (1928); L. J. Spencer in *Nature, Lond.,* 131:117–118, 876 (1933); V. E. Barnes (bibliography), *Univ. Tex. Publ. 3945,* pp. 477–582 (1940); H. C. Urey in *Proc. Nat. Acad. Sci., Wash.,* 41:27–31 (1955), *Nature, Lond.,* 179:556–557 (1957). (V. E. B.)

TELAMON (mod. Talamone), an ancient city on the coast of Etruria, 11 mi. N. of Orbetello, where the Romans defeated the invading Gauls in 225 B.C. Marius disembarked there on his return from Africa in 87 B.C., but in 82 the town was destroyed by Sulla's troops and thereafter declined. The harbour was of some importance in the middle ages. The ancient site lay on the hill of Talamonaccio. (H. H. Sd.)

TELAMON (literally "shield strap"), in Greek mythology, son of Aeacus, brother of Peleus, half-brother of Phocus and father of the greater Ajax and of Teucer. He helped Hercules in his capture of Troy (and was rewarded with the gift of the Trojan princess Hesione). When Teucer returned from the more famous later Trojan War without his half-brother, Ajax, Telamon banished him. His traditional home was on the island of Salamis where he was worshiped as a hero in historic times. But to judge by his name he was originally invented as an appropriate father for Ajax, whose chief feature in the primitive Greek epic was his colossal shield. (H. W. Pa.)

TEL AVIV-JAFFA, a twin city of Israel on the Mediterranean sea, 47 mi. N.W. of Jerusalem by road. Jaffa (Yafo, Japho, Joppa, [*q.v.*]) is an ancient port, dating back possibly to 2000 B.C.; Tel Aviv is a new creation founded in 1909 by Zionist settlers on the sand dunes north of Jaffa. In the British mandate period (1923–48) there were two municipalities, but in 1950 Jaffa

was incorporated with Tel Aviv. The combined population (1961) was 386,612. The municipalities of Ramat Gan, Holon and Bnei Brak are satellite towns; and the total population is nearly 600,000.

The principal streets of Tel Aviv are Allenby road (1918), running from the city centre toward the shore, and a series running south to north parallel with the shore, the chief being Yarkon, Eliezer Ben-Yehuda, Dizengoff, and the Rothschild boulevard. The town has constantly extended northward and nearly half the municipal area is now north of the Yarkon river, which in the mandate period was the boundary. In Arab Jaffa the main streets were Butros (Peter), close to the sea shore, and King George V avenue (now Jerusalem boulevard), laid out in 1927 to link the suburbs of the two towns. Important buildings in Jaffa are the municipality and the district government offices, damaged in the War of Independence, the Russian Orthodox church, St. Peter's monastery (Franciscan), and the English church. In Tel Aviv the Great synagogue, the chief banks and commercial buildings are in Allenby street. There are more than 400 synagogues (many are just rooms) in the municipal area.

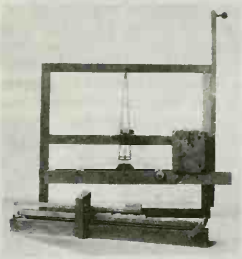
Tel Aviv is a principal cultural centre of Israel. The original nucleus of the town was the Herzlia *Gymnasium;* the original Museum of Art was the house of Meir Dizengoff, the first mayor. Beside it a new art gallery, bearing the donor's name Helena Rubinstein, was opened in 1958 in the civic centre, which also includes a modern concert hall, the Temple of Culture built in 1957, and the Habima (*q.v.*) theatre, home of the national dramatic company, opened in 1945. The Mograbi cinema in Allenby St. was formerly the principal auditorium for opera and concerts. In the northern suburb rises a vast pile, the offices of the Labour federation (Histadrut), and popularly known as the Kremlin. Near it are the centre of the Zionist Organization of America and a journalists' and writers' club. Most of the country's daily newspapers and periodicals are published in the town. The parks are small since, before the establishment of the state, Tel Aviv was congested because of the great cost of land. They include a zoo and a botanical garden. A school of law and economics, established with municipal help in the mandate period, became in 1959 a branch of the Hebrew University of Jerusalem. University colleges of arts and science were subsequently founded and form the University of Tel Aviv. The Labour federation is active in promoting adult education, and maintains two residential colleges, and an Afro-Asian institute providing courses in economic and social sciences for students from Africa and Asia. There are, too, an academy of music, several teachers' training colleges and a broadcasting station.

The main urban and interurban transport service is provided by buses run by co-operative societies. Service taxis (Sherut), in which the passenger pays for a seat, ply along the main thoroughfares and also between the towns. There are two railway stations: one in the northern outskirts for the line to Haifa; the other near the commercial centre, connecting with Lydda and Jerusalem. Tel Aviv is the centre of the road system of the whole country. Tel Aviv and Jaffa have a combined port, stretching from the mouth of the Yarkon river to the promontory of old Jaffa, and used almost entirely for freight, not for passengers. The main industry since the establishment of the state has been building. There are no heavy industries, but light industries of many kinds flourish—textile, furniture, foodstuffs, pharmaceutical. Tel Aviv is the principal banking, insurance and commercial centre of Israel. Citrus fruit from the southern region is the main export, but most of the fruit passes through Haifa. (No. B.)

TELEGONUS (literally "far-born"), in Greek mythology the son of Odysseus by Circe. In the lost *Telegonia,* the last poem of the epic cycle, Telegonus went to Ithaca in search of his father, whom he there killed unwittingly. His spear had been tipped with the point of a sting-ray, thus fulfilling the prophecy in the *Odyssey* that death would come to Odysseus "from the sea." Telegonus then married Penelope, Odysseus' widow. The legend was rehandled in a lost play of Sophocles. (H. W. Pa.)

TELEGRAPH (Gr. *tele,* "far"; *graphein,* "to write"). A system of communication at a distance through the medium of electromagnetic phenomena, in which the information transmitted is in (or may signify) written, printed or pictorial form.

The evolution of telegraphy may be traced from the signaling methods and codes of the ancient and primitive peoples to the visual system of Claude Chappe and thence with the growth of scientific knowledge to the first commercial application of electricity and magnetism to the "electric telegraph." Development in the mid-19th century provided the public with the only rapid means of communication until the advent of the telephone, which led to a continuous decline of public telegraph traffic within national boundaries during the 20th century. However, the continued development of telegraph technique evinced particularly in the teleprinter and improved transmission methods has enabled operation to be delegated to its users in certain fields. Public exchange service that gives direct printing communication between subscribers nationally and internationally is increasing rapidly. Many organizations—commercial, transportation, press and government—operate private networks for their internal business traffic. The high information density that can be carried telegraphically and the overwhelming value of the printed record provide a service that cannot be given by the telephone. *See* also FACSIMILE TRANSMISSION.



BY COURTESY OF SMITHSONIAN INSTITUTION

MORSE TELEGRAPH INSTRUMENT, 1837

This article is divided into the following sections:

I. Historic Telegraphs
II. Direct-Current Transmission Circuits
III. Telegraph Codes
IV. Machine Telegraphs; Exploitation of Circuit Capability
V. Printing Telegraphs
VI. Line Plant
VII. Transmission Theory
VIII. Telegraph Switching
IX. Submarine Telegraphy
X. Radio

I. HISTORIC TELEGRAPHS

Pith Ball and Spark.—Sir William Watson in England demonstrated in 1747 that an electric current could be transmitted through a considerable length of wire using the earth for the completion of the circuit. In 1753 the first suggestion for an electrical telegraph was made in Scotland by an anonymous writer to *Scots Magazine,* signing himself C. M., who advised using an insulated wire for each letter of the alphabet. At the receiving end of each wire a pith ball was to be suspended above a piece of paper marked with an alphabetical letter. As a charge was sent along a given wire, the ball would attract the paper beneath it, and thus words could be spelled out. He further suggested that bells might be substituted for the papers, which could be struck in turn by a ball as a charge was sent along a wire. The idea was carried out by Georges Louis Lesage in Geneva in 1774, and similar telegraph systems were suggested by Bétancourt and Lomond in 1787. Then followed other suggestions using a separate wire for each letter, by Reusser, or Reisser, in 1794, and by Francisco Salvá in 1798; each of them proposed a system of visual telegraphy by causing a spark to appear at a gap at the receiving end of each wire as desired.

Sir Francis Ronalds worked a telegraph in England in 1816 with transmitting and receiving apparatus consisting of circular plates, on which were inscribed letters and figures. In front of each plate was a disk with an aperture which allowed one letter and its corresponding figure to be visible. The plates were rotated by clockwork regulated so that the same letter was displayed simultaneously at both ends. A crude electric machine charged the line when the sender desired, and the divergence of a pair of pith balls at the receiving end indicated to the operator that the letter or figure appearing at that moment was to be recorded.

Toward the end of the 18th century, Luigi Galvani and Alessandro Volta conducted experiments which revolutionized preconceived ideas of electricity and its effects. In 1786 Galvani accidentally discovered that it was possible to cause a direct or continuous flow of current along an electrical conductor by bring-

ing two dissimilar metals into contact with a moist substance. In 1800 Volta introduced an electric battery which became known as the voltaic pile, the principles of which became embodied in the modern battery, and in that year Salvá demonstrated that voltaic currents could be used for transmitting signals.

Electrolytic Telegraphs.—An immediate result of the introduction of the voltaic pile was the discovery by William Nicholson and Sir Anthony Carlisle in 1800 that the passage of an electric current caused decomposition of liquids into their constituent elements.

Based on this principle Salvá in 1805 and Samuel Thomas von Soemmerring in 1809 introduced apparatus in which a voltaic pile was used for sending signals whose presence was indicated at the receiving end by the liberation of hydrogen. Later researches by Sir Humphry Davy on the decomposition of chemical compounds by electric current were applied in the United States to another type of telegraph by John Redman Coxe in 1816 and Harrison Gray Dyar in 1828. Both Coxe and Dyar operated a telegraph using the electric current to produce electrical decomposition at the end of the conductors. Dyar operated his telegraph line over 8 mi. on Long Island, using a single wire with ground return circuit to mark coded signals on a moving strip of litmus paper by discoloration from the received electric current.

Electromagnetic Needle.—At the beginning of the 19th century, a new set of ideas followed the introduction of the electromagnet. In 1819 Hans Christian Oersted discovered that a magnetic needle could be deflected from its normal position by passing a current through an adjacent wire and that the deflection was to the right or left, according to the direction of current. A year later Johann Salomo Christoph Schweigger found that the deflection of the needle could be increased by surrounding it with a number of turns of wire, and in 1825 William Sturgeon in England produced and named the electromagnet.

The action of the electric current on a magnet was first applied to telegraphy by André Marie Ampère in 1820, at the suggestion of Pierre Simon Laplace that the deflections of small magnets placed at the receiving ends of 26 wires could be used to indicate the letters of the alphabet. Ampère's apparatus was the pioneer of many needle instruments which came into existence between 1829 and 1841. The names of Carl Friedrich Gauss and Wilhelm Weber, Sir William Fothergill Cooke and Sir Charles Wheatstone, and the brothers Highton (Henry and Edward) are associated with some of the more important needle instruments.

In 1831 Joseph Henry (*q.v.*) constructed and successfully operated an electromagnetic signaling apparatus which consisted of a magnetized steel bar supported on a pivot in a horizontal position, which could be attracted by a magnet. When the magnet was energized by current the extremity of the bar was caused to strike a bell, and intelligence was conveyed by coded sound combinations.

Gauss and Weber conducted important experiments in 1833 on line conductors in which they demonstrated that the line conductor need not be insulated over its entire length. They used a single-needle telegraph to communicate with each other and proved that, by proper combination, five signs are sufficient for communication. In the early experiments of Cooke and Wheatstone, five wires and five needles were employed but later were reduced to one of each. In the single-needle instruments a magnetic needle was pivoted in the centre of a wire coil and intelligence was transmitted by deflecting it in accordance with a prearranged code.

The name of Samuel F. B. Morse, a U.S. artist and later professor of natural science at Yale university, is universally associated with the telegraph as result of his invention of the code bearing his name. From 1850 to the 1920s Morse was the code most generally used and, in its cable code form, is still in use. He is often credited with the invention of the telegraph on his return to the United States from a trip to Europe in 1832, during which he became acquainted with the works of Michael Faraday, but he definitely produced a telegraph instrument that was exhibited in 1837. The transmitter comprised a comb (or "portrule") with "teeth" spaced at intervals in conformity with a code. The comb was passed beneath a contact lever so that it made and broke

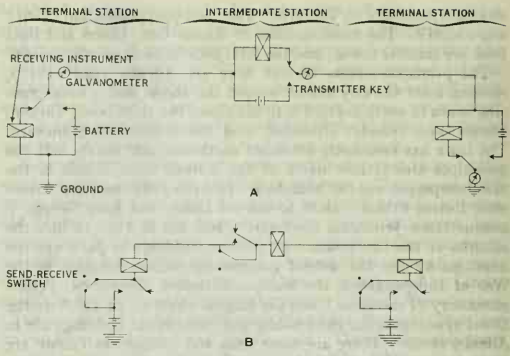


FIG. 1.—TWO SIMPLEX TRANSMISSION SCHEMES IN IDLE CONDITION: (A) SINGLE-CURRENT CIRCUIT, IN WHICH NO CURRENT IS FLOWING; (B) CLOSED-LOOP SYSTEM, IN WHICH CURRENT IS FLOWING

an electric circuit. The pulses of current passed through an electromagnet, causing it to deflect a lever carrying a pencil, thus producing an undulating line on a paper tape moved by clockwork to record the signals received.

In a practical application of these ideas Alfred Vail, an associate of Morse, discarded the "portrule" in favour of a simple hand-operated contact lever or key, and it was later found that the signals could be reliably interpreted by ear from the sound of the movements of the electromagnet armature. Recording was also discarded, not to be revived until much later. After initial government support a full-scale demonstration between Washington, D.C., and Baltimore, Md., was given in 1844, but when this support was withdrawn support was enlisted to form a company in 1845.

In England public interest in Cooke's and Wheatstone's developments was growing in the same period, and the Electric Telegraph company was formed in 1845.

II. DIRECT-CURRENT TRANSMISSION CIRCUITS

Simplex.—The elementary telegraph circuit consists of a contact device (a "key," if hand operated) that makes and breaks the connection between a source of current (*e.g.*, a battery) and the line, according to the signaling code, and the pulses of current passing to the distant end of the line actuate the receiving device; *e.g.*, a "sounder" if aural reception is required. Alternate transmission of messages in both directions along the line may be made by suitable connections of instruments, and intermediate stations may be connected.

Two general schemes have been employed: in one, during the idle condition, no current flows in the line (fig. 1[A]); in the other, the closed-loop system, current flows in the line (fig. 1[B]). In the latter, as shown, a send-receive switch must be operated before sending can take place, but an alternative connection can be used in which the free contact of the transmitter can serve this purpose automatically.

Instead of single-current (on and off), double-current (current reversal) signaling may be used with definite advantage (fig. 2). Although it requires a polarized receiver (one sensi-

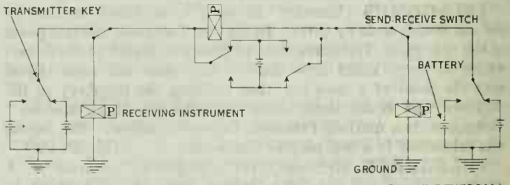
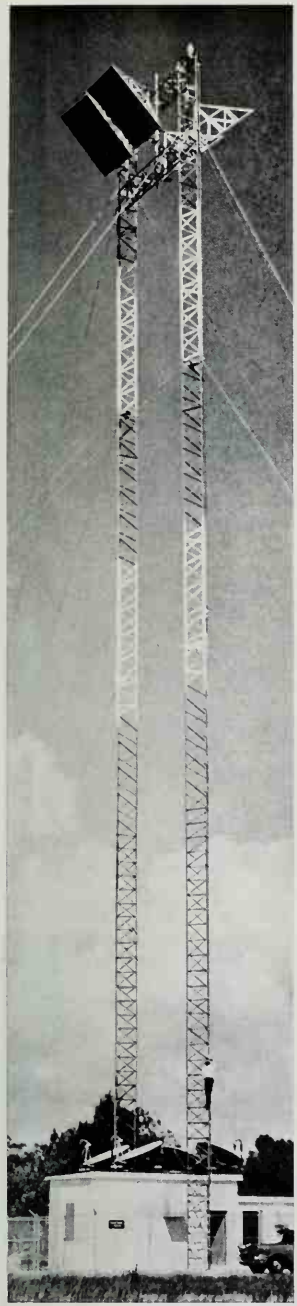
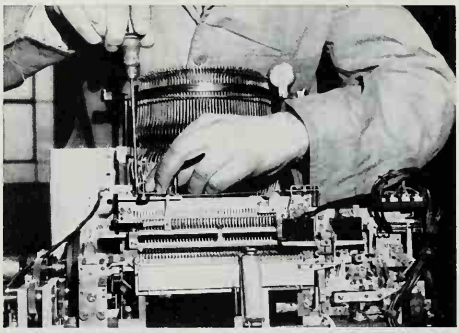


FIG. 2.—SIMPLEX CIRCUIT FOR DOUBLE-CURRENT (CURRENT-REVERSAL) SIGNALING



Radio telegraph tower; messages travel by microwave through a series of such towers placed an average of about 30 mi. apart

Racks of carrier equipment which increases capacity of telegraph circuits to 288 messages simultaneously over a single pair of wires or 2,000 telegrams over a single radio beam



Teleprinter, operated by a keyboard similar to that of a typewriter, sends signals over the telegraph circuit to a similar printer at a distant location. The message is printed instantaneously

Above, a repeater to boost the signal on a submarine cable is being lowered into the sea. *At right*, an amplifier unit to be placed in a submerged repeater is being inspected

TELEGRAPH TRANSMISSION EQUIPMENT

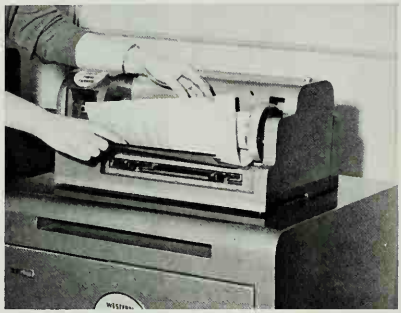
BY COURTESY OF (LEFT, TOP RIGHT) THE WESTERN UNION TELEGRAPH CO., (BOTTOM CENTRE, BOTTOM RIGHT) STANDARD TELEPHONES AND CABLES LTD., (CENTRE RIGHT) CABLE AND WIRELESS LTD.

TELEGRAPH

Compact facsimile machine used to send telegrams from a business office to the central telegraph station

Facsimile equipment used to transmit photographs

Telegrams dropped into a slot are automatically transmitted to a central telegraph office by this facsimile equipment



Coast-to-coast transmission of letters, charts, graphs and similar copy is accomplished by this facsimile equipment. The material is received as an exact picture of the original

Facsimile equipment is used by some business firms, particularly banks, to check signatures. A branch bank can have a signature verified from card files maintained miles away. The signature is transmitted as a picture

FACSIMILE EQUIPMENT

BY COURTESY OF (TOP RIGHT) CABLE AND WIRELESS LTD., (OTHERS) THE WESTERN UNION TELEGRAPH CO.

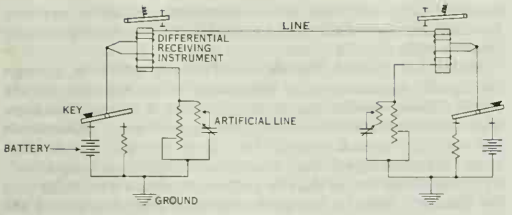


FIG. 3.—DIFFERENTIAL DUPLEX CIRCUIT

tive to current direction rather than current strength), variations in line characteristics have less effect and entail less frequent adjustment of the receiver. At terminal stations a double battery is usual, but at intermediate stations a single battery with a commutating key would be used.

Duplex.—The duplex scheme, permitting simultaneous transmission in both directions between two terminal stations, was first proposed in 1853 by Wilhelm Gintl. The principle is the interconnection of the transmitting and receiving devices with the line and a balancing network, so that the receiver is unaffected by the sending current from its own station but responds to the operation of the distant transmitter. It is usual for double-current signaling to be employed. The differential form (fig. 3) uses a receiver or relay with two identical windings connected so that the current from the transmitter divides to produce opposing magnetic effects, part flowing into the line and part into the balance or artificial line, which is adjusted so that the magnetic effects of the currents in the two paths balance to produce no resultant effect on the receiver. The changes in line current due to the action of the distant transmitter produce an unbalanced effect in the windings of the receiver, which responds accordingly. Bridge duplex (fig. 4) utilizes a receiver or relay with a single winding connected across one diagonal of a Wheatstone bridge circuit formed by two equal impedances, the line and the artificial line. When the artificial line is accurately adjusted to simulate the electrical characteristics of the line, the operation of the transmitter, which is in the other diagonal of the bridge, produces no change in electrical potential across the receiver and hence no response.

Quadruplex.—To increase further the message capacity of a single telegraph circuit, Thomas A. Edison produced a combination of double-current and modified single-current duplex, providing for the transmission of two messages in each direction simultaneously. Two receiving devices are used, one polarized and responding to the direction of current, the other nonpolarized and responding to change in current magnitude irrespective of direction. With the development of other transmission methods, this system was discarded.

In modern practice, except for submarine cable usage, direct-current transmission is employed usually only over distances up to 100 mi. Both single- and double-current simplex methods are used in different areas of the world. With the single-current method, the closed-loop system, without send-receive switches and using a return conductor in place of earth, with one current source in the loop is general. In the double-current method, the simple arrangement of two lines, one for each direction of transmission to avoid switching of transmitters and receivers and with no intermediate stations, is usual.

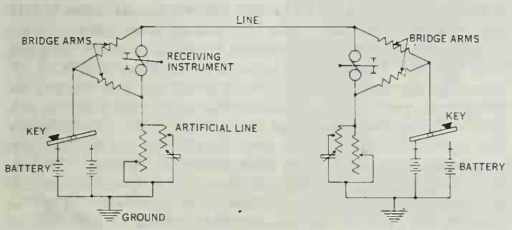


FIG. 4.—BRIDGE DUPLEX CIRCUIT

Derived Circuits.—Direct-current telegraph circuits may be derived from, and worked simultaneously with, telephone circuits by superposition. Superposing permits one ground-return telegraph circuit to be operated over the two wires forming a metallic-return telephone circuit. In a common form (fig. 5) a repeating coil or transformer is interposed at each terminal between the telephone apparatus and the line. The transformer winding which is connected to the line wires has a centre tap connected to the telegraph apparatus. By this arrangement the telegraph currents are divided, half passing over each wire; no effect is produced upon the telephone apparatus because, in dividing equally, the two halves neutralize each other's magnetic effect. The alternating currents of the speech transmission so traverse the line windings of the transformer as to cause the maximum magnetic effect and are thus repeated to the telephone apparatus.

Telegraph Repeaters.—With increasing length of a telegraph line, the received current becomes weakened so that eventually it will not operate the receiving device reliably. It was first appreciated by Edward Davy that a relay could be used to repeat signals from one line to another, the armature of an electromagnet operating contacts to send the signals anew with the original current strength. Hence by dividing a line into sections and inserting repeaters between each section, the signals can be renewed in magnitude before they deteriorate to an unusable level. In the direct-current transmission circuits described above, repeaters can be formed by making the receiving device a relay and connecting the armature or tongue to the adjacent line in place of the transmitter, the transmitting battery being connected appropriately. In the case of the closed-loop circuit, additional relays must be used in connections to prevent signals from being fed back into the line from which they were received.

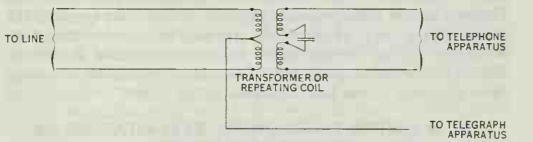


FIG. 5.—SIMPLEX CIRCUIT FOR TELEGRAPH TRANSMISSION ON TELEPHONE LINES

Distortion of signals takes place in transmission such that, although they can be restored in magnitude by a relay, the relative timing of the signal elements is disturbed and retransmission over a further section would render them liable to faulty interpretation. In machine telegraph transmission at a regular speed, regenerative repeaters can be employed to receive and retransmit the signals with all the elements restored to a perfect time relationship. This is done by automatically examining the instantaneous condition of the incoming signal at unit element intervals in correct phase and maintaining this condition on the outgoing line until the next instant of examination.

III. TELEGRAPH CODES

The best-known code is the Morse Code (fig. 6). It was the most generally used until about the mid-1920s and is still used in a modified form on some submarine cable systems and radiotelegraph circuits and in some situations where mobile stations with simple apparatus are required. The code comprises combinations of dots and dashes of one line condition (*e.g.*, current on) separated by periods of the other line condition (*e.g.*, current off); the duration of a dot is taken as one unit, that of a dash three units; the space between character components is one unit, between characters three units and between words six units (five for automatic transmission).

Early machine telegraphs used an obvious code on the basis of one pulse for A, two for B and so on, but with the development of more sophisticated systems, the advantage of codes in which all character signals are of equal length became apparent, and

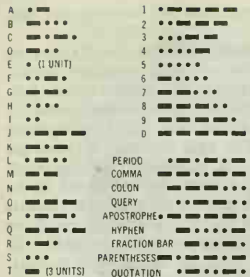


FIG. 6.—INTERNATIONAL MORSE CODE

J. M. E. Baudot applied the idea of Gauss and Weber of a five-unit code. This, with variations in the character allocations, became the most widely adopted by mid-20th century. Five units with two-condition signaling provide 32 permutations, sufficient for the Roman alphabet and (by using two permutations as case-shift signals) figures, punctuation signs and the control of certain essential mechanical functions of a printing receiver. For certain special applications, a six-unit code that avoids the use of case-shift signals is used.

Since all permutations of the five-unit code are significant, an error in the reception of any signal element will result in a wrong character being recorded. So that faulty reception can be detected, particularly after transmission over circuits liable to severe disturbance (*e.g.*, long-distance high-frequency radio), a seven-unit code was devised by J. B. Moore which permitted the detection of mutilation. In a modified form proposed by H. C. A. Van Duuren, its use rapidly spread (fig. 7). Seven units provide 128 permutations, only 35 of which are used to signify characters. Reception of any of the others indicates mutilation in transit. The 35 permutations possess the common characteristic of having three elements of one kind and four of the other.

IV. MACHINE TELEGRAPHS; EXPLOITATION OF CIRCUIT CAPABILITY

With the commercial development of the telegraph in the late 1840s it became obvious that the manual sending methods of forming the code signals directly were unable to exploit the full capabilities of the transmission lines, and attention of many inventors was turned to increasing the speed of signaling and providing more convenient means of generating and recording signals.

Wheatstone System.—The Wheatstone system (1858) employed perforated tape to operate the transmitting mechanism, which was driven by clockwork and later by electric motor. The Morse signals were received by an "inker" in which an inked wheel marked the dots and dashes on a moving tape. When first brought into use in the 1860s the system was capable of about 70 words per minute, but by the early 1900s it had been developed to 300–400 words per minute.

Initially a knowledge of Morse code was necessary for the preparation of the transmitter tape, but later keyboard perforators were devised that perforated all the holes corresponding to a character on depression of the appropriate key. Further improvements to the system included receiving reperforators that reproduced the original tape, facilitating retransmission, and the Creed printer which translated from the perforated tape to printed characters. The system required groups of operators at each end to keep it fully loaded and was supplanted by more convenient multiplex systems. Nevertheless the system had a considerable life, even persisting on some radio circuits.

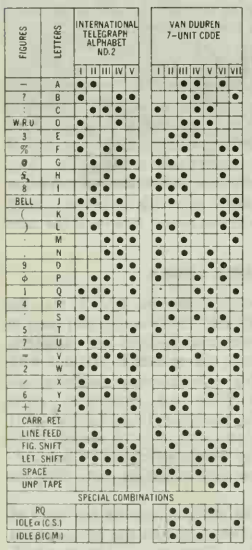


FIG. 7.—FIVE-UNIT CODE AND SEVEN-UNIT ERROR-DETECTING CODE

Multiplex (Time Division).—The alternative to a single high-speed communication is the division of the capability of the circuit among several operators, allocating a channel to each. This can be effected by allotting connection to the transmission line cyclically to each operator for a given period; *e.g.*, one character (character interlacing) or one element (element interlacing). At the receiving end of the circuit the individual receiving devices are similarly connected in cyclic order so that they record the appropriate signals.

Systems on this principle had been proposed by Wheatstone and Moses G. Farmer. A certain amount of development was done by A. J. Meyer and P. B. Delaney, but J. M. E. Baudot produced the first successful system using a five-unit code and printing receivers. The principle is illustrated in fig. 8. The "clock" determining the period and sequence per channel and code element comprises a mechanical distributor in which an arm rotating with uniform speed connects the line sequentially to segments of a conducting ring that are connected to a battery through keys set to the code permutation.

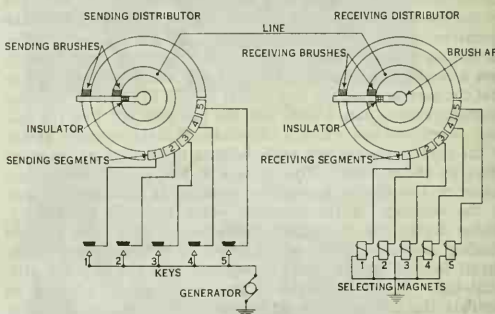


FIG. 8.—DISTRIBUTOR SYSTEM (BAUDOT) FOR MULTIPLEX TRANSMISSION

After the transmission of one character, the permutation is changed in time for the next revolution, while the other channels are being dealt with. A similar distributor at the receiving station connects the receiving devices synchronously in similar sequence but with an appropriate time lag according to the signal propagation time. Special devices are used for maintaining accurate synchronism and correct phase between sending and receiving distributors. In the original system the code keys were operated directly, but in later usage perforated tape and automatic tape readers were introduced. Speeds of up to 6×40 words per minute were achieved.

Systems based on the Baudot were used extensively on land lines until the late 1930s and are still used on submarine cables. They are also used on radiotelegraph links because synchronous operation affords a protective feature to transmission accuracy. Later developments in this field are leading to the obsolescence of the electromechanical distributor; electronic oscillators and pulse distributors, with no moving parts, provide the clock system for the sequential interconnection of the circuit elements.

Multichannel Carrier Systems (Spectrum-Division Multiplex).—Toward the end of the 19th century proposals were made to use tones of different frequency or pitch to carry telegraph code signals. Experimental systems by E. Gray and E. J. P. Mercadier proved unsatisfactory and it was not until further developments, particularly the thermionic valve, that practical systems were evolved toward the end of the 1920s. The principle of a carrier system is that, in place of a direct current, an alternating current is used to carry the code signals. This may be done by varying the amplitude of the carrier (*e.g.*, switching on and off), changing the frequency (frequency modulation, FM) or changing the phase of the carrier (PM). Each of a number of different frequencies can carry a separate telegraph signal, the

spacing of the frequencies depending upon the required speed of transmission. Each channel is controlled by an independent transmitter, and the outputs are connected to the common transmission line through electrical filters to prevent mutual interference. At the receiving station each channel is filtered out separately, and an amplifier-detector terminating in a relay reproduces a D.C. code signal for the operation of the receiving instrument.

The spacing of channels is standardized at 120 cycles per second (c.p.s.) for 66 words per minute teleprinter operation, and systems of 12, 18 and 24 channels are usual. In the United States 150- and 170-c.p.s. spacings are used and the speed is generally 60 words per minute, but later systems are capable of 100 words per minute; the number of channels varies up to 16. The range of frequencies is usually restricted to those of speech, and telephone circuits or carrier channels (on line or microwave radio) are used as the transmission path.

In Europe the majority of systems use amplitude modulation, while in the United States some of the original AM systems have been replaced by FM systems. The use of phase modulation involves some technical complexities and is not generally used.

Voice-frequency multichannel transmission by mid-20th century had become the most widely used method of providing main telegraph links, and with the growth of long-distance submarine telephone cables was extending into the field of intercontinental line communications. Multichannel systems are also applied to radiotelegraph routes. These are generally of the frequency-shift variety of modulation, which is less susceptible to the effects of noise and fading encountered on long-distance radio circuits.

V. PRINTING TELEGRAPHS

The earliest printing telegraphs were worked on the step-by-step principle in which a type wheel with characters evenly spaced around its circumference is advanced by an electromagnet and ratchet one character for each current impulse received from the line. Trains of impulses caused the type head to rotate from one character to the next one; on cessation of the pulses a paper tape was brought into contact with the inked type at the printing position. This was a development from the Wheatstone ABC in which the characters were only indicated on a dial. Such systems were used for many years for stock exchange quotation telegraphs. The major drawback was the liability of the type head to get "out of step" due to interfering currents in the line, and means had to be provided to rotate the type wheel to a known position periodically.

An improvement on this system was produced by D. E. Hughes. The type head rotated continuously in synchronism with a distributer at the sending end, and operation of a piano-type keyboard caused a current impulse to be emitted at a point in the revolution corresponding to the chosen character. This, at the receiver, caused a paper strip to be brought momentarily in contact with the type wheel as the corresponding type passed the printing point. Speeds of up to 60 words per minute could be achieved by skilled operators, and this system persisted on some European circuits until the 1930s.

The Baudot multiplex system used printing receivers working with a five-unit code. Five levers, set in alternative positions by electromagnets, respond to the element signals and, when these coincide with a series of notches in a disk carried by the continuously rotating type-head shaft, a printing hammer automatically brings paper tape in contact with the inked type.

Teleprinters.—Until the 1920s the successful printing telegraph systems involved at some point in the system continuous synchronism of sending and receiving apparatus, which meant that a skilled setting-up procedure was involved before full-speed communication could take place and that a high degree of skill, including in some cases knowledge of the code, was demanded of operators. This restricted use of the most efficient methods to main lines. Methods were pursued which could be more universally applied and would demand less skilled control, thus enabling printing telegraphs to be used on local circuits where traffic is intermittent.

To meet this requirement, telegraph machines (*i.e.*, teleprinters or teletypewriters) were developed by E. E. Kleinschmidt and the Morkrum company in the United States, Creed and company in England and Siemens-Halske in Germany. From the mid-1940s other European countries produced similar machines. The teleprinter became the commonest telegraph instrument, and hundreds of thousands were in use in the second half of the 20th century.

The teleprinter comprises a keyboard similar to that of a typewriter and a printing mechanism that produces the received message either as a continuous line of type on a tape or in page form on a continuous roll of paper. The motive power is provided by an electric motor running at a controlled speed, but the mechanisms are coupled to it by clutches that are brought into operation automatically only when required.

The depression of a key of the keyboard causes each of a set of five bars to take up one of two positions according to the five-unit code permutation representing the character and also engages the clutch of the transmitting mechanism for the period of transmission of the individual character. A cam assembly permits a contact or contacts to be operated at discrete intervals in accordance with the setting of the permutation bars to form the code signal transmitted to line. The five-unit code signal is always preceded by a start element and succeeded by a minimum period

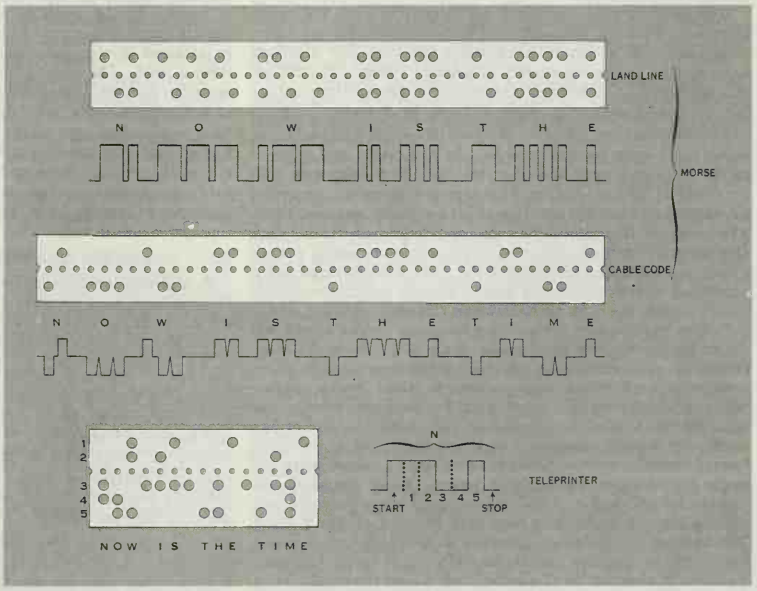


FIG. 9.—PERFORATED TAPES (ACTUAL SIZE) FOR AUTOMATIC TRANSMISSION, AND SIGNAL CURRENT WAVE FORMS

of the rest condition (1–1½ units in duration), hence this code is often termed a seven-unit start-stop code.

The received signals operate the teleprinter via an electromagnet. The first motion of the electromagnet armature engages the clutch of the mechanism, and at the mid-instants of the code elements the position of the armature is recorded by mechanical store elements, which are linked to the printing mechanism in one of several possible ways.

In one way, latches are selected to halt a rotating type head at the appropriate position; in others the positioning of slotted combination bars enable individual type bars, as in a typewriter, to be brought into action, or a rectangular type box is displaced according to the mechanical addition of the displacements of the storage elements. Since the receiver is started for the reception of each character, slight speed differences from the transmitter do not produce an accumulative phase error leading to malreception.

Teleprinters are capable of speeds of 60–100 words per minute and can be operated by competent typists. Where the utmost use of the line capability is desired, an automatic tape transmitter can be used in place of the keyboard. Receiving reperforators may be used, with or without printing, where retransmission over another circuit is necessary, to avoid rekeying the message.

VI. LINE PLANT

Bare wires were used as line conductors as early as 1816. In that year, Francis Ronalds demonstrated a telegraph system that employed a bare wire covered with glass tubing impregnated with pitch and buried in the ground. However, for reasons of economy, the early commercial telegraphs were operated with bare wires supported overhead on poles. This open-wire pole-line construction was generally used thereafter, and even after the middle of the 20th century still constituted a large portion of outside telegraph plant. Cables, which are groups of insulated wire held together by a common cover or sheath, were introduced early in the century to a limited extent, and their utilization, both aerially suspended and underground, greatly increased thereafter. Open-wire pole-line construction and suspended aerial cable are commonly used in open country, but as telegraph lines approach congested areas, underground cables are generally used. Underground cable is also used almost exclusively for local distribution within a congested area.

Overhead Lines.—Poles used for supporting wires and cables are usually of wood, although steel and concrete poles are sometimes used. In open-wire construction the bare line wires are supported on insulators, usually of glass or rubber, mounted on pins set vertically in wooden crossarms. In general, bare copper wires are used, ordinarily about 0.114 in. in diameter, weighing 210 lb. per mile, or 0.104 in. in diameter, weighing 172 lb. per mile. Other sizes are employed to a limited extent, and wires of galvanized iron, bronze and other materials have been used with considerable success. Cables, when placed on poles, are supported by suspension strands, consisting of high-strength stranded wires. The weight of the cable to be supported determines the strength of suspension strands used, the heavier cables requiring strands having a strength exceeding 18,000 lb. Cables are fastened to the suspension strands by various means, and the suspension strands are fastened to the poles by small metal clamps.

Conductors in cables designed for telegraph service are of soft drawn copper varying in diameter from 0.036 in. to 0.102 in. Ordinarily, cable conductors are separated and insulated by paper wrapped spirally about them, although cables having rubber-insulated conductors are often used for short distances. The insulated conductors are arranged in pairs which consist of two conductors twisted together or in quads consisting of two pairs twisted together, or four wires in spiral form. These pairs and quads are arranged compactly in layers. Paper insulated conductors are enclosed in a lead sheath, while rubber-covered conductors may be enclosed in lead or a covering of heavy braid. The number of conductors provided in telegraph cables varies, the largest in common use in the early 1960s being 404 pairs. The outside diameter of this cable is $2\frac{5}{8}$ in. and the weight $7\frac{3}{4}$ lb. per foot.

Underground Construction.—Where cables are placed underground within a city, some form of conduit with independent duct space for each cable is usually employed to permit the cables to be readily installed and removed. The ducts are usually of hollow tile, creosoted wood or fibre. Iron pipe may be used in special situations. All ducts are three inches or more in diameter, and the conduit is buried at least two feet deep. As the length of cable which can be pulled through a duct is limited, splicing chambers or manholes are provided at intervals of about 500 ft. These manholes are at least 3 ft. wide, 5 ft. long and 4½ ft. deep, and in many cases considerably larger sizes are required. They are usually constructed of brick or concrete with a removable cover.

Cable is pulled into conduits one length at a time, and the ends are spliced together. In splicing paper-insulated cables, a foot or more of sheath is removed from the end and a few inches of the insulation is removed from the conductors. Conductors are then joined by twisting, soldering or other means, and the joints are insulated with small paper or impregnated cotton tubes. After all conductors are joined, the splice is wrapped with strips of muslin and a lead sleeve slipped over it. This sleeve is held at each end to the cable sheath by means of wiped soldered joints.

Where cables are laid underground in open country, it is feasible to bury such cables directly in the ground without conduit protection. Large machines prepare a trench, lay the cable and cover the trench in a single operation.

Submarine Cable.—In submarine cable the conductor is made of copper, stranded or having a central wire surrounded by strips wound spirally. Over this is molded the insulation. Formerly gutta percha was almost universally employed for the insulator, but deproteinized rubber, both natural and synthetic, later competed successfully with gutta percha, while polyethylene bid fair to supplant both of them. Continuous inductance loading in the form of a thin ribbon or thread of highly permeable nickel-iron alloy wound around the conductor greatly increased the speed at which the later cables could be worked. To protect the insulation at depths where marine borers are found a brass tape is wound around it. Next comes a cushion of jute yarns upon which the spirally wound armour, or sheathing wires, rests. In the deeper portions of the ocean galvanized high-tensile steel wires are used, sometimes individually covered by tarred cotton tape or braid. As the depth decreases these are replaced by galvanized wrought-iron wires of larger size, and the armour gradually is increased in size and weight until, near the shore, a double layer is used. The added protection is necessary to prevent damage from anchors, trawler fishing gear, chafing against rocks and, in northern latitudes, ice. An outer double wrapping of tarred jute yarn or jute tape steeped in a compound of tar and pitch further preserves the armour from corrosion. The outside diameter of the ordinary deep-sea type of cable is about one inch, and the weight in air about two tons per nautical mile. The shore-end types run as large as 3½ in. in diameter and weigh as much as 30 tons per nautical mile.

VII. TRANSMISSION THEORY

In a direct-current telegraph circuit the transmitted signal can be considered as a sequence of similar rectangular pulses of potential applied at time intervals that are multiples of the duration of the shortest code element, as in the example in fig. 10(A). Analysis of the elementary pulse wave form into sine wave components by Fourier's method shows that it contains a band of frequencies ranging from zero to infinity; the relative magnitudes are indicated in fig. 10 (B). If the duration of the element pulse is T seconds (T might be 0.02 sec. in teleprinter operation), there will be no components at precisely $\frac{1}{T}$, $\frac{2}{T}$, $\frac{3}{T}$, c.p.s. When a particular pattern of pulses of duration nT seconds is continuously repeated, the components become concentrated at frequencies that are multiples of $\frac{1}{nT}$, but the relative magnitudes are still bounded by the fundamental curve. As an example, if the signal is a repeated train of Morse dots, $n = 2$ and the component frequencies are $\frac{1}{2T}$, $\frac{3}{2T}$, $\frac{5}{2T}$, . . . c.p.s.; *i.e.*, the fundamental plus the odd harmonics of the dot frequency (fig. 10[C]).

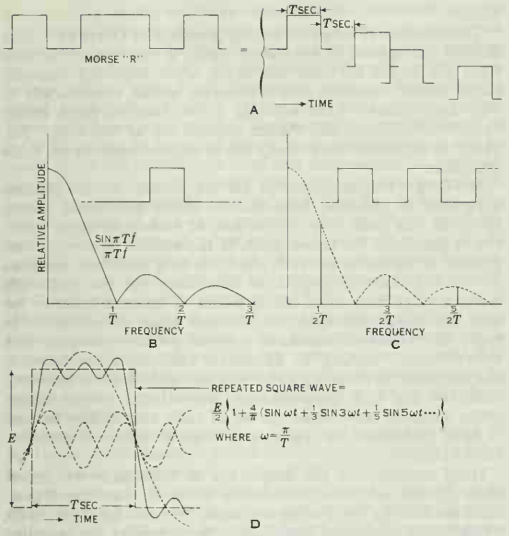


FIG. 10.—EXAMPLES OF TELEGRAPH TRANSMISSION WAVE FORMS (*see* TEXT). (E = VOLTAGE)

This may be more readily appreciated by reference to fig. 10(D), which illustrates how the combination of the first three components begins to approximate to the rectangular wave form.

To reproduce the signal wave form completely at the receiver, it would be necessary for the telegraph line to transmit all the frequency components without change of relative amplitude and phase, but in practice this is unnecessary. The energy conveying the essential message is concentrated in the lower part of the frequency spectrum, and the higher-frequency components may be discarded without destroying the message. Excellent operation can be obtained when the circuit reduces frequencies corresponding to $\frac{0.5}{T}$ c.p.s. to one-half of their original amplitude relative to the fundamental component and the response at frequencies of $\frac{3}{2T}$ c.p.s. is negligible. However, on local lines where the transmission capability is more than adequate, it is usual to transmit a frequency range including $\frac{3}{2T}$ c.p.s., to permit less critical adjustment of apparatus.

In carrier transmission, there is a similar distribution of signal energy with respect to frequency in side bands symmetrically located about the carrier frequency, the lower side band being a mirror image of the upper one. Hence the frequency band width required for this form of transmission is double that for direct-current transmission.

The energy of the signal is dissipated in transmission to an extent dependent on the resistance, capacitance, inductance and leakance of the line. The effects of capacitance and inductance increase with frequency but have some opposing characteristics. The inherent characteristics of telegraph cables are such that the resistance and capacitance preponderate, so by artificially increasing the inductance (*i.e.*, "loading" the cable), by the insertion of inductor coils at uniform distances or by wrapping the conductor with magnetic material, the dissipation of the higher frequencies can be reduced and a higher speed of operation obtained. Even with this, the relative magnitudes of the various frequency components have to be modified to the correct proportions by inserting at the sending or receiving end (or both) of the line correcting devices or equalizers with frequency responses that compensate for those of the line.

The transmission path may be subject to interfering currents from various sources; *e.g.*, induction from other telegraph systems, power supply systems, stray currents from traction systems or varying earth potentials due to changes in the earth's field. The magnitude of interfering currents limits the speed of operation on a long cable because the higher frequencies required are severely reduced in magnitude and the signals become too deformed to be interpreted correctly.

Telegraph Distortion.—The perfect telegraph signal changes its condition according to the code at predetermined intervals of time. In the course of transmission it is restricted in frequency range, either purposely to prevent the higher frequencies from interfering with other channels in a multichannel system or incidentally by the line characteristics, so that the transitions from one condition to the other are not abrupt but occur in a finite time. The receiving instrument or relay is operated from one condition to the other when the received current attains certain values. If there are no interfering currents, the transmitting path is conveying the currents of various frequencies in the correct relationships and the relay is adjusted to the correct operating values, the signal will be perfectly reproduced (assuming a perfect transmitter) with the transitions taking place at the prescribed intervals. But if interference is present these currents will increase, or decrease, the signal received and the relay operating currents will be attained at other instants. Similar displacements in time of the relay operations will ensue if the relay is maladjusted or the line is incapable of transmitting the required range of frequencies in their proper relationship. If the telegraph distortion (time aberration of the signal) is too great, the margin (maximum acceptable distortion) of the receiver will be exceeded and confusion with other code signals will ensue.

VIII. TELEGRAPH SWITCHING

Communication between a large number of stations requires either direct lines from each to every other one or means of transferring messages from one line to another at central points. Stations can be grouped around different centres and distant traffic concentrated to use intercentre lines at higher efficiency. In the early days of telegraphy, the only means of handling messages between widely separated points was to record them at intermediate points and retransmit them on the next stage of the route. About the beginning of the 20th century exchanges were brought into use in large cities for the manual interconnection by means of switchboards of local lines using hand-speed Morse, but the diversity of methods and speeds used on long-distance lines dictated the continuance of the employment of operators at long-distance centres, which persisted at places dealing with international traffic to the late 1950s.

The successful introduction of telegraph exchanges awaited the development of a standardized telegraph machine and cheaper telegraph circuits, which eventuated as the teleprinter and multichannel carrier systems.

Teleprinter Manual Switching.—In the early 1930s a few European countries introduced subscriber exchange service, Telex, for teleprinters, using single-channel voice-frequency signaling, over the public telephone networks. In the second part of the decade this service was extended internationally. During World War II the development of teleprinter communication proceeded rapidly on both sides; large manually switched networks were put into operation using separate telegraph circuits comprising multichannel systems on main routes and direct-current signaling on local circuits. In the United States the extensive exchange system (TWX) was based on a line system using derived direct-current circuits and multichannel carrier.

The switchboard closely resembles a telephone manual switchboard in which lines terminate in jacks (sockets) that are interconnected as desired by flexible conductor cords terminating in plugs. The operator, of course, communicates with the subscribers by teleprinter.

Teleprinter Automatic Switching.—The desire for further reduction of human intervention in telegraphs led to automatic switching on principles similar to those for telephony. Systems were designed during the mid-1930s and in fact introduced to a limited extent in Germany before World War II. Introduction elsewhere in Europe was delayed until the late 1940s and '50s. In some countries the operator is provided with a dial, like that of

an automatic telephone, for setting up a connection; in others the called number is sent from the keyboard, and apparatus at the exchange converts the signals to pulses for operating the switches. Special signals are sent back for indications of "engaged," "line out of order," "number unobtainable," etc., to be printed. On receipt of a call the teleprinter is automatically started and sends its own number automatically to confirm the connection to the caller.

Reperforator Switching.—The systems of switching so far mentioned provide both-way communication (but not necessarily simultaneously) by direct line connection between two stations. However, flexibility in message handling can be achieved by semi-automatic or automatic means in which the message is received as perforated tape in the exchange and then retransmitted over an appropriate outgoing line to the recipient or the next switching centre en route. The idea of tape relay systems led in the late 1930s to the development of the printing reperforator by which the message is simultaneously typed on the tape as it is perforated. In the first systems, operators read the address from the tape which was torn off on completion of the message and transferred it to the appropriate position in the office, where it was sent onward with an automatic transmitter. This system was further developed so that the receiving operator could place the message tape in one of an adjacent group of autotransmitters and connect that automatically to the required outgoing line by pressing the appropriate one of a group of buttons controlling relay switching apparatus. By the mid-1950s fully automatic tape relay was achieved. In this, the originating operator sends route code letters at the head of the message. The code letters are sensed automatically following reperforation and interpreted by a director that automatically sets up connections through the exchange and causes the message to be retransmitted at high speed to a reperforator and associated retransmitter, from where it is sent out on the next section of its route. If lines are busy at any stage in the process, the message is kept stored in the perforated tape.

Applications.—In the early 1960s the use of switching systems was extensive and extending rapidly to reduce the labour costs of human handling. Europe generally appeared to favour the line-switching technique for the public services, both Telex and ordinary telegrams, and automation proceeded. Private networks operated by public utilities and business concerns for communication between head and branch offices and factories, etc., employed various of the systems described.

United States practice followed two distinct patterns. The subscriber service (TWX) was, necessarily, a line-switching system, but the public telegraph service (Western Union Telegraph Company) developed on a message-switching basis, due largely to the economic relationship between costs of station apparatus and transmission channels where great interurban distances are involved. There were also many large private networks.

Military organizations of nations with overseas bases operated world-wide systems in which message switching was employed. Trunk circuits were carried by cable and radio.

IX. SUBMARINE TELEGRAPHY

Except for a few river and harbour cables, the first submarine cable was that laid in 1850 between England and France. It was broken shortly after communication had been established but was successfully followed the next year by another, between Dover and Calais. The first transatlantic attempt in 1857 was a failure when the cable broke at 2,000 fathoms' depth during laying and could not be recovered. However, the Atlantic was spanned in 1858 between Ireland and Newfoundland. After about three months' use the cable broke in deep water and had to be abandoned. It was 1866 before the first permanently successful transatlantic cable was laid. In the same year another cable, partially laid in 1865, was also completed. These two, with which the U.S. financier Cyrus Field and the British scientist Lord Kelvin were closely associated, were the only transatlantic cables until 1869, when the French laid a cable. In the following 60 years, the earth-encircling telegraph cable network became fully established, and no new long-distance routes were established afterward.

The principle of inductive loading proposed by Oliver Heaviside in 1887 was applied to some short cables at the turn of the century, but it was not until the 1920s, when improved magnetic alloys became available, that it could be applied economically to very long cables. Even so, only a few long telegraph cables (across the Atlantic and Pacific oceans) are of this type. Advances in utilization were chiefly due to improvements in the cable apparatus.

Sending.—Originally, except for the attempt to use printing telegraphs on the first cross-channel cable, hand-speed Morse operation was used. Due to the lack of knowledge about methods of equalizing the transmission of the various frequency components of signals in the early days, the long and short elements of the code suffered appreciable distortion, and the cable-code method which uses equal length pulses of opposite polarity for dots and dashes was used. It was also found advantageous to apply short reversed periods of opposite polarity following each code element ("curbing"). Except for relatively short cables, it was not found satisfactory to use five-unit code until it was introduced by the U.S. companies on transatlantic loaded cables. Other operators maintained the use of cable code chiefly because of large investment in extensive systems of chains of interconnected cables.

Hand sending with the double key in cable code was superseded by the automatic transmitter modified from the Wheatstone machine by two French engineers, Belz and Brahic, in 1879, which used a perforated tape. Later, as techniques for reception improved and higher speeds became possible, multiplex operation was introduced.

Receiving.—Due to the characteristics of long submarine cables and the requirement to extract the greatest speed from an expensive channel of communication, the received currents are very feeble, often effectively only of the order of ten microamperes. This necessitates much more sensitive apparatus than that employed on land lines. The earliest specialized instrument was the mirror galvanometer of W. Thomson (later Lord Kelvin) by which the signals were read as deflections of a reflected spot of light on a scale; this was used on the first Atlantic cable. In 1867 Thomson produced the siphon recorder, in which an ink siphon is harnessed to the moving coil of the galvanometer so that the signals are recorded as an undulating line on paper tape. With the development of sensitive relays more robust recording instruments could be employed and receiving reperforators appeared about 1900. This greatly facilitated retransmission at intermediate stations. In the early 1920s the invention of the Creed Morse printer provided the next step of translation from the perforated tape. About ten years later a direct printer which could cope with the various-length cable-code signals completed the development of cable-code operation.

The higher speeds made possible by the loaded cable led to the use of five-unit code on certain routes to facilitate repetition of signals from land-line multiplex systems. The higher capability of the cable was difficult to realize reliably with the existing electromagnetic relays which decrease in sensitivity as they are adjusted for higher speed, and so thermionic valve terminal amplifiers were introduced to amplify the signals for use with more robust high-speed relays to feed the multiplex apparatus.

Transmission and Repeaters.—Simplex operation was used in the early days of submarine cables, but the desirability of duplexing was immediately felt. Early attempts were not very successful, because the weak received signals made it imperative to achieve accurate and stable artificial lines for balancing the cable, since the residual unbalance introduced interference from the sent signals at the same end. Improvements in the manufacture and adjustment of artificial cables made duplex operation usual on unloaded cables, but no generally satisfactory means of producing a balance for loaded cables was found. However, one special cable, laid in 1928 between Nova Scotia and the Azores, had graded loading that tapered to zero at the ends and was successfully duplexed.

The possible speed of operation of a given type of cable terminating at a particular station will be inversely proportional to the

square of its length, so where convenient it is desirable to split long routes into sections, locating repeater stations on islands on transoceanic routes where the signals can be completely regenerated. The ultimate limit in speed is determined by the interference induced by natural phenomena, such as changes in the magnetic and electric fields of the earth and electric storms. These effects are manifested in shallow water near the shore, so if the signal can be amplified before it reaches this area, it can be raised above the interference level and faster operation obtained. The technique of the submerged telegraph repeater was first employed by the Western Union Telegraph company in 1950. The repeater consisted of a valve amplifier with equalizing networks, contained in a watertight steel case, the necessary power for its operation being supplied from the shore via the cable. Such repeaters may be located 60–300 mi. from shore at depths of 250–1,250 fathoms. Such a repeatered cable can be operated only in the simplex fashion, but the speed may be increased three- or fourfold over unrepeatered duplex. This method is best employed on routes where there are at least two cables to provide both-way communication, but repeaters can be inserted in both ends and be switched in or out of circuit from the shore to reverse the direction of working.

Submerged telegraph repeaters have been applied to many of the transatlantic telegraph cables by U.S. companies and to a few in the British commonwealth cable system.

Typical speeds of operation of long-distance unloaded cables range from 250 to 600 characters per minute duplex. Loaded cables about 2,000 mi. long are capable of 2,500 characters per minute simplex, and repeatered unloaded cables attain similar speeds.

X. RADIO

Radiotelegraphy is essentially carrier telegraphy at relatively high frequency. The radio carrier may be directly modulated by the telegraph signal in any of the usual modes (*i.e.,* amplitude or frequency) or by a subcarrier system of one or more channels (*i.e.,* the telegraph signals may be borne over a line to the radio transmitting station by a voice-frequency carrier that is then sent by a radiotelephony transmitter). Radio propagation over long distances is subject to severe disturbances that cause the signal to fade and introduce interference by atmospheric noise, which results in distortion and sometimes mutilation of the signal. To minimize these effects, diversity reception may be employed. Diversity can be of several forms: time diversity in which the message is repeated twice and the most likely version of the message accepted in the case of disagreement; space diversity in which two receiving aerials separated by a considerable distance feed a special receiver that selects the stronger signal; or frequency diversity in which the telegraph signal is sent simultaneously on two different frequencies.

Teleprinter transmission can be more seriously impaired than synchronous transmission, because if the start or stop elements of character signals are incorrectly received, the mechanism is likely to become out of step and successive characters to be incorrectly printed. Therefore, on regular fixed traffic routes, synchronous time-division multiplex is often employed and as a further protection a seven-unit error-detecting code may be adopted. When a return channel is available a system of automatic error-correction devised by Van Duuren is used. In this, the detection of an error causes a special signal to be sent back to the transmitting office to request the automatic repetition of the previous three characters of the message, which are automatically stored in the apparatus; meanwhile, the output of the receiving apparatus is blocked until the originally faulty signal has been received again. If this is then apparently correct, through transmission is resumed; if not, the automatic repetition cycle is repeated. In practice, the error rate in the printed message may be reduced by a factor in the order of 100. Such systems can be connected in teleprinter networks; the teleprinter signal is received at the sending terminal station on a reperforator, and the tape from this is accepted by the multiplex apparatus, which automatically translates it into the special seven-unit code. At the receiving end of the multiplex the signals, after being checked for conformity, are automatically translated back into teleprinter code signals.

In some countries, particularly where interurban distances are great and the terrain rugged, microwave radio-relay systems are being introduced as more economical than long-distance carrier telephone cable, and channels in such systems can be used to bear multichannel telegraph systems. *See* also Index references under "Telegraph" in the Index.

BIBLIOGRAPHY.—C. Bright, *The Story of the Atlantic Cable* (1903); F. J. Brown, *Cable and Wireless Communications of the World,* 2nd ed. rev. (1930); D. H. Cameron, *Submarine Telegraphy* (1927); J. A. Fleming, *Propagation of Electric Currents in Telephone and Telegraph Conductors,* 4th ed. rev. (1927); H. H. Harrison, *Printing Telegraph Systems and Mechanisms* (1923); T. E. Herbert, *Telegraphy,* 5th ed. rev. (1930); H. W. Malcolm, *The Theory of the Submarine Telegraph and Telephone Cable* (1917); D. McNicol, *American Telegraph Practice* (1913); J. Z. Millar, "Preview of the Western Union System Radio Beam Telegraphy," *J. Franklin Inst.,* pt. 1, 241:316–326; pt. 2, 242: 23–40 (1946); H. W. Pendry, *Baudot Printing Telegraph System,* 2nd ed. rev. (1920); Edward Lind Morse (ed.), *Samuel F. B. Morse, His Letters and Journals,* 2 vol. (1914); G. A. Schreiner, *Cables and Wireless and Their Role in the Foreign Relations of the United States* (1924); T. P. Shaffner, *The Telegraph Manual* (1859); H. G. Sellars, "A Brief Chronology for Students of Telegraphs, Telephones and Posts," *Teleg. Teleph. J.* (Nov. 1927 *et seq.*); W. D. Weaver, *Catalogue of the Wheeler Gift of Books, Pamphlets and Periodicals in the Library of the American Institute of Electrical Engineers* (1909); Western Union Committee on Technical Publications, *American Telegraphy After 100 Years* (1944); R. Appleyard, *Pioneers of Electrical Communication* (1930); A. F. Harlow, *Old Wires and New Waves* (1936); W. T. Perkins, *Modern Telegraph Systems and Equipment* (1946); R. L. Thompson, *Wiring a Continent* (1947). *See* also Western Union Engineering Department textbooks, prepared for the American School of Correspondence, Chicago; Federal Communications Commission *Annual Reports; Bell System Technical Journal; Post Office Electrical Engineers' Journal; Proceedings* of the Institute of Radio Engineers; *Transactions* of the American Institute of Electrical Engineers; *Western Union Technical Review;* J. W. Freebody, *Telegraphy* (1959); K. Henney (ed.), *Radio Engineering Handbook,* 5th ed. (1959); H. Fülling, *Fernschreib-Übertragungstechnik* (1957); E. Rossberg and H. Korta, *Fernschreib-Vermittlungstechnik* (1959); R. Roquet, *Théorie et technique de la transmission télégraphique* (1954); D. Faugeras, *Appareils et installations télégraphiques* (1955); F. Schiweck, *Fernschreib-Technik,* 2nd ed. (1944); *Journal* and *Proceedings* of the Institution of Electrical Engineers. (W. P. MA.; L. K. W.)

TELEMACHUS, in Greek legend, son of Odysseus and Penelope. When he reached manhood he visited Pylos and Sparta to make inquiries about his father. On his return, he found that Odysseus had reached home before him. Then father and son, aided by Eumaeus and Philoetius, slew the suitors of Penelope (*see* ODYSSEUS). According to later tradition he married Circe (or Calypso) after Odysseus' death.

TELEMANN, GEORG PHILIPP (1681–1767), German composer, was born on March 14, 1681, at Magdeburg, Ger. He received no formal musical training, but he studied published works of leading composers and became an accomplished arranger and choirmaster. In 1700 he entered the University of Leipzig, where he wrote operas for the Leipzig theatre and was commissioned to write original compositions for the Thomas church. In 1704 he was appointed organist of the New church and also became musical director to a prince at Sorau, Ger. From 1708 through 1711 he held the positions of concertmaster at Eisenach, where he later was musical director, and of choirmaster at the Church of St. Catherine in Frankfurt am Main.

Telemann was musical director of the Barfüsser church from 1712 until 1721, and also at the court of the prince of Bayreuth. During his later years he was cantor of the Johanneum and musical director of five Hamburg churches. He died in Hamburg on June 25, 1767.

TELEMARK, *fylke* (county) of southern Norway. Area 5,905 sq.mi. Pop. (1960) 149,828. It covers the uplands and fjelds of the southward projection of the country, and contains several lakes, as Nordsjö, Bandaksvand, Tinsjö, Mjösvand and Totakvand.

The first two are connected by the Bandaks canal, giving access from the port of Skien to Dalen. Rjukanfos waterfall (334 ft.) provides power for nitric acid works, and Ulefos, near the Bandak canal, for pulp mills.

TELEOLOGY means explanation by reference to some purpose or end (Greek *telos,* "end"; *teleios,* "complete"); also described as final causation, in contrast with explanation by reference to efficient causes only. Human conduct, insofar as it is considered rational, is generally explained with reference to ends pursued or alleged to be pursued; and human thought tends to explain the behaviour of other things in nature on this analogy, either as of themselves pursuing ends, or as designed to fulfil a purpose devised by a Mind external to nature. The most celebrated account of final causation was given by Aristotle when he laid down that for a full explanation of anything it was necessary to take account not only of the "material," the "formal," and the "efficient" causes, but also of the purpose for which the thing existed or was produced, namely its "final" cause. With the rise of modern science in the 16th and 17th centuries, interest was directed to mechanistic explanations of natural phenomena with reference to efficient causes; if teleological explanations were used they took the form, not (as in Aristotelian teleology) of saying that things develop toward the realization of ends internal to their own natures, but of looking even on biological organisms as machines ingeniously devised by an Intelligent Being. William Paley gave classical expression to this kind of teleology (*Natural Theology, or Evidences of the Existence and Attributes of the Deity Collected From the Appearance of Nature,* London, 1802).

In the 19th century controversy centred on whether the phenomena of growth, regeneration and reproduction characteristic of living organisms could be explained in purely mechanistic terms. The "vitalism" of Hans Driesch (1867–1941), according to which some entelechy (*q.v.*) must be postulated in an organism, found little support after his death. There remains, however, the question whether biological processes can be explained purely in physicochemical terms ("mechanistic" explanation), or whether the problems of structure, function and organization necessitate some kind of teleology.

See C. D. Broad, *The Mind and Its Place in Nature* (1925); J. Woodger, *Biological Principles* (1929). (D. M. E.)

TELEOST, any typically bony fish belonging to the group Teleostei. In the classification issued in 1846 by Johannes Müller, the geologically modern bony fishes were arranged in three subclasses, Dipneusti, Ganoidei and Teleostei, the last including the great majority of the fishes, all those with a homocercal tail fin. Subsequent work, especially on fossil fishes, has shown that the ganoids are not a natural group, some being related to very primitive fishes, others, namely the Holostei, which include the recent gars, *Amia* and *Lepidosteus,* approaching the teleosts.

The extinct Pholidophoridae and Oligopleuridae are exactly intermediate in structure between the Holostei and Teleostei, and have been included by some authorities in the one group, by others in the other. This led C. T. Regan to combine the Holostei and Teleostei to form one subclass, termed Neopterygii. *See* FISHES.

TELEOSTOME, common name given by some authorities to the members of the Teleostomi, ganoid and bony fishes, either with or (usually) without the lungfishes. *See* FISHES.

TELEPATHY, communication of information from one organism to another without mediation of the recognized channels of sense. *See* PARAPSYCHOLOGY.

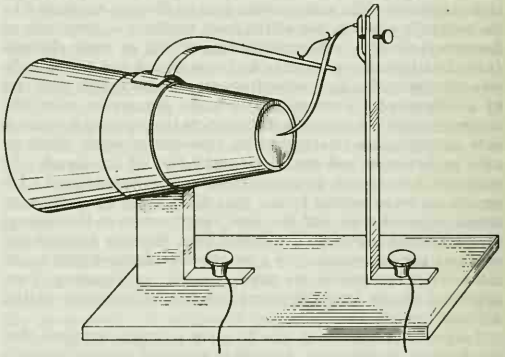
TELEPHONE. The term telephone (from the Greek roots, *tele,* "far," and *phoné,* "sound") was first used to describe any apparatus for conveying sounds to a distant point. Specifically, the word was applied as early as 1796 to a megaphone, and not long afterward to a speaking tube. Subsequently the name string telephone was given to the device invented long before by Robert Hooke (1667), in which vibrations in a diaphragm caused by voice or sound waves are transmitted mechanically along a string or wire to a similar diaphragm which reproduces the sound. Still later, devices employing electric currents to reproduce at a distance the mere pitch of musical sounds were called telephones. Nowadays, however, this name is assigned almost exclusively to apparatus for reproducing articulate speech and other sounds at a distance through the medium of electric waves. The term telephony covers the entire art and practice of electrical speech transmission, including the many systems, accessories and operating methods used for this purpose.

This article is outlined as follows:

I. Invention
II. Introduction and Growth
1. United States
2. Great Britain
3. Germany
4. France
5. Switzerland
6. Scandinavia
7. Belgium and the Netherlands
8. Austria, Italy and Spain
9. U.S.S.R. and Satellites
10. Canada
11. Latin America
12. Other Countries
13. Telephone Statistics
14. ITU and CCIF
III. Technical Development
1. Station Instrumentalities
2. Telephone Lines
3. Phantoms
4. Loading
5. Electron Tubes and Repeaters
6. Carrier Systems
7. Radio
8. Submarine Telephony
9. Information Theory
10. Switching
11. Associated Services
IV. Trends

I. INVENTION

Like most inventions, the telephone drew heavily upon previous work and had scarcely appeared before notable improvements were made. Among the pioneer contributors in this field, the most outstanding was Alexander Graham Bell, who invented, and patented



BY COURTESY OF BELL TELEPHONE LABORATORIES, INC.

FIG. 1.—TRANSMITTER OF PHILIPP REIS, 1861 OR 1862

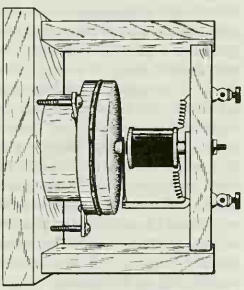
in 1876, the first telephone capable of practical use. As early as 1874 he had conceived the correct principle of telephone transmission, which he later stated as follows: "If I could make a current of electricity vary in intensity precisely as the air varies in density during the production of sound, I should be able to transmit speech telegraphically."

This conception of an undulatory current corresponding to a speech wave formed the foundation for the entire telephonic art. Earlier workers, notably C. G. Page in the United States (1837) and Charles Bourseul in France (1854), had devised methods employing the make-and-break principle of the telegraph for transmitting the pitch of sounds, but not articulate speech, to a distant point.

Philipp Reis came closer. Working at Frankfurt, Ger., in 1860 and subsequent years, he devised an apparatus using at the transmitting end a diaphragm structure something like that of the human ear to control an electric current (fig. 1). At the other end of the circuit this produced audible tones by controlling the mag-

netization of a needle whose changes of length, in accordance with the magnetostrictive effect discovered by Page, vibrated a sounding board. Later, an electromagnet was substituted for the needle. Reis's mechanism, like others before and after, included the important feature of deriving the electrical power for transmitting sound by making the sound power control the current from a battery.

Reis had some comprehension of the requirement for electrical transmission of speech, for he noted in his memoir on telephony the need "to set up vibrations whose curves are like those of any given tone or combination of tones" (S. P. Thompson translation). Though his apparatus served primarily to reproduce tones by the make-and-break scheme, he did, by extremely delicate adjustment, succeed in reproducing articulate sounds quite imperfectly. Reis seems not to have realized, however, that this success resulted because his apparatus could, over a very narrow range of speech volume, operate on the principle of changing an electric current in accordance with the voice wave by varying a loose contact, in this case a spring contact between platinum electrodes. His understanding of his own apparatus is indicated by his statement that "each sound vibration effects an opening and a closing of the current." About 20 years later, the German patent office after careful investigation decided that Reis's instrument was not a "speaking telephone."



BY COURTESY OF BELL TELEPHONE LABORATORIES, INC.

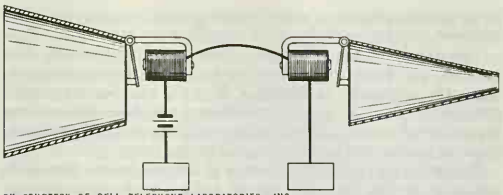
FIG. 2.—"GALLOWS FRAME" TRANSMITTER OF ALEXANDER GRAHAM BELL, JUNE 3, 1875

In the following years attempts were made by other workers (*e.g.*, the Italians A. Meucci and I. Manzetti) but without full realization of the requirements for articulate speech.

Bell's approach was different. In the summer of 1874 the idea of the "electric speaking telephone" became complete in his mind. He described to his father a form of apparatus consisting of a strip of iron attached to a membrane which, when actuated by the voice, would vibrate in front of an electromagnet, thus inducing an undulatory electric current theoretically capable of transmitting speech. At the receiving end a similar device could be used in reverse to reproduce the voice. But Bell doubted that the current generated by the voice would be strong enough to be useful, and for almost a year he made no attempt to construct the apparatus.

On June 2, 1875, while working in Boston on multiplex telegraph apparatus, Bell heard over an electric wire a sound corresponding to the twang of a steel spring at the other end. Recognizing this as a manifestation of the undulatory current principle, he gave his assistant, Thomas A. Watson, instructions for embodying it in a model (fig. 2) of a telephone. The transmitter and receiver were of the electromagnetic type described a year before, and in between was a circuit which included an electric battery. This apparatus transmitted speech sounds the next day, June 3. Bell filed his application for a U.S. patent on Feb. 14, 1876 (fig. 3). Further experiments produced an instrument (fig. 4) which on March 10, 1876, transmitted the first complete sentence: "Mr. Watson, come here; I want you."

A few hours after Bell filed his application for patent, Elisha Gray filed a caveat (*i.e.*, a notice of intent to perfect his ideas and file a patent application within three months) for an electric telephone. Gray described a "liquid transmitter," somewhat similar to one patented by Thomas A. Edison for telegraphy in 1873. In Gray's transmitter a voice-actuated diaphragm varied the electrical resistance, and hence the current, by changing the depth of immersion of a rod in water. Bell in his application mentioned the possibility of a similar liquid transmitter, and later used it for his historic summons to Watson. The variable-resistance principle, which subsequently, in the form of a variable carbon contact,



BY COURTESY OF BELL TELEPHONE LABORATORIES, INC.

FIG. 3.—ELECTROMAGNETIC TRANSMITTER (LEFT) AND RECEIVER (RIGHT) ILLUSTRATED IN BELL'S FIRST TELEPHONE PATENT, FILED FEB. 14, 1876

proved vital to telephone transmission, makes it possible to obtain an electric wave which is an amplified copy of the sound wave. Like Bell's receiver, Gray's was of the electromagnetic type, similar to one he had patented in Great Britain in July 1874 and in the U.S. in July 1875.

In view of Bell's prior filing of a patent application, the patent for the telephone was issued to him on March 7, 1876. Gray's status as to the invention of the telephone is best set forth in his own words, written to Bell on March 5, 1877: "I do not claim even the credit of inventing it." The claims of Gray, Daniel Drawbaugh and others were subsequently threshed out in prolonged litigation, involving about 600 separate suits, which finally resulted in Bell's patent's being upheld in a divided vote by the supreme court of the United States.

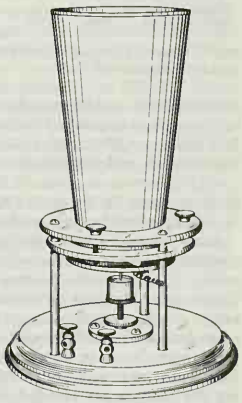
Bell's first transmitter employing electromagnetic induction, while valid enough in theory, delivered such feeble electrical currents as to be inadequate for general application. The liquid transmitter afforded considerable improvement, but this too had drawbacks. The final essential element for a satisfactory working telephone was the variable-contact carbon transmitter, due in large measure to Thomas A. Edison, as discussed below.

II. INTRODUCTION AND GROWTH

1. United States.—The development of the telephone business in the United States was undertaken by a group of Bell's backers, under the leadership of Thomas Sanders and Gardiner G. Hubbard. They began by renting or lending telephones in pairs to individuals for local communication. The instruments were extremely crude. Connection between them was made by a circuit consisting of a single iron wire with ground return; transmission, which was uncertain and poor at best, was possible for only a few miles. Initially there were no switchboards to interconnect a number of users. These came into being in 1877 and 1878.

It was on March 25, 1878, that Bell made a bold prediction that became a charter for the founders of the telephone business:

> It is conceivable that cables of telephone wires could be laid underground, or suspended overhead, communicating by branch wires with private dwellings, country houses, shops, manufactories, etc., etc., uniting them through the main cable with a central office where the wires could be connected as desired, establishing direct communication between any two places in the city. Such a plan as this, though impracticable at the present moment, will, I firmly believe, be the outcome of the introduction of the telephone to the public. Not only so, but I believe, in the future, wires will unite the head offices of the Telephone Company in different cities, and a man in one part of the country may communicate by word of mouth with another in a distant place.
>
> I am aware that such ideas may appear to you Utopian. . . . Believing, however, as I do that such a scheme will be the ultimate result of the telephone to the public, I will impress upon you all the advisabil-



BY COURTESY OF BELL TELEPHONE LABORATORIES, INC.

FIG. 4.—LIQUID TRANSMITTER USED BY BELL ON MARCH 10, 1876

ity of keeping this end in view, that all present arrangements of the telephone may be eventually realized in this grand system. . . .

The owners of the telephone patent early incorporated their business, and funds were raised for its progressive development, under the leadership of Theodore N. Vail, who became general manager in 1878. It was recognized that telephone performance is a matter of mutual concern to users, and the practice was established of leasing telephones instead of selling them.

Within ten years after the issuance of the Bell patent, the organization of the Bell Telephone system had assumed something close to its present form. The local systems were gradually brought together into regional companies operating throughout a state or several states. The systems of these regional companies were linked together by long-distance circuits operated by the American Telephone and Telegraph company. This company, through ownership of stock in the regional companies, became the parent company of the Bell system.

It soon became evident that standardization of equipment and centralized research on improvements are essential to telephone progress. Accordingly, the Western Electric company was acquired in 1882 as chief manufacturer and supplier for the Bell system, as well as to conduct research and development. In 1925 the Bell Telephone laboratories was organized to take over the expanding research and development activities for the system. The 1,800,000 owners of shares in the American Telephone and Telegraph company in the early 1960s constituted the largest number of public owners for any one corporation in the world.

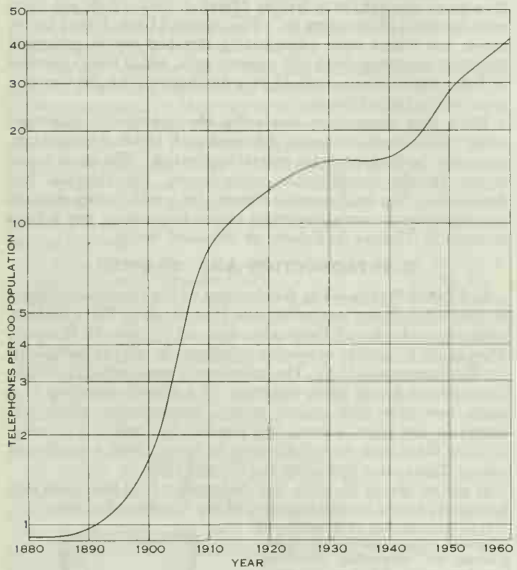


FIG. 5.—GRAPH SHOWING TELEPHONE DEVELOPMENT IN THE U.S. (VERTICAL SCALE LOGARITHMIC)

After the expiration of the basic telephone patents, many independent telephone companies, not affiliated with the Bell system, sprang up all over the country. Competition became so intense that in many localities there were two companies sharing the business. As this became an increasing source of inconvenience and expense to the public, the service was unified through acquisitions and mergers so as to leave a single company, either Bell or non-Bell, operating in each area. Provision was made for interconnecting the facilities of non-Bell companies with those of the Bell system, thus making possible the interconnection of nearly all telephones in the United States, as well as connections to the rest of the world. Service and rates are regulated by state utility commissions and the Federal Communications commission.

There were about 98,000,000 telephones in the U.S. in the early 1960s, or more than one telephone for every three persons in the country. About 82% of these telephones were served by the Bell system, comprising the American Telephone and Telegraph company and its 20 operating subsidiaries, and 2 associated but non-controlled companies (the Southern New England Telephone company and the Cincinnati & Suburban Bell Telephone company). The remaining telephones were served by about 4,100 independent companies and additional thousands of rural or farmer lines, virtually all of which connect with the Bell system. The total investment in telephone plant and equipment was more than $22,000,000,000, of which 86% belonged to the Bell system. Telephone traffic averaged about 240,000,000 conversations daily. About 800,000 persons were employed by the telephone industry, including more than 100,000 employed by the manufacturers of telephone equipment and about 10,000 by the Bell Telephone laboratories.

2. Great Britain.—Bell visited England and Scotland on his wedding trip in 1878 in the hope of developing a demand for the telephone. Despite the able support of Lord Kelvin, Sir William Preece and others, he aroused little public interest. He did, however, demonstrate his invention to Queen Victoria, who asked to purchase a pair of telephones and instead was presented with two instruments done in ivory. Stimulated by this royal recognition, the first telephone exchange was opened in London in 1879 with seven or eight subscribers. Several telephone companies were organized in various parts of Great Britain, but in 1880 the British courts held that the telephone system was legally a telegraph system under an antecedent law which made the telegraph a government monopoly under the postmaster general. The government officials, reluctant to assume the risks involved in developing this new form of communication, issued licences on a royalty basis to several private companies, which were later consolidated into a single company. A few municipal telephone systems were also established under licence.

As the potentialities of telephone communication began to be appreciated, the government gradually took over the service. In 1896 the post office purchased the long-distance lines, and in 1902 it began establishing in London its own local telephone exchanges which were interconnected with those of the privately owned company. Finally, on Jan. 1, 1912, the post office acquired all the private telephone properties. Since then the post office has operated practically all telephones in Great Britain. When the Irish Free State was established, the British post office transferred to the Free State government its telephone system in southern Ireland.

In the early 1960s there were about 7,500,000 telephones in Great Britain, which was more than in any other European country, and represented 1 telephone per 8 persons. The British post office maintains a large research and development organization. Telephone apparatus is supplied by private companies that manufacture to post-office specifications.

3. Germany.—In Germany the telephone was a government monopoly from the beginning. Heinrich von Stephan, postmaster general and manager of the imperial telegraphs, when he learned about Bell's invention from an article in the *Scientific American*, ordered models for trial. After successful experiments at distances up to 90 mi. (Berlin-Magdeburg), he suggested on Nov. 9, 1877, to Bismarck, the imperial chancellor, that the telephone be used as an adjunct to telegraph service in rural post offices where there was not sufficient traffic to justify a trained telegraph operator. Within two years 800 villages were thus connected.

So it came about that in Germany the first public use of the telephone was for long-distance communication, telephones being found only in government post offices. The situation was therefore the reverse of that in the United States and Britain, where local exchanges came first. Later, as public demand forced the issue, exchange service was gradually introduced, starting in the larger German cities. In the German states of Bavaria and Württemberg, the telephone systems were operated by the state administrations until 1920, when they were transferred to the German post office.

In the early 1960s the Federal Republic of Germany had the second largest telephone system in Europe, with about 5,400,000 telephones or about 1 telephone per 12 persons. The German Democratic Republic had 1,200,000 telephones, or about one telephone per 17 persons.

4. France.—In France the telephone was first exhibited at the Paris exhibition in 1878, where it attracted little interest. The next year, however, the French telegraph officials, unwilling themselves to pioneer in this new field, granted concessions to several private companies which later consolidated into the Soeiété Générale des Téléphones. This company initiated public telephone service in Paris in 1881 and subsequently in other cities. Starting in 1883, the French government established exchanges in various centres. In 1889 the government took over the entire private system, and after that time operated the telephone service. With about 3,900,000 telephones (one-fourth of these in Paris) at the beginning of the 1960s, the French telephone system was third in size among those of Europe.

5. Switzerland.—In Switzerland a privately owned telephone system was established under government concession in Zürich in 1880, while the government itself opened exchanges in Berne and Basel in 1881. Thereafter the government proceeded rapidly to establish new exchanges, and in 1886 purchased the Zürich system. After that time the Swiss government operated all telephones. At the beginning of the 1960s there were about 1,500,000 telephones, amounting to about one telephone for every four persons.

6. Scandinavia.—The Scandinavian countries have achieved a high degree of telephone development. In Sweden, the International Bell Telephone company of New York opened exchanges in Stockholm and Göteborg in 1880, and not long after in Malmö and elsewhere. In 1883 a competing telephone company was set up in Stockholm under the enterprising leadership of H. T. Cedergren. Co-operative telephone associations were established in Göteborg and in many rural communities. After a strongly competitive phase, in which the government participated, virtually all telephones were taken over by the state during the period 1890–1923. In the early 1960s the Swedish Telecommunication administration operated about 2,600,000 telephones. The telephone density, approximately one telephone for every three persons, was higher than that of any country except the United States.

The Swedish administration not only maintains a development organization, as do most other countries in Europe, but is unique among European telephone administrations in possessing factories for production of telephone equipment.

In Norway the International Bell Telephone company in 1880 secured franchises for Oslo and Drammen. The next year local companies established exchanges in several cities and a competing system in Oslo. Within ten years, largely through local enterprise, the telephone came to hold much the same place in Norwegian rural life that it does in the sparsely settled districts of the U.S. Ultimately the state acquired the more important local systems, and at the beginning of the 1960s about 89% of the country's 700,000 telephones were government-operated.

In Denmark, likewise, the telephone was introduced by private enterprise, the government interposing no serious difficulties. Gradually numerous small systems were consolidated into several relatively large organizations. The principal one is the Copenhagen Telephone company, a stock company in which the government owns a controlling interest. Nearly half of Denmark's 1,100,000 telephones are in the city of Copenhagen. Denmark had at the start of the 1960s about one telephone for every five persons.

7. Belgium and the Netherlands.—In Belgium telephone exchanges were first established in various cities by private concessionaires. In 1896, however, the telephone system became a complete government monopoly, and has remained so. At the beginning of the 1960s there were about 1,100,000 telephones. Private grants formed the initial pattern in the Netherlands also. As competition with the government telegraphs became apparent, the private telephone systems were integrated into a single government system, which at the beginning of the 1960s comprised about 1,500,000 telephones.

8. Austria, Italy and Spain.—Austria followed the typical European course, starting with private companies which later were bought in so as to create a government monopoly. There were about 650,000 telephones at the beginning of the 1960s, or about one telephone for every 12 persons.

In Italy the story is somewhat different. To begin with, concessions were granted to private companies, and in a number of important cities competitive situations arose. As the disadvantages of competition manifested themselves, consolidation was effected either voluntarily or by the dictates of local authorities. Partly as a result of onerous regulations, telephone development failed to keep pace with that in other European countries. In 1925 the structure of Italian telephone service assumed essentially the form that obtains today. Five concessionary companies operate local and toll service, each within one of the five zones into which Italy has been divided for this purpose. These companies are privately operated, although the government participates in the ownership. The government operates long-distance interconnecting service as well as the land-line portion of international service. Growth after World War II was rapid, and at the beginning of the 1960s Italy had about 3,500,000 telephones.

In 1924 Spain granted a concession to a subsidiary of the International Telephone and Telegraph company of New York to provide a nation-wide telephone system to supersede the previous government-owned and privately owned systems, a limited development. There were operating difficulties during the Spanish civil war, and in 1945 the government purchased the International Telephone and Telegraph operating interest. At the beginning of the 1960s, about 1,600,000 telephones were operated by a private company in which the government held a controlling interest.

9. U.S.S.R. and Satellites.—The Soviet ministry of communications announced that there were 3,558,000 telephones in the U.S.S.R. at the end of 1957. No figures were available for the satellite countries.

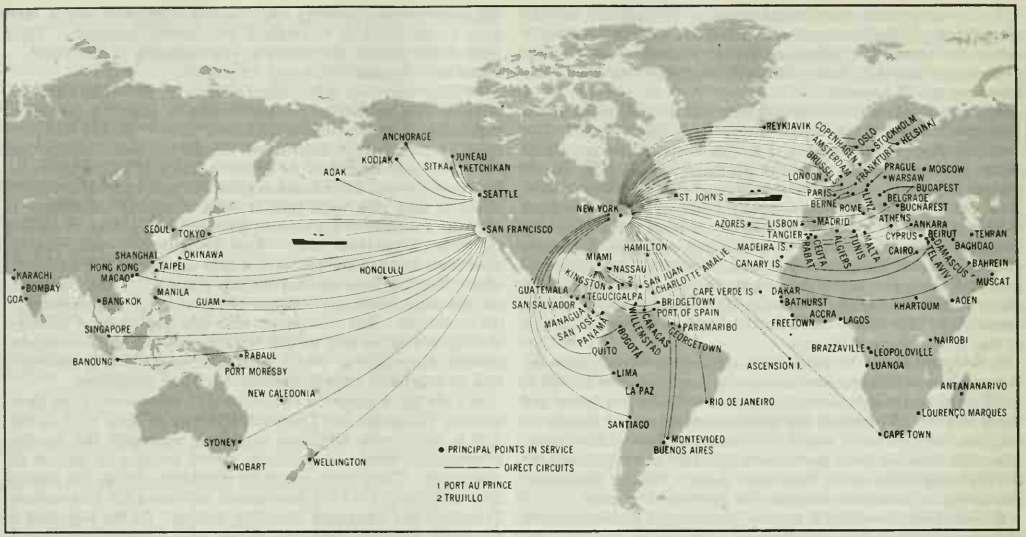
10. Canada.—Canada, divided by natural barriers and with most of its population concentrated near the long southern border, nevertheless in the early 1960s ranked next after the United States and Sweden in density of development, having one telephone for every four persons and a total of about 5,400,000 telephones. Moreover, Canadians led the world in use of the telephone, making about 510 calls annually per person, compared with 480 for the U.S. and 450 for Sweden. Most of the telephones in Canada are administered by seven private companies, which were banded together as the Trans-Canada system to furnish through service.

11. Latin America.—Early development of telephony in Latin America was slow for various reasons. In many cases foreign capital was obtained through private concessions, but certain countries where this was done followed the European plan of subsequent transfer to government ownership.

At the beginning of the 1960s Argentina, with about 1,300,000 telephones, led the Latin-American countries in telephone development. Brazil, with about 950,000, came next. Telephone systems of some size in Latin America were operated by subsidiaries of certain companies located in foreign countries, as follows: International Telephone and Telegraph corporation, in Brazil, Chile, Peru, Cuba and Puerto Rico; the L. M. Ericsson company, in Argentina and Peru; Cable and Wireless, Ltd., in Peru. Mexico's largest system was owned jointly by the I.T.&T. corporation and the L. M. Ericsson company.

12. Other Countries.—Telephone development in the rest of the world may be summarized with the statement that while some sort of telephone service is to be found almost everywhere, systems of substantial size are limited to Japan (about 4,800,000 telephones), Australia (2,100,000), South Africa (900,000), New Zealand (700,000) and Finland (600,000). Asia, with the lowest average level of telephone development, exceeded all the other continents in percentage growth during the five-year period ending with 1956.

13. Telephone Statistics.—At the beginning of the 1960s there were approximately 130,000,000 telephones in the world, 55% of these in the United States. Almost 70% of the world's telephones were privately owned, the remainder government-owned. Major U.S. overseas telephone connections are shown in fig. 6.



BY COURTESY OF BELL TELEPHONE LABORATORIES, INC.

FIG. 6.—U.S. OVERSEAS TELEPHONE CIRCUITS

Among the world's cities, Washington, D.C., led in density of telephone development, with 71.4 telephones per 100 population. Others ranking high among cities of more than 100,000 population included: Los Angeles, 61.3; San Francisco, 60.6; Basel, Switz., 59.1; Stockholm, 54.8; Berne, Switz., 54.1; Geneva, Switz., 53.5; Hartford, Conn., 56.3; Pasadena, Calif., 57.7; Wilmington, Del., 54.6; and Denver, Colo., 51.6.

14. ITU and CCIF.—Much of the progress in international telephone communication may be attributed to the work of international co-ordinating bodies. This work now centres in the Union Internationale des Télécommunications (UIT), called in English the International Telecommunication union, abbreviated ITU. This organization, now an agency of the United Nations, was founded in 1865 as the Union Télégraphique Internationale. The purposes of this organization as set forth in its charter are: "To maintain and extend international co-operation for the improvement and rational use of telecommunication; to promote the development of technical facilities and their most efficient operation, in order to improve the efficiency of telecommunication services, increase their usefulness, and make them, as far as possible, generally available; to harmonize the actions of nations in the attainment of those common ends." The ITU has a membership in about 90 countries and territories. The activities of the ITU and its constituent bodies are financed by contributions from member governments.

For many years there were three permanent and essentially autonomous organs of the ITU: (1) the Comité Consultatif International Téléphonique (International Telephone Consultative committee), abbreviated CCIF, founded in 1924; (2) the Comité Consultatif International Télégraphique (International Telegraph Consultative committee), abbreviated CCIT, founded in 1925; and (3) the Comité Consultatif International des Radio Communications (International Radio Consultative committee), abbreviated CCIR, founded in 1927.

For implementing its work, each of these international consultative committees organized a number of study groups covering different phases of activity. A plenary assembly distributes questions to be investigated by the study groups. These work by correspondence or by meeting and submit their findings to the succeeding plenary assembly, which studies and discusses them and makes recommendations. Ordinarily the recommendations take the form of directives or rules which, without being mandatory, are usually observed by the technical and operating services of the government administrations and private operating agencies of the countries belonging to the union. The director of the consultative committee co-ordinates the work of the study groups, of the plenary assembly and of the committee as a whole.

The work of these committees has emphasized the importance of good engineering and transmission standards to world-wide communication service and has gone far in bringing about the introduction of such standards. In particular, exhaustive consideration has been given to problems of transmission and interference, frequency bands and allocations, standards of telephone performance and questions of suitable rates and classes of service, all of which has contributed to effective international long-distance telephony. On Jan. 1, 1957, the CCIF and CCIT were merged into a single committee known as the CCITT (Comité Consultatif International Téléphonique et Télégraphique).

III. TECHNICAL DEVELOPMENT

The first telephones were extremely crude; it was barely possible to talk over them. The problems of transmitting speech for substantial distances were not understood, or even visualized. Nor was there the slightest comprehension of the complications involved in interconnecting large numbers of telephones.

One indispensable phase in the development of telephony was the progressive improvement of telephone station instruments as to quality and loudness of speech, as well as convenience and cost. Concurrently, it was necessary to develop telephone lines and circuits that could transmit speech currents reliably, without appreciable impairment or interference, and economically, for short or long distances. These two advances are sometimes grouped under the single term transmission development. Through such development, telephone conversation, except under adverse conditions, has become practically effortless.

Another essential was the development of switching mechanisms whereby any two telephone instruments in a large system could be connected together certainly, rapidly and economically. This part of the advance is referred to as switching development. Progress in these areas came about gradually in different countries, along generally similar but not identical lines, and with, in many instances, substantial time differences.

1. Station Instrumentalities.—The station set at the premises of a telephone customer normally consists of a transmitter,

which converts the speech waves in the air into their electrical replicas; a receiver, which performs the reverse operation of converting the incoming electrical waves into sound waves; a transformer (formerly called an induction coil) designed to increase the effectiveness of the transmitter and to permit full duplex operation (*i.e.*, use of a single pair of wires for speech in both directions, without requiring a switch to change from talking to listening or vice versa); a bell or equivalent summoning device; a switch hook to control connection of the set to the customer's line; and various associated items. In automatic switching systems the station set also includes a device, such as a dial, for generating signals to actuate the switching mechanism.

Bell's original electromagnetic transmitter served also as the receiver, the same instrument being held alternately to the mouth and the ear. An important forward step was the invention by Émile Berliner in 1877, and by Thomas A. Edison later in the same year, of different forms of transmitters utilizing variable-contact resistance between two solid electrodes connected in a battery circuit. Berliner employed metallic electrodes, while Edison used semiconducting materials, particularly plumbago (*i.e.*, graphite). Berliner also introduced the induction coil into the transmitter circuit in 1877.

The principle of the microphone contact (*see* MICROPHONE), which underlay both the Berliner and Edison transmitters, was elucidated by David E. Hughes in England in 1878. By virtue of this principle, energy from a battery is controlled by the voice waves so that the telephone transmitter acts as an amplifier as well as a converter of sound waves. The first recorded recognition of amplification through a microphonic contact seems to be found in a German patent of Robert Lüdtge, issued Jan. 12, 1878.

A transmitter in which granular carbon was used for the variable contact was invented in 1878 by Henry Hunnings, an English clergyman. Edison invented a transmitter whose carbon granules were obtained from anthracite coal, and Anthony C. White in 1890 invented the solid back transmitter, using a "button" of granular carbon placed between a fixed electrode and a diaphragm-actuated movable one. White's transmitter incorporated all the basic features of the modern telephone transmitter, though great improvements in efficiency (up to a thousandfold amplification of speech sounds), naturalness, resistance to aging and other features have resulted from continuing research and development. Extensive studies of the properties of speech and hearing have contributed to this result.

Modern telephone receivers utilize the same basic principles as those found in Bell's original instrument. Fundamentally, the receiver consists of a permanent magnet having pole pieces wound with coils of insulated fine wire, and a diaphragm driven by magnetic material which is supported near the pole pieces. Speech currents passing through the coils vary the attraction of the permanent magnet for the diaphragm, causing it to vibrate and produce sound waves. Through the years the design of the electromagnetic system has been continuously improved to provide better talking qualities. In a new type of receiver, introduced in the United States in 1951, the diaphragm, consisting of a central cone and ring-shaped armature, is driven as a piston to obtain efficient response over a wide frequency range.

In early telephone sets a hand-cranked generator, or so-called magneto, was used for signaling the operator, and local dry batteries supplied power to the transmitter. A centralized battery arrangement for signaling was invented in 1880 and one for both talking and signaling in 1886. With the gradual introduction of this common battery system, magneto telephone sets were largely superseded.

The first suggestions for mounting a telephone transmitter and receiver on a common handle, thus forming what is now known as a handset, were made by two Englishmen, C. A. McEvoy and G. E. Pritchett, in 1877. Starting in 1878, a type of handset devised by Robert G. Brown was used by boy operators in the Gold and Stock telephone exchange in New York city. Handsets for customers' use were introduced in France about 1882 and spread quickly in Europe. Initial difficulties in meeting satisfactory transmission standards with the handset arrangement were gradually overcome, and most modern instruments are of this type.

Methods of connecting the transmitter and receiver to the line likewise have been gradually improved. Starting with work by George A. Campbell, arrangements known as antiside-tone circuits have been developed whereby most of the electrical energy generated in the transmitter is directed toward the distant station, with a minimum entering the speaker's receiver.

Apart from the basic telephone instrument, special customer equipment and arrangements are provided to meet individual requirements of various kinds. Telephone sets are available which permit the user to hold a conversation without lifting the receiver, a small microphone being used to pick up the voice and a small loud-speaker to reproduce the incoming speech. There are loud gongs or other signaling devices for noisy locations, amplifiers for persons with subnormal hearing, automatic telephone-answering devices, arrangements for recording conversations, loud-speakers, etc.

Telephone pay stations installed in public or semipublic locations provide a coin telephone set which includes, in addition to the usual station equipment, a coin collector having one or more slots designed to accept legitimate coins and reject slugs or spurious coins. The coins, in passing down their respective slots, strike distinctive gongs whose tones permit an operator to supervise the deposits. Many coin collectors are arranged to hold the coins in suspension in a hopper, with means provided whereby the suspended coins can be collected or refunded, depending on whether or not the desired connection is completed.

Key telephone systems, affording greater flexibility of telephone usage than a single station set, are frequently used in business offices and residences. Combinations of push-button, turn-button or lever-type keys, installed on a desk or table or mounted integrally in the base of a telephone set, can be arranged to perform a variety of functions, particularly where two or more lines and a number of stations are involved. Thus a telephone may be connected to any of several lines, a call may be held on one line while conversation proceeds on another, etc.

2. Telephone Lines.—Telephone circuits are furnished principally by wire lines, although radio is used to a considerable extent. Telephone lines comprise a network of wires which interconnect individual telephone stations, central offices and communities. These wire lines are of two forms—cable and open wire. The usual type of cable consists of insulated copper wires, twisted together and usually covered with a protecting sheath. Open-wire lines consist of bare wires, generally of copper, fastened to insulators which are supported on poles at some distance above ground, commonly on pins in crossarms.

Following previous telegraph practice, the first telephone circuits utilized a single overhead wire, usually iron, with ground return. An important early improvement was the development by Thomas B. Doolittle in 1877 of hard-drawn copper wire that gave good tensile strength together with improved electrical conductivity. The advantages of a two-wire or so-called metallic circuit in reducing noise and interference were soon realized. It was found also that by transposing the wires (*i.e.*, interchanging their positions) a number of circuits could be carried on a single pole line without excessive interaction which would permit the conversation on one circuit to be heard on another. These several features made possible a successful open-wire telephone line between Boston and New York in 1885.

Early in telephone history, as the number of open-wire circuits strung on poles and rooftops in the large cities began to reach the point of impracticability, methods of compacting the lines in overhead or underground cable were tried. At first the wires were placed in pipes and sealed against moisture with oil, paraffin or asphaltum. Means were soon developed whereby lead, heated to plasticity, could be extruded over a core of conductors. The introduction of dry paper as insulation for the conductors completed the foundation for the modern telephone cable.

Research and development covering many materials and processes made it possible to increase the number of pairs in a full-sized local cable from a maximum of 50 in the year 1888 to more than 2,000 in the second half of the 20th century. This increase was

accomplished largely by reducing the size of the copper wires, from 25 lb. per wire mile in 1888 to 4 lb. Following World War II, new types of cable sheath were developed in order to reduce cost and lessen dependence on lead supply. One form uses thin layers of aluminum and steel covered with polyethylene.

In a typical urban or suburban installation, insulated wires from each customer's premises extend to a distribution cable which in turn connects to a feeder cable leading to the central office. In densely built areas the feeders are generally placed in underground ducts; elsewhere, overhead. Circuits between central offices, known as trunks, are provided in trunk cables which usually employ larger wires than those leading to customer stations.

The network of lines which interconnect approximately 75,000 communities in the U.S. included in the early 1960s approximately 35,000,000 mi. of wire on 260,000 mi. of route. About 93% of the wires were in cable, the remainder open wire. For the rest of the world, the percentage of long-distance facilities furnished by cable ranges from nearly 100% in a few countries to a very low figure in the less populated regions. Since cable is virtually stormproof, it affords much greater reliability than does open wire.

A type of long-distance cable common in the U.S. employs conductors 36 mils in diameter, weighing 20 lb. to the mile (no. 19 AWG gauge). In making this cable, two paper-insulated wires are first twisted together to form a pair; then a quad is formed by twisting two pairs together. This twisting aids in the prevention of cross talk (*i.e.,* overhearing) between different pairs. In European toll cables, four conductors are usually twisted together to form a spiral four or star quad. In either case the quads are grouped together and enclosed by the sheath. Such a cable may be supported aerially or placed in an underground duct, or, with a suitable protective covering of jute or steel tape, may be buried in the ground. A special type of cable used for coaxial systems is discussed under *Carrier Systems* below.

3. Phantoms.—Around 1900 the principle of "phantoming" two pairs of wires was introduced. The original idea of phantoming was devised by Frank Jacob in 1882 and the present method by John J. Carty in 1886. The phantom arrangement makes it possible to derive from two pairs of wires a total of three telephone circuits. The additional circuit, called the phantom, is obtained by using the two wires of each pair in parallel as a side of the phantom. To render this scheme practical, efficient balanced transformers were needed for the end connections, and transposition arrangements were required to keep the cross talk between the three circuits within tolerable bounds. The phantom principle came to play an important part in both open-wire and cable plant. Because the presence of phantoms makes high-frequency transmission difficult, the advent of carrier systems greatly curtailed the use of phantoms for long-distance circuits.

4. Loading.—As speech currents pass along a line their strength decreases (a process referred to as attenuation), so that after some distance they become too weak to actuate a receiver properly. Studies by A. Vaschy (1889) and Oliver Heaviside (1893) developed the theoretical possibility of improving the transmission efficiency of telephone lines by artificially increasing their inductance. Various investigators speculated on the practicability of approximating the beneficial effect of an increase in uniformly distributed inductance by introducing in the line concentrated or lumped inductance in the form of low-resistance loading coils. Finally, in 1899, Michael I. Pupin and George A. Campbell (working independently, Pupin having a slight priority) discovered that the key to the problem was the spacing of the loading coils. By providing at least pi (*i.e.,* about 3.14) loading coils per wave length at the highest frequency to be transmitted, a substantial reduction in the attenuation of the speech waves is obtained. Thus a cable circuit may, by means of coil loading, be made to transmit telephonic currents as efficiently as a nonloaded one whose conductors weigh many times as much.

Another way of adding inductance to a cable circuit was proposed by Carl Emil Krarup of Copenhagen, his idea being to wind helically along the copper conductor a fine wire of soft iron. A cable with this continuous loading was placed between Elsinore and Helsingborg in 1902. While continuously loaded cables have proved of some importance for submarine applications, coil loading has been almost universally preferred for loaded land cables.

By 1913 coil loading made it possible to extend the useful range of open-wire circuits to approximately 2,000 mi., and to employ underground cable to connect Washington, D.C., and Boston via New York city. After electron tube amplifiers became available, open-wire loading was largely abandoned, but loading is still extensively applied on cable circuits used for voice frequencies, and especially on local trunk circuits.

Satisfactory loading requires that the loading coils have very low energy losses in their magnetic cores and copper windings. The toroidal-shaped cores of modern loading coils make use of a powdered magnetic alloy whose particles are individually insulated, compressed under high pressure and then usually heat treated to develop optimum magnetic properties. Progressive improvement of magnetic core materials has brought about large reductions in the size and cost of loading coils, as well as improved performance. A core material commonly used in the U.S. consists of an alloy of nickel, iron and molybdenum known as molybdenum permalloy, discovered by Gustav W. Elmen. In Europe powdered iron cores of somewhat larger size are employed.

5. Electron Tubes and Repeaters.—For years the range of telephony was severely limited by loss of energy due to dissipation along wire lines, or to spreading in the case of radio waves. The idea of inserting one or more repeaters in a telephone line for the purpose of reinforcing or amplifying the telephonic currents from some local source of energy is almost as old as the telephone itself, but many years elapsed before the quest for a satisfactory repeater achieved success. In a so-called mechanical repeater, tried in 1904, inertia of the moving parts was found to present inherent limitations. Subsequently H. E. Shreeve developed a repeater employing carbon-contact amplification which was capable of practical use. The real solution to the amplification problem, however, was found in the device invented by Lee De Forest in 1906, which he called the audion, and which is now known as a three-electrode vacuum tube or electron tube, or in England as a valve. In its original form as used in radiotelegraphy, this tube was unsuited for telephone purposes (*see* ELECTRON TUBE; AMPLIFIERS). Research by Harold D. Arnold, Irving Langmuir and others showed that a major requirement for adequate performance in an amplifier was the creation of a high degree of vacuum inside the tube envelope. By 1914 satisfactory high vacuum tubes were produced. Using telephone repeaters with amplification supplied by vacuum tubes, telephone service between New York and San Francisco was inaugurated in 1915.

Continuing development work brought about many improvements in vacuum tube repeaters. Efficient circuits and auxiliary equipment were devised for utilizing the amplifier element and associating it with the line circuits. The life of the repeater tube most commonly used in 1917 was about 1,000 hours, whereas standard tubes of a type introduced in 1935 have a life of about 90,000 hours, equivalent to more than ten years of continuous operation. In addition, the power required to heat the tube filaments was reduced to one-tenth that required in 1917.

With repeaters available, transmission defects of different kinds became increasingly apparent. The lines, particularly if loaded, were found to introduce severe distortion by reason of differences in the transmission efficiency and transmission velocity at different frequencies. Furthermore, large variations in transmission loss resulted from changes in the electrical resistance of cable conductors with temperature, or from changes in the leakage of open-wire conductors. Extensive developments in the field of network theory made it possible to design equalizers with frequency characteristics that compensate accurately for the line distortion. Variations in line loss are automatically counteracted by transmission regulators. Devices known as companders, in which the amplitude of speech syllables serves to compress the range of speech volumes at the transmitting end and to introduce a corresponding expansion at the receiving end, have proved beneficial in reducing the effect of line noise on both wire and radio circuits.

Another unwanted effect on telephone circuits is the presence of electrical echoes due to discontinuities in the line. These, when

converted into sound, may disturb both talker and listener. The echoes become more annoying the longer they are delayed and hence are of greatest concern on long circuits. They are controlled by restricting their occurrence, by reducing the line delay or by applying devices called echo suppressors to prevent their reaching the telephone users.

The vacuum tube provided for the first time the means for precise measurement of these and other transmission effects, and thus made it possible to establish a firm foundation for the transmission art. Vacuum tubes connected as generators of electric oscillations supply the range of frequencies needed for communication measurement. Other vacuum tubes amplify the extremely weak currents involved in telephone transmission to a level at which they can be conveniently measured.

Following the three-element tube, many different types of vacuum tubes were developed for various applications. These include tubes with four or five elements for providing high and stable amplification, klystrons and traveling wave tubes for operation at extremely high frequencies, phototubes responsive to light, etc. Millions of cold-cathode, gas-filled electron tubes are used in telephone handsets attached to multiparty lines to permit ringing only the desired one out of four or more stations on a line. The normal life of such tubes is estimated to be hundreds of years.

6. Carrier Systems.—It is common practice on long-distance routes to multiply the circuit capacity by the application of carrier telephone systems. These are called carrier systems because of the underlying principle, known as modulation, whereby the voice frequencies modulate a higher frequency current which "carries" the voice currents. A voice wave is composed of undulations with a frequency of occurrence ranging from about 200 to 3,000 per second or, as it is commonly expressed, from 200 to 3,000 cycles per second (c.p.s.). In the modulation process these frequencies are transposed to a higher frequency range (*e.g.*, 10,200 to 13,000 c.p.s.) for transmission over the line. By using different carrier frequency bands a number of conversations can be sent simultaneously over one transmission path. At the receiving end the carrier frequency bands are separated by electric networks called filters (invented by G. A. Campbell in 1915), and the original voice frequencies are recovered by an inverse process known as demodulation. Carrier systems make extensive use of vacuum tubes—as amplifiers, as oscillators for generating carrier frequencies and sometimes as modulators for shifting bands of frequencies from one range to another.

The art of multiplex carrier telephony grew out of the harmonic telegraph systems associated with the names of Gray, Bell, E. Mercadier and others. The extension of the carrier principle to telephony, together with the use of electrical resonance instead of mechanical resonance for selecting the carrier frequencies, was accomplished in 1891 by two Frenchmen, Maurice Hutin and Maurice Leblanc. During the period 1908–11, demonstrations of carrier techniques were conducted by Ernst Ruhmer in Germany and Maj. Gen. George O. Squier in the U.S.

Following experiments between Toledo, O., and South Bend, Ind., in 1917, the first commercial application of the carrier principle was made in 1918 on an open-wire line between Baltimore, Md., and Pittsburgh, Pa., giving four additional telephone circuits on a pair of wires. With modern carrier techniques, as many as 16 telephone circuits are derived from a single open-wire pair. Beginning in 1934 in Germany, a system which yielded one extra telephone channel on a lightly loaded cable pair was extensively applied in Europe.

Application of substantial numbers of carrier channels to cables necessitates many more repeaters because of the higher attenuation of the cable pairs, so that minor imperfections in each amplifier become quite serious. The solution to this problem was provided by Harold S. Black when he invented the negative feedback amplifier in 1927. The complete development of the underlying theory came in subsequent work by Harry Nyquist. Though difficult in theory, the idea itself is quite simple. A part of the amplifier output is fed back to the input in such a way as to give practically distortionless amplification, together with almost complete absence of variation in amplification as a result of power supply variation and tube aging. The improvement thus obtained may be a thousandfold or more.

The negative feedback principle is now applied almost universally to amplifiers used for any purpose. This principle formed the basis for the introduction in 1937, between Toledo and South Bend, of a 12-channel carrier system operating on nonloaded cable pairs. Today carrier systems are quite generally used on long-distance cable routes, yielding circuits with excellent transmission properties for distances up to thousands of miles. Important in this connection has been continued progress in the control of cross talk and noise. Cable carrier systems in the U.S. are arranged to derive 12 to 24 telephone circuits from two nonloaded 19-gauge pairs. Similar systems, some of them providing much larger numbers of channels, are used in Europe and elsewhere.

With the aid of a special type of cable conductor, called a coaxial unit or merely a coaxial, carrier techniques have been greatly expanded. A coaxial consists essentially of a copper tube, commonly about the size of a lead pencil, with a wire centrally supported inside. A full-size cable may contain as many as eight such coaxials, plus a number of conventional pairs of wire. By applying amplifiers at close intervals, four to eight miles, a very wide band of frequencies can be transmitted over a single coaxial.

The transmission properties of a coaxial circuit were considered by various 19th-century workers, especially Lord Kelvin and Alexander Russell. But it was a far journey from these studies to a wide-band transmission system suitable for long-distance multichannel telephony and television. The first such coaxial systems were applied in 1936 between New York city and Philadelphia and between Berlin and Leipzig. More highly developed systems are now widely used in both the U.S. and Europe.

A type of coaxial system introduced between Minneapolis, Minn., and Stevens Point, Wis., in 1939 utilizes a frequency band nearly three megacycles (*i.e.*, 3,000,000 cycles) wide, with two coaxials giving either 600 two-way telephone channels or a television circuit in each direction. Techniques were further augmented in a system, first applied in 1952, which provides a frequency band nearly eight megacycles wide, so that two coaxials can provide either (1) 1,800 telephone channels; or (2) 600 telephone channels plus a 4.2-mc. television circuit in each direction. Since a long coaxial system may have 1,000 or more repeaters connected in tandem, each one providing a ten-thousandfold amplification, the utmost perfection is necessary in the performance of each repeater.

The minimum distance for which carrier systems prove economical depends on the cost of the terminal apparatus. In recent types of systems the economical distance has been greatly reduced, so that carrier can be used for short-haul toll circuits, interoffice trunks and rural circuits.

7. Radio.—Radio, originally employed for telegraphy, has become an important instrument for telephone purposes. Both the theoretical and the utilitarian aspects of radio transmission are treated elsewhere (*see* BROADCASTING; RADIO).

The same basic principle of modulation used in carrier systems is required to shift the telephone signals to the desired radio frequency. There are several different kinds of modulation that may be used in either radio or wire systems. Simplest of these is amplitude modulation, in which the amplitude of a modulating wave (*e.g.*, a speech wave) controls the amplitude of a sine wave carrier. The first complete analysis of amplitude modulation was made by John R. Carson, who showed that in order to convey the intelligence only one of the bands of frequencies generated in the process need be transmitted. Because of its efficient use of frequency space, Carson's single side band method is widely employed.

Another common form is frequency modulation, wherein the instantaneous frequency of a sine wave carrier is varied in proportion to the amplitude of the modulating wave. This method makes it possible by sacrificing frequency space to gain an advantage in the ratio of signal to noise.

Still another form is pulse modulation, in which the carrier consists, not of a sine wave, but of a series of pulses whose amplitude, duration, position or mere presence may be controlled so as to convey the message. In pulse code modulation, originated by

A. H. Reeves in 1939, successive quantized samples of the modulating wave produce corresponding code patterns of pulses. This provides, at the expense of band width, low vulnerability to noise and interference and adaptability to repeated regeneration of the signals without distortion. For multichannel transmission, the pulses corresponding to different channels are interleaved.

The advantages of being able to telephone without a wire connection were obvious from the beginnings of radio. Many early radio experimenters, notably R. A. Fessenden and De Forest in America and Quirino Majorana, Giuseppe Vanni and V. Poulsen in Europe, succeeded in transmitting speech over distances of a few miles. Several essentials for practical application were lacking, however; to wit, a practicable generator of continuous high-frequency waves, a means for modulating these waves in accordance with speech and a receiving amplifier for revivifying the waves after enfeeblement in transit. In the main, it was the vacuum tube that provided the solution for all these problems.

In 1915, the same year that telephone service across the U.S. was begun, intelligible speech was experimentally transmitted by radio from Arlington, Va., to Hawaii and to Paris. In 1927 the first commercial overseas radiotelephone circuit was opened between the U.S. and England. Service from Berlin to Buenos Aires, Arg., was begun in 1928, from England to South America and Australia in 1930, and service between the U.S. and South America, Central America, the Hawaiian Islands, the Philippine Islands, the Netherlands Indies and Japan during the years 1930 to 1934. In the early 1960s radiotelephone service was available between all principal countries not connected by wire and in addition was used in many instances to supplement wire facilities. Most of the overseas radiotelephone circuits were in the short-wave range from 3 to 30 mc., where long-distance transmission is accomplished by reflecting the waves from the ionosphere. The performance of the radio circuits, although fairly satisfactory, was subject to some difficulty due to atmospheric disturbances.

Radio systems that make use of scattering from the tropospheric layer in the atmosphere can transmit wide frequency bands capable of handling a number of telephone channels to distances "beyond the horizon." A system of this type was placed in service between Miami and Havana in 1957.

Starting with service to the steamship "Leviathan" in 1929, radiotelephone service has been extended to many large oceangoing ships. Thousands of smaller ships in coastal waters, on lakes, in harbours and on rivers are equipped for connection to shore telephone stations by radio. Commercial radiotelephone service between the public telephone system and motor vehicles in cities and on highways began in the U.S. in 1946. By the early 1960s there were about 18,000 such mobile stations. In addition, about 500,000 other vehicles were provided with radiotelephone sets for private services such as those of air lines, police, taxicabs, etc., and there were 16 mobile radio stations for public service on trains and airplanes. Radio systems were used also for bridging short water gaps and crossing other difficult terrain. Portable radiotelephone equipment was employed in emergencies such as floods or hurricanes to bridge gaps in the regular toll circuits until normal service could be restored. Radio paging systems were available in which a small receiving set carried on the person was used to summon an individual to the nearest telephone in order to receive a call.

Repeatered radio systems for obtaining substantial numbers of long-distance telephone circuits, as well as broad frequency bands for television transmission, are used in many parts of the world. These commonly transmit radio waves of thousands of megacycles, known as microwaves, over line-of-sight paths. As early as 1934 a microwave radio system developed by A. G. Clavier and others was employed for transmission across the English channel. The large-scale development during World War II of microwave techniques for radar use provided a powerful stimulus in the postwar development of microwave communication systems. A microwave radio-relay (*i.e.*, repeatered) system was placed in service between New York city and Boston in 1947, and one across the United States in 1951. This latter system provides six radio channels in each direction, and each radio channel affords either 600 telephone circuits or a single television circuit.

Concentration of large numbers of circuits on a single route by means of carrier systems, coaxial systems or radio-relay systems yields large economic advantages through sharing common elements of cost including right of way, line conductors, installation, line maintenance, radio towers, power supply, etc. Combining the requirements for both telephony and television on a common route affords further economy.

8. Submarine Telephony.—Submarine cables for transmitting telegraph signals antedated the invention of the telephone. Many years went by, however, before long submarine cables suitable for telephony became practical. As early as 1891 a cable containing four wires was laid under the English channel, between St. Margaret's bay, Eng., and Sangatte, Fr. During the next five years several cables of similar construction, suitable only for shallow water and relatively short distances, were placed under the channel.

When cable is laid in deep water, the high pressure necessitates a different type of cable, the usual construction for deepwater telephone cables consisting of a central conductor surrounded by insulation around which is a return conductor. The first such cables were laid in 1921 between Key West, Fla., and Havana, Cuba, a distance of more than 100 mi. Each of these cables provided a single voice circuit and four telegraph circuits. Two years later the first long submarine cables adapted for carrier telephone transmission were placed across the 40-mi. stretch between the California coast and Santa Catalina Island.

For spanning greater distances with the wide frequency band required to accommodate a number of telephone channels, intermediate repeaters are needed. The first submerged telephone repeaters were applied between Anglesey, Wales, and the Isle of Man in 1943 and between Lowestoft, Eng., and Borkum, Ger., in 1946. These repeaters were adapted only for shallow-water operation. Repeaters designed for deepwater use were included in a pair of cables placed in 1950 between Key West and Havana, the two cables giving a total of 24 telephone circuits.

To handle the expanding requirements for transatlantic telephone service, the first transoceanic telephone cable system, between Clarenville, Nfd., and Oban, Scot., with an extension from Newfoundland to Nova Scotia and a radio-relay link to Portland, Me., was placed in service in 1956. This was a joint project of the American Telephone and Telegraph company, the British post office and the Canadian Overseas Telecommunication corporation. The two cables on the main crossing had submerged repeaters at approximately 40-mi. intervals to provide a total of 36 telephone circuits. The vacuum tube repeaters were designed to operate continuously and flawlessly with no attention for at least 20 years, at depths up to 2,000 fathoms. Similar submarine systems between Port Angeles, Wash., and Ketchikan, Alsk., and between California and Hawaii are in service.

9. Information Theory.—Progress in telephony and allied arts was greatly furthered by the formulation of a comprehensive theory underlying the communication of information. This theory, variously referred to as information theory (*q.v.*) or communication theory, resulted from studies by Claude E. Shannon, Norbert Wiener and others, following earlier work by R. V. L. Hartley. It states in essence that information of the kind contained in messages transmitted over communication systems is measurable; as a consequence, the loss of information caused by unpredictable perturbations (noise) introduced during transmission can be evaluated. The basic unit of measure is a "yes-or-no" choice, which is called a bit (short for binary digit). All information can for communication purposes be expressed as, or encoded into, sequences of on-or-off (*i.e.*, binary) pulses. The development of the basic mathematical theory made it possible to treat transmission factors such as band width, noise, distortion and the like, and the relations among them, in quantitative terms. This facilitated the comparison of different transmission systems—for example, systems using different forms of modulation, different coding, etc. The theory also established upper limits for the rate at which information can be transmitted over systems of different kinds.

10. Switching.—Establishing a connection between any two telephones out of a large group is a complicated process. Even in

the simplest situation, where both telephones are served by the same switchboard, it is necessary to: (1) observe that a customer wishes to make a call; (2) connect the operator (or switching mechanism) to his line; (3) determine what telephone he wishes to be connected with; (4) select a speech path between them which is not already in use; (5) determine whether the wanted telephone is idle or busy; (6) if idle, ring the bell, or if busy, inform the calling customer; (7) determine when the call has ended; and (8) restore all equipment to its quiescent state, in readiness for other calls. If the telephones are served by different switchboards, perhaps in widely separated cities, it is also necessary to determine what switchboard serves the wanted telephone and how it may be reached.

From the beginning, the importance of the switching function in telephony was recognized, but the means initially employed were primitive. The first telephone switching arrangement was provided by the Holmes Electric company at Boston in 1877, by using for telephone connections in the daytime the line wires and plug-and-block connectors employed for a burglar alarm system at night. A commercial telephone switchboard placed in operation at New Haven, Conn., in 1878 served 21 stations on eight grounded lines. In contrast, a modern central office may serve as many as 100,000 telephones on 50,000 lines.

The New Haven exchange, and other primitive ones, provided a separate switch for each combination of lines that might have to be connected together. For eight lines this required only 28 switches; but the number of switches increased much more rapidly than the number of customers and quickly became impracticable. With 100 lines, for example, 4,950 switches would have been required. To get around this difficulty, the cord circuit was introduced in 1880. Each line was terminated on the switchboard in a socket (called a jack), and a number of short flexible circuits (called cords) with a plug on each end were also provided. Two lines could thus be interconnected by inserting the two ends of a cord in the appropriate jacks. This was an efficient system as long as the number of calls passing through a switchboard could be handled by a single operator.

Another early invention, which permitted much larger volumes of traffic to be handled efficiently on a manual basis, was the multiple system devised by Leroy B. Firman in 1878. With this, each customer's line is connected to a number of jacks, placed at suitable intervals along the switchboard, so that one is within the reach of every operator.

With the introduction of the multiple system, it became necessary to develop a busy test to determine whether or not a line with which connection was desired was already in use through a connection made at some other part of the switchboard. In a manual switchboard the operator performs this test by touching the tip of a connecting plug to a conducting sleeve forming part of the jack of the desired line. If the line is busy, a click is heard in the operator's receiver.

In 1895 the magnetically operated drop signals which informed the operator of a customer's desire for service were first replaced by incandescent lamps. These were more reliable and smaller than the drop signals and permitted a much more compact arrangement of the subscriber's jacks.

There was continuous improvement to adapt manual switchboards to the growing volume and complexity of telephone traffic and to modern standards of workers' comfort and of customers' convenience. There also were improvements in the associated circuits. But the design of all subsequent manual switchboards was based upon the plug-ended cord, the multiple, the busy test and lamp signals.

The idea of fully automatic switching also appeared quite early. In fact the first patent for an automatic switching system was issued to Daniel Connolly, T. A. Connolly and T. J. McTighe in 1879, only two years after the first primitive switchboard in Boston. This system never achieved commercial success. In 1889 Almon B. Strowger invented an automatic system which was installed at La Porte, Ind., in 1892, the first commercial automatic exchange in the world. The system was subsequently developed by Alexander E. Keith and other engineers of the Automatic Electric company into a form which is still extensively used throughout the world under the names step-by-step or Strowger system. In this system the switching mechanisms are operated directly by pulses generated at the customer's instrument. Originally the customer operated a push button to produce the switching pulses, but this and other types of calling devices gradually gave way to the dial mechanism now in general use. This was invented in 1896 by A. E. Keith, C. J. Erickson and John Erickson of the Automatic Electric company.

The basic mechanism around which the Strowger system is built is the Strowger switch, which is sometimes called a connector and sometimes a selector according to the use to which it is put. In complete form this consists essentially of two parts: a ten-by-ten array of terminals (the bank) arranged in a cylindrical arc; and a movable switch (the brush) which is translated along the axis of the cylinder by one ratchet mechanism and rotated about it by another, so that it can be brought to the position of any one of the 100 terminals. Each ratchet mechanism is driven by an electromagnet, which can respond to the pulses produced by a telephone dial.

Many other switching systems have been invented, notably in the United States, Germany and Sweden, and a number of them have been put into successful commercial operation. Some of these have operated on the step-by-step principle but have used different types of apparatus. Others have operated on the principle of common control, in which pulses are stored for a short time in a device which then controls the switches either directly or through some intermediate mechanism.

The distinguishing feature of such systems is that the common-control mechanisms (known as registers, senders, translators, directors, markers, etc.) are not assigned to the customer for the duration of his call but are used only as long as they are needed and are then free to serve other customers. Thus each serves many calls per hour, and few are required; it is therefore practicable to provide complicated devices which can perform a variety of useful functions but which would be too expensive to assign for the duration of the call. Systems using common control have great flexibility and efficiency in the use of trunk groups, and are especially advantageous for large exchanges and for automatic routing of long-distance calls.

The earliest common-control systems were developed by the Western Electric company, primarily to meet the needs of large metropolitan centres. In a program beginning in 1906, two systems were developed alongside one another, and both proved to be successful. One, called the rotary system, was first put into commercial service in England in 1914. The other, called the panel system, went into commercial use at Newark, N.J., in January of the following year. Common-control principles were also applied to the step-by-step switch by the Automatic Electric company, and the resulting system was adopted by the British post office in 1922 for use in London. The most recent common-control systems have been crossbar systems, of which several commercial types have been developed by the Bell system in the United States, by the Swedish Telecommunication administration and the L. M. Ericsson company in Sweden and by subsidiary companies of the International Telephone and Telegraph company in Belgium and Germany.

The most important concepts in the evolution of the modern types of crossbar exchange were probably translation, the sender, the marker, the crossbar switch and the principle of call-back operation.

Translation, invented by E. C. Molina of the American Telephone and Telegraph company in 1906, makes it possible to convert incoming dial pulses from decimal to nondecimal form and thus affords flexibility and efficiency in the use of trunk groups.

The sender, first used in the rotary system, is essentially an automatic mechanism which generates new dialing signals, either in the code given by the translator or in other appropriate codes.

The basic function of the marker is to make a preliminary test of several alternative paths to a wanted destination through an array of switches, before any of the switches is closed, so as to avoid the possibility of encountering a busy switch after part of

the switching operation has been performed. It was invented by N. G. Palmgren and G. A. Betulander of Sweden in 1912 and was first used commercially in a system manufactured by the Relay Automatic company of England from 1915 to 1920. As used in crossbar systems of the Bell system, it has been developed into a complex assemblage which serves as the basic control element for the entire switching operation. Among other things, it tests the circuits before connections are established; it seeks out alternate paths when needed; and it reports trouble conditions which may be encountered in its preliminary tests. Since it can examine a large number of trunk circuits practically simultaneously, it uses them with great efficiency.

The crossbar switch is essentially a multiple relay structure affording fast operation and reliable contacts of precious metal. Unlike other switches mentioned above, the moving parts have little inertia and move through small distances. It was first conceived in America by Homer J. Roberts of Automatic Electric company in 1901, and later patented in separate forms by John G. Roberts and John N. Reynolds, both of Western Electric company. The first satisfactory mechanical design, worked out by Palmgren and Betulander in Sweden in 1919, was used by the Swedish telephone authorities in a commercial exchanger in 1926. Aside from a small-scale trial installation in Stockholm in 1919, which was subsequently abandoned, this was the first public use of the crossbar switch. The system in which it was used, however, and all other crossbar exchanges in Sweden until about the middle of the century, operated on step-by-step principles. Most of the crossbar switches used in America and other countries follow the Swedish design.

Call back is a principle of operation, invented by Edson L. Erwin in 1938, which has been effectively used in crossbar systems. When a customer originates a call, the register stores not only the wanted number but also the identity of the calling telephone, which is determined automatically. The connection with the calling subscriber is then disconnected, and an entirely new connection established, from a favourable point within the exchange, to both the calling and wanted telephones.

All the automatic switching systems described were built around the electromagnetic relay and other electromagnetically operated devices. The relay employs magnetic attraction produced by an electric current to move a magnetic armature and thus open or close electrical contacts so as to perform switching or other operations. This device dates back to the early part of the 19th century, and in a form much cruder than the present was the basis for the development of telegraphy. The name relay results from the early use of the device to pass along telegraph signals after enfeeblement in transmission. As switching systems developed, relays came to be employed for a wide variety of purposes, including multiple contacts, slow or fast operation or release, marginal operation, etc. Modern switching systems require approximately 1,000 relay operations to establish a single telephone connection, and some systems have as many as ten relays per subscriber. Hence, reliability in relays is extremely important. Less than one failure in 40 years of service is a normal criterion for satisfactory operation.

As automatic switching systems were improved, their application was extended until, by the early 1960s, 77% of the world's telephones were automatic, as compared with 15% only three decades earlier. This metamorphosis occurred not only because automatic operation is faster, more accurate and more economical than manual service, but more basically because in many areas the enormous number of operators necessary to support the rapid telephone growth would have far exceeded the possible supply.

In America the crossbar system, designed originally for metropolitan areas, presented so many advantages as compared with earlier automatic systems that its basic principles were soon applied in other fields. Crossbar tandem equipment, for example, is the modern version of the tandem principle developed originally for manual systems. Through the tandem scheme, traffic between offices on opposite sides of a large metropolitan area is handled through one or more intermediate (tandem) offices which act as clearinghouses for these relatively small amounts of traffic, handling them more efficiently than if direct paths were provided and thus reducing the number of trunk circuits required.

Another type of crossbar system is specially adapted to the automatic switching of long-distance circuits. With this, connections are made on a two-path or four-wire basis, a separate path being used for each direction of transmission so as to obtain superior transmission performance on connections comprising a number of circuit links in tandem.

The first crossbar switching system in the long-distance field, introduced in Philadelphia in 1943, enabled operators but not customers to dial long-distance calls. This advance was followed by another which permitted direct dialing by customers without the aid of an operator. The switching mechanism selects the route to the distant telephone. Should it find all direct circuits busy, it explores in succession as many as five alternate routes, and establishes the connection along the least circuitous one available. All this is done automatically. Necessary information regarding direct and alternate routes to the destination is permanently available in a translator, which supplies it to the control circuits as required.

Arrangements for customer dialing of at least some intercity calls were provided in all the principal countries of the world by the mid-1950s. In North America a complete program was worked out for eventually handling substantially all calls in this way. The initial installation of equipment to permit nation-wide dialing by customers was placed in service at Englewood, N.J., in 1951. With this equipment, customers were able to dial directly about 11,000,000 other customers in selected areas as far away as San Francisco. They dial a three-digit area code followed by the digits of the directory listing of the called number. By the early 1960s this type of service had been extended to more than 500 originating locations reaching about 40,000,000 customers.

Closely related to automatic switching is the automatic recording of data for preparing the customer's bill. The simplest means for this purpose, used in either manual or automatic offices, is an electromechanical counter, known as a message register, which records the number of calls made by a customer. In a more elaborate arrangement, referred to as multiunit registration, the register can be operated more than once for a single call, the number of operations depending on the distance and duration of the call. Each customer is billed on a bulk basis for all of the "message units" totaled on his register. This plan is widely used in America and almost universally elsewhere.

Another method of charging is "automatic ticketing," first introduced in Belgium. With this method, automatic equipment prints for each call a ticket similar to one that might be prepared by an operator.

In America, where distances are great and the tariff structure complicated, a highly versatile type of message registration system, called automatic message accounting (usually abbreviated to AMA), was developed by the mid-1950s. In this system the information needed for billing calls is recorded in the form of coded holes in paper tape. Automatic processing of the tapes in an accounting centre yields bills with any desired amount of detail. The AMA system is extensively used in association with direct distance dialing.

A big step in the improvement of telephone service in outlying areas resulted from automatizing the telephone switching in very small communities. Community dial offices, either of the step-by-step or all-relay type, were provided in many localities. These function with no attendance except for occasional maintenance visits.

Another phase of telephone switching is found in switchboards, known as private branch exchanges, which are located on the customer's premises. The private branch exchange (abbreviated PBX) serves both for interconnecting the station sets of the customer's establishment and for connection over trunk lines to the central office. The smaller PBX's can be placed on a desk and operated by a person who may perform other duties as well. Large PBX's resemble central office switchboards and may be of either the manual or automatic type.

By the early 1960s exploratory developments were well under

way in the principal communications laboratories of many countries with the ultimate objective of replacing electromechanical switching systems with electronic ones. A small beginning had been made in some units that are partly electronic and partly electromechanical. Among these may be mentioned the mechanoelectric system developed by the International Telephone and Telegraph company in Belgium and first installed at Ski, Nor., in 1954. The Bell system also announced that a fully electronic system was under development.

11. Associated Services.—Much of the plant required for telephone service is well adapted to the provision of other types of communication service, so that economies usually can be effected by designing the plant to handle the combined requirements of different services. The earliest of such associations was with the telegraph. Direct-current telegraph channels superposed on the telephone wires were leased to private customers and to the companies furnishing message telegraph service, and were employed also for telephone line maintenance. The introduction of carrier transmission methods made it possible to derive as many as 18 telegraph channels from one telephone circuit. Such channels serve both for teletypewriter exchange service, which is a switched service analogous to telephone service, and also for private-line teletypewriter service.

Satisfactory picture transmission requires a frequency band of about the same width as that for telephone conversation. Special networks are provided for picture service, and occasional use is made of regular telephone circuits for this purpose. Telephone circuits are employed also to render facsimile and other services. High-speed transmission of information in the form of pulses is being increasingly applied for both commercial and military purposes. Generally speaking, the chain networks used for the radio broadcasting of either sound or television programs are provided in conjunction with the telephone plant. For good quality of music reproduction, a frequency band somewhat wider than that adequate for speech is required. Television broadcasting necessitates a very wide frequency band, equivalent to that required for about 1,000 telephone circuits.

Television circuits are derived, commonly in association with large numbers of telephone channels, from coaxial or microwave systems. Wire broadcasting systems, in which sound programs are conveyed by wire directly to the customers' premises, are in limited use.

IV. TRENDS

Technical trends likely to be present in telephony over the years to come were foreshadowed by various developments in progress in the early 1960s. Foremost in the area of new art was the transistor (*q.v.*), invented by W. H. Brattain and John Bardeen and first announced in 1948. This is a three-electrode amplifying device employing a semiconductor such as germanium or silicon. An improved form, known as the junction transistor, was invented later by William Shockley. The transistor can perform many of the functions previously assigned to vacuum tubes but is far more efficient because it requires no power to heat a cathode. Furthermore, it operates with much lower electrode voltages and is much smaller than a vacuum tube. Because of the transistor's low power dissipation and low voltage requirements, other electronic components associated with it can be miniaturized. The over-all result is that the size, weight and power consumption of apparatus employing transistors usually can be reduced to a small fraction of that for equivalent apparatus using vacuum tubes.

Transistors therefore are finding large and varied applications in almost every field of electronics, and particularly in telephony and associated services, including not only places where vacuum tubes were previously used but many new situations as well. Thus transistors promised economies through more extensive use of amplification and carrier techniques in the local telephone plant.

In contrast with relays and other electromagnetic devices which require a minimum of several thousandths of a second to operate, electronic devices such as the transistor may be used to perform switching operations with a speed of the order of a few millionths of a second. Besides being faster, the electronic devices also are capable of performing more complex operations involving choices based on conditions existing at the instant. This means that no longer must a large complement of apparatus be set aside to serve a customer during the entire period of his call, or even a large part of it. The extremely high speed was expected to afford substantial economies in new switching systems through centralizing of functions and time sharing of the apparatus used for establishing connections. The common apparatus would serve a large number of customers in such rapid succession that each one received the equivalent of continuous and exclusive service. Reduction of space and power requirements would yield further economy. Altogether, the new art of solid-state electronics provided the basis for a revolution in both switching and transmission technology.

Research on the use of hollow wave-guide conductors as transmission lines promised new long-distance systems that would transmit extremely wide bands of frequencies, of the order of thousands of megacycles, to provide a great multiplicity of communication channels. Intercontinental submarine cables with amplifiers capable of handling large numbers of telephone channels promised to be of increasing importance to world communication. Experiments with radio communication via artificial satellites indicate that this method offers potentialities for providing wide transmission bands for communication across oceans. Such bands might be used for television or other visual presentations, or for telephone or data communication. There might be (1) a number of relatively low-altitude satellites (2,000–3,000 mi. high) each serving as a passive reflector, or (2) one or more satellites equipped for use as an active relay station. There is special interest in an active satellite placed about 22,300 mi. above the equator with an orbital velocity enabling it to remain above the same spot on Earth.

As to telephone service in general, there were, in the early 1960s, no indications of saturation in demand, even in highly developed areas. Substantial further growth in number of telephones and their utilization was therefore in prospect. Large expansion in services associated with telephony could also be envisioned. *See* also Index references under "Telephone" in the Index volume.

BIBLIOGRAPHY.—For an account of Elisha Gray's invention *see* Lloyd W. Taylor, "The Untold Story of the Telephone," *American Physics Teacher,* 5:243–251 (1937). General bibliographical references are as follows: Count T. A. L. du Moncel, *The Telephone, Microphone and Phonograph* (1879); S. P. Thompson, *Philipp Reis: Inventor of the Telephone* (1883); G. B. Prescott, *Bell's Electric Speaking Telephone* (1884); H. N. Casson, *The History of the Telephone* (1910); F. L. Rhodes, *The Beginnings of Telephony* (1929); William Aitken, *Who Invented the Telephone?* (1939); A. W. Page, *The Bell Telephone System* (1941); A. B. Clark, "The Development of Telephony in the United States," *Trans. Amer. Inst. Elect. Engrs.,* vol. 71, p. 348 (1952); F. E. Terman, *Electronic and Radio Engineering,* 4th ed. (1955); M. J. Kelly, Sir Gordon Radley, G. W. Gilman and R. J. Halsey, "A Transatlantic Telephone Cable," *Communication and Electronics,* vol. 17, p. 124 (March 1955); R. L. Peek, Jr., and H. N. Wagar, *Switching Relay Design* (1955). (E. I. G.)

TELEPHOTOGRAPHY: *see* FACSIMILE TRANSMISSION.

TELESCOPE, an optical instrument employed to view distant objects (Gr. *tele,* "far," *skopein,* "to see").

Early History.—The telescope was invented in Holland about 1608. The credit has been attributed variously to three individuals, Hans Lippershey and Zacharias Jansen, spectacle makers in Middelburg, and James Metius of Alkmaar (brother of Adrian Metius, the mathematician). The common story is that Lippershey, happening one day while holding a spectacle lens in either hand to direct them toward the steeple of a neighbouring church, was astonished on looking through the nearer lens to find that the weathercock appeared nearer. Accordingly he fitted the lenses in a tube to preserve their relative distance and thus constructed the first telescope.

Telescopes were made in considerable numbers and found their way over Europe soon after their invention. Galileo states that, happening to be in Venice about May 1609, he heard that a Belgian had invented a perspective instrument for making objects appear nearer and larger. The day after his return to Padua he made his first telescope by fitting a convex lens in one end of a leaden tube and a concave lens in the other end (*see* LENS). A few days afterward, having succeeded in making a better telescope,

he took it to Venice, where he communicated the details to the public, and presented the instrument itself to the doge Leonardo Donato. The senate, in return, settled him for life in his lectureship at Padua and doubled his salary, which previously had been 500 florins. Galileo devoted much of his time to improving the telescope. He conquered the difficulties of grinding and polishing the lenses, and soon succeeded in producing telescopes of greatly increased power. His first telescope magnified 3 diameters; he soon made instruments magnifying 8 diameters, and finally one magnifying 33 diameters.

With this last instrument Galileo discovered in 1610 the satellites of Jupiter, and soon afterward the spots on the sun, the phases of Venus, and the hills and valleys on the moon. He demonstrated the revolution of the satellites of Jupiter round the planet, and gave rough predictions of their configurations, proved the rotation of the sun on its axis, established the general truth of the Copernican system as compared with that of Ptolemy, and fairly routed the fanciful dogmas of the philosophers. These brilliant achievements, together with the immense improvement of the instrument under the hands of Galileo, overshadowed in a great degree the credit due to the original discoverer, and led to the universal adoption of the name of Galilean telescope for the form of instrument invented by Lippershey.

In the Galilean telescope the object glass is a convex lens and the eyepiece concave (*see* fig. 1). Johann Kepler was the first to explain the theory and some of the practical advantages of a convex eyepiece (*see* fig. 2) in his *Dioptrice* (1611). The first person who actually constructed a telescope of this form was the Jesuit, Christoph Scheiner, who gives a description of it in his *Rosa Ursina* (1630). William Gascoigne pointed out one great advantage of the form of telescope suggested by Kepler, viz., the visibility of the image of a distant object simultaneously with that of a small material object placed in the common focus of the two lenses. This led to his invention of the micrometer and his application of telescopic sights to astronomical instruments of precision. It was not till about the middle of the 17th century, however, that Kepler's telescope came into general use, and then, not so much because of the advantages pointed out by Gascoigne, but because its field of view was much larger than that of the Galilean telescope. The first powerful telescopes of this construction were made by Christiaan Huygens, after much labour, assisted by his brother. With one of these, of 12-ft. focal length, he discovered the brightest of Saturn's satellites (Titan) in 1655, and in 1659 he published his *Systema Saturnium*, in which was given for the first time a true explanation of Saturn's ring, founded on observations made with the same instrument. Because of the imperfections of the single objective lens it became essential to increase the focal length when high magnifying power was required. G. D. Cassini discovered Saturn's fifth satellite (Rhea) in 1672 with a telescope of 35 ft. and the third and fourth satellites in 1684 with telescopes made by M. Campani of 100- and 136-ft. focal length. Huygens states that he and his brother made object glasses of 170- and 210-ft. focal length, and he presented one of 123 ft. to the Royal Society of London. Adrien Auzout (d. 1691) and others are said to have made telescopes of from 300- to 600-ft. focus, but it does not appear that they were ever able to use them in practical observations. James Bradley, on Dec. 27, 1722, measured the diameter of Venus with a telescope whose object glass had a focal length of 212¼ ft. In these very long telescopes no tube was employed. They were termed aerial telescopes.

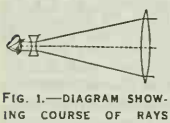


FIG. 1.—DIAGRAM SHOWING COURSE OF RAYS WHEN THE EYE LENS IS CONCAVE (OR NEGATIVE)

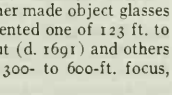


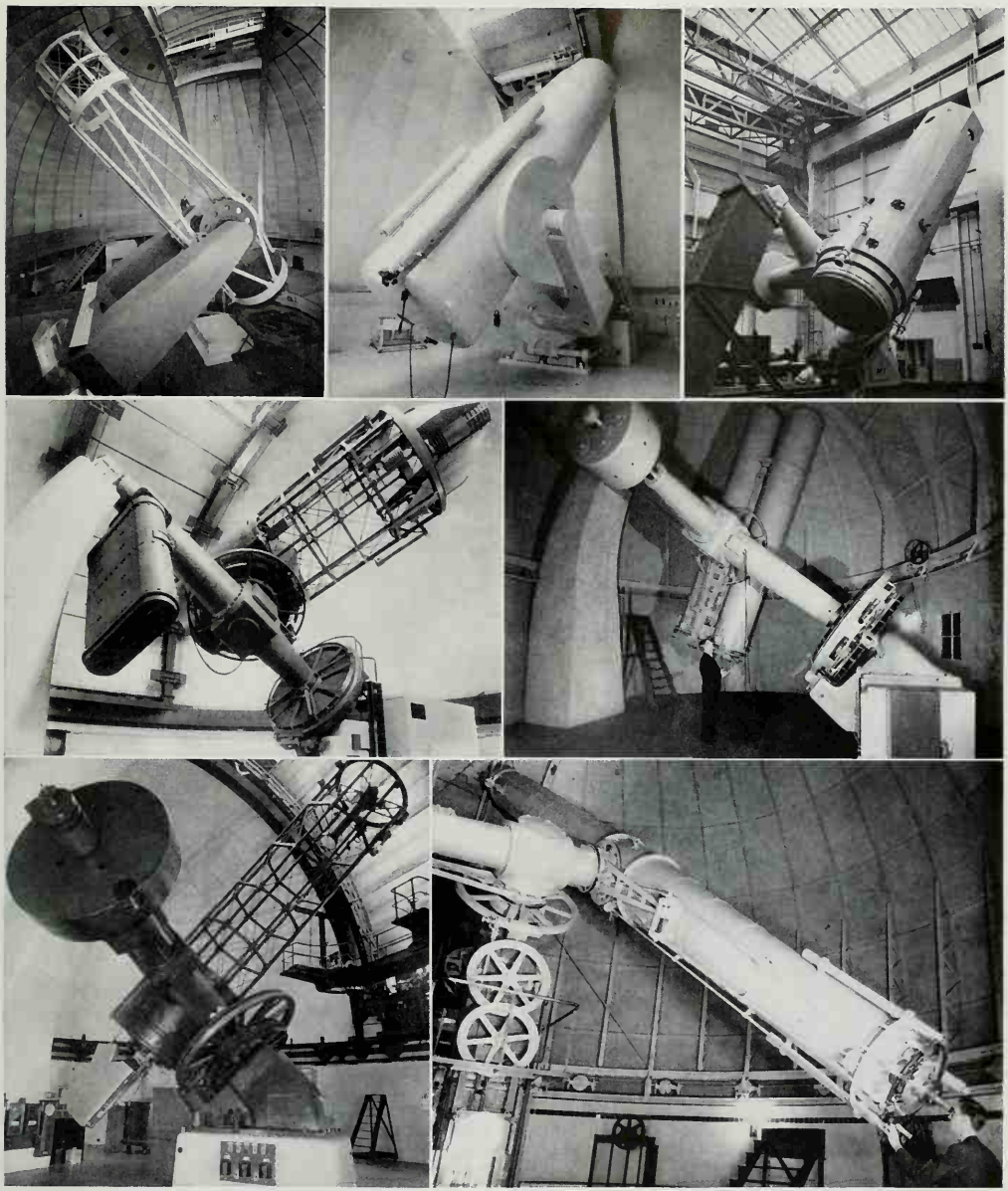
FIG. 2.—DIAGRAM SHOWING COURSE OF RAYS WHEN THE EYE LENS IS CONVEX (OR POSITIVE)

Early Reflecting Telescopes.—It was not until mid-18th century that these unwieldy instruments were supplanted by the achromatic telescope. Meanwhile the refracting type of telescope had a rival in the reflecting telescope invented by Sir Isaac Newton. It was, in fact, Newton who discovered what was the trouble with the refractor, which led to the need for excessive length. It had been supposed that the only imperfection in the image arose from the error known as spherical aberration, and the efforts of opticians were concentrated on devising lenses of suitable forms of curvature to correct this. In 1666 Newton discovered the different refrangibility of light of different colours, and he soon perceived that the fault of the refracting telescope was that the light of different colours followed different paths; so that if, for example, the telescope was focused sharply for blue light—the green image would be altogether out of focus and blurred. He overhastily concluded from rough experiments (*Optics*, book i. part ii, proposition 3) "that all refracting substances diverge the prismatic colours in a constant proportion to their mean refraction." If this were true no combination of refracting substances could bend the path of the light without introducing colour, and therefore no improvement could be expected in the refracting telescope. He therefore turned his attention to the construction of reflectors. The form later known as the Gregorian reflector had been proposed by James Gregory in 1663; but he had not succeeded in constructing the instrument practically.

Newton, after much experiment, selected an alloy of tin and copper for his specula, and he devised means for grinding and polishing them. He did not attempt the formation of a parabolic figure on account of the probable mechanical difficulties, and he had besides satisfied himself that the chromatic and not the spherical aberration formed the chief fault of previous telescopes. Newton's first telescope so far realized his expectations that he could see with its aid the satellites of Jupiter and the horns of Venus. Encouraged by this success, he made a second telescope of 6⅛-in. focal length, with a magnifying power of 38 diameters, which he presented to the Royal society in Dec. 1671. A third form of reflecting telescope was devised in 1672 by N. Cassegrain. No further practical advance appears to have been made in the design or construction of the instrument till the year 1723, when John Hadley (best known as the inventor of the sextant) presented to the Royal society a reflecting telescope of the Newtonian construction, with a metallic speculum of 6-in. aperture and 62⅝-in. focal length, having eyepieces magnifying up to 230 diameters. The instrument was examined by James Pound and Bradley, the former of whom reported upon it in *Philosophical Transactions*, 1723.

Bradley and Samuel Molyneux, having been instructed by Hadley in his methods of polishing specula, succeeded in producing some telescopes of considerable power, one of which had a focal length of eight feet; and, Molyneux having communicated these methods to Scarlet and Hearn, two London opticians, the manufacture of telescopes as a matter of business was commenced by them. However, it was reserved for James Short of Edinburgh to give practical effect to Gregory's original idea. Born at Edinburgh in 1710 and originally educated for the church, Short attracted the attention of Colin Maclaurin, professor of mathematics at the university, who permitted him about 1732 to make use of his rooms in the college buildings for experiments in the construction of telescopes. In Short's first telescopes the specula were of glass, as suggested by Gregory, but he afterward used metallic specula only, and succeeded in giving to them true parabolic and elliptic figures.

Achromatic Telescopes.—The historical sequence of events now brings us to the discovery of the achromatic telescope. The first person who succeeded in making achromatic refracting telescopes seems to have been Chester Moor Hall, of Essex, Eng. He argued that the different humours of the human eye so refract rays of light as to produce an image on the retina which is free from colour, and he reasonably concluded that it might be possible to produce a like result by combining lenses composed of different refracting media. After devoting some time to the inquiry he found that by combining lenses formed of different kinds of glass the effect of the unequal refrangibility of light was corrected, and in 1733 he succeeded in constructing telescopes which exhibited objects free from colour. The principal development of the achromatic refractor is due to John Dollond who invented it independently (*Phil. Trans.*, 1758). In principle his object glasses

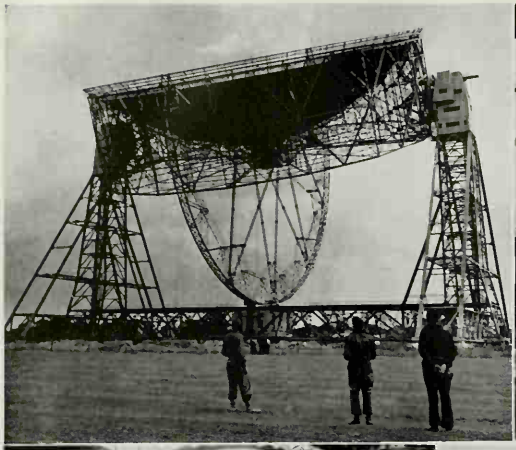


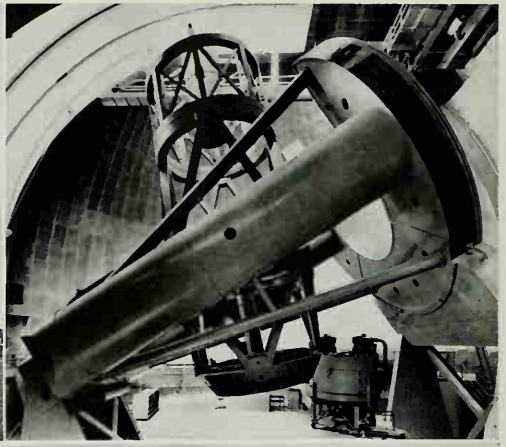
BY COURTESY OF (TOP LEFT, CENTRE RIGHT) LICK OBSERVATORY, (TOP CENTRE) MOUNT WILSON AND PALOMAR OBSERVATORIES, (TOP RIGHT) SIR HOWARD GRUBB PARSONS AND CO., LTD., (CENTRE LEFT) YERKES OBSERVATORY, (BOTTOM LEFT) DOMINION ASTROPHYSICAL OBSERVATORY, (BOTTOM RIGHT) U.S. NAVY

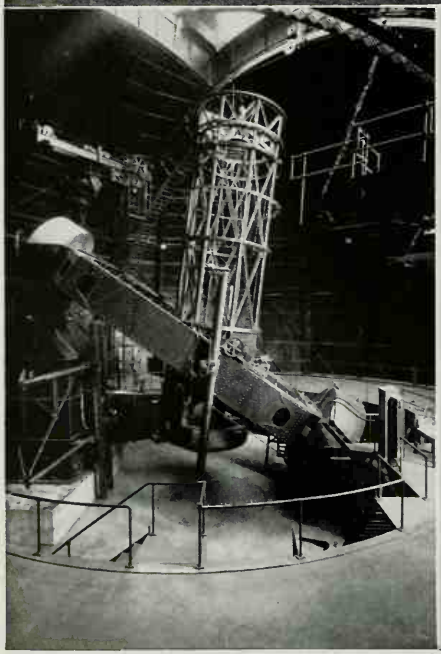
TELESCOPES IN LARGE OBSERVATORIES

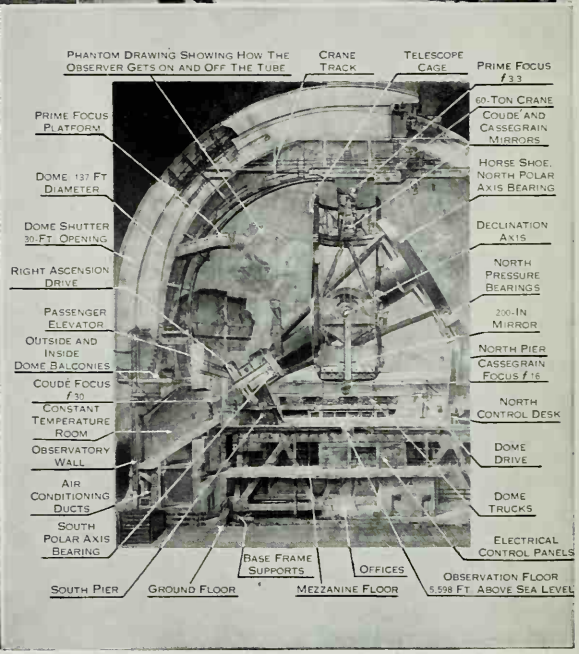
Top left: 120-in. reflector, Lick observatory, Mt. Hamilton, Calif.
Top centre: 48-in. Schmidt telescope, Palomar observatory, Calif.
Top right: 75-in. reflector, Astrophysical observatory, near St. Michel, France
Centre left: 82-in. reflector, McDonald observatory, Ft. Davis, Tex.
Centre right: 20-in. astrographic telescope, Lick observatory
Bottom left: 72-in. reflector, Dominion Astrophysical observatory, Victoria, B.C.
Bottom right: 26-in. refractor, U.S. Naval observatory, Washington, D.C.

TELESCOPE









BY COURTESY OF (TOP LEFT) BRITISH INFORMATION SERVICES, (TOP RIGHT, BOTTOM LEFT) MOUNT WILSON AND PALOMAR OBSERVATORIES, (BOTTOM RIGHT) R. W. PORTER AND THE CALIFORNIA INSTITUTE OF TECHNOLOGY

THREE OF THE WORLD'S LARGEST TELESCOPES

Top left: Radio telescope at Jodrell Bank Experimental station at the University of Manchester, Eng. The telescope, completed in 1957, has a 250-ft. mirror aperture

Top right: The 200-in. Hale reflector telescope at Palomar observatory, Calif.

Bottom left: The 100-in. Hooker telescope with Cassegrain spectrograph, Mount Wilson observatory, Calif.

Bottom right: Sectional drawing of the 200-in. Hale telescope. The drawing shows the telescope cut away along a vertical plane through the polar axis, with explanatory notes giving essential features of the instrument. The tube weighs about 140 tons, the entire instrument 500 tons, and the rotating dome 1,000 tons

were of the pattern mainly used at the present day, viz., convex lens of crown glass combined with a concave lens of flint glass. The concave lens is of less power than the convex, so that the combination converges the light as a single convex lens would do; but the flint glass having much wider difference of refractive index for light of different colours, is able to correct the colour dispersion introduced by the more powerful crown lens. A triple objective, consisting of two convex lenses of crown glass with a concave lens of flint glass between them, was introduced in 1765 by Peter, son of John Dollond. This type is also employed in some modern telescopes. The difficulty of making large transparent and homogenous disks of glass at first limited the use of these achromatic lenses to small telescopes for terrestrial use.

Later Improvements.—Late in the 18th century William Herschel greatly advanced the art of making mirrors of speculum metal, a hard brittle alloy having a composition of one part tin and two parts copper and a reflectivity of about 60%. Herschel's telescopes were much in demand for they were of better quality and larger than any previously made. His two most famous telescopes were the 20-ft. of 19-in. aperture and the great 40-ft. of 48-in. aperture completed in 1789. The next step in the improvement of the reflector was made by Lord Rosse who built a series of telescopes culminating in the Parsonstown reflector of 72-in. aperture in 1845. The refracting telescope was brought back into favour in the first quarter of the 19th century after Joseph von Fraunhofer, who had learned from the younger Guinand how to make glass, succeeded in casting several disks of glass of excellent quality and in making lenses from them, computing the curves himself. The desire for greater magnification and light-gathering power led to the construction of lenses of larger and larger aperture. The U.S. optician Alvan Clark was especially successful in figuring large lenses and produced a long series of instruments culminating in the 36-in. of the Lick observatory and the 40-in. of the Yerkes observatory, which was completed in 1897. Disks of optical glass of so great a size, free from strain, striations, veins and bubbles and other troublesome defects, are extremely difficult to cast and anneal. The great thickness of glass required for these very large lenses absorbs an appreciable fraction of the light and reduces their effectiveness as light collectors. These and other reasons resulting from the shift in observing techniques discussed below caused a change back to the use of reflectors early in the 20th century.

In the meantime many of the disadvantages of the speculum metal mirrors had been eliminated by the development of the silver-on-glass mirror. The discovery of a simple process by which a thin film of silver could be deposited on a glass surface was made by Justus von Liebig; the application of this discovery to glass mirrors and the methods of making and testing such mirrors was carried out by Jean Foucault in the 1860s. A freshly prepared silver film has a high reflectivity exceeding 90% throughout the visual spectrum. For wave lengths shorter than the blue the reflectivity falls at first slowly and then rapidly to a value of about 4% in the ultraviolet at 3,200 Å.

In the 1930s a process was developed by John Strong for coating glass mirrors with a film of aluminum by a vacuum distillation process. On exposure to the air the surface of the aluminum film rapidly oxidizes, giving a hard and durable reflecting layer. The reflectivity of the aluminum film is nearly as high as that of a freshly deposited silver film in the visual and does not have the large drop in reflectivity in the ultraviolet. The aluminum film deteriorates much less rapidly than the silver, requiring replacement every five or ten years while the silver must be renewed every few months. With these advantages aluminum almost completely replaced silver for coating large astronomical mirrors.

In order that the mirror shall not bend under its own weight and destroy the precision of the optical surface the glass disk must be very thick, usually about one-eighth of its diameter. Furthermore, to reduce the distortion of the mirror by changes in temperature, the present practice is to use mirrors of fused quartz or of a glass, such as Pyrex, that has a low coefficient of thermal expansion. The disk is ground to the rough shape required and then figured to the exact paraboloid or other surface specified. Since the final figure must not depart from the theoretical surface by more than a very few millionths of an inch this is a long slow process; short periods of polishing being interspersed with careful optical tests to determine where the next polishing should be done and how much glass should be removed at each point.

After a precise surface has been achieved a layer of silver or aluminum is deposited to complete the mirror. The layer of metal is so thin that it does not affect the accuracy of the surface. When the surface tarnishes, the old film can be removed chemically and a new film deposited without injuring the figure of the glass.

(D. Gr.; A. S. E.; H. S. Js.; A. B. Ml.; I. S. B.)

THEORY OF THE TELESCOPE

Magnifying Power.—The telescope consists essentially of two parts, the objective and the eyepiece. The light first passes the objective which forms an image, similar to that appearing on the ground glass of a camera. In small terrestrial telescopes the objective is an achromatic lens; in astronomical telescopes it may be either a lens or a concave mirror. The eyepiece is a short focus lens, like a small hand magnifier, which is used to look at and magnify the image formed by the objective. The magnifying power of a telescope is equal to the focal length of the objective divided by that of the eyepiece.

Image Brightness.—In addition to the magnifying power other factors that determine the performance of a telescope are image brightness, resolving power and field of view. In general a beam of light whose diameter is equal to that of the objective enters the telescope. The diameter of the beam leaving the telescope and entering the eye is equal to the objective diameter divided by the magnification. If the diameter of this beam, or emergent pencil as it is called, is greater than the diameter of the pupil of the eye, part of the light fails to enter the eye and is wasted. On the other hand, if the emergent pencil is much smaller than the pupil of the eye the apparent brightness of the image will be much lower than that of the object and the image will appear dark and indistinct.

The diameter of the pupil of the eye, however, varies through a large range depending on the illumination. Thus in full daylight the diameter of the pupil is 1.5 to 2 mm. while in darkness it expands to 8 or 10 mm. It is for this reason that telescopes designed primarily for nighttime observations have an exit pupil of at least 7 mm.; *e.g.*, the 50-mm., 7-power binocular used widely by the U.S. navy. An objective aperture of about one-fourth this would be adequate for daytime observation with the same power. In many general-purpose binoculars, however, a compromise emergent pencil of 3 to 5 mm. diameter is used as in the case of the 6- or 8-power glass with a 25- or 30-mm. objective.

Resolving Power.—The resolving power of a telescope is defined as the inverse of the smallest angle by which two points must be separated in order that they appear double in the telescope. From the wave theory of light it can easily be proved that the resolving power is proportional to the aperture of the telescope divided by the wave length of the light. If the magnification is increased to too high a value relative to the aperture of the objective the object appears blurred and no additional details can be detected. In general, no gain in observable detail is achieved by pushing magnification appreciably higher than the aperture of the objective expressed in millimetres.

Field of View.—The field of view of a telescope is defined as the angular diameter of an object that just fills the field that can be seen at one time. The apparent field of view is the angle subtended by the image of this object at the eye and is equal to the field of view times the magnification. The size of this apparent field of view is fixed primarily by the design of the eyepiece. The negative lens used as an eyepiece in the Galilean telescope gives a very small apparent field, ranging from 30° in a low-power opera glass to 5° in a telescope of high power. For this reason this type of eyepiece is used only in very low-power opera glasses and in cheap field glasses. The eyepiece consisting of a single positive lens suggested by Kepler yields a larger field but the definition is poor except at the centre of the field.

The first important improvement in eyepieces was made by

Huygens who introduced a design having two lenses separated by a distance equal to the average focal length of the component lenses. The Huygenian eyepiece which gives an apparent field of about 45°, is still used, especially for the lower powers. The best instruments of the 20th century use the orthoscopic eyepiece which consists of four lenses cemented in two groups and which yields critical definition even in the higher powers over an apparent field of 40° to 50°. Wide-field binoculars and some military instruments use the Erfle eyepiece which consists of up to six glass elements and yields an apparent field of 65° to 70°.

Erectness of the Image.—For terrestrial observations the question of the erectness of the image must be considered. The negative eyepiece of the Galilean telescope yields an erect image. All of the other eyepieces mentioned above give inverted images if used alone with the objective. Erection of the image may be attained for the other types of eyepiece by the introduction of a third lens system between the objective and the eyepiece. This is the procedure used in the common draw tube telescope, although the introduction of the additional lens limits somewhat the field of good definition. The best instruments erect the image with two erecting prisms commonly of the Porro type (*see* BINOCULAR INSTRUMENT). This type permits the effective use of even the wide-angle Erfle eyepiece mentioned above.

ASTRONOMICAL TELESCOPES

Photographic and Photoelectric Observation.—By the end of the 19th century the photographic plate had been developed to the point that it had many advantages over the eye for astronomical observations. The first of these is greater sensitivity for the study of faint objects such as nebulae or galaxies. Thus the photographic plate records fainter and fainter objects the longer it is exposed, while the eye sees nothing additional after the first glance. A few minutes exposure of the plate records everything the eye can see with the same instrument but the exposure may be continued for hours. No eye has ever seen the detail recorded on most astronomical photographs of nebulae, galaxies and other faint objects.

A second advantage of the photograph is the permanent record which it provides and which may be studied at leisure in the laboratory. The information recorded on a single plate may require months of measurement to evaluate. Such a photograph is often of value years later to determine the initial state of a star that afterward explodes into a nova, or to locate the position of a planet that later calculations show passed the region covered by the plate at the time of its exposure.

Finally, modern photographic plates are sensitive to a much wider range of wave lengths than the eye.

Many other astronomical problems require the precise measurement of the brightness of an object with a photoelectric photometer or the analysis of its light with a spectrograph. For this purpose the objective of the telescope is used like a condensing lens or burning glass to concentrate the light of a star or other object on the aperture of the photometer or slit of the spectrograph.

Because of these developments visual observations are practically never made with the large modern telescopes used for fundamental research in astronomy. Strictly speaking these instruments are not telescopes at all but cameras or light-condensing systems. Most such telescopes have provision for the insertion of an eyepiece in place of the plateholder and this is occasionally done to view the field and make certain the telescope is pointed at the desired object before starting a long exposure. Another eyepiece is often mounted beside the plateholder to observe a guide star just outside the field being photographed. This is checked from time to time during the exposure and the position of the telescope adjusted to compensate for slight errors in the telescope drive. However, since practically all actual observations are made photographically or photoelectrically the telescopes are designed to give them maximum effectiveness as cameras or light condensers rather than as visual telescopes.

(*See* PHOTOGRAPHY, CELESTIAL; PHOTOMETRY, CELESTIAL; SPECTROSCOPY, ASTRONOMICAL.)

Change in Design.—This shift in the method of observation early in the 20th century resulted in a radical change in telescope design. The photography of very faint objects requires an objective of great speed, that is, one with a large ratio of aperture to focal length, *f*. Many problems require achromatism for the full range of plate sensitivity from the ultraviolet to the infrared. The two-lens objective of the 19th-century telescope with its lack of achromatism except for a narrow range of wave lengths in the visual and its aperture of *f*/15 to *f*/20 is quite inadequate for this method of observation. The concave paraboloidal mirror, however, is completely achromatic and gives critical definition at apertures as large as *f*/3. As a result practically all large telescopes constructed after 1900 were of the reflector type using concave mirrors with apertures between *f*/3 and *f*/6. (*See* fig. 3–5.)

The reflecting telescope introduced several new problems of design. Thus a concave mirror (AA in fig. 3–5) focuses the image at a point (F) in the centre of the incoming beam of light. This is usually known as the prime focus. Except in one or two of the largest telescopes an observer placed at this focus to load plates and manipulate the telescope would obstruct a very large fraction of the incoming light. Most reflecting telescopes, therefore, have a small 45° plane mirror (BB in fig. 3) mounted a short distance inside the focus to reflect the image to a point just outside the incoming beam. Since this arrangement was originally proposed by Newton, this is known as the Newtonian focus.

For some purposes a longer focal length and therefore smaller ratio of aperture to focal length may be desirable. This is obtained by placing slightly inside the focus a convex mirror (BB

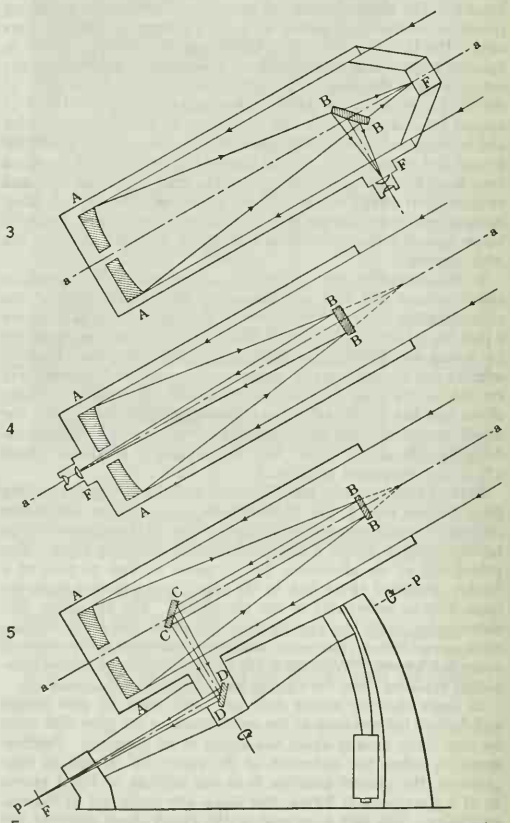


FIG. 3–5.—(3) PRIME FOCUS AND NEWTONIAN TELESCOPE; (4) CASSEGRAIN TELESCOPE; (5) COUDÉ TELESCOPE

in fig. 4) which reflects the light back through a hole in the centre of the main mirror and forms the image just behind this mirror. Because of the telephoto effect of this convex mirror, the focal length is increased by a factor of 3 to 8, and the ratio of aperture to focal length is decreased by the same factor. From the inventor of this arrangement the focus is known as the Cassegrain focus.

Finally in the coudé (French, "elbow") arrangement the light from a Cassegrain mirror is reflected down the polar axis by one or more plane mirrors (BB, CC and DD in fig. 5). In this case the focus, usually a short distance below the south polar bearing, remains fixed as the telescope is turned toward various parts of the sky. This arrangement can therefore be used to condense light on the slit of a spectrograph or other instrument too large to be mounted on the telescope.

Wide-Field Cameras.—For many investigations it is necessary to photograph large areas of the sky at one exposure. This requires a field of good definition much larger than that provided by the simple achromat of the refracting telescope or by the paraboloidal reflector, especially when this is used at the low focal ratios needed for the photography of faint objects. For these studies complicated three- or four-lens systems, very similar to the anastigmat lenses for ordinary photography, were originally designed. These have been constructed with apertures up to 20 in.

In 1931 Bernhard Schmidt of the Hamburg observatory, Bergedorf, Ger., invented an entirely new optical system for use as a wide-angle camera. It consists of a concave spherical mirror and a thin correcting plate placed at the centre of curvature of the mirror. The very superior corrections of this system compared with that of any lens system made it possible to construct wide-angle cameras operating at apertures greater than $f/1$. Others have been constructed with an aperture as great as 53 in. The chief problem in the manufacture of the Schmidt camera is the corrector plate that has to be hand figured to a complicated fourth degree curve which is much more difficult than the spherical surfaces of the ordinary lens.

Mounting.—The mirror and photographic plate, spectrograph or photometer must be mounted on a structure that can be pointed to objects at any place in the sky. This structure must also be provided with a drive mechanism so that once pointed at an object, the telescope can be turned to follow the object as it moves across the sky from east to west because of the earth's rotation. In order that no blurring of the image occurs it is necessary that this structure be very rigid so that it does not deform as it changes direction during an exposure of many minutes or even hours.

In typical mountings designed for modern telescopes, the main mirror is mounted at the lower end of the structure while for direct photography the plate is placed at either the prime or Newtonian focus near the upper end. In the Cassegrain or coudé arrangements the convex mirror is at the upper end.

In general, a structure can be pointed in all directions if two mutually perpendicular axes of rotation are provided. In very small visual telescopes these axes are often made vertical and horizontal in what is known as an altazimuth mount. For telescopes that are to be driven to follow the object across the sky, the equatorial mount is used. In this mount one of the axes, known as the polar axis, is made accurately parallel to the axis of the earth. The axis perpendicular to this is known as the declination axis. This type of mounting has the great advantage that any object can be followed from east to west by driving the polar axis at the uniform rate of one revolution in 24 hours.

Three different types of equatorial mounting were developed. (1) The English off-axis mount in which the polar axis is a relatively small tube with two bearings, one at the north and the other at the south. The telescope tube is mounted on one side with a counterweight on the other side of this polar axis. (2) The fork type has the telescope mounted between the tines of a fork projecting out from the two polar axis bearings both of which are at the south. (3) The yoke mounting holds the telescope tube symmetrically in a yoke, the polar axis bearings of which are at the north and south. This type provides the most rigid support for the telescope and is therefore used for the largest instruments. Its disadvantage is that either the telescope cannot be pointed to the north pole or the north bearing must be made so large that the telescope can be turned down into it.

The Hale 200-in. Telescope at Palomar Observatory.—As of the 1960s the largest telescope constructed was the 200-in. Hale reflector on Palomar mountain in California. The main mirror is 200 in. in diameter and 24 in. thick and is ground and figured to give a focal length of 55 ft. To reduce thermal distortion during temperature changes it was cast from Pyrex glass with a ribbed structure and when finished weighed 14½ tons. In order to avoid flexure as the mirror is turned into different directions it is floated on 36 elaborate lever support systems each of which applies a force to the mirror just sufficient to compensate for the pull of gravity on the section of the mirror assigned to it. The mirror and its support system are mounted at the bottom of the telescope tube which is a rigid girder structure about 60 ft. long and with the mirror weighs 140 tons. At the centre of the upper end of the telescope tube is the prime focus cage in which provision is made for mounting the plateholder, photometers and small spectrographs. While observing at this focus the astronomer rides in the cage, six feet in diameter, which provides ample working space but intercepts only 13% of the light approaching the main mirror. Immediately below the observer's cage is a second cage in which are mounted the various secondary mirrors for the Cassegrain and coudé arrangements. When needed these mirrors are tipped forward into operating position by remote control.

The telescope tube is mounted between the arms of a yoke on declination bearings which permit the telescope to turn in a north-south direction. The yoke in turn is supported at both the north and south ends on the polar axis bearings which allow rotation of both telescope tube and yoke about an axis parallel to the earth's axis. The north polar bearing is made in the shape of an open horseshoe, the outer bearing surface having a diameter of 46 ft. This allows the upper end of the telescope tube to turn down to the centre of the bearing and thereby point at the north pole. The total weight of the telescope tube and yoke which turns about the polar axis is over 500 tons. Both north and south bearings are of the pressure-oil-pad type and have so low a friction that a ½-h.p. motor is sufficient to drive the telescope when following a star.

At the Cassegrain focus the focal length is 267 ft. and at the coudé focus it is 500 ft. In addition to the photoelectric photometers and small dispersion spectrograph for use at the prime focus, a large spectrograph is available for use at the coudé focus. This has a collimator of 30-ft. focal length and five Schmidt cameras of aperture 12 in. and focal lengths from 8 in. to 12 ft.

The telescope is housed in a dome 135 ft. high and 137 ft. in diameter. The dome is rotated with a drive which automatically keeps the shutter opening in the dome lined up with the telescope at all times. The rotating part of the dome weighs 1,000 tons. (*See* also OBSERVATORY [ASTRONOMICAL].)

Large Telescopes.—The largest telescopes in operation in the mid 1960s are listed in the table. They are all reflectors, since no refractor with an aperture greater than 30 in. was built after the 19th century.

Large Telescopes

Name	Location	Aperture (in inches)	Focal length of primary	Date completed
Hale	Palomar Mtn., Calif.	200	660	1948
Lick	Mt. Hamilton, Calif.	120	600	1959
Crimean	Crimea, U.S.S.R.	102	395	1960
Hooker	Mt. Wilson, Calif.	100	508	1917
Kitt Peak	Kitt Peak, Ariz.	84	220	1963
McDonald	Mt. Locke, Tex.	82	324	1939
Haute Provence	St. Michel, France	77	385	1958

See also BINOCULAR INSTRUMENT; TELESCOPE, RADIO; TRANSIT CIRCLE and ZENITH TELESCOPE for treatment of special types.

BIBLIOGRAPHY.—H. C. King, *The History of the Telescope* (1955); for construction details and theory: G. R. Miczaika and W. M. Sinton, *The Tools of the Astronomer* (1961); G. P. Kuiper and B. M. Middlehurst (eds.), *Telescopes* (1960). Specific telescopes are described in the publications of the observatories operating them. (I. S. B.)

TELESCOPE, RADIO. After World War II, the applica-

tion of radio techniques to the study of extraterrestrial objects contributed greatly to our knowledge of the universe. The growth of radio astronomy was such that the number of radio observatories promised to equal the number of optical observatories, and the number of astronomers making observations in the radio spectrum to equal that of the astronomers working in the visual range. However, the radio telescope will never replace the optical telescope; each has its place in the study of different phenomena, and in many cases the observations are complementary.

The range of wave lengths over which radio observations of astronomical value can be made covers 11 octaves from about 1 cm. to 30 m. The short-wave-length limit is due to absorption in the lower atmosphere of the earth and the long-wave-length limit is set by absorption in the ionized layers of the upper atmosphere at heights above 100 km. Both limits vary somewhat due, for instance, to the water vapour content near the earth's surface and the influence of solar activity on the upper atmosphere.

The radio waves from celestial objects have three broadly distinct origins. One type is thermal radiation, in which the intensity of the radiation is simply related to the temperature of the object, whether it be solid (as for example a planet) or gaseous (as for example the outer envelope of the sun or an emission nebula surrounding a hot star). A second type appears to be generally associated with tenuous gas in violent motion. Examples of this type are the radiation caused by streams of particles shot through the sun's atmosphere at very high speed, the radiation from the remnants of exploded stars or supernovae and the radiation produced by the interaction of the interstellar gas of two galaxies in collision. The third type is the line radiation from the cold neutral hydrogen in interstellar space. This hydrogen is the raw material from which stars are formed and which supplies their energy through nuclear reactions. This material can only be studied by means of the radio emission which is concentrated in a narrow band of wave lengths about 21 cm. The actual wave length on which the radiation is received from some local region of space depends on the relative motion between that region and the earth; thus the measurement of the wave length provides a method of determining velocities.

A complete radio installation consists of two parts, the telescope or antenna which collects the radiation and a receiver which amplifies and processes it in some way. The functions of the two parts are not always completely separate. In most systems the direction in which a telescope is pointed is achieved by mechanical movement of the whole structure as in the case of optical telescopes. In some radio telescopes, however, the direction is governed by changes of electrical phase or wave length in the associated electronic equipment.

The receiving equipment can have many forms and functions. It may, for example, be designed to analyze the spectrum of a burst of radiation from the sun lasting only a fraction of a second or it may serve to integrate for several hours to produce a detectable signal from the weak radiation of some distant galaxy.

The over-all performance of an astronomical instrument can be expressed in terms of three criteria: (1) the sensitivity which depends on the collecting area and the efficiency of the processing device; (2) the ability to control the range of wave lengths accepted; and (3) the resolving power. Radio instruments are in general more efficient than optical in their ability to process the radiation for presentation. Very little of the signal is lost between collection by the telescope and presentation, unlike the case of an optical telescope and photographic plate where about nine-tenths of the incoming light is lost in mirror reflections and the plate. An optical telescope-photographic plate combination has the advantage that it can accept a considerably greater range of wave lengths relative to the mean wave length than any radio instrument. Radio instruments, on the other hand, can be made to accept a considerably smaller range than any optical combination. The wide spectral range of the optical instrument is of value in obtaining high sensitivity in studies of continuous radiation; radio instruments of small spectral range are of value in studies of line radiation and in any observation where a precise definition of wave lengths is required.

In the third criterion of resolving power, any radio telescope falls a long way short of optical instruments. Resolving power which determines the ability of an instrument to distinguish apart two close objects, or to define precisely the position or size of a source of radiation, depends on the ratio of the wave length to the aperture of the telescope. A two-inch telescope in the optical range has a resolving power of 10^{-5}, or about 2″ of arc. To produce the equivalent resolving power at the short-wave-length end of the radio spectrum would require an aperture of one and a half miles. The earth would hardly be large enough to accommodate a similar instrument for a wave length of 30 m. The 50-ft. diameter radio telescope at the Naval Research laboratory in Washington, D.C., has at present the highest resolving power of any single radio telescope. Its highly accurate surface permits observations at wave lengths as low as eight millimetres where the resolving power is some 3′ of arc.

As a consequence of the low resolving power, the radio astronomer has to be satisfied in general with a relatively ill-defined or blurred picture of the universe. Much of his efforts are devoted to obtaining high resolving power in specific cases by trick methods, and it is here that his tight control over the electronic processing of the radiation shows to advantage.

The Simple Interferometer.—One of these methods is interferometry. The simple interferometer (*q.v.*) consists of two antenna elements connected together by cables or other means to a control receiver. The individual elements may consist of a half-wave hertzian dipole, with or without reflector, a Yagi (parasitic array), a helix or a parabolic reflector. The choice depends partially on the wave-length range, the collecting area and the desired polarization. Parallel rays from a source effectively at infinity arrive at the two antenna elements and therefore at the receiver with a certain path difference which changes with the angle between the direction of the source and the normal to the base line (fig. 1). If the path difference is an integral number of wave lengths the two signals interfere constructively to give twice the signal from a single element. If the path difference is an odd number of half wave lengths the two signals interfere destructively to give a zero result. As the earth rotates, a radio source passes through alternate maxima and minima of the interference pattern and a sinusoidal variation is recorded at the output of the receiver (fig. 2[B]).

Interference techniques played a valuable part in the early studies of the sun and radio stars, enabling the experimenter to distinguish relatively weak point sources in a large total background of noise from all directions on the celestial sphere (and the receiver itself) with relatively simple antenna systems. In Australia interferometers were used which consisted of a single antenna situated on a cliff edge and its mirror image in the sea. The resolving power of an interferometer depends on the ratio of the wave length to twice the base line or distance between the sepa-

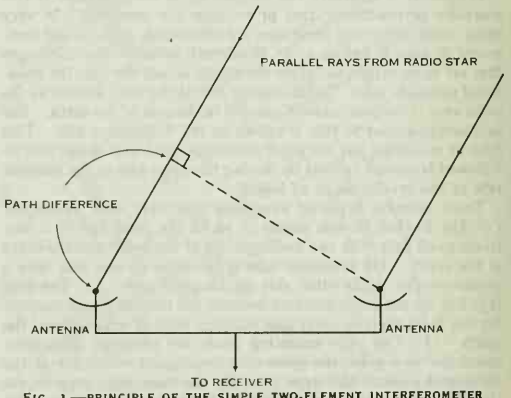


FIG. 1.—PRINCIPLE OF THE SIMPLE TWO-ELEMENT INTERFEROMETER

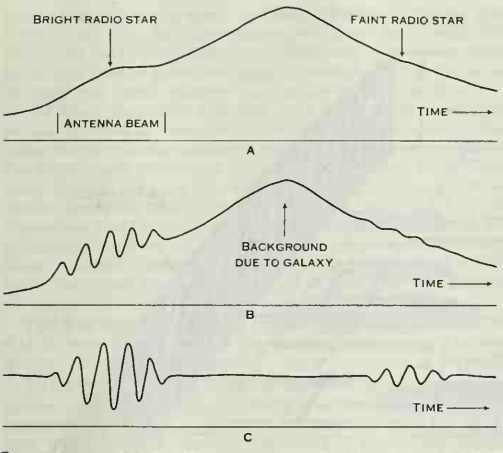


FIG. 2.—TYPICAL OBSERVATION WITH (A) A SINGLE ANTENNA ELEMENT RECORDING THE PASSAGE OF RADIO STARS AND THE GALACTIC BACKGROUND ACROSS ITS BEAM; (B) AN INTERFEROMETER CONSISTING OF TWO ANTENNA ELEMENTS; (C) A TWO-ELEMENT INTERFEROMETER AND THE PHASE SWITCHING TECHNIQUE. ONLY THE RADIO STARS ARE RECORDED

rate elements. With the use of interferometers the positions of some radio stars have been determined with accuracies of a fraction of a minute of arc. Determination of such positions involves the measurement of the time of the interference pattern maxima and minima. Measurements have also been made of the angular diameters of sunspots and radio stars. Such measurements require observations of the object with interferometers of different base lines.

The visibility or amplitude of the interference pattern quickly approaches zero as the angular separation of the interference fringes approach the diameter of the source. For some objects, base lines of up to 30 mi. have been necessary and in such cases radio links are used to transmit the signals to a central receiver.

Multielement Interferometers and Crosses.—More sophisticated than the simple interferometer are the multielement interferometers, which are analogous to optical diffraction gratings. The addition of further antenna elements has the effect of suppressing some of the lobes or fringes of the two-element pattern. An early example of this type of interferometer is the W. Christiansen antenna used to obtain the radio brightness distribution over the surface of the sun. This array has 32 elements. Its reception pattern consists of a number of knife-edge beams through which the sun passes in turn. The beams are about 3′ of arc wide and are separated by about 2°.

One of the problems in the early interferometers was that the interference pattern due to the passage of a weak source through the fringes was observed superimposed on a varying general background caused by the passage of sections of the galaxy across the primary acceptance pattern of the individual elements (fig. 2[B]). This problem was solved by M. Ryle and the Cambridge radio astronomers who devised the phase switching technique. Here the fringe pattern is rapidly switched half a fringe width and the difference between the outputs of the receiver for the two fringe positions recorded. In this way background effects which are virtually the same in the two switched positions are eliminated (fig. 2[C]). At the same time Ryle pointed out the value of using interferometer elements of different physical size and beam pattern where interference can only occur in areas which are common to the two primary beams.

The next logical step was made by B. Y. Mills in Australia in the design of antenna which bears his name—the Mills cross. The original cross, designed for a wave length of about four metres, consists of two arms aligned east-west and north-south. Each arm has 250 individual dipole and reflector elements which combine to give an acceptance pattern about $\frac{1}{2}°$ wide in one plane (fig. 3). The transmission lines from the two arms of the cross are alternately switched in and out of phase at the input to the receiver and the difference recorded at the receiver output. This recorded output only contains signals from objects in the area common to the two beams and in effect produces a pencil beam (shown in black in fig. 3).

The beam can be directed in either plane by changing the phases of the individual dipoles in one of the two arms. The Mills cross is a radio telescope which has a collecting surface equal only to its physical area, but has a resolving power almost that of fully filled square array of sides equal to the lengths of the arms.

Mills crosses have been built in a variety of wave lengths between 20 cm. and 15 m.

There are a number of variations on the original cross principle, including one which combines the advantages of the Christiansen grating and the Mills cross. This instrument has a resolving power in both directions of about 3′ of arc and can make a televisionlike scan of the sun in a matter of an hour. The Cambridge group has devised an ingenious system which they call aperture synthesis, applicable to both simple crosses and interferometric crosses. Here observations are made with a single complete cross arm and a smaller antenna element which occupies in turn various positions in the missing arm. The results are combined in an electronic computer to produce the equivalent observations to a full cross in all the possible beam positions.

Radio telescopes of the interferometric and cross type have proved extremely valuable for specific investigations, particularly at the longer wave lengths. However, they are unsatisfactory as general purpose instruments; the advantages of ease of construction and low cost are offset by disadvantages such as the restriction to a small wave length range and change of beam shape with beam position.

For many investigations, for example the study of the distribution of the continuous and 21-cm. line radiation from the galaxy, a reflector is necessary. Where very small signals are to be measured, long integration times are required and this means the reflector must be steerable.

Steerable Reflectors.—The designer of a large reflecting radio telescope faces at least one problem that is not present in the case of an optical telescope and that is the wind loading. Radio telescopes are far too large to be enclosed in protective domes and thus must be built to withstand the destructive loads imposed by high winds and ice.

In order to minimize the wind load on drive gears, etc., it is

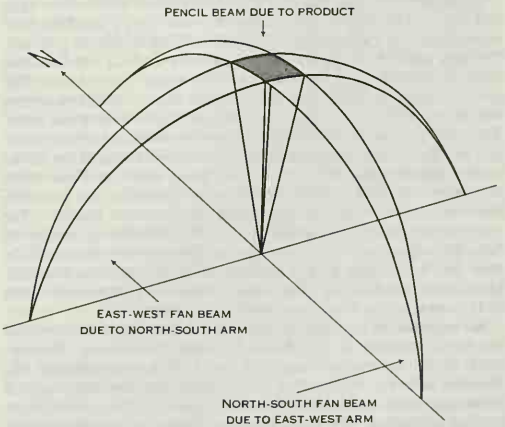
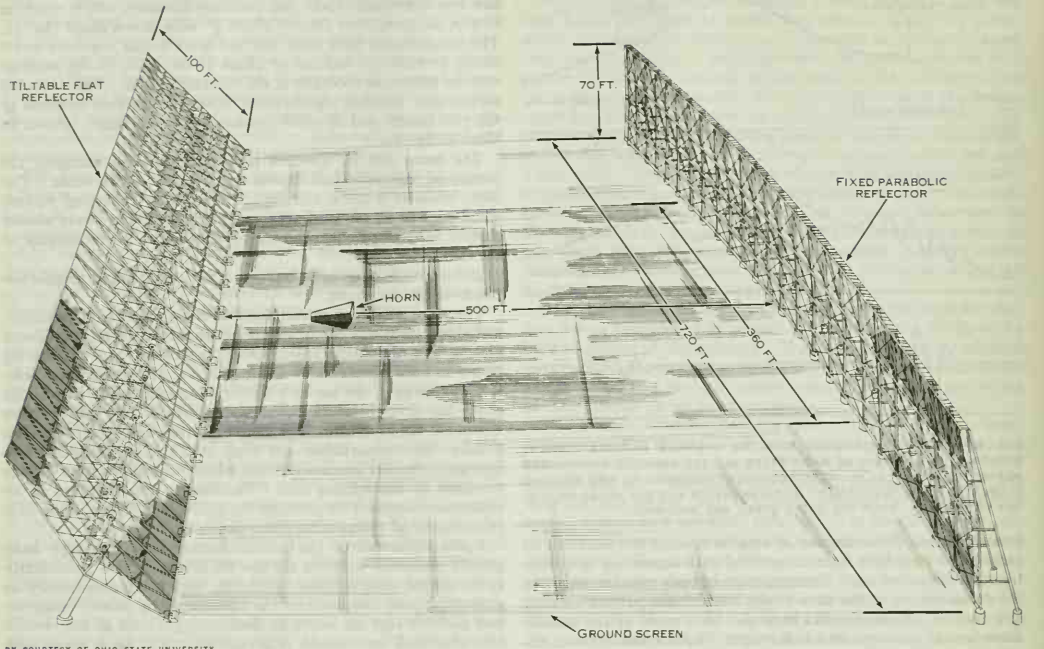


FIG. 3.—PRINCIPLE OF THE MILLS CROSS RADIO TELESCOPE
The pencil beam indicated by the shaded cone is the product of the fan beams of the east-west and north-south arms



BY COURTESY OF OHIO STATE UNIVERSITY

FIG. 4.—DRAWING OF A RADIO TELESCOPE WITH COMBINATION OF FIXED AND TILTABLE REFLECTORS

desirable to ensure that the wind force passes through the intersection of the two axes.

In an equatorial mount, such as is used for optical telescopes, it is not possible to obtain full coverage of the sky without cantilevering the declination axis from the polar axis. This means a double set of counterweights and a high wind load on the hour angle drive gear. For this reason where full sky coverage is required, an altazimuth mounting is often used in association with an analogue computer to convert celestial co-ordinates to altazimuth co-ordinates. An example of the altazimuth mounting is that of the 250-ft. reflector at Manchester, Eng. This instrument is the largest steerable reflector and was constructed between 1950 and 1957. The weight of the reflector, which is $\frac{1}{2}$ in. thick steel sheet supported on a network of radial and ring trusses, is 700 tons. The two supporting towers which carry the altitude trunions and drive mechanism and the base carriage weigh 1,200 tons. The drive in azimuth is via four of the eight bogies (weight-carrying wheels) which ride on a 300-ft. diameter double circular rail track. This reflector can be used successfully at wave lengths longer than 50 cm. Below this the deviations from true parabolic shape in the surface produce deterioration in the beam width and the effective collecting area. Such deterioration generally sets in when the surface errors exceed one-eighth of the wave length. The Manchester design is unusual in that the reflector is supported from its outer edge. This was done to provide easy access to the feed and first stages of the receiver with the reflector inverted. Most other reflectors on altazimuth mounts are supported closer to the apex, as are those on equatorial mounts.

An example of a steerable reflector on an equatorial mount is the 60-ft. reflector at the Harvard college observatory. In contrast to the enormous weight of the Manchester reflector, the Harvard reflector weighs four tons. There are two other notable differences—the surface and the feed support. The surface is expanded aluminum with diamond-shaped holes approximately one inch by one-half inch, but it behaves as a perfect reflector down to a wave length of about ten centimetres. The perforated surface has the advantage that the wind load is considerably less than that of a solid sheet. The feed support is in the form of a tripod with the individual legs meeting in a ring just beyond the focus. Such a support produces secondary lobes at the side of the main beam pattern, but leaves a cleaner working space near the focus than does a central mast. The central mast support produces some shielding of the reflector near the vertex. Other reflectors have bipod or quadrupod feed supports. The Harvard 60-ft. reflector is one of about 12 steerable reflectors in the 50 to 100-ft. size range in operation or under construction in the late 1950s. Other large ones proposed are a 140-ft. in the United States and a 210-ft. in Australia.

There are, in addition, a number of telescopes whose proper classification lies between the fully steerable reflector and the large array. These consist of fixed reflectors in which limited movement of the beam can be achieved by a motion of the feed or a secondary reflector. An interesting example of this type is the Ohio state telescope in which the beam can be directed in altitude by tilting the large reflecting plane or over a small range in azimuth by moving the feed horn (*see* fig. 4).

See also RADIO ASTRONOMY; OBSERVATORY (ASTRONOMICAL).

(J. G. BN.)

TELESIO, BERNARDINO (1508–1588), Italian philosopher and scientist, the head of the great south Italian movement which protested against the accepted authority of abstract reason, was born of noble parentage at Cosenza near Naples in 1508. After studying at Milan, Rome and Padua, he began his attack upon the medieval Aristotelianism which then flourished in Padua and Bologna. He lectured at Naples and finally founded the academy of Cosenza. His great work, *De natura rerum iuxta propria principia*, appeared in 1565 and was followed by a large number of scientific and philosophical works of subsidiary importance. After his death at Cosenza in 1588 his books were placed on the Index.

Telesio proposed an inquiry into the data given by the senses, from which he held that all true knowledge comes. Instead of

postulating matter and form, he based existence on matter and force. This force has two opposing elements: heat, which expands, and cold, which contracts. These two processes account for all the diverse forms and types of existence, while the mass on which the force operates remains the same. The harmony of the whole consists in this, that each separate thing develops in and for itself in accordance with its own nature while at the same time its motion benefits the rest. The obvious defects of this theory were pointed out at the time by Telesio's pupil Francesco Patrizzi, viz., (1) that the senses alone cannot apprehend matter itself; (2) that it is not clear how the multiplicity of phenomena could result from these two forces; and (3) that he adduced no evidence to substantiate the existence of these two forces. Telesio's system is, however, a forerunner of all subsequent empiricism and marks the period of transition from authority and reason to experiment and individual responsibility.

TELESPHORUS, SAINT, pope from *c.* 125 to *c.* 136, is said to have been a Greek. St. Irenaeus calls him a glorious martyr. His feast day is Jan. 5. (J. V. Hn.)

TELEVISION, the electrical transmission of pictures in motion and the simultaneous electrical transmission of the accompanying sounds. A discussion of the theories upon which such transmission is based will be found in the articles Radio and Radio Receiver; the electrical transmission of still photographs is discussed in Facsimile Transmission. This article is divided into the following parts:

I. Introduction
II. History of Television
III. Fundamentals of Picture Transmission and Reception
  A. The Television Picture Signal
  B. Television Camera Tubes
    1. Iconoscope
    2. Orthicon
    3. Image Orthicon
    4. Vidicon
    5. Image Dissector
    6. Flying Spot Scanner
  C. Video Amplification
  D. The Television Channel
  E. The Television Transmitter
  F. The Television Receiver
  G. Television Picture Tubes
IV. Compatible Colour Television
  A. Colour Camera and Picture Tube
  B. Colour Transmitter and Receiver
V. Television Recording
VI. Television Broadcasting
    1. Studio Programs
    2. Film Broadcasting
    3. Outside Broadcasts
    4. Networks
    5. Transatlantic Broadcasts
VII. Closed-Circuit Television
    1. Equipment
    2. Applications in Education
    3. Industrial Applications
    4. Commercial Applications

I. INTRODUCTION

The purpose of a television system is to extend the senses of vision and hearing beyond their natural limits. Television systems are designed, therefore, to embrace the essential capabilities of these senses, with appropriate compromises between the quality of the reproduction and the costs involved. The aspects of natural vision which must be considered in a television system include the ability of the human eye to distinguish the brightnesses, colours, details, sizes, shapes and positions of the objects in the scene before it. The aspects of hearing include the ability of the ear to distinguish the pitch, loudness and distribution of sounds. The television system must also be designed to override, within reasonable limits, the effects of interference and to minimize visual and aural distortions in the transmission and reproduction processes. The particular compromises adopted for public television service, *i.e.*, by broadcast means, are embodied in the television standards adopted and enforced by the responsible government agency in each country.

Television technology must deal with the fact that human vision employs many hundreds of thousands of separate electrical circuits, in the optic nerve from the retina to the brain, to convey simultaneously in two dimensions the whole content of the scene on which the eye is focused, whereas in electrical communications it is feasible to employ only one such circuit (*i.e.*, the broadcast channel) to connect the transmitter and the receiver. This fundamental disparity is overcome in television by a process of image analysis and synthesis, whereby the scene to be televised is first translated into an electrical image, and the latter is then broken up into an orderly sequence of electrical impulses which are sent over the channel one after the other. At the receiver the impulses are translated back into a corresponding sequence of lights and shadows and these are reassembled in their correct positions on the viewing screen.

This sequential reproduction of visual images is feasible only because the visual sense displays persistence; that is, the brain retains the impression of illumination for about 0.1 sec. after the source of light is removed from the eye. If, therefore, the process of image synthesis occurs within less than 0.1 sec., the eye is unaware that the picture is being reassembled piecemeal, and the viewing screen appears as if its whole surface were continuously illuminated. By the same token, it is then possible to recreate more than ten complete pictures per second and to simulate thereby the motion of the scene.

In practice, to depict rapid motion smoothly, it is customary to transmit from 25 to 30 complete pictures per second. To provide detail sufficient to accommodate a wide range of subject matter, it is customary to analyze each picture into 100,000 or more elementary details. This implies that the rate at which these details are transmitted over the television system exceeds 2,000,000 per second. To provide a system suitable for public use and yet capable of such speed has required the full resources of modern electronic technology.

II. HISTORY OF TELEVISION

The development of television was marked by an early appreciation of its basic concepts, followed by a delay of several decades in which apparatus was gradually improved to such a point that the principles could be demonstrated. The basic principle of translating light into its electrical counterpart was discovered in 1873 by an English telegrapher, May, who found that the electrical resistance of the element selenium was substantially reduced under illumination. The photoconductive action of selenium is so sluggish that the early television systems could not use this phenomenon and it was not until 1950 that a photoconductive television camera tube, the Vidicon, was fully developed. W. Hallwachs in 1888 first demonstrated the photoelectric effect, by which electrons are emitted substantially instantaneously from an illuminated surface. This effect was first put to use in a crude television system in 1923, when J. L. Baird in England and C. F. Jenkins in the United States demonstrated the transmission of crude black-and-white silhouettes in motion. Baird and Jenkins were successful in 1925 in demonstrating electrically transmitted moving pictures in half tones, and this is generally conceded to be the start of television as an operational technology. But the televised pictures were so crude that they hardly qualified as an extension of the sense of sight, and it was not until 1936 that a public television service, capable of supporting a regular program of entertainment, was inaugurated in London by the British Broadcasting corporation.

The earliest proposals for television systems suggested that the action of the eye be imitated by setting up a mosaic of light-sensitive elements, as in the retina, with an individual telegraph circuit attached to each element, as in the optic nerve. The picture was to be reproduced by a similar mosaic of electrically operated shutters, actuated individually by the telegraph circuits. To secure a picture of detail equal to that of modern television by this method would require over 100,000 circuits between transmitter and receiver, an obviously impractical arrangement. The multiple-circuit mosaic system has never been demonstrated as a television system, although the principle has been used in animated electric signs to produce crude pictures in motion.

The concept of analyzing and synthesizing an image so that it could be transmitted over a single communication channel was first announced in 1884 in a German patent by P. Nipkow. Variations of Nipkow's scheme were used by Baird and Jenkins in 1923–25 and it deserves an important place in the history of television. As illustrated in fig. 1, the image to be televised is focused on a rotating disk with square apertures arranged in a spiral around its edge.

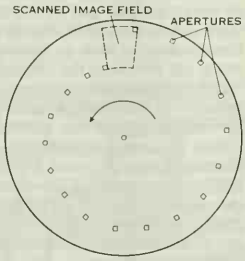


FIG. 1.—NIPKOW DISK FOR AN 18-LINE PICTURE

As the disk rotates, the outermost aperture traces out a line across the top of the image, and the light passing through the aperture varies in direct proportion to the light and shadows of that line of the image as they are sequentially uncovered by the aperture. When the outermost aperture has passed over the image, the next inner aperture traces out another line, parallel to and just below the line just traced. The changes in the light passing through the second aperture represent, in sequence, the lights and shadows present in the image along this second line. As the disk continues to rotate, successive lines are traced out, one beneath the other, until the whole area of the image has been covered, one line at a time.

In this way, the lights and shadows of the image are sequentially explored in an orderly manner. The light passing through the apertures enters a photoelectric cell which translates the sequence of light values into a corresponding sequence of electrical values. These are transmitted over a single electric circuit to the receiver.

At the receiver, the electrical impulses cause light to be produced by a lamp (such as a gas-discharge lamp) capable of reproducing the sequence of light values. The varying light from the lamp is projected on the surface of a disk, similar to that at the transmitter. The receiver disk must rotate in precise synchronism with the disk at the transmitter, that is, the position of each aperture, relative to the light source, must at each instant of time be precisely the same as the position of the respective aperture of the transmitter disk relative to the scene to be transmitted. Thus, as the outermost aperture of the disk at the transmitter passes over the upper line of the scene, the corresponding aperture in the receiver disk allows light to pass to the observer, and the lights and shadows of the upper line in the image are thus recreated. As the disks continue to revolve in synchronism, the receiver reproduces, line by line, the same sequence of lights and shadows as are uncovered by the apertures of the transmitter disk. If the whole area of the image is covered within a few hundredths of a second, the persistence of vision of the observer comes into play and the full lengths of all the lines appear to be illuminated simultaneously. The eye thus sees an area of illumination, when viewing the receiver disk, although in point of fact light is emerging from only one aperture at a time. The receiver disk thus reassembles the lights and shadows of the transmitted sequence in their proper positions and the image is reproduced.

The process is repeated once during each rotation of the disks. If the disks rotate at, say, 30 revolutions per second, 30 complete still pictures are reproduced each second. Any motion in the image is thus broken up into increments of $\frac{1}{30}$-sec. duration. The succession of such small motions cannot be resolved by the eye, and the illusion of smooth motion is thus created. This principle of rapid, sequential exploration of light values and synchronous reassembly of corresponding light values underlies all modern television technology.

The Nipkow scheme and its many mechanical variants failed to provide a satisfactory television system because of the extraordinary demands of the sense of sight with respect to pictorial detail. To provide a pictorial setting for a wide variety of subject matter, the image must be capable of being analyzed into at least 100,000 (preferably 200,000) separate lights and shadows (picture elements). This means that the picture must be divided into several hundred lines (approximately the square root of the number of picture elements) and that each line must contain several hundred individually identifiable halftone light values. The Nipkow disk must then contain several hundred precisely located apertures, and the receiver light source must be capable of assuming several hundred different light values during the approximately 0.00005 sec. that is available for the tracing of each line. Neither requirement is beyond the limits of modern technology (a nonmechanical variant of the Nipkow system, the flying spot scanner, is in fact used in modern practice for film transmissions). The difficulty is that the light passing through each aperture of the Nipkow disk is then limited to roughly $\frac{1}{100,000}$ of the total light present in the scene, and the scheme is therefore effective only when the scene is very brilliantly illuminated (as in sunlit outdoor scenes or in motion-picture projection apparatus). In modern television, mechanical image analysis has been replaced by electronic exploration of the lights and shadows of the image. Such electronic apparatus can operate at any required speed and, most important, it can "store" the effect of light so as to increase the sensitivity of the television camera by about 100,000 times.

The long delay between the Nipkow proposal and Baird and Jenkins' practical demonstrations was occasioned largely by the necessity of amplifying the weak electrical impulses produced by the photocell at the transmitter to a strength sufficient to operate the lamp at the receiver. The invention of the thermionic amplifier tube by Lee De Forest in 1907 provided the basic means, but substantial additional work was necessary to secure the high-speed amplification needed for even crude television transmissions.

For a decade after the first halftone transmissions in 1925, great effort was expended to produce images of fine detail from scenes under normal illumination, and experimental broadcasts were conducted by Baird, Jenkins, E. F. W. Alexanderson and others, using image analysis of 60 to 120 lines per picture at about 20 pictures per second. When public broadcasting began in London in 1936, one of the two systems used was a highly refined mechanical system developed by Baird, producing images of 240 lines at 24 pictures per second. But within a few months the mechanical methods were abandoned in favour of electronic analysis at 405 lines and 25 pictures per second, the values still in use in the British television system. Experimental broadcasting by the electronic method also began in New York in 1936, but officially authorized transmissions for the public did not begin in the United States until July 1941.

The conceptual base of the modern electronic television system was laid down in 1907 by a Russian, Boris Rosing, and an Englishman, A. A. Campbell-Swinton, who independently suggested the use of cathode rays (electron beams) for reconstitution of the image at the receiver. A cathode-ray tube had been developed as early as 1897 by K. F. Braun to trace curves on a phosphor surface.

In 1911 Campbell-Swinton proposed, for a television camera, the use of a photosensitive mosaic which would accumulate electric charge and be discharged by a moving cathode ray. This important concept, without which modern television could hardly exist, was not reduced to practice until 1925 when the American V. K. Zworykin (who had studied under Rosing) applied for his patent on the iconoscope. The iconoscope is still used in film telecasting and its lineal descendants, the orthicon and the image orthicon (see *Television Camera Tubes*, below), are used for nearly all telecasting of "live" subject matter.

Prior to 1932, experimental television transmissions were broadcast on frequencies near the standard sound broadcasting band, in the vicinity of two megacycles, since the crude 60-line picture transmissions of those days could be accommodated in channels not dissimilar to those used for sound broadcasting. In the early 1930s it became clear that much finer image detail was needed to support a program service, and at that time the electronic devices (iconoscope camera tube and cathode-ray picture tube) became available to produce higher definition in the pictures. It was then

recognized that the two megacycle frequencies were not suited for the more refined methods of reproduction, and a wholly new region of the radio spectrum was explored for television. This region extends upward from about 50 mc. in the so-called very-high-frequency (VHF) band. By 1936 all activity on the two megacycle band had ceased in favour of the VHF channels, which provide the necessary channel space for transmission of highly detailed pictures. The VHF channels continue in use (extending from 54 to 216 mc., channels 2 through 13) and, in fact, form the backbone of television service throughout the world. Additional channels in the ultrahigh-frequency (UHF) band (from 470 to 890 mc., channels 14 through 83) were allocated to television broadcasting in the United States.

The period 1936–40 was marked in the United States by a vigorous debate on the precise form which public television service should take, particularly how detailed the pictures should be to support a wide variety of program fare. The first "all-electronic" system to be demonstrated in the U.S. transmitted pictures by radio relay between New York city and Camden, N.J., with one intermediate relay point at Arney's Mount, N.J. These pictures were analyzed into 240 lines and sent at 24 pictures per second. By 1934 the number of lines per picture had been raised to 343 and the picture rate to 30 per second. By 1943 an "industry standard" of 441-line pictures was suggested by the trade association of the radio industry and was used experimentally. Finally, in 1940, a committee of experts, the first National Television System committee (N.T.S.C.), recommended transmission at 525 lines per picture, 30 pictures per second. These values, with other specifications recommended by the N.T.S.C., were adopted by the Federal Communications commission in 1941 and subsequently served as the basis of U.S. television service. The only other national television standard then in use was the 405-line picture, sent at 25 pictures per second, of the British Isles. Shortly thereafter, transmissions at 441-lines began in Paris; after World War II this standard was raised to 819 lines. Beginning in 1949 a group of European countries, led by the Dutch and the Swiss, adopted standards generally patterned after the U.S. system, with pictures analyzed into 625 lines, transmitted at a rate of 25 pictures per second. The latter standards are widely used, except in North America, most of South America and Japan, which adhered to the 525-line system.

When wartime restrictions governing the manufacture of receivers were removed in 1946, the stage was set in the U.S. for a rapid growth of the television broadcasting industry. By 1949 there were 1,000,000 receivers in use; the 10,000,000 mark was passed in 1951 and the 50,000,000 mark in 1959. About half that number were in use in the 50 other nations which by then had established regular program services.

The development of colour television paralleled that of the black-and-white variety. It was early appreciated that it was necessary to transmit and superimpose three images in the respective primary colours to achieve the desired result. A German patent issued in 1904 contained the earliest definitive proposal. Reduction to practice was delayed while the operative problems of monochrome television were being solved. The first demonstration of television in colour was offered by Baird in 1928, five years after his first monochrome demonstration. Mechanical scanning was used, with a Nipkow disk having three sets of apertures, one for each primary colour. In 1929 the first colour television demonstration in the U.S. was made by H. E. Ives and his colleagues, who transmitted 50-line colour images at about 18 pictures per second over wires between New York and Washington, D.C.

The subsequent refinement of the art to a form suitable for public colour program service took place in the United States, culminating in the compatible colour system introduced in 1954. Prior to this, mechanical colour television methods were brought to an advanced state by Peter Goldmark, who first demonstrated in 1940 in New York a mechanical system of televising images in full colour, with 343 lines per picture and complete sequential coverage of the image in three colours at a rate of 20 pictures per second. Following the war, Goldmark further refined this system. In 1949 he proposed to the Federal Communications commission that colour television be officially authorized with images of 405 lines, transmitted sequentially with three-colour coverage at a rate of 24 pictures per second.

In 1950 the commission adopted this recommendation and authorized transmissions under these standards began in New York in 1951. This colour system was incompatible with the then-existing monochrome service; that is, the 10,000,000 receivers then in the hands of the public could not reproduce the colour transmissions in monochrome. This fact, together with the limitations on design and manufacture imposed by the Korean war, forced abandonment of the sequential colour system late in 1951.

Meanwhile, the second National Television System committee was convened to study the means of transmitting colour television in compatible fashion. The basis of compatible colour television was laid down in 1948–49 by various workers in the U.S. television industry, and the task of the second N.T.S.C. was to develop and select the precise methods which would comprise the public service. The N.T.S.C. system was perfected in 1953 and recommended to the FCC in that year. Beginning in 1954, authorized public service began under standards identical to the N.T.S.C. recommendations. Subsequently, the N.T.S.C. system was the subject of study by European television engineers. The majority opinion was that the European colour service, when authorized, should follow the principles of the N.T.S.C. system.

The compatible colour system (*see* below) is a remarkable technical tour de force which, prior to 1946, would have been flatly rejected by experts as impossible of practical attainment. The basic principle is the combination of two image transmissions, one carrying the fine details in black and white (to which monochrome receivers respond), the other of coarser structure, carrying the colour information. This second component has no appreciable effect on monochrome sets. Colour receivers respond to it, in concert with the monochrome transmission. Since the ability of the eye to perceive detail is most acute when viewing white light, the black-and-white component of the colour image carries the impression of fine detail and the superimposed coarse colour information does not substantially alter the sharpness of the resulting colour picture.

Colour receivers are intrinsically more complicated than black-and-white receivers, and hence more expensive to purchase and maintain in good order. This fact, plus the fact that the essential content of the compatible colour programs is available (in monochrome) to owners of black-and-white receivers, delayed full acceptance of the colour service by the public. The technical basis of the compatible colour service is sound, however, and it seemed certain that the economic barriers would be removed with further development of receiver circuits and components.

III. FUNDAMENTALS OF PICTURE TRANSMISSION AND RECEPTION

The quality and quantity of television service are limited fundamentally by the rate at which it is feasible to transmit the picture information over the television channel. As already noted, in modern practice the televised image must be capable of being dissected, within a few hundredths of a second, into more than 100,000 picture elements. This implies that the electrical impulses corresponding to the picture elements must pass through the channel at a rate as high as several million per second. Moreover, since the picture content may vary from simple close-up shots having little fine detail to comprehensive distant scenes in which the limiting detail of the system comes into play, the actual rate of transmitting the picture information varies from time to time, from a few impulses per second to several million per second. The television channel must be capable, therefore, of handling information over a continuous band of frequencies several million cycles wide. This is testimony to the extraordinary comprehension of the sense of sight. Hearing is comparatively crude. The ear is satisfied by impulses which can be carried over a channel only 10,000 cycles wide.

In the United States, the television channel occupies a width of 6 mc. (6,000,000 cycles per second) in the radio spectrum.

This is 600 times as wide as the channel used by each standard sound broadcasting station. In fact, each television station uses nearly six times as much spectrum space as all the commercial sound broadcasting channels combined. Since each television station must occupy so must spectrum space, few channels are available in a given locality. Moreover, the quantity of service is in conflict with the quality of reproduction. If the detail of the television image is to be increased, other parameters of the transmission being unchanged, the channel width must be increased proportionately, and this decreases the number of channels which can be accommodated in the spectrum. This fundamental conflict between quality of transmission and number of available channels dictates that the quality of reproduction shall just satisfy the typical viewer under the normal viewing conditions of the home. Any excess of performance beyond this would ultimately result in a restriction of program choice.

The first requirement to be met in image analysis is that the reproduced picture shall not flicker, since flicker induces severe visual fatigue. Flicker becomes more evident as the brightness of the picture increases. If flicker is to be unobjectionable at brightnesses suitable for home viewing during daylight as well as evening hours (25 to 100 foot-lamberts), the successive illuminations of the picture screen should occur not less than 50 times per second. This is approximately twice the rate of picture repetition needed for smooth reproduction of motion. To avoid flicker, therefore, twice as much channel space is needed as would suffice to depict motion. The same disparity occurs in motion-picture practice, in which satisfactory performance with respect to flicker would require twice as much film as would be necessary for smooth simulation of motion.

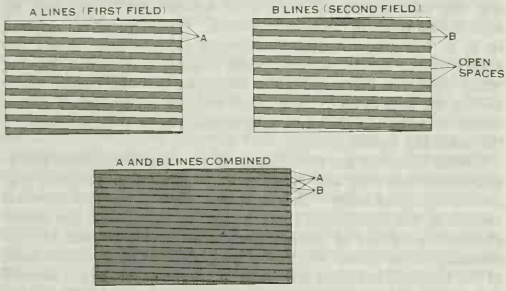


FIG. 2.—COMBINATION OF TWO SEPARATE SCANNING CYCLES (FIELDS) DURING EACH TRANSMISSION PERIOD PRODUCES COMPLETE PICTURE

A way around this difficulty has been found, in motion pictures as well as in television, by projecting each picture twice. In motion pictures, the projector interposes a shutter briefly between film and lens while a single frame of the film is being projected. In television, each image is analyzed and synthesized in two sets of spaced lines, one of which fits successively within the spaces of the other (fig. 2). Thus the picture area is illuminated twice during each complete picture transmission although each line in the image is present only once during that time. This technique is feasible because the eye is comparatively insensitive to flicker when the variation of light is confined to a small part of the field of view. Hence flicker of the individual lines is not evident. If the eye did not have this fortunate property, many towns possessing two television channels would have only one.

It is thus possible to avoid flicker and simulate rapid motion by a picture rate of about 25 per second, with two screen illuminations per picture. The precise value of the picture-repetition rate used in a given region has been chosen by reference to the electric power frequency which predominates in that region. In Europe, where 50-cycle power is the rule, the television picture rate is 25 per second (50 screen illuminations per second). In North America, the picture rate is 30 per second (60 screen illuminations per second) to match the 60-cycle power which predominates on that continent.

A precise match between power frequency and the screen illumination rate is not necessary, and in fact seldom occurs over the area of an extensive network broadcast. A disparity of more than one or two cycles between the two values, however, would require additional costs in receivers to avoid picture distortion. The higher picture-transmission rate of North America allows the pictures there to be about five times as bright as those in Europe for the same susceptibility to flicker, but this advantage is offset by a 20% reduction in picture detail for equal utilization of the channel.

The second aspect of performance to be met in a television system is the detailed structure of the image. A printed halftone engraving may possess several million halftone dots per square foot of area. Such reproductions are intended for minute inspection and the dot structure must not be apparent to the unaided eye even at close range. Such fine detail would be a costly waste in television, since the television picture is viewed at comparatively long range (5 to 15 ft. in the typical home setting) and the picture area is not greater than 2 sq.ft. Under these conditions, a picture structure of about 200,000 halftone elements is a suitable compromise. This detail is about equal to that provided by 16-mm. home motion pictures and substantially exceeds that of the 8-mm. film widely used for amateur cinematography.

The physiological basis of this compromise lies in the fact that the normal eye, under conditions typical of television viewing, can resolve pictorial details if the angle they subtend at the eye is not less than two minutes of arc. This implies that the structure of 200,000 elements in a picture one foot high can just be resolved at a distance of about ten feet. The structure of the picture may be objectionably evident at short range, *e.g.*, while tuning the receiver, but it is inappropriate to require the system to assume the heavy costs of transmitting detail which would be used by a small part of the audience for a small part of the viewing time.

The third item to be selected in image analysis is the shape of the picture. This has been standardized universally as a rectangle whose width is one-third longer than its height. This 4 by 3 ratio (aspect ratio) was originally chosen to match the dimensions of standard 35-mm. motion-picture film (prior to the advent of the wide-screen cinema) in the interest of televising film without waste of frame area. The width of the screen rectangle is greater than its height, as in the proscenium of a theatre, to accommodate the horizontal motion which predominates in virtually all televised events.

The fourth determination in image analysis is the path over which the image structure is explored at the camera and reconstituted on the receiver screen. In the original Nipkow disk, this path was a series of arcs of circles (fig. 1). In modern television, the pattern is a series of parallel straight lines, each progressing from left to right, the lines following in sequence from top to bottom of the picture frame. The exploration of the image structure proceeds at a constant speed along each line, since this provides uniform loading of the transmission channel under the demands of a given structural detail, no matter where in the frame the detail lies. The line-by-line, left-to-right, top-to-bottom dissection and reconstitution of television images is known as scanning, from its similarity to the progression of the line of vision in reading a page of printed matter. The agent which disassembles or reassembles the light values along each line is called the scanning spot, and the path it follows is the scanning pattern or raster.

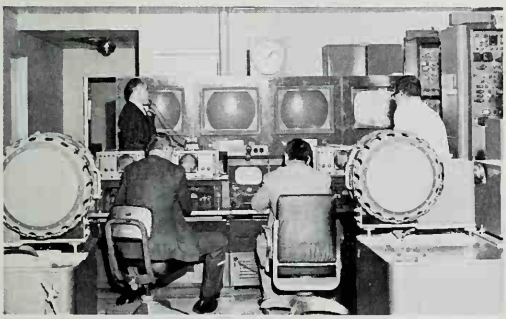
The geometry of the modern scanning pattern is shown in fig. 2. It consists of two sets of lines. One set, marked A, is scanned first and the lines are so laid down that an equal empty space is maintained between lines. The second set, marked B, is laid down after the first, and is so positioned that its lines fall precisely in the empty spaces of set A. The area of the image is thus scanned twice, but each point in the area is passed over once. This is known as interlaced scanning, and is used in all the television broadcast services of the world. Each set of alternate lines is known as a scanning field; the two sets together, comprising the

A television studio monitoring booth. The program director, technical director and lighting director oversee the on-the-air picture on the large monitor and preview the pictures to follow on the smaller monitors

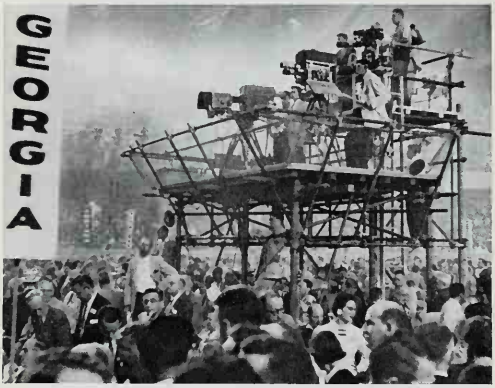
A television show in production on the set, showing banks of floodlights, suspended microphones and three colour television cameras. The centre camera is mounted on a crane dolly for vertical movement. The circular ring on the camera at left is used for changing the direction of the camera

A film camera chain, used in the adaptation of photographic film for television transmission. It consists of a 16-mm. projector (1) and 35-mm. projector (3) for filmed programs and motion picture film; a slide projector (2) for station identifications, commercials and test patterns; a multiolexer (4) for combinations of the three; and a monitor (5) for previewing the pictures

TELEVISION BROADCASTING



Colour film camera chain video control room. Engineers view the pictures on the monitors (below clock) and, guided by the oscilloscopes directly in front of them, make technical corrections



A battery of television cameras in use at a U.S. political convention

PHOTOGRAPHS, (TOP LEFT) COLUMBIA BROADCASTING SYSTEM, (CENTRE AND BOTTOM LEFT) WNBQ, NATIONAL BROADCASTING COMPANY, CHICAGO, (TOP RIGHT, BOTTOM RIGHT) NATIONAL BROADCASTING COMPANY, NEW YORK

TELEVISION

Six cameras designed to televise the moon from the Ranger spacecraft; they transmit at the rate of one full-scan picture every $2\frac{1}{2}$ seconds and one partial-scan every 2.10 second

Closed-circuit television finds wide application in schools, making it possible for the special skills of teachers to be shared by larger audiences of students and other classroom teachers

APPLICATIONS OF TELEVISION

An eight-inch diameter transistorized camera being used to probe underground conditions to aid in rescuing trapped miners

An experimental camera designed to transmit pictures of astronauts en route to the moon

Police officers directing and co-ordinating traffic by means of closed-circuit television combined with two-way radio dispatch facilities

PHOTOGRAPHS: (TOP LEFT, BOTTOM LEFT) RADIO CORPORATION OF AMERICA. (TOP RIGHT) BOARD OF EDUCATION, CITY OF CHICAGO. (CENTRE RIGHT, BOTTOM RIGHT) MOTOROLA INC.

whole scanning pattern, are known as a scanning frame. The repetition rate of field scanning is standardized in accordance with the frequency of electric power, as noted above, at either 50

Television Systems of the World

Region or country	Number of lines per frame	Number of pictures per second	Maximum detail (picture elements per frame)	Available picture band width (mc.)	Channel band width (mc.)
England	405	25	130,000	3	5
North America, South America, Japan	525	30	150,000	4	6
Europe, Australia, Africa, Eurasia	625	25	210,000	5*	7*
France and French dependencies	819	25	440,000	10.4	14

*Six and eight megacycles respectively in U.S.S.R.

fields per second or 60 fields per second. The corresponding rates of frame scanning are 25 and 30 frames per second. The localities in which various scanning standards are used are listed in the table.

The total number of lines in the scanning pattern has been set to provide a maximum pictorial detail of the order of 200,000 picture elements. Since the frame area is 4 units wide by 3 units high, this figure implies a pattern of about 520 picture elements in its width (along each line) and 390 elements in its height (across the lines). This latter figure would imply a scanning pattern of about 400 lines, were it not for the fact that many of the picture details, falling in random positions on the scanning pattern, lie partly on two lines and hence require two lines for accurate reproduction. Scanning patterns are designed, therefore, to possess about 40% more lines than the number of picture elements to be reproduced in the vertical direction. Actual values in use in television broadcasting in various regions are 405 lines, 525 lines, 625 lines and 819 lines per frame, as listed in the table. These values have been chosen to suit the frequency band of the channel actually assigned in the respective geographical regions.

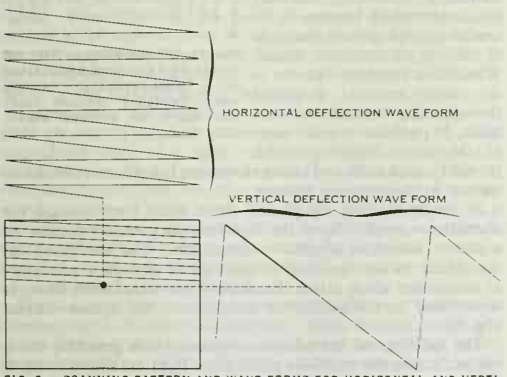


FIG. 3.—SCANNING PATTERN AND WAVE FORMS FOR HORIZONTAL AND VERTICAL DEFLECTION IN SEQUENTIAL PARALLEL-LINE SCANNING

The scanning spot is made to follow the paths shown in fig. 2 by imparting two repetitive motions to the spot simultaneously (fig. 3). One is a horizontally directed back-and-forth motion. The spot is moved at constant speed from left to right and then returned as rapidly as possible, while extinguished and inactive, from right to left.

Simultaneously, a vertical motion is imparted to the spot, moving it comparatively slowly from top to bottom of the frame. This motion spreads out the more rapid left-to-right scans, forming the first field scan of alternate lines and empty spaces. When the bottom of the frame is reached, the spot moves vertically upward as rapidly as possible, while extinguished and inactive. The next top-to-bottom motion then spreads out the horizontal line scans so that they fall in the empty spaces of the previously scanned field. Precise interlacing of the successive field scans is facilitated if the total number of lines in the frame is an odd number. All the numbers of lines used in ordinary transmissions (*see* table) were so chosen.

FIG. 4.—ACTIVE AND INACTIVE PORTIONS OF SCANNING PATTERN

The return of the scanning spot from right to left and from bottom to top of the frame, during which it is inactive, consumes time which cannot be devoted to transmitting picture information. This time is used to transmit synchronizing control signals which keep the scanning process at the receiver in step with that at the transmitter. The time lost during retracing of the spot proportionately reduces the actual number of picture elements which can be reproduced. In the 525-line scanning pattern used in the U.S., about 15% of each line is lost in the return motion and about 35 out of the 525 lines are blanked out while the spot returns from bottom to top of two successive fields. The scanning area actually in use for reproduction of the picture contains a maximum of about 435 picture elements along each line and has 490 active lines, capable of reproducing 350 picture elements in the vertical direction. The frame can accommodate at maximum, therefore, about 350×435 or 152,000 picture elements. The detail-transmission capacity of the scanning patterns of other systems is listed in the table.

The time taken by the scanning spot to move over the active portion of each scanning line is of the order of .000005 sec. (microseconds; μ sec.). In the U.S. system, 525 lines are transmitted in $\frac{1}{30}$ sec., which is equivalent to about 64 μ sec. per line. Up to 15% of this time is consumed in the horizontal retrace motion of the spot, leaving 54 μ sec. for active reproduction of as many as 435 picture elements in each line. This represents a maximum rate of $435 \div 54 \times 10^{-6} \cong 8{,}000{,}000$ picture elements per second. Two picture elements can be approximately represented by one cycle of the transmission signal wave. The signal must therefore be capable of carrying components as high as four megacycles. The U.S. six-megacycle television channel provides a sufficient band of frequencies for this (plus an additional two megacycles for transmission of the sound program, to protect against interference and to meet the requirements of vestigial side-band transmission [see *The Television Channel*, below]).

The relation between number of lines, number of pictures per second and the frequency band needed to transmit the picture information is

$$F = \tfrac{1}{2}R = KL^2P$$

where F is the frequency band in cycles per second which must be transmitted, R is the rate of scanning picture elements per second, L is the number of lines per picture, P the number of pictures per second and K a factor (approximately 0.5) which takes into account the spot retrace losses, the aspect ratio and vertical resolution losses due to line straddling.

The relationship between the ideal and actual scanning patterns is shown in fig. 4. The part shown as darker is lost during the spot retraces. The remaining area of the pattern is actively employed in analyzing and synthesizing the picture information and is adjusted to have the 4 by 3 dimensions of the standard aspect ratio. In practice, some of the active area may also be hidden behind the decorative mask which surrounds the picture tube of the receiver, as shown by the dashed line.

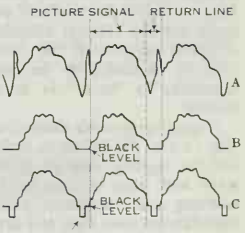


FIG. 5.—VIDEO SIGNAL

(A) Signal as obtained from pickup; (B) after blanking out return-line signal and inserting black level; (C) after adding synchronizing signal

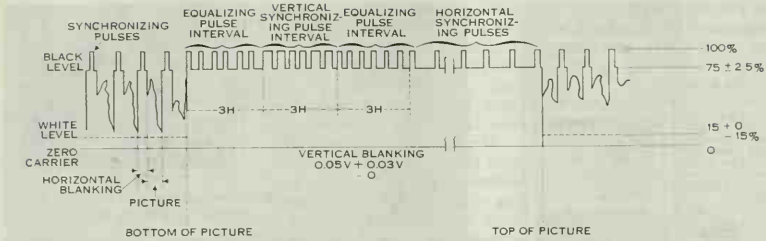


FIG. 6.—WAVE FORM OF SYNCHRONIZING SIGNAL: H, LINE-SCANNING PERIOD; V, FIELD-SCANNING PERIOD

A. The Television Picture Signal

The translation of the televised scene into its electrical counterpart results in a sequence of electrical impulses known as the television picture signal. This can be represented graphically as a signal wave form in which the electrical values (voltage or current) are plotted vertically against time, plotted horizontally. Corresponding to the electrical values and time, respectively, are the brightness of the image at each point on the scanning line and the position on the line of the point in question. These quantities are shown in fig. 5.

The television signal wave form is a continuous sequence of electrical values, corresponding with the brightnesses along each line, interspersed with blank periods representing the times during which the inactive scanning spot retraces from the end of one line to the beginning of the next. A blank interval also occurs when the scanning spot, having reached the bottom of the frame, retraces to the top.

Superimposed on the blank intervals are additional impulses, known as synchronization signals. The horizontal synchronization signal is a small impulse which occurs shortly after the completion of each scanning line. Its purpose is to cause the scanning spots at the transmitter and receiver to retrace to the next line at precisely the same instant. The vertical synchronization signal is a serrated series of impulses (fig. 6) which occurs shortly after the scanning spot has reached the bottom of the frame, causing the receiver and transmitter spots to retrace to the top in synchronism.

Since the signal wave form embodies all the picture information to be transmitted from camera to receiver screen, the television system must deliver the wave form to the picture tube of each receiver as accurately and as free from blemishes as possible. Unfortunately, virtually every item of equipment in the system (amplifiers, cables, transmitter, the transmitting antenna, the radio wave in space, the receiving antenna and the receiver circuits) conspires to distort the wave form or permits it to be contaminated by "noise" or interference.

Among the possible distortions (fig. 7) are: (1) failure to maintain the rapidity with which the wave form rises or falls as the scanning spot crosses a sharp boundary between light and dark areas of the image, producing a loss of detail or smear in the reproduced image; (2) the introduction of overshoots which cause excessively harsh outlines; and (3) failure to maintain the average value of the wave form over extended periods, which causes the image as a whole to be too bright or too dark.

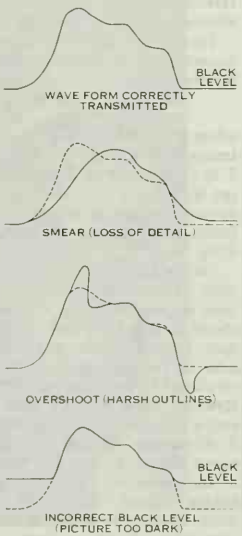


FIG. 7.—TRANSMISSION DEFECTS IN PICTURE SIGNAL WAVE FORMS

Throughout the system, amplifiers must be used to keep the television signal strong relative to the random electrical currents (noise) which are everywhere present. These random currents, caused by thermally induced motions of electrons in the circuits, cause interference having a speckled appearance, known as "snow." Pictures received from distant stations are subject to this form of interference, since the radio wave is then so weak that it cannot override the random currents in the receiving antenna. Interference of a striated type may be caused by the signals of stations other than that to which the receiver is tuned. Care in the design of the receiver tuner and amplifier circuits is necessary to minimize such interference, and channels must be allocated to neighbouring communities at sufficient geographical separations and frequency intervals to protect the local service.

The television signal wave form can be conveyed within transmitting and receiving apparatus, and over short distances of wires and cable, in the form shown in fig. 5. For transmission over greater distances, however, as in network connections and over the air, the signal wave form must be imposed on a high-frequency carrier wave. The broadcast transmitter, for example, is actually a generator of very-high-frequency or ultrahigh-frequency alternating current; *i.e.*, current which reverses its direction several tens or hundreds of millions of times per second. When such a current (known as the carrier current) is passed through the transmitting antenna, it produces a radio wave of the same frequency which travels through space and induces a weaker, but otherwise identical, current in the receiving antenna.

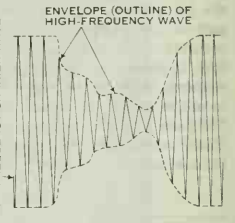


FIG. 8.—TRANSMISSION OF PICTURE SIGNAL BY AMPLITUDE MODULATION OF A HIGH-FREQUENCY CARRIER WAVE; ENVELOPE (BROKEN LINE) HAS SHAPE OF PICTURE SIGNAL WAVE FORM

At the transmitter the picture signal wave form changes the strength, or amplitude, of the high-frequency carrier current by a process known as amplitude modulation. The alternations of the carrier current are thus constrained to take on a succession of amplitudes which match the shape of the signal wave form, as is revealed in the outline or envelope of the carrier current (fig. 8).

The functions of the television system which generate, transmit and utilize the television signal wave form are shown in fig. 9. The scene to be televised is focused by a lens on a camera tube, within the camera. A scanning spot within the camera tube is passed over the image in the scanning pattern of fig. 3, its motion being controlled by the synchronization generator. The latter equipment also renders the scanning spot inactive during its retrace motions and generates the synchronization signals mentioned above. The camera produces the signal wave form, but not the synchronizing signals. The latter are thereupon added to the signal and the complete signal wave form (fig. 5 and 6) is established. The signal is then amplified as required, by amplifiers capable of operating over the wide band of frequencies (from a few cycles per second to 4,000,000 or 5,000,000 cycles per second) necessary to preserve the precise shape of the wave form. In a transcontinental network, the signal may pass through more than 100 such amplifiers in tandem before it finally reaches the transmitter. There the wave form is amplified to a high power level.

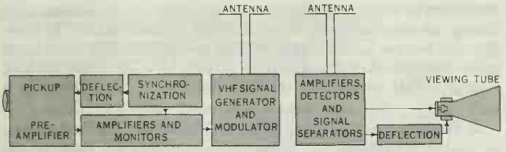


FIG. 9.—BLOCK DIAGRAM OF (A) TRANSMITTER, (B) RECEIVER

Concurrently, the carrier current is generated by other circuits of the transmitter. The signal wave form and carrier current are then combined in the modulator, which imposes the wave form on the carrier (fig. 8). The modulated carrier current may then be further amplified (typically to 10,000 or more watts) and finally passed through the transmitter antenna, which radiates the television radio wave. A similar current is induced in any receiving antenna within the service area but the current is enormously weakened by the outwardly diverging travel of the wave from the transmitter. Reception of excellent quality is possible when the power picked up by the receiving antenna is only 0.00000001 w.

The transmitter antenna must be as high and in as exposed a location as possible, since the radio waves tend to be intercepted by solid objects, including the earth's surface at the horizon. Reception beyond the horizon is possible, but the signal at such distances becomes rapidly weaker as it passes to the limit of the service area.

In the receiver, the current from the antenna is passed through a tuner, which separates it from other currents of different carrier frequencies which may be present from other stations. The desired current is then amplified in several successive amplifier stages, while still in the amplitude-modulated form. Passage through a detector follows. The picture signal wave form is thereby recovered and is passed through still another amplifier before being sent to the picture tube. The picture screen displays a scanning spot whose brightness is controlled by the picture signal. Simultaneously, the signal is applied to synchronizing circuits which abstract from it the synchronizing signals. These are used to control the motion of the scanning spot so that it traces out the scanning pattern in precise synchronism with the motion of the scanning spot in the camera.

Transmission and reception of the accompanying sound program are accomplished by the apparatus shown in fig. 10. The microphone produces a signal wave form which represents the sound pressure changes impinging on it. This wave form is amplified, transmitted over the network and finally fed to the sound transmitter. The latter, like the picture transmitter, is a generator of high-frequency carrier current, but of a frequency slightly different from that used to carry the picture information (for U.S. channel 2, the picture carrier is frequency 55.25 mc. while the sound carrier is 59.75 mc.).

For the sound transmission, however, a different modulation process is used, whereby the frequency, rather than the amplitude, of the carrier is changed by the sound wave form. This is frequency modulation (FM). The use of different forms of modulation for picture and sound makes for substantial simplification of receiver design. At the receiver, the sound carrier wave is tuned in simultaneously with the picture carrier wave, and the two carrier currents are amplified in concert. The sound wave form is finally recovered in a frequency detector and is passed, after further amplification, to the loud-speaker.

The amount of amplification conferred on the picture and sound currents by a typical television receiver is extremely large. When tuned to a station at a distance of 50 mi., the power picked up by the antenna is typically 10^{-11} w., whereas the signals fed to picture tube and loud-speaker, after amplification, are of the order of 1 w. That is, the receiver produces a faithful amplification of the order of 100,000,000,000 times.

B. Television Camera Tubes

The television camera tube is an electronic device which converts an optical image into a sequence of electrical impulses; *i.e.*, it produces the television picture signal. Four types of television camera tube are in wide use: the image orthicon, the orthicon, the iconoscope and the Vidicon. These devices differ in detail but have in common the following elements: (1) a photosensitive surface, on which the scene to be televised is focused by a lens, which converts each light value into a corresponding value of electric charge and forms thereby an electrical image; (2) a means of storing the electrical image, that is, causing the charge to accumulate during the interval of several hundredths of a second between successive scannings of each line in the image; (3) an electron beam, formed in an electron gun and deflected over the electrical image, following the scanning pattern of fig. 2; and (4) a mechanism for producing an electrical current or voltage which is proportional at every instant to the electric charge accumulated at the point passed over by the electron beam at that instant. The latter is the picture signal represented in the wave form of fig. 5.

Means are also provided for extinguishing the scanning beam during its retrace motions, so that no television signal is then generated. Two other television pickup devices, which do not store the electrical image, are the flying spot scanner and the image dissector.

1. Iconoscope.—The iconoscope is obsolescent, since it is not as sensitive as the image orthicon and its images are subject to uneven shading and flare, but it is used in televising motion-picture film. Here it serves to introduce the concepts of electron image storage and scanning in simple form. The iconoscope is housed in a dipper-shaped vacuum-tight glass envelope, as shown in fig. 11. Within the wide end is a flat sheet of mica. A uniform metallic coating (the signal plate) is placed on the rear surface of this sheet, away from the image. The front surface of the mica is covered with a mosaic comprised of many hundreds of thousands of tiny globules of silver.

During the manufacture of the tube the mosaic is treated with cesium vapour and oxygen, so each globule has a surface of the oxides of silver and cesium. This combination of elements possesses a low electronic work function, that is, electrons are readily liberated, by the photoelectric effect, when light falls on the surface. Since the globules are insulated from each other and from the signal plate by the mica, the loss of electrons under illumination causes the globules to assume and hold a positive charge, the charge on each globule being proportional to the strength of the illumination falling on it and to the time it has been illuminated.

When an optical image is focused on the mosaic, the whole surface assumes a distribution of positive charge which matches

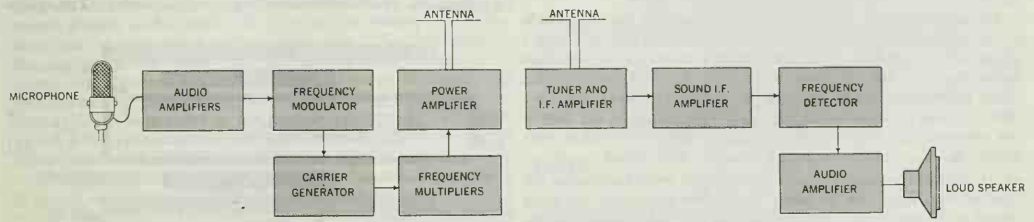


FIG. 10.—BLOCK DIAGRAM OF SOUND TRANSMISSION AND RECEPTION EQUIPMENT

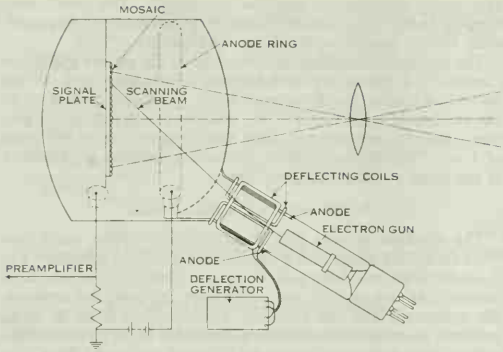


FIG. 11.—ICONOSCOPE CAMERA TUBE

the distribution of light in the image, and the amount of charge at each point steadily increases until the scanning spot passes over the globule at that point.

The scanning spot of the iconoscope is formed by a narrow beam of electrons, shot out of an electron gun in the side arm of the tube. On its way to the mosaic this beam passes within two sets of electromagnet coils. Currents like those of fig. 3 are passed through these coils, causing the beam to be deflected horizontally at a rapid rate and vertically at a slower rate. The extent of the horizontal motion is adjusted from top to bottom of the mosaic, so that the pattern traced out by the electron beam on the mosaic is a rectangular pattern like that in fig. 2.

The electrons in the beam are accelerated to a relatively high velocity (the accelerating potential in the electron gun is of the order of 1,000 v.). When the beam hits a globule, therefore, it transfers a substantial amount of energy to the atoms of the globule and the globule surface gives off electrons (additional to those freed by the optical image) by virtue of the impact. This effect is known as secondary electron emission.

The number of secondary electrons thus liberated depends on the electrical potential which the globule in question has assumed under the prior influence of the light falling on it. As each globule is passed over by the beam, therefore, it undergoes a sudden change in electrical potential, the amount of the change being proportional to the light falling on it. The change in potential of the globule is transferred through the mica support to the signal plate behind it, the globule and plate forming in effect the plates of an electrical condenser. Thus, as the beam passes in succession over the globules lying along a given scanning line, the signal plate assumes a succession of voltages (the picture signal) which match the corresponding succession of light values along that line. The signal plate is connected to an amplifier, external to the iconoscope, which increases the strength of the picture signal.

The secondary electrons liberated by the impact of the scanning beam are in large part collected by a positively charged anode ring (fig. 11), but some of the electrons fall back on the mosaic in a random pattern, thus destroying to some degree the charge pattern produced by the optical image. This produces uneven shading and flare in the reproduced image, as previously mentioned.

The phenomenon of charge storage, by which the magnitude of the electrical image is continually increased between successive scannings of each line, is of the utmost significance in television technology. It will be remembered that the Nipkow disk (and other nonstorage television pickup devices) employs only the light which is present at a given point in the image at the instant the scanning spot passes over that point. Since in modern television the area of the scanning spot is only about $\frac{1}{200,000}$ of the area of the scanning pattern, only this small fraction of the light of the image can be used. But when the image charge is stored in increasing amount for the full interval between successive scannings of a given point, the accumulated charge is then theoretically about 200,000 times that which can be accumulated during the time the beam moves through its own width. In practice, the storage action makes the iconoscope about 100,000 times more sensitive than an equivalent camera tube operating without storage. This means that the light needed by the iconoscope is only $\frac{1}{100,000}$ that needed by a nonstorage tube.

2. Orthicon.—The orthicon camera tube was the first successor to the iconoscope. A refined version (the CPS Emitron described below) is used in England for studio work. This tube (fig. 12) is housed in a cylindrical glass structure, on one end of which the optical image is focused on a mosaic. This mosaic is similar to that of the iconoscope, but is made up of a precise array of squares of photosensitive material, forming by evaporation on a transparent support. A transparent metal coating on the reverse side serves as the signal plate.

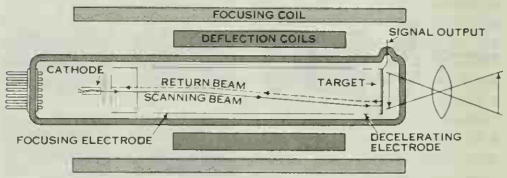
The optical image is focused on the signal plate, passing through it and the support to the mosaic, where it creates and stores a positive electrical charge image, in the manner of the iconoscope. The electron beam is formed in an electron gun at the other end of the tube and is directed toward the mosaic and deflected by electromagnet coils surrounding the tube so that it passes over the mosaic in the standard interlaced scanning pattern. The electrons are accelerated by a low voltage, so that they do not liberate any substantial number of secondary electrons. Instead, the negative electrons in the beam land on the mosaic squares, neutralizing the stored positive charge and ultimately building up a negative charge. The electrons in the beam are then repelled and return to the opposite end of the tube where they are removed by a positive collecting electrode which surrounds the electron gun. The change in electrical potential, undergone by each mosaic element as its charge is neutralized, is transferred by capacitive action, as in the iconoscope, to the signal plate. The latter thereby undergoes a succession of voltage changes which constitute the picture signal and which are passed to amplifier stages external to the tube.

The orthicon is unstable if subjected to high light levels. The CPS Emitron minimizes this difficulty by care in the design of the mosaic structure, but the tube operates best under closely controlled illumination and hence is used primarily in studios rather than in outdoor situations.

3. Image Orthicon.—The image orthicon is, by common consent, the most highly developed of the television camera tubes and is perhaps the most remarkable electronic device in existence. In its most refined form, its sensitivity to light is phenomenally great. It can respond to light levels far below those capable of exposing motion-picture film or affecting the eye itself (the "film speed" rating of such tubes is of the order of ASA 10,000). Less sensitive image orthicon tubes are used in broadcasting practice; they serve to produce satisfactory image quality under any light levels likely to be encountered in studio and outdoor scenes.

The image orthicon is similar to the orthicon but includes an additional electrical-imaging process and contains a high-gain amplifier based on the phenomenon of electron multiplication. It is used for virtually all live-scene telecasting in the United States, indoors and out.

The image orthicon (fig. 13) is housed in a cylindrical glass envelope having an enlarged section on one end, which is closed by an optically flat glass plate on the inside of which is deposited a continuous photosensitive coating (photocathode). The optical



BY COURTESY OF RADIO CORPORATION OF AMERICA

FIG. 12.—ORTHICON CAMERA TUBE

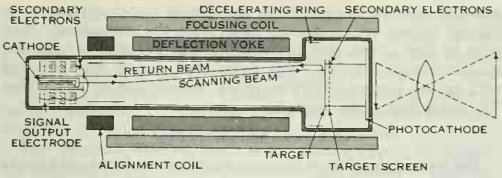


FIG. 13.—IMAGE ORTHICON CAMERA TUBE

image is focused through the glass support onto the coating, where it liberates electrons by photoelectric emission, the emission at each point matching the amount of light in the image at that point.

The liberated electrons stream away from the photosensitive surface, through the enlarged cylindrical portion of the tube, until they encounter the target electrode. This is a piece of very thin glass of uniform thickness. The streams of electrons induce secondary emission of electrons from the surface of the glass, and the latter are collected by a fine mesh screen (500 or more meshes to the inch). This lies parallel and close to the glass target; the streams of image electrons pass through it on their way to the target.

Since the glass target is very thin, its longitudinal electrical resistance, through its thickness, is small (resistance varies directly with the length of the conducting path). The changes in electrical potential induced by secondary emission are transferred through this low resistance to the opposite face of the target, where they are scanned by the electron beam. Conversely, the lateral resistance, across the face of the thin glass target, is very great (since electrical resistance varies inversely with the cross section of the conduction path). Consequently any tendency of the charge pattern to dissipate itself laterally is arrested, and the fine structure of the electrical image is preserved throughout the scanning interval. Moreover, as the streams of electrons from the photocathode continue to fall on the target glass, the intensity of the electrical image continually increases; that is, storage of charge occurs.

Both the photocathode and the target are continuous surfaces and in themselves contribute no structural limit to the detail of the image. The fine mesh screen which collects the secondary electrons from the target imposes the limit on detail, and consequently the holes in the screen must be substantially smaller than the scanning spot. Actually, the screen possesses nearly 1,000,000 holes in an area of 1.5 sq.in., compared with the roughly 200,000 picture elements into which the image is dissected.

The stored charge image on the reverse side of the target is scanned by an electron beam, which, like that of the orthicon, is accelerated at lower voltage and is actually decelerated before hitting the target, so that the scanning electrons have a velocity corresponding to voltage from zero to a few volts. Consequently, the secondary emission caused by scanning produces no deleterious effects.

The electron beam is deflected in the standard interlaced scanning pattern over the electrical image by magnetic fields produced by coils external to the tube. When the beam electrons hit the target, they neutralize the stored positive charge, as in the orthicon, and when equilibrium is reached, further electrons may not land on that spot. Thereafter, the electrons in the beam are turned around at the target and return to the vicinity of the electron gun at the other end of the tube. At this juncture the action differs from that of the orthicon. Instead of taking the picture signal from a signal plate at the target, the return beam of electrons is used. The number of electrons returning from each point in the scanning pattern is conditioned by the prior positive charge neutralization occurring at that point. Consequently the return beam constitutes a picture signal current.

When the return beam reaches the opposite end of the tube it is passed through an amplifier structure which contributes greatly to the sensitivity of the tube. This structure comprises several metal pin-wheel structures. The surfaces of the pin-wheel vanes are treated to have low work function, and hence emit secondary electrons copiously when the return-beam electrons hit them. The pin-wheel shape causes the secondary electrons liberated at the first vane to be deflected to the next vane in the same structure, where still further emission of secondary electrons occurs. The magnified stream then passes to the next pin-wheel structure to the rear, where the process repeats and the electron stream is cumulatively multiplied.

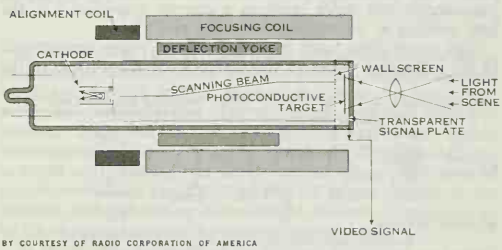
Passage through this electron multiplier structure increases the strength of the television picture signal by several thousand times (in addition to the effect provided by storage at the target) with little addition of random noise current. The image orthicon thereby achieves a sensitivity greater than that of any other continuously registering optical device. The amplified current at the final stage of the electron multiplier is passed out of the tube to a conventional vacuum-tube amplifier.

4. Vidicon.—The Vidicon camera tube, first used in the early 1950s, was the first camera device to employ the phenomenon of photoconductivity. In its early form it was sluggish (objects in motion were reproduced with noticeable smear) and hence it was limited to industrial applications. Later versions were free of this defect and the Vidicon came to enjoy some use in broadcasting, in monochrome and colour service. The tube elements (fig. 14) are contained in a cylindrical glass envelope, substantially smaller than the image orthicon and hence more adaptable to portable cameras. At one end, a transparent metallic coating serves as a signal plate. Deposited directly on the signal plate is a photoconductive material (a complex compound of selenium) whose electrical resistance is high in the dark but progressively lowers as the amount of light increases.

The optical image is focused on the end of the tube and passes through the signal plate to the photoconductive layer, where the light induces a pattern of varying conductivity which matches the distribution of brightness in the optical image. The conduction paths through the layer allow positive charge from the signal plate (which is maintained at a positive voltage) to pass through the layer, and this current continues to flow during the interval between scans. Charge storage thus occurs, and an electrical charge image is built up on the rear surface of the photoconductor.

At the other end of the tube an electron beam is formed and passed through magnetic fields which deflect it, in the standard interlaced scanning pattern, over the rear surface of the photoconducting layer. The beam electrons neutralize the positive charge on each point in the electrical image and the resulting change in potential is transferred by capacitive action to the signal plate, from which the television signal is derived. There is no structural limit within the photoconductive layer itself, that is, the electron beam scanning spot imposes the only limit on the image detail.

It is possible, therefore, to derive an image of broadcast quality (200,000 picture elements) from a photosensitive area no larger than 0.2 sq.in. This permits the use of small and comparatively inexpensive lenses, with correspondingly large depth of focus for a given lens opening. The internal structure of the tube is very simple and this, with its small size, makes it adaptable to a wide range of camera arrangements in broadcasting, industrial and military applications.



BY COURTESY OF RADIO CORPORATION OF AMERICA

FIG. 14.—VIDICON CAMERA TUBE

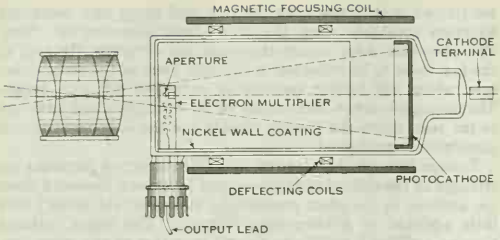


FIG. 15.—IMAGE DISSECTOR CAMERA TUBE

5. Image Dissector.—The image dissector, originally developed by P. T. Farnsworth, was used in early experimental broadcasting, but has been replaced by the storage-type camera tubes described above. The tube, in cylindrical form (fig. 15), has a photocathode surface at one end on which the optical image is focused. Electrons are thereby liberated photoelectrically and stream from the photocathode toward the opposite end of the tube, the number of electrons arising from each point being proportional to the light falling on that point. An electron image is thereby produced, and the whole image is moved bodily by magnetic scanning fields passed through the tube from external coils.

At the other end of the tube is a closed metal finger with an aperture at its end. As the electron image is deflected past this aperture, electrons from the image enter the aperture. As the image passes, the electrons entering are representative of the succession of lights and shadows along each line in the scanning pattern. The electrons entering the aperture are, in effect, the television signal current. Within the finger is a succession of electron-multiplying surfaces which increase the current by successive secondary emission. The magnified current derived at the last surface is taken from the tube and subsequently amplified by vacuum tubes.

Since the current selected from the electron image by the aperture is limited to that liberated from the photocathode at the instant the aperture passes each point, no storage occurs, so the dissector requires many thousands of times as much light as the storage tubes described above to produce an image of acceptable quality. The electronic action of the tube is extremely straightforward, however, and it can produce an image of very high quality, free from incidental blemishes, if sufficient light is available. It is used only in industrial and military systems.

6. Flying Spot Scanner.—Another form of television pickup device, used in televising film transparencies, is the flying spot scanner (fig. 16). The light source is a cathode-ray tube in which a beam of electrons, deflected in the standard scanning pattern, impinges on a fluorescent phosphor surface. The beam produces thereby a spot of light which moves in the scanning pattern. The light from this spot is focused optically on the surface of the photographic film transparency to be televised. As the image of the spot moves, it traces out a scanning line across the film, and the amount of light emerging from the other side of the film at each point is determined by the transparency of the film at that point. The emerging light is collected and caused to enter a photoelectric cell which produces a current proportional to the light entering it.

This current thus takes on a succession of values proportional to the successive values of transparency along each line in the scanning pattern, that is, it is the picture signal current. No storage action occurs, so the light from the cathode-ray tube must be very intense and the optical design very efficient to secure noise-free reproduction. The flying spot system may be used with motion-picture film if an optical immobilizer is used. (See *Television Recording*, below.)

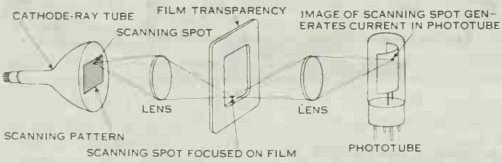


FIG. 16.—FLYING SPOT CAMERA SYSTEM

C. Video Amplification

The picture signal wave form produced by the television camera is initially very weak, of the order of a few hundredths of a volt, and is therefore ready prey to the random currents (noise) present in the succeeding transmission circuits. Immediately upon leaving the camera, therefore, the signal must be strengthened, and this process must be repeated at intervals throughout the transmission to the picture tube at the receiver. The strengthening process, known as video amplification, occurs in electron tubes with associated circuit elements which permit the extremely wide band of frequencies comprising the signal to be magnified without relative discrimination. The picture signal is applied to the grid (control electrode) of the electron tube, which governs the passage of electrons from the negative, electron-emitting element (cathode) to the positive, electron-collecting element (anode). Since the electrons pass from cathode to anode of an electron tube in a few thousandths of a millionth of a second, the grid is able to exercise control over the electron current several hundred million times per second, that is, at rates considerably above the 4,000,000 cycles per second (c.p.s.) upper frequency limit of the picture signal. The video amplifier must, however, provide uniform amplification throughout the band of frequencies from a lower limit near zero cycles per second to this upper limit. The amount of amplification tends to fall near the upper frequency limit, since it takes a finite time for the electron current of the tube to charge the electrical capacitance associated with the tube electrodes and the circuit elements. This tendency may be counteracted by inserting in the anode circuit of the amplifier a small inductance coil which resonates with the capacitance at a frequency above the upper limit of the picture signal spectrum.

At the other end of the spectrum, near zero frequency, the amplification also tends to fall off as a result of incomplete coupling between successive amplifier stages. This effect may be counteracted by care in proportioning the resistive and capacitive elements in the coupling connections, which makes possible uniform amplification down to about 30 c.p.s. However, it is impractical to reach zero frequency (direct current), since a succession of direct-current coupled amplifiers displays unmanageable instability. This difficulty is circumvented by arranging for the envelope of the broadcast carrier current to have a fixed amplitude at a level corresponding to black (absence of light in the televised scene). In this way the average brightness of the image may be preserved and recovered.

Electron tube video amplifiers can be designed to have a remarkably comprehensive range, amplifying a picture signal comprising any combination of frequency components lying between 30 and 4,000,000 c.p.s. This implies not merely that the amplitudes of all such components are uniformly increased, but that their relative times of occurrence are strictly preserved. Only if these requirements are met can the fine structure, as well as the gross outline, of the picture signal wave form be preserved from the camera to the receiver picture tube.

D. The Television Channel

When the band of frequencies in the picture signal is imposed on the high-frequency broadcast carrier current in the modulator of the transmitter, as previously described (fig. 8), two bands of frequencies are produced above and below the carrier frequency. Thus, if the picture signal occupies a band of frequencies extending to four megacycles, two side bands are produced which extend four megacycles downward and four megacycles upward from the carrier frequency. The side bands are identical in frequency content, that is, both carry the complete picture signal information. One of the side bands is superfluous and, if transmitted, would wastefully consume space in the broadcast spectrum. Therefore,

the major portion of one of the side bands is removed, and the other is transmitted in full. Complete removal of the superfluous side band is possible, but this would complicate receiver design. Hence, a vestige of the unwanted side band is retained to serve the over-all economy of the system. This technique is known as vestigial side-band transmission. It is now universally employed in all the television broadcast systems of the world.

The television channel contains the picture carrier frequency, one complete picture side band and a vestigial portion of the other picture side band. In addition, the carrier for the sound transmission and its side bands is included within the channel. Since the band of frequencies needed to convey the sound is very much narrower than that needed for the picture, it is feasible to include both sound-carrier side bands. To avoid mutual interference between sound and picture, the picture and sound side bands must not overlap. Moreover, some space must be allowed at the edge of the channel to avoid interference with the transmissions of stations occupying the adjacent channels.

These requirements are met in the television channel shown in fig. 17. Each contains the following bands: 4.5 mc. for the fully transmitted picture side band, 1.25 mc. for the vestige of the other picture side band, 0.15 mc. for the sound carrier and its two side bands and the remaining 0.1 mc. to guard against overlap between picture and sound and between channels. In compatible colour transmissions an additional colour-information signal, known as the chrominance subcarrier, is included within the fully transmitted picture side band (fig. 23). A special technique (frequency interlacing, described under *Compatible Colour Television*, below) avoids interference between the two signals which then occupy a portion of the channel in common.

The television channels of the U.S. are assigned in the following segments of the spectrum: channels 2, 3 and 4, 54–72 mc.; 5 and 6, 76–88 mc.; 7 through 13, 174–216 mc.; and the UHF channels, 14 through 83, 470–890 mc. These channels are allocated to communities according to a master plan established and administered by the Federal Communications commission. No more than seven channels are provided in any one area (this number is allocated only to New York and Los Angeles). Many smaller cities must be content with one or two channels.

E. The Television Transmitter

The task of the television transmitter is to produce the picture and sound carriers, and their side bands, in conformity with the arrangement of the channel assigned to it. The electronic equipment which performs these functions is shown in fig. 18. The central element is the synchronizing generator, which controls the timing of the scanning of the camera and which produces the synchronizing impulses. It applies three control functions to the camera: timing for horizontal (line-by-line) deflection of the scanning beam; timing for vertical deflection (field-by-field); and the blanking impulses which extinguish the scanning spot during its retrace motions. The camera, under the control of the synchronizing generator, produces the picture signal without the synchronizing impulses. The picture signal is passed first through a preamplifier within the camera housing and then through additional amplifiers in the control room of the studio. In one of these amplifiers, the synchronizing impulses from the synchronizing generator are added to the picture signal. At this point, it is

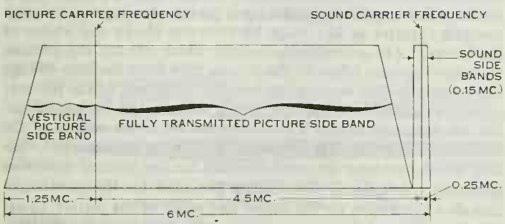


FIG. 17.—ALLOCATION OF TELEVISION CHANNEL FOR MONOCHROME BROADCASTING

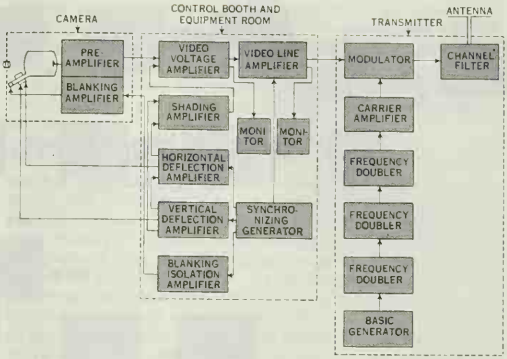


FIG. 18.—DETAILED BLOCK DIAGRAM OF MONOCHROME TRANSMITTER

customary to inspect the quality of the reproduced image by feeding the signal to a monitor which displays the image on a picture tube. The monitor screen is viewed by the program director and operating personnel. Any shortcomings are corrected by instructions telephoned to the camera operator or by adjustment of video controls in the control room. Simultaneously the audio controls are used to adjust the sound signal.

The adjusted picture and sound signals are fed to the network which interconnects stations or, in local broadcasts, to cables leading directly to the transmitter. The latter is usually located at some distance from the studio to permit placing the transmitting antenna in as favourable a location as possible with respect to the intended service area.

The transmitter performs two functions: generation of the carrier currents for picture and sound, and imposition of the sound and picture signals on the respective carrier currents. Precautions are taken to assure that the frequencies of the carrier waves have precisely the values assigned to the channel in use. Small wafers of quartz, which vibrate under electrical forces by virtue of the piezoelectric effect, are used as the basic sources of the carrier currents. The frequency at which quartz can vibrate is lower than that of the intended radio waves, so the current is passed through a succession of frequency-doubling circuits until it reaches the required value. The resulting carrier currents have frequencies accurate to roughly one part in 200,000.

The carrier current and the picture signal are combined in the modulator, which modulates the envelope of the carrier as previously described (fig. 8). The modulated carrier contains two identical side bands. The major portion of one of these (that occupying frequencies below the carrier) is removed by a wave filter, and the remaining part, plus the picture carrier current and the fully transmitted side-band currents, are passed to the transmitting antenna. A similar arrangement (carrier-current generator, frequency multipliers and frequency modulator) produces the sound carrier current and its side bands. These are also fed to the transmitting antenna, which is designed to direct the radiated waves toward the surface of the earth and to minimize radiation toward the sky.

F. The Television Receiver

At the receiver (fig. 19), the sound and picture carrier waves pass through the rods of the receiving antenna, inducing currents which are identical in form to those flowing in the transmitter aerial, but much weaker. The carrier currents and their side bands are conducted from the antenna to the receiver by a lead-in transmission line, typically a ½-in. ribbon of polyethylene plastic in which are embedded two parallel copper wires. This form of transmission line is capable of passing the carrier currents to the receiver, without relative discrimination between frequencies, on all the channels to which the receiver may be tuned.

At the antenna terminals of the receiver, the picture and sound signals are at their weakest, so particular care must be taken to

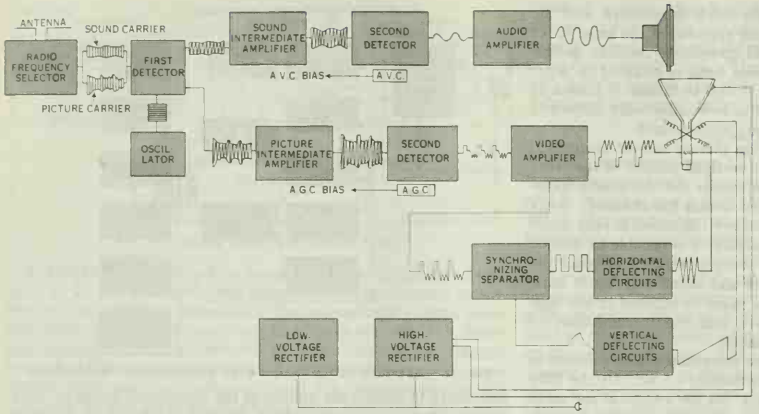


FIG. 19.—BLOCK DIAGRAM OF MONOCHROME RECEIVER, SHOWING WAVE FORM BETWEEN COMPONENTS

control noise at this point. The first circuit in the receiver is a radio-frequency amplifier, particularly designed for low-noise amplification. The channel-switching mechanism (tuner) of the receiver connects one of several individual circuits to this amplifier, each tuned to its respective channel. This amplifier magnifies the voltages of the incoming picture and sound carriers and side bands in the desired channel by about ten times and it discriminates by a like amount against the transmissions of stations on other channels.

From the radio-frequency amplifier, the signals are passed to the superheterodyne mixer, also a part of the tuning mechanism. This is a circuit that transposes the frequencies of the sound and picture carriers to values better suited to the subsequent amplification processes. The transposed frequencies (intermediate frequencies) remain the same, no matter what channel the receiver is tuned to. In typical receivers they are located in the band from 41 to 47 mc.

Since the tuning of the subsequent amplifiers (intermediate-frequency amplifiers) need not be changed as the channel is switched, they can be adjusted for maximum performance in the 41–47-mc. range. Three or four stages of such amplification are used in tandem, increasing the voltage of the picture and sound carriers by a maximum of 25 to 35 times per stage, representing an over-all maximum amplification of the order of 10,000 times. The amplification of these intermediate-frequency stages is automatically adjusted by a process known as automatic gain control in accordance with the strength of the signal, full amplification being accorded to a weak signal, less to a strong signal. After passage through the intermediate amplifiers, the sound and picture carriers and their side bands reach a relatively fixed level of about one volt, whereas the signal levels applied to the antenna terminals may vary, depending on the distance of the station and other factors, from a few millionths to a few tenths of a volt.

From the last intermediate amplifier stage, the carriers and side bands are passed to another circuit, known as the video detector. This removes one-half of each alternation of the signal currents, converting the alternating current into pulsating direct current. An averaging circuit or filter then forms, from the direct-current pulsations (1) a picture signal, which is a close replica of the picture signal produced by the camera and synchronizing generator in the transmitter, and (2) a frequency-modulated sound signal whose variations in frequency match the variations in sound pressure which impinged on the studio microphone. Thus far, the picture and sound signals are handled simultaneously in each amplifier and detector circuit. Both may also be passed through a video amplifier which increases their voltage by about ten times.

At this point the picture and sound signals separate. Another video amplifier passes the picture signal to the picture tube, where it controls the brightness of the scanning spot. Simultaneously, the picture signal wave form is passed to a circuit (the synchronizing signal separator) which selects the blank periods in the wave form (fig. 5 and 6) and the superimposed synchronizing signals.

Finally, the sound signal is passed through a frequency detector (discriminator or ratio detector) which converts the frequency modulations into pulsations representative of the sound pressure variations. This current, an accurate replica of the current produced by the studio microphone, is passed through one or two additional audio-frequency amplifier stages to the loud-speaker, which reproduces the sound waves in conformity to those in the studio.

Meanwhile, back at the synchronizing signal separator, the synchronizing signals are passed through filters which separate the horizontal synchronizing impulses (fig. 5) from the serrated vertical synchronizing impulses (fig. 6). The synchronizing impulses are then passed, respectively, to the horizontal and vertical deflection generators. These produce the currents that flow through the electromagnet deflection coils, causing the scanning spot of the picture tube to be deflected across the viewing screen in the standard scanning pattern.

The synchronizing signals control the timing of these currents, causing the scanning spot of the receiver to move synchronously with that of the camera. Pains are taken in the synchronizing action of the receiver to prevent the scanning motion from being disturbed by interference, since such disturbances would displace the picture elements from their correct positions in the pattern. It is one of the achievements of the modern television system that the scanning pattern is stably reproduced even when the picture signal is so infested with noise that the picture elements themselves are barely recognizable.

In addition to the amplifiers, detectors and deflection generators described above, the receiver contains two power-converting circuits. One of these (the low-voltage power supply) converts the alternating current from the power line into the direct current needed for the vacuum tube circuits; the other (high-voltage power supply) produces the high voltage, typically 15,000 to 18,000 v., needed to create the scanning spot in the picture tube.

The receiver controls commonly provided for adjustment by the viewer are: (1) the channel switch, which connects the required circuits to the radio-frequency amplifier and superheterodyne mixer to amplify and convert the sound and picture carriers of the desired channel; (2) a fine tuning control which precisely adjusts the superheterodyne mixer so that the response of the tuner is exactly centred on the channel in use; (3) a contrast control which adjusts the voltage level reached by the picture signal in the video amplifiers, producing a picture having more or less contrast (greater or less range between the blacks and whites of the image); (4) a brightness control which adjusts the average amount of current taken by the picture tube from the high-voltage power supply, thus varying the over-all brightness of the picture; (5) a horizontal-hold control, which adjusts the horizontal deflection generator so that it conforms exactly to the control of the horizontal synchronizing impulses; and (6) a vertical-hold control, which performs the same function for the vertical deflection generator. Additional adjustments provided for the serviceman include adjustments for the height and width of the picture (magnitudes of the deflection currents) and adjustments (linearity controls) to secure precisely uniform rates of horizontal and ver-

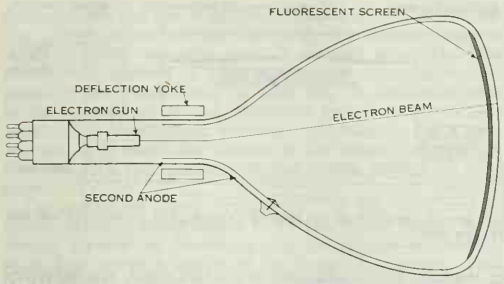


FIG. 20.—BASIC ELEMENTS OF MONOCHROME PICTURE TUBE

tical scanning so as to avoid distortion of the shapes of the objects in the picture.

G. Television Picture Tubes

The television picture tube is illustrated in fig. 20. The tube is a highly evacuated, funnel-shaped structure. The viewing screen is located inside the tube face, a curved glass plate which closes the wide end of the funnel. The screen itself is composed of fluorescent materials, a mixture of two complex chemical compounds such as silver-activated zinc sulfide and silver-activated zinc cadmium sulfide. These materials, known as phosphors, glow with blue and yellow light, respectively, under the impact of high-speed electrons. The phosphors are mixed, in a fine dispersion, in such proportion that the combination of yellow and blue light produces white light of slightly bluish cast. A water suspension of these materials is settled on the inside of the face plate of the tube during manufacture, and this coating is overlaid with a film of aluminum sufficiently thin to permit the bombarding electrons to pass without hindrance. The aluminum provides a mirror surface which prevents the backward-emitted light from being lost in the interior of the tube and reflects it forward to the viewer.

At the opposite end of the tube is the electron gun (fig. 21), a cylindrical metal structure which generates and directs a stream of free electrons, the electron beam. At the rear of the electron gun is the cathode, a flat metal support covered with oxides of barium and strontium. These oxides have a low electronic work function; when heated by the heater coil behind the metal support they liberate electrons. In the absence of electric attraction, the free electrons form a cloud immediately in front of the oxide surface.

Directly in front of the cathode is the cylindrical sleeve of the gun, which is made electrically positive with respect to the cathode. The positively charged sleeve (first anode) draws the negative electrons away from the cloud and they move down the sleeve toward the viewing screen at the opposite end of the tube. They are intercepted, however, by a flat disk (the control electrode) having a small circular aperture at its centre. Some of the moving electrons pass through the aperture; others are held back.

The television picture signal is applied, from the video amplifier previously mentioned, between the control electrode and the cathode. During those portions of the picture signal wave form which make the potential of the control electrode less negative, more electrons are permitted to pass through the control aperture,

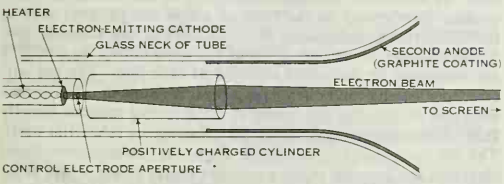


FIG. 21.—ELEMENTS OF PICTURE TUBE ELECTRON GUN

whereas during the more negative portions of the wave, fewer electrons pass. The brightness control applies a steady (but adjustable) voltage between control electrode and cathode. This voltage determines the average number of electrons passing through the aperture, whereas the picture signal causes the number of electrons passing the aperture to exceed or fall below the average.

From the aperture, the controlled stream of electrons passes down the anode sleeve into the glass neck of the tube. Inside the latter is a graphite coating, which extends throughout the funnel of the tube and connects to the aluminum back coating of the phosphor screen. The full value of positive high voltage (typically 15,000 v.) is applied to this coating and it therefore attracts the electrons from the sleeve along the neck into the funnel and toward the screen of the tube. As the electrons emerge from the sleeve and come under the influence of the graphite coating, each electron experiences a force which directs it toward the centre of the viewing screen. The electron beam is thus brought to focus on the screen and the light there produced is the scanning spot. Additional focusing forces may also be provided by an adjustable permanent magnet surrounding the neck of the tube, which introduces magnetic lines of force parallel to the direction of the electron beam. The scanning spot must be intrinsically very brilliant since (by virtue of the integrating property of the eye) the light in the spot is effectively spread out over the whole area of the screen during scanning.

Scanning is accomplished, as previously outlined, by two sets of electromagnet coils wound on a core of magnetic material. These coils must be precisely designed to preserve the focus of the scanning spot no matter where it falls on the screen, and the magnetic fields they produce must be so distributed that the deflections across each line and from line to line occur at uniform velocities.

Deflection of the beam occurs by virtue of the fact that an electron in motion through a magnetic field experiences a force at right angles both to its direction of motion and to the direction of the magnetic lines of force. The deflecting magnetic field is passed through the neck of the tube at right angles to the electron-beam direction and the beam incurs a force tending to change its direction at right angles to its motion, the amount of the force being proportional to the amount of current flowing in the deflecting coils.

To cause uniform motion along each line, the current in the horizontal deflection coil, initially negative, becomes steadily smaller, reaching zero when the spot passes the centre of the line and then increasing in the positive direction until the end of the line is reached. The current is then reversed and very rapidly goes through the reverse sequence of values, bringing the scanning spot to the beginning of the next line. The rapid rate of change of current during the retrace motions causes pulses of a high voltage to appear across the circuit feeding current to the coil, and the succession of these pulses, smoothed into direct current by a rectifier tube, serves as the high-voltage power supply.

A similar action in the vertical deflection coils produces the vertical scanning motion. The two sets of deflection coils are combined in a structure known as the deflection yoke which surrounds the neck of the picture tube, directly at the junction of the neck with the funnel section.

Picture tubes vary in screen size from 8 to 27 in., as measured diagonally across the tube face. The 17-in. size is most widely used in portable receivers, the 21-in. size in table and console model receivers. The design trend has called for greater angles of deflection, with correspondingly wider funnel sections and shallower over-all depth from electron gun to viewing screen. The increase in deflection angle from 55° in the first (1946) models to 110° in models produced in the early 1960s required corresponding refinement of the deflection system because of the higher deflection currents required and the greater tendency of the scanning spot to go out of focus at the edges of the screen.

IV. COMPATIBLE COLOUR TELEVISION

The technique of compatible colour television comprises two

transmissions. One of these, the luminance transmission, employs methods essentially identical to those of the monochrome television system. The second, the chrominance transmission, has virtually no effect on monochrome receivers. When used with the luminance transmission in a colour receiver, however, it produces an image in full colour.

The luminance-chrominance method of representing colour values is one of several alternative ways in which coloured light may be analyzed and synthesized. This method is particularly appropriate in a television system since it produces a compatible signal which can serve both black-and-white and colour receivers by one and the same broadcast. Historically, compatibility was of great importance because it allowed colour transmissions to be introduced without obsolescence of the many millions of monochrome receivers in use. In a larger sense, the luminance-chrominance method of colour transmission is advantageous because it utilizes the limited channels of the radio spectrum more efficiently than other colour transmission methods.

To create the luminance-chrominance values, it is necessary (in the present state of the art) first to analyze each colour in the scene into primary colours. Coloured light may thus be analyzed by passing the light through three coloured filters, typically red, green and blue. The amounts of light passing through each filter, plus a description of the colour transmission properties of the filters, serve uniquely to characterize the coloured light.

The fact that virtually the whole range of colours may be synthesized from only three primary colours is essentially a description of the process by which the eye and mind of the observer recognize and distinguish colours. This process, like visual persistence, is a very fortunate property of vision, since it permits a simple three-part specification to represent any of the 10,000 or more colours and brightnesses which may be distinguished by human vision. If vision were dependent on the energy *v.* wave length relationship (the physical method of specification), it is doubtful if colour reproduction could be incorporated in any mass-communication system.

By transforming the primary-colour values, it is possible to specify any coloured light by three other numbers: (1) its luminance (brightness or "brilliance"); (2) its hue (the redness, orangeness, blueness or greenness, etc., of the light); and (3) its saturation (its vivid *v.* pastel quality). If the intended luminance value of each point in the scanning pattern is transmitted by the methods of the monochrome television system, it is only necessary

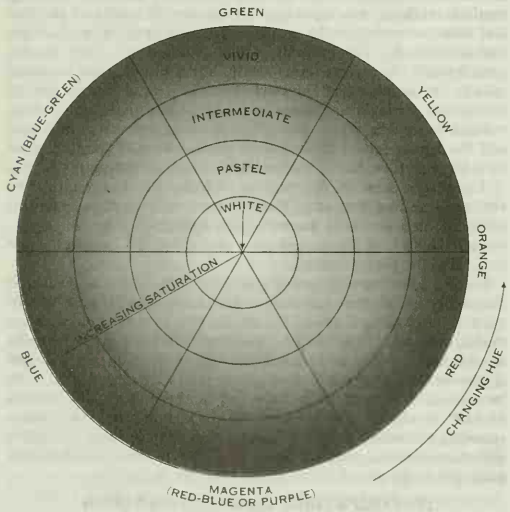


FIG. 22.—HUE AND SATURATION (CHROMINANCE) VALUES ARRANGED IN POLAR CO-ORDINATES

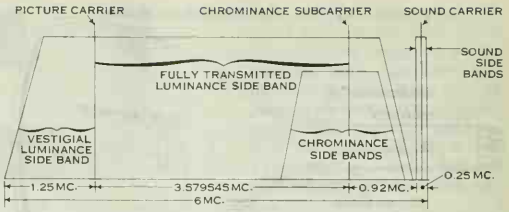


FIG. 23.—ALLOCATION OF TELEVISION CHANNEL FOR COMPATIBLE COLOUR TRANSMISSION

to transmit, via an additional two-valued signal, supplementary information giving the hue and saturation of the intended colour at the respective points in the scanning pattern.

Chrominance, defined as that part of the colour specification remaining when the luminance is removed, represents two independent quantities, hue and saturation. These may be represented graphically in polar co-ordinates (fig. 22) with saturation as the radius and hue as the angle. The centre of the diagram represents white light (the colour of zero saturation). Points on the outermost circle represent the most saturated colours, arranged counterclockwise in the order in which they appear in the spectrum from red to blue. Points on any radius of the diagram represent all colours of the same hue, the saturation becoming less (the colour less vivid or more pastel) as the point approaches the central "white point." A diagram of this type is the basis of the international standard system of colour specification.

In the compatible colour television system, the chrominance signal is an alternating current of precisely specified frequency (3.579545 ± 0.000010 mc.), the precision permitting its accurate recovery at the receiver even in the presence of severe noise or interference. Any change in the amplitude of its alternations at any instant corresponds to a change in the saturation of the colours being passed over by the scanning spot at that instant, whereas a shift in time of its alternations (a change in the phase angle of the alternations) similarly corresponds to a shift in the hue.

As the different saturations and hues along each scanning line are successively uncovered by scanning in the camera, the amplitude and phase, respectively, of the chrominance signal change accordingly. The chrominance signal is thereby simultaneously modulated in amplitude and in phase. This doubly modulated signal is imposed on the picture signal carrier current, along with the luminance signal.

The television channel, when occupied by such a compatible colour transmission, appears as shown in fig. 23 (compare this with the channel when occupied by a black-and-white transmission, fig. 17). The chrominance signal takes the form of a subcarrier located precisely 3.579545 mc. from the picture carrier frequency. The picture carrier is thus simultaneously amplitude modulated by the luminance signal to represent changes in the intended luminance, and by the chrominance subcarrier which, in turn, is amplitude modulated to represent changes in the intended saturation and phase modulated to represent changes in the intended hue.

When compatible colour transmissions are received on a monochrome receiver, the receiver treats the chrominance subcarrier as though it were a part of the intended monochrome transmission. If steps were not taken to prevent it, the subcarrier would produce interference in the form of a fine dot pattern; that is, the subcarrier would change the brightness of the scanning spot 3,580,000 times per second, or about 190 times during the 0.000053 sec. consumed by the active scanning of each line. The dot pattern, fortunately, can be rendered almost invisible in monochrome reception by deriving the timing of the scanning motions directly from the source that establishes the chrominance subcarrier itself. The dot pattern of interference from the chrominance signal therefore can be made to have opposite effects on successive scannings of the pattern; that is, a point brightened by the dot inter-

ference on one line scan is darkened an equal amount on the next scan of that line, and the net effect of the interference, integrated in the eye over successive scans, is virtually zero. Thus, the monochrome receiver in effect ignores the chrominance component of the transmission. It deals with the luminance signal in the conventional manner, producing from it a monochrome image. This monochrome rendition, incidentally, is not a compromise; it is essentially identical to the image that would be produced by a monochrome television system viewing the same scene.

The channel for colour transmissions (fig. 23), when used by colour receivers, would appear to be affected by mutual interference between the luminance and chrominance components, since these occupy a portion of the channel in common. Such interference is avoided by the fact that the chrominance subcarrier component is rigidly timed to the scanning motions. The luminance signal, as it occupies the channel, is actually concentrated in a multitude of small spectrum segments, by virtue of the periodicities associated with the scanning process. Between these segments are empty channel spaces of approximately equal size. The chrominance signal, arising from the same scanning process, is similarly concentrated. Hence it is possible to place the chrominance channel segments within the empty spaces between the luminance segments, provided that the two sets of segments have a precisely fixed frequency relationship. The necessary relationship is provided by the direct control by the subcarrier of the timing of the scanning motions. This intersegmentation is known as frequency interlacing. It is one of the fundamentals of the compatible colour system, and without it the superposition of colour information on a channel originally devised for monochrome transmissions would not be feasible.

When a colour receiver is tuned to the transmission represented in fig. 23, the picture signal is recovered in a video detector in the usual manner. An amplifier stage, tuned to the 3.58 mc. chrominance frequency, then selects the chrominance from the picture signal and passes it to a synchronous detector which recovers independently the amplitude-modulated and phase-modulated components. The synchronous detection process recovers the phase-modulated component by measuring the signal alternations against a reference timing signal transmitted during a part of the retrace motion between scanning lines. The detector is also simultaneously responsive to the amplitude-modulated components, in the manner of the video detector in monochrome reception.

A. Colour Camera and Picture Tube

The terminals of the colour television system, the camera and picture tube, analyze and synthesize, respectively, the colour quantities present at each point in the scanning pattern. The television camera in its current state of development does not produce the luminance, hue and saturation values directly. Rather it produces three picture signals, representative of the amounts of three primary colours present at each point in the image pattern. From these signals the luminance and chrominance components are then derived by manipulation in electronic circuits.

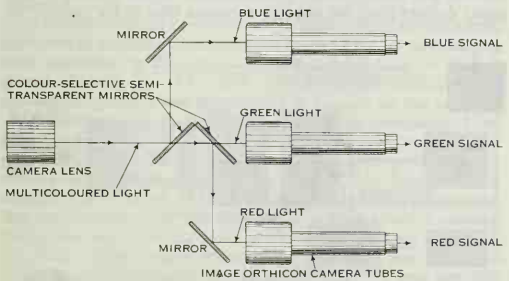
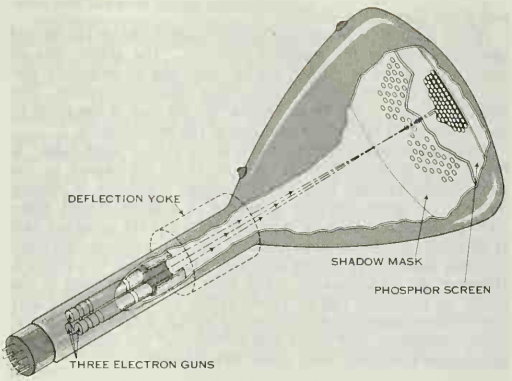


FIG. 24.—SCHEMATIC DIAGRAM SHOWING SELECTIVE SEPARATION OF LIGHT IN COLOUR CAMERA



BY COURTESY OF RADIO CORPORATION OF AMERICA

FIG. 25.—COLOUR PICTURE TUBE EMPLOYING THE SHADOW-MASK PRINCIPLE

The typical colour television camera (fig. 24) contains three image orthicon tubes, with an optical system which casts an identical image on the sensitive surface of each tube. The optics comprise a lens and four mirrors which reflect the image rays from the lens onto the three camera tubes, as shown. Two of the mirrors are of a colour-selective type (dichroic mirror) which reflects light of one colour and transmits the remaining colours. The mirrors, plus colour filters which perfect their colour-selective action, direct a blue image to the first tube, a green image to the second tube and a red image to the third tube.

The deflection systems of the three camera tubes are arranged to produce essentially identical scanning patterns in each tube, so the picture signals developed by the respective tubes represent images of the same geometrical shape, differing only in colour. The respective primary-colour signals are passed through video preamplifiers associated with each camera tube and emerge from the camera as separate entities. If compatibility were not required, and if sufficient space were available in the radio spectrum, the three signals might then be transmitted to the receiver over three separate channels. Actually, as mentioned above, the primary-colour signals are recast into luminance and chrominance components, and at the receiver the latter are converted back into primary-colour signals before application to the colour picture tube.

The colour picture tube (fig. 25) contains three electron guns which produce three separate electron beams. These are deflected simultaneously in the standard scanning pattern over the viewing screen. One of the beams is controlled by the red primary-colour signal and produces a red image. The second beam produces a blue image and the third a green image.

To permit three primary-colour images to be formed simultaneously, the screen is composed of three sets of individual phosphor dots, which glow respectively in the three different colours and which are uniformly interspersed over the screen (fig. 25). The "red" beam impinges only on the red-glowing phosphor dots, being prevented from hitting the other two colours, with similar restriction of the other two beams to the blue and green dots respectively.

The sorting out of the three beams so that they produce images of only the intended primary colour is performed by a mask which lies directly behind the phosphor screen. This mask contains about 200,000 precisely located holes, each accurately aligned with three different coloured phosphor dots on the screen in front of it. Electrons from the three guns pass together through each hole, but at slightly different angles. The angles are such that the electrons arising from the gun controlled by the red primary-colour signal fall only on the red dots, being prevented from hitting the blue and green dots by the shadowing action of the mask. Similarly, the "blue" and "green" electrons fall only on the blue

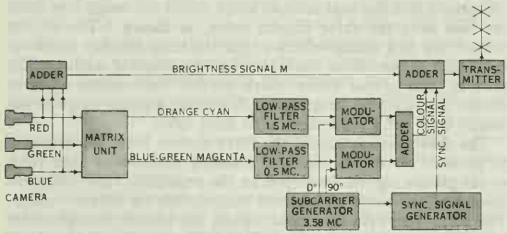
and green dots, respectively.

In this manner three separate primary-colour images are formed simultaneously, as the scanning proceeds. The colour dots of which each image is formed are so small and so uniformly dispersed that the eye does not detect their separate presence. The primary colours in the three images thereby mix in the mind of the viewer and a full-colour rendition of the image results.

This type of colour tube is known as the shadow-mask tube. It has several shortcomings: (1) the electrons intercepted by the mask cannot produce light and the image brightness is thereby limited; (2) great precision is needed to achieve correct alignment of the electron beams, the mask holes and the phosphor dots at all points in the scanning pattern; and (3) precisely congruent scanning patterns, as among the three beams, must be produced. However, no other type of colour tube had been offered to the public by the early 1960s. Other colour tubes, employing only one electron beam which sequentially excites stripes of primary-colour phosphors, and hence free of many of the alignment and congruence problems, were under development.

B. Colour Transmitter and Receiver

Fig. 26 shows the essential elements of the colour transmitter, as they differ from the monochrome transmitter shown in fig. 18. Immediately following the colour camera is the colour coder which converts the primary-colour signals into the luminance and chrominance signals. The luminance signal is formed simply by applying the primary-colour signals to an electronic addition circuit which adds the values of the three signals at each point along the respective picture signal wave form. Since white light results from the addition, in appropriate proportions, of the primary colours, the resulting sum signal represents the black-and-white (luminance) version of the colour image. The luminance signal thus formed is subtracted individually, in three electronic subtraction circuits, from the original primary-colour signals, and the colour-difference signals are then further combined to produce two signals, which represent points on the orange-cyan and green-magenta axes of the chrominance diagram (fig. 22). These signals are then applied simultaneously to a modulator, where they are mixed with the chrominance subcarrier signal. The latter is thereby amplitude modulated in accordance with the saturation values and phase modulated in accordance with the hues. The luminance and chrominance components are then combined to form the over-all colour picture signal (fig. 27), which is then carried to the transmitter over the network or studio-transmitter link.



BY COURTESY OF RADIO CORPORATION OF AMERICA

FIG. 26.—BLOCK DIAGRAM OF COLOUR TRANSMITTER

The chrominance subcarrier is generated in a precise electronic oscillator at the standard value of 3.579545 mc. Samples of this subcarrier are injected into the signal wave form during the blank period between line scans just after the horizontal synchronizing pulses. These samples are employed in the receiver to control the synchronous detector, mentioned above. Finally, the horizontal and vertical deflection currents, which produce the scanning in the three camera tubes, are formed in a scanning generator whose timing is controlled by the chrominance subcarrier. This common timing of deflection and chrominance transmission produces the dot-interference cancellation in monochrome reception, noted above.

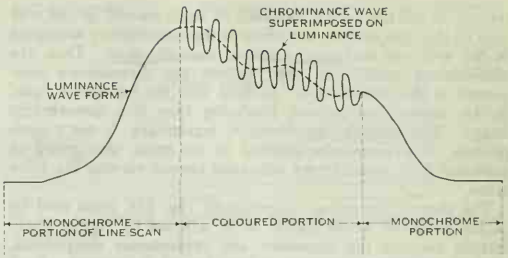


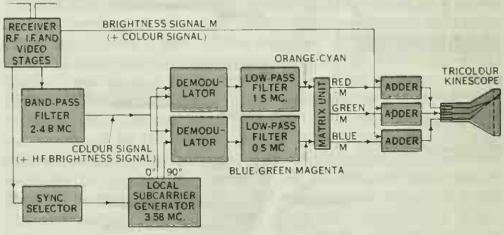
FIG. 27.—PICTURE SIGNAL WAVE FORM FOR COLOUR TRANSMISSION

At this stage, the colour picture signal can be handled in the same manner as the monochrome signal, except that throughout the network and in the transmitter special precautions must be taken to transmit the chrominance subcarrier at its correct amplitude, with its amplitude and phase modulations preserved against incidental distortions.

The colour receiver (fig. 28) contains a tuner and intermediate-frequency amplifiers similar to those used in monochrome receivers, but specially designed to preserve the chrominance subcarrier during its passage through these stages. The colour receiver also has power supplies, sound reception circuits and synchronization and deflection current generators similar to those used in monochrome reception.

The colour reproduction processes are divided into the luminance and chrominance functions. A video detector develops the luminance component and applies it through video amplifiers simultaneously to all three electron guns of the picture tube. This part of the signal thereby activates all three primary-colour images, simultaneously and identically, in the fixed proportion needed to produce white light. When tuned to monochrome broadcasts, the colour receiver produces a monochrome image by this mechanism, the chrominance component then being absent.

When the receiver is tuned to a colour broadcast, the chrominance subcarrier component appears in the output of the video detector and it is thereupon operated on in circuits which ultimately recover the primary-colour signals originally produced by the colour camera. The recovery starts in the synchronous detector. This is followed by circuits which perform the inverse operations of the addition and subtraction circuits at the transmitter. The end result of this manipulation is the production of three colour-difference signals which represent, respectively, the difference between the luminance signal, already applied to all three electron guns of the picture tube, and the primary-colour signals. The three colour-difference signals are applied separately to the respective electron guns, where each reduces the strength of the corresponding electron beam to change the white light, which would otherwise be produced, to the intended colour for each point in the scanning line. The net control signal applied to each electron gun bears a direct correspondence to the primary colour signal derived from the respective camera tube at the studio. In



BY COURTESY OF RADIO CORPORATION OF AMERICA

FIG. 28.—BLOCK DIAGRAM OF COLOUR RECEIVER

this manner, the three primary-colour signals are transmitted as though three separate channels had been used.

The colour receiver controls to be adjusted by the viewer (in addition to the tuning, brightness, contrast and synchronizing controls present in monochrome receivers) are the hue control and the saturation control. The hue control adjusts the timing of the synchronous detector. It sets the reference phase against which the phase modulation of the chrominance subcarrier is measured, and hence shifts all the hues in the reproduced image. The hue control is usually adjusted by reference to the colour of the flesh of the performers, since unnatural flesh tone is readily recognized even if the viewer has no knowledge of the intended hues in other parts of the scene. The saturation control adjusts the magnitudes of the colour-difference signals applied to the electron guns of the picture tube. If these magnitudes are reduced to zero, by turning the saturation control to the "off" position, no colour difference action occurs and the reproduction occurs in monochrome. As the saturation control is advanced, the colour differences become more accentuated, and the colours become progressively more vivid. This control is adjusted to suit the personal preference of the viewer. Natural appearance of flesh tone is an appropriate index.

V. TELEVISION RECORDING

Recording of television programs on photographic film or magnetic tape is an important technique, not only to preserve a permanent record of a live-scene program for subsequent rebroadcast, but also to perfect a performance, prior to its initial presentation, by the photographing and editing methods of the motion-picture industry. Recorded segments are often used in a "live" studio program when it is impractical to set up the desired scene in the studio. Transmission of programs from film originally produced for motion-picture theatres is also a major activity in broadcasting.

Television film projectors fall into two classes, continuous and intermittent, according to the type of film motion. In the continuous projector, a scanning spot from a flying spot camera tube is passed through a rotating optical system, known as an immobilizer, which focuses the scanning spot on the motion-picture film. As the film moves continuously through the projector, the immobilizer causes the scanning pattern as a whole to follow the motion of the film frame, so that there is no relative motion between pattern and frame. The light passes through the film to a phototube where the light, modified by the transmissibility of the film at each point, produces the picture signal. As one film frame moves out of the range of the immobilizer, the next moves into range and there is a condition of overlap between successive scanning patterns.

The optics are so arranged that the amount of light in the spot focused on the film is constant at all times and in all positions. This constancy permits the film to be moved at any desired speed, while the pattern scans at the standard rate of 25 or 30 pictures per second. The film is actually moved at the standard rate for motion pictures, 24 frames per second, so the speed of objects and pitch of the accompanying sounds (picked up from the sound track by conventional methods) are reproduced at the intended values.

In the intermittent projector, which more nearly resembles the type used in theatre projection, each frame of film is momentarily held stationary in the projector while a brief flash of light is passed through it. The light (which passes simultaneously through all parts of the film frame) is focused on the sensitive surface of a storage-type camera, such as the iconoscope. The light flashes are timed to occur during the intervals between successive field scans, that is, while the extinguished scanning spot is moving from bottom to top of the frame. The light is strong enough to produce an intense electrical image in the tube during this brief period. The electrical image is stored and is scanned during the next scanning field, producing the picture signal for that field. Light is again admitted between fields and the stored image scanned thereupon by the second field. When one film frame has been thus scanned, it is pulled down by a claw mechanism and the next frame takes its place.

In Europe and other localities where the television scanning rate is 25 picture scans per second, the custom is to operate intermittent projectors also at 25 frames per second, or about 4% faster than the intended film projection rate of 24 frames per second. The corresponding increases in speed of motion and sound pitch are not so great as to introduce unacceptable degradations of the performance. In the U.S. and other areas where television scanning occurs at 30 frames per second, it is not feasible to run the film projector at 30 film frames per second, since this would introduce speed and pitch errors of 25%. Fortunately, a small common factor, 6, relates the scan rate of 30 and the film projection rate of 24 frames per second. That is, four film frames consume the same time ($\frac{4}{24} = \frac{1}{6}$ sec.) as five scanning patterns ($\frac{5}{30} = \frac{1}{6}$ sec.). Thus, if four film frames pass through the projector while five picture scans are completed, both the film motion and the scanning proceed at the standard rates. The two functions are kept in step by holding one film frame for three scanning frames, the next frame for two scans, the next for three scans and so on.

Recording of television programs on film involves similar techniques, but in reverse. The television image to be recorded is reproduced at high brilliance on a high-quality monitor picture tube having a phosphor producing highly actinic light. The television image is focused optically on film in a special form of motion-picture camera, which in effect discards six television scans per second, thus yielding a recording at the standard film rate of 24 frames per second. The extra time thus made available is used to move the film from frame to frame in the camera. The aberrations and halation of the high-brightness picture tube, compounded with the imperfections of scanning, film exposure and processing, make the quality of television film records (kinescopes or "kines") substantially poorer than that of films made by direct photography of the same subject matter. The degradation is evident on home receivers during rebroadcast of kinescope recordings.

Most television program recording is done on magnetic tape. This technique produces rebroadcast programs which, when the equipment is in good adjustment, are virtually indistinguishable from the original broadcast. The magnetic tape, 2 in. wide, is moved at 15 in. per second past a magnetic recording head which rotates across the width of the tape, imposing on it a series of crosswise magnetic line patterns which correspond to the lines of the scanning pattern. To keep the pickup and tape motions closely identical during recording and playback, index signals are imposed during recording and these are used to control the drive motors during playback. The television signal is recorded by frequency modulation on a carrier signal, thus avoiding the difficult problem of direct recording at the extremely low frequencies (30 c.p.s. and less) present in the picture signal itself.

Magnetic video recording is widely used in broadcast operations, not only because of its high quality but because it is immediately available (without subsequent processing) for playback. In the U.S. many of the network programs broadcast in and west of Chicago are derived from magnetic recordings, stored for an hour or more after their release along the eastern seaboard to allow for the difference in time in western cities. Recording of colour programs is also possible on magnetic tape, but in the early 1960s was still subject to somewhat lower quality than black-and-white recording, since the necessary precision of tape transport and pickup rotation is greater in the colour system. *See* also TAPE RECORDING, MAGNETIC.

VI. TELEVISION BROADCASTING

The methods and equipment employed in television broadcasting follow the basic principles outlined above, but differ in detail according to the nature of the program source; *i.e.*, the studio, film or a point remote from the studio.

1. Studio Programs.—Television studios, like those for sound broadcasting, are designed with acoustical properties which produce a "live" sound without noticeable reverberation. The studio must also possess a comprehensive system of lighting, capable of creating uniform illumination of several hundred to a thousand

foot-candles for monochrome productions.

For colour programs, two to three times as much light is required to overcome the losses in the mirrors and filters within the camera. In addition to general lighting, localized lighting fixtures ("scoops" and spotlights) are used to emphasize the centre of interest, for modeling of facial contours and the like. A ventilating system capable of removing the heat generated by the lighting system without noise must be provided and air conditioning may be needed if the lighting system is to be used for extended periods in the summer months.

The cameras and microphones are mounted so that they can follow the action on the set. The camera usually is carried on a dolly equipped with smooth-rolling wheels and is mounted on an elevating and training stand which permits the camera operator to achieve the desired angle and position relative to the action. The microphone is mounted on a boom or long arm which can position the microphone above the action, just out of view of the camera. The floor crew (camera operator, a man to assist the operator in moving the camera dolly, the microphone boom operator and the director's floor assistant) are equipped with earphones through which instructions are relayed from the control room.

Studio scenery for television productions is generally simpler than that used in motion-picture production. Since the production occurs in "real time" the cameras must be strictly co-ordinated with the action and must move from one set to another, on the floor of a single studio, as the presentation proceeds. In live broadcasts, retakes cannot be made, and close co-ordination of camera and microphone action with the script must be achieved in rehearsal.

Control of the camera is achieved in part by telephoned instructions from the control room, but the main reliance for correct composition and focus is placed on the cameraman, who views the picture picked up by his camera on an electronic view finder. This is a compact picture monitor connected to the camera tube. The camera is usually equipped with a lens turret containing as many as four lenses of different focal lengths. The turret is rotated by the camera operator (while the camera is not in action) to bring into play the lens having the focal length appropriate for the desired area of the set. The lenses are similar to those used in 35-mm. professional motion-picture cameras, with maximum apertures of the order of f/4.5 to f/3.5. Small apertures are used wherever possible to secure the necessary depth of focus.

2. Film Broadcasting.—To permit uninterrupted programing, two film projectors are needed, one of which is being readied for the next reel while the other is on the air. A single camera tube may accept images from either of the projectors by use of a mirror set at 45° to the respective optic axes (fig. 29). The picture and sound signals produced by film equipment are viewed and auditioned on monitors and are adjusted by video- and audio-control operators, as in studio productions. When an iconoscope camera tube is used, the video-control operator must adjust special controls to minimize the flare and uneven shading to which this type of tube is subject. Televising colour film requires that the coloured image be broken up into three primary-colour images, as in the studio camera for colour productions (fig. 24). Vidicon camera tubes or flying spot cameras are used.

3. Outside Broadcasts.—To televise sporting events and other programs at locations remote from the studio, portable equipment is used. This includes cameras, microphones and several units which house the control and switching system, monitors and the synchronization generator. The portable synchronization unit, which performs all the functions of the more elaborate generators permanently installed in the studio building, must meet rigid standards since it must correctly control the scanning of millions of receivers and is often dependent on uncertain power sources at remote locations.

Remote broadcast signals may be fed directly into a network connection, but more usually they are transmitted to the associated station by a microwave radio system. This comprises a low-power transmitter operating at a frequency in the vicinity of 2,000 or 7,000 mc., using frequency modulation for the picture transmission and a receiver at the studio or transmitting station.

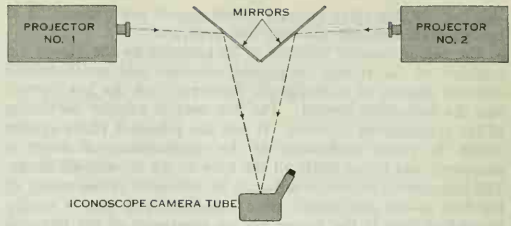


FIG. 29.—ARRANGEMENT FOR USING EITHER OF TWO FILM PROJECTORS WITH ONE CAMERA TUBE

At these extremely high frequencies, a highly directive radio beam may be produced by reflecting the radio waves from a paraboloid surface ("dish" antenna) of modest size. The radio energy is thereby so concentrated that a transmitter power of less than one watt is capable of covering distances of 10 to 20 mi. so long as the path from transmitter to receiver is unobstructed.

4. Networks.—National distribution of major television programs is achieved through a comprehensive common-carrier network system. Three types of facilities are used. Within a given city, special telephone lines ("video pair" circuits) are provided; these are specially equalized to provide uniform transmission over the wide band of frequencies occupied by the picture signal.

For intercity connections, coaxial cables were once widely used, although these were largely replaced by the microwave relay system. The coaxial cable consists of two electrical conductors, a hollow sheath and a wire located concentrically within it. This configuration of conductors is capable of transmitting a wide band of frequencies without discrimination over a distance of several miles. At intervals of about eight miles, video amplifier repeaters are used to restore the signal energy lost in the preceding section of cable. The early coaxial cable network in the U.S. was limited to a maximum frequency range of about 2.7 mc., substantially less than the 4-mc. maximum actually present in the picture signal. Later coaxial cable installations, with repeater amplifiers spaced every four miles, could handle a maximum frequency range of nearly 8 mc.

Major reliance for television network transmission in the U.S. is placed on the microwave relay system. This consists of a large number of relay stations (over 100 are needed to span the country) operating in the vicinity of 4,000 mc. Highly directive transmitting and receiving antennas, mounted on towers in exposed locations at intervals of about 25 mi., receive and retransmit to the next relay station a frequency-modulated picture signal.

5. Transoceanic Broadcasts.—The initial attempt to interconnect the television networks of Europe and North America came in 1962, when the American Telephone and Telegraph company used its artificial satellite, called Telstar, to relay television signals between Andover, Me., Goonhilly Downs, Cornwall, Eng., and Pleumeur-Bodou, Brittany, France. The first transmission, of a purely experimental nature, originated in the U.S. on July 10, 1962, and this was followed the next day by transmissions to the U.S. from France and England; the first colour transmission occurred on July 16. Reception was limited to about 15 min., the period during which the satellite was within sight of the sending and receiving stations. To maintain continous transmissions, the planners of the system proposed using a series of satellites, so that at least one would always be in position to relay signals. An alternate proposal called for a single relay satellite in a "stationary" orbit, so adjusted that it would remain always above the same point on the surface of the Earth. The first public demonstration of this system was on Oct. 10, 1964, when television coverage of the opening ceremonies of the Olympic games was relayed from Tokyo to North America via a Syncom satellite positioned above the Pacific ocean.

VII. CLOSED-CIRCUIT TELEVISION

The term closed-circuit television, which originally meant a

local system consisting of a camera and several monitor receivers connected by a coaxial cable, came to refer to any system, however extensive, in which the televised subject matter is intended for the organization that owns or has hired the system for its exclusive use, as against broadcasting or other general use by the public. Such systems are widely used in industry and commerce and for educational purposes. They range from a single camera and monitor installed in a steam plant to monitor flame condition or water level to a nationwide network hired on a single-occasion basis (*e.g.*, by an automobile manufacturer to present information on new models to dealers assembled in theatres having projection television equipment). Coaxial cable is used in local installations, but microwave relay systems identical to those used in broadcasting may be employed for large closed-circuit hookups.

Since a closed-circuit system is entirely under the control of its proprietor and is not intended to serve the general public, it is possible to employ scanning standards which differ from those used in broadcasting. Advantage of this freedom in system design is taken in the so-called slow-scan closed-circuit system, which transmits signature records and balances in banks, for example. In the typical slow-scan system, the pictures may be scanned at a rate as slow as one picture every three seconds, with a scanning pattern of 180 lines per picture. The maximum frequency in the picture signal then reaches only 8,000 c.p.s., which can be transmitted inexpensively; *i.e.*, over ordinary telephone lines. In the majority of closed-circuit systems for industry, education and theater projection, however, the image-quality requirements are essentially identical to those of broadcasting, so the broadcast standards are used.

1. Equipment.—Equipment for closed-circuit television is similar to but generally less elaborate than that used for broadcast purposes. The Vidicon camera tube is used in industrial television almost to the exclusion of other types, since it is less expensive than the other types and more suitable for compact cameras. In educational television, however, image orthicon tubes and associated equipment very similar to broadcast-station units are widely used. Since closed-circuit installations in industry are seldom operated by television specialists, care must be taken in equipment design to make operation and maintenance simple. An important industrial use is in viewing remote or hazardous events in which it is inconvenient or unsafe to have an operator directly in attendance at the camera. For such purposes, controls for remote operation of the camera, including adjustment of lens aperture and focus, are provided.

2. Applications in Education.—Educational television is often conducted on a public-broadcast basis in cities where a channel is available for this purpose and where the expense of the transmitting station can be borne. The simpler and less costly closed-circuit system found increasing favour with educators, however, particularly in larger institutions and in medical schools, where the whole installation can be confined to one building or a few neighbouring buildings.

In education, television's most impressive use is extending the range of a gifted teacher beyond a single classroom. When a broadcast system is available, pupils in many schools may be taught simultaneously by a particularly able teacher prior to a more personal instruction and discussion of the subject matter in individual classes with a local teacher in attendance. While motion pictures may also be used in this way, the televised image commands the pupils' attention to a far greater extent than does film projection. Even when film is used as a program source, the psychological impression on the student is heightened.

Television is also an important teaching aid in bringing to a number of students simultaneously the details of an experiment or operation which otherwise could be viewed only by a single person or small group. Examples include viewing through a microscope and bringing medical students directly into visual contact with operating procedures without interfering with those performing the operation. Use of closed-circuit television for instruction in surgical procedures is widespread; elaborate colour-television installations have been made in medical schools and hospitals, colour being an important adjunct in delineating diagnosis and procedure. An important advantage of television in such instruction is the fact that the final outcome is not known as the demonstration proceeds and this lends an air of expectancy which is lacking when colour motion pictures are used for the same purpose.

3. Industrial Applications.—One of the first uses of closed-circuit television in industry was an installation of boilerwater-level monitors at the Hell Gate electric generating station in New York, made directly after World War II. This application later was extended to the monitoring of flames in burners and smoke issuing from stacks. Here the justification for television, as against photographic methods, is the need for immediate information without the delay of film processing. Other applications involve the safe handling of nuclear materials, jet fuels and other hazardous substances. A stereoscopic form of television, comprising two complete systems which present two images separately to the two eyes of the operator, has been used in conjunction with precise remote-handling equipment for packing radioactive substances, disposing of radioactive wastes and the like. The ability of the television camera to operate in situations unfriendly to human observers (as outside the fuselage of an airplane in flight, to observe operation of landing gear during engineering tests) also accounts for increasing usage in industry.

The radiation hazard brought industrial television into play in X-ray technology. A fluoroscope or other device for translating X-ray intensities into visible light is viewed by a sensitive television camera and the image brought out at a safe distance from the exposure site. Here, again, motion-picture photography might be used, but the television system permits immediate observation without processing delays and avoids the cost of film and processing which might, in the course of a year, outweigh the cost of a simple television system.

Television systems may also be used for surveillance, particularly to assist guards in patrolling restricted areas. Another surveillance application is in reading numbers on freight cars in railway switchyards to aid the yardmaster in making up trains.

4. Commercial Applications.—The primary application of closed-circuit television in commerce, as contrasted with industry, is remote inspection of records in banks and other institutions. In one installation, four branches of a bank were connected with 12 mi. of coaxial cable to permit any teller to call for the signature card or the current balance of any depositor from a central bookkeeping department. Although the initial and maintenance costs of such a system were substantial, the speed and accuracy with which doubtful accounts may be checked justified the use of the system. Extension of this technique to remote reading of blueprints and other records depended upon lowering the cost of interconnecting cables and increasing the detail of images so that, for example, a full 8½ by 11-in. page of typewritten material could be read on a monitor without error.

Techniques akin to those of television are finding increased use in other commercial and industrial fields. One application is a scanning technique whereby printed characters are "read" and identified by comparing the character with each of a series of standard templates or matrices. This may be done by storing the video signal arising from scanning in a memory array in the form of voltages which may be compared with a similar array of voltages in a matrix corresponding to each letter or number to be recognized. Either mechanical or electronic scanning may be used in this technique. *See* also Index references under "Television" in the Index volume.

BIBLIOGRAPHY.—H. A. Chinn, *Television Broadcasting* (1953); R. Bretz, *Techniques of Television Production* (1953); D. G. Fink, *Television Engineering*, 2nd ed. (1952), (ed.); *Television Engineering Handbook* (1957); J. W. Wentworth, *Color Television Engineering* (1955); V. K. Zworykin and G. A. Morton, *Television: the Electronics of Image Transmission*, 2nd ed. (1954). (D. G. F.)

TELFORD, THOMAS (1757–1834), Scottish architect and civil engineer renowned for his building of roads and bridges, was born near Westerkirk in Eskdale, Dumfriesshire, on Aug. 9, 1757, the son of a shepherd. After apprenticeship to a stonemason in Langholm, he worked as a journeyman mason in Eskdale and Edinburgh until 1782, when he traveled to London and was employed in building Somerset house under Sir William Chambers.

At this time his ambition was to become an architect and his self-education was directed to this end. After a brief period of employment at Portsmouth his patron, William Pulteney, obtained for him the post of surveyor of Shropshire in 1786. While so engaged, Telford built churches at Bridgnorth and Madeley, rebuilt the county jail, excavated the ruins of Uriconium (mod. Wroxeter) and constructed many bridges, including three over the river Severn at Montford, Buildwas and Bewdley, the second being of iron.

In 1793 Telford was appointed agent and engineer to the Ellesmere Canal company and forsook architecture for engineering. His two great aqueducts which carry this canal over the Ceiriog and Dee valleys at Chirk and Pontcysyllte showed him to be a master of the structural use of cast iron and brought him national fame. In 1803 he was employed by the government on a development scheme for the Scottish Highlands, where he was responsible for the Caledonian canal, harbour works at Aberdeen, Dundee and elsewhere and the building of over 900 mi. of roads, including many bridges. Telford was next engaged on the improvement of the roads from Chester and Shrewsbury to Holyhead in the course of which he built his two famous suspension bridges over the river Conway and the Menai straits. The threat of railway competition induced canal companies to employ him to carry out improvement works which included a new tunnel at Harecastle, Staffordshire, on the Trent and Mersey canal, the improvement of the Birmingham canal and the construction of a new canal from Wolverhampton to Nantwich. Among Telford's other works were the Gotha canal in Sweden, St. Katherine's dock, London, roads in the Scottish lowlands and the bridges over the Severn at Tewkesbury and Gloucester. On the invitation of the founders, Telford became first president of the Institution of Civil Engineers, the growth of which was largely due to his efforts and prestige. He died in London on Sept. 2, 1834.

BIBLIOGRAPHY.—*Life of Thomas Telford, Written by Himself . . . With a Folio Atlas* (1838); S. Smiles, *Life of Thomas Telford* (1867); A. Gibb, *The Story of Telford: the Rise of Civil Engineering* (1935); L. T. C. Rolt, *Thomas Telford* (1958). (L. T. C. R.)

TELL, WILLIAM. The story of William Tell's skill in shooting at and striking the apple which had been placed on the head of his little son by order of Gessler, the tyrannical Austrian bailiff of Uri, is so closely bound up with the legendary history of the origin of the Swiss confederation that they must be considered together. Both appear first in the 15th century, probably as results of the war for the Toggenburg inheritance (1436–50); for the intense hatred of Austria, increased by its support of the claims of Zürich, favoured the circulation of stories which assumed that Swiss freedom was of immemorial antiquity, while, as the war was largely a struggle between the civic and rural elements in the confederation, the notion that the (rural) Schwyzers were of Scandinavian descent at once separated them from and raised them above the German inhabitants of the towns.

The Tell story is first found in a ballad, the first nine stanzas of which (containing the story) were certainly written before 1474. It is probably to this that Melchior Russ of Lucerne (who began his *Chronicle* in 1482) refers when he excuses himself from giving the story. He narrates how Tell then stirred up his friends against the governor, who seized him and took him by boat to his castle on Lake Lucerne. A storm arose and Tell, on account of his great strength, was given the rudder, on his promise to bring the boat to land. He steered it toward a shelf of rock, called in Russ's time Tell's Platte, sprang on shore, shot the bailiff dead with his crossbow and, returning to Uri, stirred up the strife which ended in the battle of Morgarten. In these two accounts, which form the basis of the Uri version of the origin of the confederation, it is Tell who is the leader. We first hear of the cruelties of the Austrian bailiffs in the forest districts in the *Bernese Chronicle* of Conrad Justinger (1420), who makes no allusion to Tell. The Tell story and the "atrocities" story are first found combined in a manuscript, the *White Book of Sarnen*. They are contained in a short chronicle written between 1467 and 1476, probably about 1470, and based on oral tradition.

The task of smoothing away inconsistencies and rounding off the tale was accomplished by Giles Tschudi (*q.v.*), whose recension was closely followed by Johannes von Müller in his *History of the Confederation* (1780). The final recension of Tschudi's *Chronicle* (1734–36), which differs in many particulars from the original draft preserved at Zürich, tells how Albert of Austria, in order to deprive the forest lands of their ancient freedom, sent bailiffs (among them Gessler) to Uri and Schwyz. Their tyranny resulted in a rising, planned at the Rütli, on Nov. 8, 1307, and led by Werner von Stauffacher of Schwyz, Walter Fürst of Uri, Arnold von Melchthal in Unterwalden, each with ten companions, among whom was William Tell, to expel the oppressors. On Nov. 18 the Tell incident takes place (described according to the *White Book* version), and on Jan. 1, 1308, the general rising. Tschudi thus finally settled the date, which had before varied from 1260 to 1334. He distorts the historical circumstances. In his first draft he speaks of the bailiff as Gryssler —the usual name up to his time, except in the *White Book* and in Stumpff's *Chronicle* of 1548—but in his final recension he calls him Gessler, knowing that this was a real name. Later writers added a few more particulars. Johannes von Müller (1780) described the oath at the Rütli by the three (Tell not being counted in), and threw Tschudi's version into a literary form. Schiller's play (1804) gave the tale of a world-wide renown.

The general result of the researches of various students, J. E. Kopp, Vischer, Rochholz and others, has been to show that a mythological marksman and an impossible bailiff bearing the name of a real family have been joined with confused and distorted reminiscences of the events of 1245–47, in which the names of many real persons have been inserted and many unauthenticated acts attributed to them. T. von Liebenau has, however, shown (in an article reprinted from the *Katholische Schweizerblätter* in the *Bollettino Storico della Svizzera Italiana* for 1899) that in 1283 the Emperor Rudolf of Habsburg gave the right of receiving the tolls for escort over the St. Gotthard pass to his sons, the dukes of Austria. The levying of these tolls gave rise to various disputes between the men of Uri and the bailiffs of the dukes of Austria and by 1319 (if not already in 1309) the claim to levy them was given up. These facts show (what could not hitherto be proved) that at the time when legend places the rising of Uri, Tell exploit, etc., the dukes of Austria really had disputes with Uri.

The alleged proofs of the existence of a real William Tell in Uri in the 14th century break down hopelessly: (1) The entries in the parish registers are forged. (2) As to the Tell chapels: (*a*) That in the "hollow way" near Küssnacht was not known to Melchior Russ and is first mentioned by Tschudi (1572). (*b*) That on Tell's Platte is first mentioned in 1504. The document which alleges that this chapel was built by order of a *Landsgemeinde* held in 1388, at which 114 men were present who had been personally acquainted with Tell, was never heard of till 1759. The procession in boats to the place where the chapel stands may be very old, but is not connected with Tell till about 1582. (*c*) The chapel at Bürglen was founded in 1582.

In general *see* two excellent works by Franz Heinemann, *Tell Iconographie*, reproductions, with text, of the chief representations of Tell in art from 1507 onward (Lucerne, 1902), and *Tell Bibliographie* (including that of Schiller's play), published in 1908 at Bern. (W. A. B. C.; X.)

TELL EL AMARNA (TALL EL AMARNA), the modern name given to the collection of ruins and rock tombs in upper Egypt, on the east bank of the Nile, 58 mi. by river below Asyut and 190 mi. above Cairo. The ruins are those of Akhetaton, the city built about 1375 B.C. by Ikhnaton (Amenhotep IV) as the new capital of his kingdom (in place of Thebes) when he abandoned the worship of Amon and devoted himself to that of the Aton; *i.e.*, the solar disk. Shortly after Ikhnaton's death the court returned to Thebes, and the city, after an existence of hardly more than 15 years, was abandoned (*see* IKHNATON).

Tell el Amarna is of great interest not only because of its association with the revolutionary Pharaoh Ikhnaton but also because it is one of the few city sites of ancient Egypt that has been carefully excavated. The first scientific clearing of the site was undertaken by Sir Flinders Petrie in 1891–92, and in succeeding years both British and German archaeologists unearthed the ruins of the city, which extended along the Nile for a distance of five miles.

Because Ikhnaton chose a virgin site for his capital and because of the relatively short duration of its occupancy the excavators could reconstruct an accurate picture of the layout of the city.

When Ikhnaton ordered the founding of the city in the fourth year of his reign, he had stelae (inscribed tablets) carved in the cliffs at the end of the desert to demarcate the city limits. Although the city as a whole displays little in the way of town planning, the major official quarters at its centre were carefully organized. Running north and south throughout the entire length of Akhetaton was the Royal road, on either side of which were the principal buildings. The largest of these was the Great temple of the Aton, the enclosure wall of which was nearly ½ mi. long and 300 yd. broad. The temple consisted primarily of a series of walled courts separated by towered gateways leading eastward to the main sanctuary, which was completely in the open air. An elevated platform in the centre of this final court served as the high altar to the sun-god. Throughout the length of the temple and also outside its walls were numerous offering tables of stone and mud brick. The open character of the cult of the Aton stood in contrast to that of the Theban Amon, whose worship had been carried on in the dimly lighted innermost chambers of his temple.

South of the Great temple were situated the buildings for Ikhnaton's use. On the east side of the Royal road was the commodious residence of the royal family, provided with terraced gardens, while directly across the road was the official palace, where the affairs of state were carried on. Both these buildings, as indeed most of the dwellings at Tell el Amarna, were constructed principally of mud brick, baked in the sun. Stone was sparingly used for such architectural elements as columns and doorways. Many of the rooms of these complex royal buildings had walls, floors and ceilings painted in a lively naturalistic style. Such scenes as those depicting the intimate family life of Ikhnaton from his living quarters are among the masterpieces of Egyptian art. A bridge, spanning the Royal road, joined the two buildings and provided a place where the king might appear publicly to reward his nobles with gifts of gold. A second smaller temple to the Aton, similar in plan to the Great temple, lay immediately to the south of the king's residence.

Two other buildings from the central city are important as sources of major archaeological finds. In the house of the sculptor Thutmose were discovered many of the masterworks of Amarna sculpture, notably the portraits of Queen Nefertiti. In 1887 the chance diggings of a peasant woman in the records office brought to light a store of approximately 300 clay tablets inscribed in the cuneiform script. From these tablets, written by the vassal princes of the Egyptian empire in Syria and Palestine, as well as by the kings of Babylonia, Assyria and Mitanni, to Amenhotep III and his son Ikhnaton, it is possible to trace the decline of the Egyptian empire of the 18th dynasty. Ikhnaton, who was deeply preoccupied with religious matters at Akhetaton, had a corresponding lack of interest in foreign affairs.

Unlike Thebes, where crowded conditions had forced the construction of multistoried houses, the dwellings at Akhetaton had only one floor. The nobles possessed spacious villas with landscaped gardens. A small shrine open to the sky for the worship of the Aton was one of the distinctive features of such an estate. Generally the bedrooms and other rooms surrounded a central living room, the roof of which was higher than that of the rest of the house, thus permitting clerestory lighting and ventilation. The workers were lodged in simple cottages lined next to one another in rows.

Toward the end of Ikhnaton's reign there is evidence of his desire to seek a possible reconciliation with the old priestly hierarchy of Thebes. At about the same time a rift between Ikhnaton and his wife, Queen Nefertiti, developed; the queen retired to the north part of Tell el Amarna, where a special palace was constructed for her use. The archaeological evidence from this quarter suggests that she was an ardent supporter of the Aton religion.

In the desert hills to the east were hewn the tombs of the important officials. These tombs resemble those at Thebes in having an exterior forecourt, an entrance hall sometimes provided with columns, and a second hall at the rear of which was a small room containing a statue of the tomb owner. These chambers were accessible for the funerary cult; the burial shaft generally descends from the floor of the first hall, but in none of the tombs has any burial been discovered. Although the work in decorating the tomb chapels was often hastily and incompletely carried out, the painted reliefs of these tombs have been a major source of information regarding the daily life and religion of Akhetaton. The representations of various buildings of Tell el Amarna that appear on the walls have been valuable aids to the excavators in interpreting the often meagre architectural traces surviving in the actual city, particularly of the Great temple, which was razed to the ground by the opponents of the Aton religion after Ikhnaton's death. In these tombs the emphasis is upon the activities of the owner in Akhetaton rather than the life after death. In a number of the tombs have been found versions of the hymn to the Aton, in which is described the universality of the sun-god as creator and sustainer of all life. The reliefs are in an unconventional, often exaggerated style, tending to reflect a momentary and emotional aspect rather than the static mood that had characterized Egyptian art previously. The tomb of Ikhnaton and his family is situated in the side of a dry watercourse about 4 mi. E. of the city. It contains an unprecedented scene of the royal family in mourning over the death of the princess Meketaton, who was buried in a section of the tomb. There is no certain evidence, however, that either Ikhnaton or his wife was ever interred there.

The return of the court to Thebes after Ikhnaton's death must have been rather sudden, since houses under construction in the northern sector of the city were abandoned before completion. The reform in religion and art that Ikhnaton had sought to effect ended in failure, for the number of faithful adherents was relatively small. Even at Tell el Amarna itself in the homes of the workers were discovered figures and names of the gods of the older tradition. After the Pharaoh died, a number of his retainers found it more convenient to adapt themselves to life at Thebes rather than support a lost cause at Akhetaton.

BIBLIOGRAPHY.—J. D. S. Pendlebury, *Tell el-Amarna* (1935); *The City of Akhenaten*, part iii, 2 vol., by J. D. S. Pendlebury *et al.* (1951); N. de G. Davies, *The Rock Tombs of El Amarna*, 6 vol. (1903 *et seq.*); C. F. Pfeiffer, *Tell el Amarna and the Bible* (1963).

(E. F. W.)

TELLICHERRY, a seaport of Kerala state, India, on the Southern railway between Cannanore and the settlement of Mahe. Population (1961) 44,763. The town is built upon a group of wooded hills running down to the sea, and is protected by a natural breakwater of rock. It is the seat of the Government Brennan college, connected with Madras university. Ships are able to anchor 2 mi. from the shore. The port has a pier and sea wall and can be used during the monsoon. Tellicherry was at one time defended by a strong mud wall and the old fort still stands north of the town. The East India company established a factory there in 1683 for the pepper and cardamom trade. For two years (1780–82) the town withstood a siege by Hyder's general, and in the subsequent wars with Mysore, Tellicherry was the base for the ascent of the Ghats.

TELLURIUM, a chemical element and a brittle, silvery-white metallic substance. It is the third member of the sulfur family (symbol Te, atomic number 52, atomic weight 127.61).

In 1782 F. J. Müller von Reichenstein, in examining Transylvanian gold minerals, suspected the presence of a new element in *Weisgolderz* and called it *metallum problematicum* or *aurum paradoxum*. In 1798 M. H. Klaproth extracted the element from the crude mineral and established its identity, giving the first differentiation between the new element and antimony with which it had been confused. He named the element tellurium. It was not until 1832 that J. J. Berzelius made a careful study of the chemistry of the element.

Occurrence.—Although widely distributed, the estimated percentage in the earth's crust is only 10^{-9}. Rarely found as native tellurium, it commonly occurs as tellurides of copper, lead, silver, gold, iron and bismuth. Typical minerals are sylvanite, $(AuAg)Te_4$; calaverite, $AuTe_2$; muthmannite, $(AuAg)Te$; krennerite, $(AuTe_2)$; petzite, (Ag_3AuTe_2); tetradymite, Bi_2Te_2S; rickardite, Cu_4Te_3; melonite, $NiTe_2$; and altaite, PbTe. Less common are the

tellurites, telluric ochre, TeO_2, durdenite, $Fe_2(TeO_3)_3.4H_2O$, and tellurates as montanite, $Bi(OH)_2.TeO_4$, and ferrotellurate, $FeTeO_4$. The tellurides occur with the sulfides of copper, silver, gold and nickel. Red tellurium sulfur of Japan contains 99.76% S, 0.17% Te, 0.06% Se and 0.01% As. Tellurium minerals are found in the United States, Canada, Mexico, South America, Sweden, central Europe, Northern Rhodesia, the U.S.S.R. and Japan. The chief sources of tellurium are the slimes from copper and lead refineries and flue dusts from processing of telluride gold ores.

Uses.—Although comparatively large supplies of tellurium are available, no single use has been developed which creates large demand for this element. Early attempts to use tellurium in metallurgical products seemed to indicate that the addition of tellurium to the alloys of aluminum, tin, certain steels and cast iron gave undesirable properties. Apparently too much was added, for it has been reported that small concentrations of tellurium in aluminum alloys increases ductility, in tin alloys increases the hardness and tensile strength and in stainless steels increases the machinability. Appreciable increase in the depth of chill in iron (*see* CAST-IRON) can be obtained by addition of minute amounts of tellurium. The addition of 0.05% Te to lead increases its resistance to corrosion. In manganese-magnesium alloys tellurium increases resistance to corrosion. Tellurium has been used as a brightener in certain plating baths. A beryllium-copper-tellurium alloy has been reported. Selenium-copper and tellurium-copper alloys used to make arc tips for electric contacts possess antiwelding properties. Selenium and tellurium in copper or copper-rich alloys improve machinability without producing "hot shortness." Both are used in preference to sulfur. In tin-rich alloys like Babbitt metal the addition of tellurium refines the grain structure, aids casting and improves physical properties at high temperatures. Tellurium in hydrogenation and cracking catalysts, used for conversion of heavy oils to motor fuels, reduces carbon formation. The addition of 0.1% to 2% tellurium in rubber compounding improves resistance to heat, oxidation and abrasion and increases resiliency. Some organic tellurium compounds have germicidal properties, and diethyl telluride possesses antiknock characteristics. This latter use has not been important since it is accompanied by a disagreeable odour. Zinc tellurite has been patented as an antifouling paint. Tellurium has been used to a limited extent in colouring glass or porcelain, large colloidal particles giving a blue colour, small particles a brown colour and telluride a red to red-violet colour. Acid solutions of dioxide have been used as a dip for silver ornaments, giving a "platinum finish" to the metal. Soluble tellurium compounds are used in toning baths in photography. As little as 1 μm (micromole) of tellurium taken into the human system gives the breath a penetrating garlicline odour, which may appear in less than half an hour after the ingestion of the tellurium. Although not well documented, 2 g. of tellurium as sodium tellurite may be a lethal dose. The use of colloidal tellurium as a fungicide and insecticide in wood preservation has been reported. This had not become common practice by mid-century.

In the 1940s and 1950s a variety of articles were published in technical journals on such topics as effect of tellurites on pyruvic dehydrogenase, catalyst in Kjeldahl digestion, effect on electrodes for the electrolysis of sodium chloride, grain growth of chromium steel, effect on hydrogen absorption by steel, effect on iron carbonyl, tellurites as poisons for enzymes, solution of tellurium mirrors by free radicals and the catalytic hydrogenation of shale oils. However, no new use was reported which would contribute to large-scale consumption of tellurium or its compounds.

Methods of Production.—The crude copper anodes used in the electrolytic refining of copper contain a high percentage of selenium and tellurium which was present in the ore. Upon electrolysis of the copper, the selenium and tellurium and other valuable elements (gold and silver) settle in the electrolyte as anode mud. Although procedures for the recovery of tellurium are not identical in this and other types of residues, the general principles include fusion with sodium carbonate and nitrate or extraction with sodium hydroxide, neutralization of the water solution of sodium tellurite to form precipitated tellurium dioxide and, finally, dry reduction of the tellurium dioxide by means of a mixture of flour and borax to form molten tellurium. An alternative method includes solution of the dioxide in hydrochloric acid and reduction with sulfur dioxide to form precipitated tellurium. Tellurium appears on the market in the form of bars, as ground crystalline tellurium or as dry precipitated powder. Some tellurium has been redissolved in nitric acid (specific gravity 1.2) and crystallized as basic nitrate.

Properties.—Tellurium is a complex element, having eight stable isotopes with mass numbers, arranged in order of decreasing abundance, as follows: 130, 128, 126, 125, 124, 123, 122 and 120. There are 14 artificially produced unstable isotopes, 4 of which have nuclear isomers. Unstable isomers of two of the stable isotopes have been produced. The United States Atomic Energy commission listed the following isotopes as available for distribution: Te^{127} (90 days), Te^{129} (32 days), Te^{131} (1.25 days). There are six electrons ($5s^2$, $5p^4$) in the outermost main level.

Tellurium has a density of 6.24; it melts at 449.8° C. and boils at 1,390° C.; its hardness is 2.3 (Mohs). It crystallizes in hexagonal-rhombohedral prisms and is isomorphous with the stable gray modification of selenium. Its allotropy is less well defined than that of either selenium or sulfur. It was presumed that amorphous and crystalline modifications of tellurium existed. X-ray defraction has shown that precipitated tellurium is not amorphous but crystalline (hexagonal). Colloidal preparations are easily made by reducing dilute solutions of telluric acid with hydrazine or sulfurous acid, with or without protective colloids. Alpha particle bombardment of tellurium produces Xe^{133}. The vapour of tellurium at the boiling point is diatomic Te_2. Tellurium vapour lamps containing also mercury provide almost a continuous spectrum which approaches the characteristics of daylight. Transmission is in the red (7,200–5,950 Å) 14.5%, yellow (5,950–5,650 Å) 22.5%, green (5,650–4,950 Å) 56% and blue (4,950–4,000 Å) 7%.

Tellurium occurs in Group VI of the periodic system with oxygen, sulfur and selenium. In the second long series of 18 elements, it lies between antimony and iodine. Although strongly electronegative, tellurium is less so than selenium, sulfur or oxygen. Tellurium exhibits valence states of -2, $+2$, $+4$ and $+6$. These are illustrated by (-2) hydrogen telluride and the tellurides; ($+2$) tellurium monoxide and tellurium dichloride; ($+4$) tellurium dioxide, tellurous acid, the tellurites and the tetra halides; ($+6$) tellurium trioxide, telluric acid, the tellurates and the hexafluorides.

Compounds.—Hydrogen telluride (H_2Te) is prepared by the action of hydrochloric acid on aluminum or magnesium telluride or by action of water on aluminum telluride (Al_2Te_3). Pure H_2Te is reasonably stable, but dry or moist reacts readily with oxygen to form elemental tellurium and water. Both melting and boiling points of H_2Te are lower than those of water and H_2Te is less polar. Its solubility in water is at least 0.09 moles per litre at 760 mm. pressure. Hydrogen telluride is both a stronger reducing agent and a stronger acid than either hydrogen sulfide or hydrogen selenide. The respective constants for the primary ionization of H_2S, H_2Se and H_2Te are 1.15×10^{-7}, 1.88×10^{-4} and 2.27×10^{-3}. Like H_2S and H_2Se, H_2Te has a very disagreeable odour and is very toxic. Sodium telluride (Na_2Te) is prepared by a direct combination of sodium and tellurium in liquid ammonia; an appropriate excess of tellurium dissolves the precipitated Na_2Te and yields a polytelluride (viz., $Na_2Te_2 - Na_2Te_6$). Hydrogen telluride may be added to aqueous solutions of sodium hydroxide to form NaHTe or Na_2Te. Air or oxygen must be excluded since oxidation will cause complete precipitation of tellurium. The $+2$ valence is illustrated by tellurium monoxide (TeO) and tellurium dichloride ($TeCl_2$). The former is prepared by heating tellurium sulfite in vacuum at 200° C., the products being tellurium monoxide, TeO, and sulfur dioxide, SO_2. Tellurium dichloride is prepared by the direct union of chlorine and tellurium, an excess of chlorine yielding the tetrachloride, $TeCl_4$($+4$ valence). Similar iodides and bromides have been prepared. All are solids at 25° C. The tetrachloride and the tetrabromide melt at 214° and 380° and boil at 414° and 421° C. respectively. The chlorides or bromides react rapidly with water according to the following type reaction: $2TeCl_2 + 2H_2O = TeO_2 + 4H^+ + 4Cl^-$. Although

only slightly soluble in water, tellurous acid, H_2TeO_3, and its salts are known. Ionization constants have been reported for tellurous acid; $K_1 = 2 \times 10^{-3}$ and $K_2 = 10^{-8}$. Dissolved in strong acids, TeO_2 forms the positive ion $(TeOOH)^+$. The alkali tellurites are subject to air oxidation at elevated temperatures.

Tellurium dioxide can be prepared by heating the trioxide above its melting point but is most commonly prepared by the ignition or neutralization of tellurium basic nitrate, $2TeO_2.HNO_3$. The hexafluoride, telluric acid and the tellurates illustrate the +6 valence for tellurium. Contrasted with the other halides, the hexafluoride has a much lower melting and boiling point, −37.8° and −38.9° C. respectively, and a much higher heat of formation, 315,000 cal. Orthotelluric acid (H_6TeO_6) is easily prepared by the chromic acid oxidation of tellurium basic nitrate or the hydrogen peroxide oxidation of an acid solution of tellurium dioxide. A solubility curve for telluric acid shows $H_2TeO_4.6H_2O$ the stable form below 10° C. and $H_2TeO_4.2H_2O$ the form crystallized above 10° C. Salts of the hypothetical acids pyrotelluric, $H_2Te_2O_7$, and tetratelluric, $H_2Te_4O_{13}$, are known. Orthotelluric acid loses water upon heating. Conversion to tellurium trioxide, TeO_3, is complete at 300° to 360° C. The conversion to metatelluric acid, H_2TeO_4, is considered doubtful. If orthotelluric acid is heated in a closed tube at 140° C., a viscous liquid, allotelluric acid, $(H_2TeO_4)_n$, $n = c.\ 2.9$, is formed. Presumably the transition point between orthotelluric and allotelluric acid is 131° C. Allotelluric acid is a relatively strong acid and exhibits properties quite different from those of the ortho acid. It is soluble in alcohol and separates sulfur from sodium thiosulfate. Upon standing, aqueous solutions hydrolyze and form orthotelluric acid, H_6TeO_6. Two new acids have been reported, $H_6TeCl_6.2H_2O$, citron yellow, and H_6TeBr_6, red. Both tellurites and tellurates form co-ordination compounds such as ammonium ditelluratohexamolybdate, potassium chromotellurate and ammonium triarsenatotellurate.

Organic compounds vary in complexity from the relatively simple alkyl tellurides to the more complex cyclic compounds which contain tellurium in the ring. There are simple aliphatic compounds of the types, R_2TeX_2, R_3TeX and $RTeX_3$. Among the aromatic derivatives there are types R_2Te, R_2Te_2, $RTeR'$. J. A. N. Friend gave an excellent review of organic compounds of tellurium and their preparation.

Analytical.—Like selenium, tellurium is easily precipitated from dilute hydrochloric acid solutions by reducing agents such as sulfur dioxide or hydrazine hydrochloride. Two common reactions are used for the separation of tellurium from selenium; red selenium is precipitated from strong hydrochloric acid solutions, by sulfur dioxide, tellurium is not. Upon separation of the selenium by filtration, the solution is diluted and the tellurium precipitated by further addition of sulfur dioxide. The fusion of selenium and tellurium with potassium cyanide forms potassium selenocyanate, KCNSe, and potassium telluride, K_2Te. Bubbling air through a water solution of this mixture precipitates elemental tellurium. The addition of potassium iodide to a dilute acid solution of tellurium precipitates tellurium tetraiodide, TeI_4, which is soluble in an excess of potassium iodide, forming K_2TeI_6. Tellurium is commonly weighed as the element. Tellurous acid may be oxidized to tellurate by potassium permanganate. Tellurates are reduced quantitatively by titanous chloride.

BIBLIOGRAPHY.—R. Abegg, F. Auerbach and I. Koppel, *Handbuch der Anorganischen Chemie*, vol. iv, with a chapter on the colloid chemistry of tellurium (1927); B. S. Hopkins, *Chapters in the Chemistry of the Less Familiar Elements* (1940); Victor Lenher, *Miscellaneous Papers;* U.S. Atomic Energy Commission, Isotope Division, no. 4 (1951); Joseph William Mellor, *Comprehensive Treatise on Inorganic and Theoretical Chemistry* (1922–37); Fritz Ephraim, *Inorganic Chemistry*, 4th ed. rev., ed. by P. C. L. Thorne and E. R. Roberts (1943); U.S. Bureau of Mines, *The Minerals Yearbook* (with yearly supplements); Don M. Yost and Horace Russell, *Systematic Inorganic Chemistry of the Fifth-and-Sixth-Group Nonmetallic Elements* (1944). John Albert Newton Friend, *Textbook of Inorganic Chemistry*, vol. xi, part iv (1937); Charles Palache, Harry Berman and Clifford Frondel, *System of Mineralogy of James Dwight Dana and Edward Salisbury Dana* (1944). (V. W. M.)

TELUGU, one of the five great Dravidian languages. The Telugu-speaking peoples occupy an area from north of Madras city northwest to Bellary, where Telugu meets Kanarese, and northeast to near Orissa. Telugu can be characterized as the only descendant of the Andhra dialect of Old Dravidian.

TEMBU (Ama-Tembu), a powerful Kaffir tribe which has given its name to Tembuland, a division of Cape Colony. They are traditionally descended from Tembu, elder brother of Xosa.

TEMPE, VALE OF, the ancient name of the gorge, through which the Peneus river (mod. Salambria) reaches the sea from the plain of Lower Thessaly. Greek legend attributed this chasm to Poseidon; before it was riven, between Olympus and Ossa, Thessaly was a lake. It is about 4½ mi. long, and toward the middle of the pass, where the rocks are highest, the precipices in the direction of Olympus fall steeply to the stream; but those which descend from Ossa are the loftiest, for they rise 1,500 ft. from the valley. Tempe was sacred to Apollo. Every ninth year a sacred mission visited the valley to pluck laurel for the chaplets for the Pythian games.

TEMPERA is a kind of painting executed with paints that consist of pigment ground in a medium that is miscible with water; a few moments after application the paint is sufficiently resistant to water to allow overpainting with more colour. Thin, translucent layers of paint produce a clear, luminous effect and the colour tones of the successive brushstrokes blend optically. When the paint is skillfully built up, it will retain its luminosity, as well as the effect of its undercoats, as an opaque or semiopaque coating. Tempera is distinguished from the other aqueous paints which are either transparent (*see* WATER-COLOUR PAINTING) or solidly opaque (gouache, casein [*qq.v.*], poster colours). Tempera may be applied in thin, transparent washes or in heavy, opaque touches when required, but these are not typical of the work as a whole.

Tempera paints have their distinctive properties because they are emulsions. An emulsion is a stable mixture of an oily ingredient (oils, fats, resins, waxes) and a thick, gummy or gluelike (colloidal) aqueous ingredient. The proverbial rule that oil and water do not mix does not apply when the oil is of an emulsifiable type and there is a good emulsifier in the water solution. The oily ingredient of an emulsion is dispersed in the aqueous solution in myriads of tiny round droplets. The binding ingredients in water colour, gouache and poster colours are not emulsions; they are simple gluey solutions.

The word tempera originally came from the verb "temper," *i.e.,* to bring to a desired consistency; dry pigments were made usable by "tempering" them with a fluid binding and adhesive vehicle. The Italians distinguished such painting from fresco painting (*q.v.*), the colours for which contained no binder. The artist was said to "temper" his colours with egg or with oil. Eventually, after the rise of oil painting, the use of the word tempera became confined to its present meaning. In some circles, especially among artists who work solely for reproduction, "tempera" has been indiscriminately used to denote any opaque water medium and the labels on some aqueous colours in tubes and jars bear this name, but this usage is incorrect and confusing.

TECHNIQUE

The standard tempera vehicle is a natural emulsion, egg yolk, thinned with water. The yolk of hen's eggs contains a fatty substance (egg oil) emulsified with a gummy substance (albumen) and a few minor ingredients, chief among which is lecithin, a powerful emulsifier and stabilizer. The man-made emulsions of more recent origin are variants of the original egg vehicle. Their purpose is to create some special property of manipulation or some effect for an individual requirement; they aim at duplicating or closely resembling the good properties of straight egg-yolk tempera, such as its tough, leathery permanent paint film.

In preparing the paint, the yolk is separated from the white and is dried by being rolled on a paper towel, and its sac is punctured; a pure yolk is thus obtained, without any white. The ingredients of pure egg yolk and their proportions govern the whole technique of egg tempera—its behaviour, its effects and its longevity. Egg white was never used in easel painting; it is a relatively weak and impermanent paint binder. The yellow colour of the yolk is imperceptible when pigmented; moreover, it is a fugi-

tive colouring matter (carotene) which soon disappears on exposure. The pigments (the same as those used in water colour) are ground to paste consistency with water and kept in screw cap jars. Prior to each day's work the artist mixes with egg yolk as much of a colour as he needs. He paints with fine red sable watercolour brushes; a skilled technique of brushwork is gained from experience. The three basic points of control are the amount of water used, the loading of the brush and the manner of stroking. It is a deliberate, controlled method, adaptable to carefully planned

FROM RALPH MAYER, "THE PAINTER'S CRAFT" (1948), REPRODUCED BY PERMISSION OF D. VAN NOSTRAND COMPANY, INC.

METHOD FOR EXTRACTING PURE EGG YOLK FROM ITS SAC. HAVING BEEN SEPARATED FROM THE WHITE AND DRIED, THE YOLK SAC IS READY TO BE PUNCTURED

work rather than to spontaneous painting. A finished drawing (cartoon), in which all the artist's creative problems are worked out, is made first; this is traced on the ground. The artist's efforts can then be devoted to the technique and the development of colour and painterly effects. The special ground for tempera painting is a rigid wood or wallboard panel coated with gesso, which is a brilliant white, ivory smooth, fully absorbent preparation made of precipitated chalk and rabbitskin glue.

Variants of the pure egg-yolk vehicle have been developed to widen its use among contemporary artists, although earlier painters chose to develop a superlative skill in the handling of a somewhat difficult and intractible material rather than to make the material itself amenable to a more easily acquired skill. Among the man-made emulsions a popular one is prepared by shaking or homogenizing whole egg with linseed oil; this vehicle is preferred by many painters to that of pure yolk. Another is gum tempera, the standard recipe for which contains gum arabic or cherry gum emulsified with a blend of dammar varnish and stand oil. This vehicle is also well-liked by numerous painters; it can produce an almost impasto effect when desired. Wax temperas are made from a group of experimental recipes and have the reputation of producing delicate, sensitive effects. Casein colours are opaque water paints, not correctly classified as tempera. Modern tempera paintings are sometimes varnished or overpainted with thin, transparent oil glazes to produce full, deep-toned results, or they are left unglazed for blond effects. Unvarnished tempera paintings, especially the pure egg-yolk variety, will assume a pleasing, satiny sheen when polished with cotton wool.

CASEIN PAINTING

For more robust, solid or spontaneous effects than are obtainable with tempera, artists have used colours ground in a solution of casein (*q.v.*). In the form of homemade curd made from soured skim milk, it has been a traditional adhesive and binder for more than eight centuries. Refined, pure, powdered casein which can be dissolved with ammonia has been used by artists for easel and mural paintings since the latter 19th and early 20th centuries. More recently, ready-made casein paints in tubes have come into very wide use to create effects that approach those of oil paintings. Casein painting permits the use of bristle brushes and a moderate impasto, like oil painting, but not the fusion of tones. It is preferred by some because of speedy drying and mat effects. When dry, the paint becomes water-resistant to a considerable degree. Casein paintings may be varnished to further resemble oil paintings, and they are frequently glazed or overpainted with oil colours. Proper technique calls for the use of rigid boards or panels; casein is too brittle for canvas.

HISTORY

No positive date is assigned to the first use of egg yolk as a binder of colours; it is so readily available that it may have served the most primitive cultures. The earliest European forerunners of a controlled egg-tempera medium are found among the religious paintings of the Byzantine era. By the 13th century a well-developed, standard technique on wood panels emerged in Italy; the works of Duccio (*c.* 1255–1319) in Siena and Cimabue (*c.* 1240–*c.* 1302) in Florence are outstanding examples. The medium served well when panel painting moved to a higher plane of artistic and technical sophistication as in the tempera paintings of Giotto (1267?–1337) and the Lorenzettis of 14th century Siena. Anonymity and routine following of symbolism gave way to a more individual, personal or inventive manner; the madonnas and saints were depicted as actual people. The Italian schools thereafter developed until their style of painting culminated in a refinement and perfection which it is not possible to excel with pure egg tempera. A number of painters exhibit this high stage of development; examples are the Venetian Carlo Crivelli (1430–1494) and, perhaps the most accomplished of all, the Florentine Sandro Botticelli (1444–1510). With the changing demands of the times, art forms evolved from the generally dry and deliberate linear styles that were admirably served by tempera, to a more flowing, intensely coloured, spontaneous and juicier kind of painting, with smooth blending of colours and tones. Artists necessarily turned to the use of oily materials. (*See* OIL PAINTING, TECHNIQUE OF).

The earliest examples usually cited in historical accounts of oil painting are actually tempera paintings with thin, transparent oil glazes; such pictures are perhaps more accurately classified as tempera painting in its ultimate form. But before oil painting supplanted it, tempera was used for a group of paintings that constituted a transitional stage; the medium was still aqueous tempera, but with properties that could scarcely have been duplicated by strict adherence to the pure egg-yolk technique as used from Giotto to Botticelli. In the paintings of this transitional group there was fusion of tones and a choice of scenes and subjects that challenged the potential of egg yolk and presaged oil-painting technique. Perhaps a little oily ingredient was incorporated into the paint in some manner; it seems to require some such alteration to duplicate the effects of Carpaccio (1460–1522) for instance, or of Giovanni Bellini (*c.* 1430–1516), who eventually did employ oil paint. Use of old techniques continued long after the introduction of new ones, so the rapidly changing practices of the 15th century cannot be attributed to specific dates.

In the following centuries tempera was practised sporadically; it never became a totally defunct technique. In the mid-19th century its literature was studied intensively, although academically for the most part, for its relationship to oil painting. Egg tempera is completely and satisfactorily documented and the technique can be easily reconstructed from writings and from hundreds of examples that have survived in fine condition. One prominent source of information is *The Book of the Art of Cennino Cennini*, written in the 14th or early 15th century by a master painter; it is a complete treatise on artists' techniques of the Gothic and early Renaissance periods. Cennini's idol was Giotto, and the book stems directly from Giotto's studio. Besides tempera it covers fresco, oil paints for decorative work and a number of studio practices. It also contains observations on painting, the proper personal behaviour and decorum of artists, the minutiae of artists' materials and procedures (*e.g.*, "the yolk of a town hen's egg is paler than those of country hens"). Artists in Europe and the U.S. in the latter part of the 19th and the early 20th centuries

"Madonna and Child" by Giotto (Giotto di Bondone, 1267?–1337), Florentine. An example of the basic, standard egg yolk tempera technique

"The Presentation in the Temple" by a 15th-century Byzantine painter. This painting follows an old tradition of a circumscribed painting in a primitive tempera technique

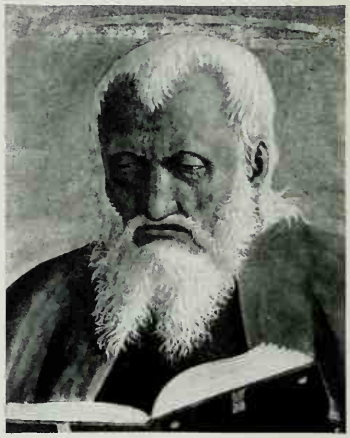
EARLY DEVELOPMENTS IN TEMPERA PAINTING

"The Calling of the Apostles Peter and Andrew" by Duccio (Duccio di Buoninsegna, c. 1255–c. 1318), Sienese. An example of the beginnings of the controlled tempera technique

Detail of St. Francis from "St. Francis in Ecstasy" by Giovanni Bellini (c. 1430–1516), Venetian. Egg tempera figure in a landscape. Bellini was active during the transition of tempera into the beginning of oil painting

BY COURTESY OF (TOP LEFT, BOTTOM LEFT) NATIONAL GALLERY OF ART, WASHINGTON, D.C.; SAMUEL H. KRESS COLLECTION, (TOP RIGHT) THE METROPOLITAN MUSEUM OF ART, N.Y.; BEQUEST OF LIZZIE P. BLISS, 1931, (BOTTOM RIGHT) © THE FRICK COLLECTION, N.Y.

TEMPERA



LATER IMPROVEMENTS IN TEMPERA TECHNIQUE

Top left: "Madonna and Child" by Carlo Crivelli (c. 1430/35–c. 1493/95), Venetian. Pure egg tempera technique developed to the utmost degree of skill and refinement

Right: Three examples of tempera portraiture: (*top*) "Portrait of a Young Man" by Antonello da Messina (c. 1430–79), Venetian school. Tempera with thin, transparent oil or varnish glazes; (*centre*) detail of the head from "An Apostle" (St. John or St. Simon) by Piero della Francesca (c. 1420–92), Italian (Umbrian). A tempera of facile manipulative qualities beyond those of pure egg yolk; (*bottom*) "Portrait of a Lady" by Hans Memling (c. 1430–94), Flemish. Meticulous, jewellike tempera painting with oil and varnish glazes; a highly developed technique

Bottom left: "Christina's World" by Andrew Wyeth (1917–), U.S. A contemporary adaptation of classic tempera technique

BY COURTESY OF (TOP LEFT) THE METROPOLITAN MUSEUM OF ART, N.Y.; THE JULES S. BACHE COLLECTION, 1949. (TOP RIGHT) THE METROPOLITAN MUSEUM OF ART; BEQUEST OF BENJAMIN ALTMAN, 1913. (CENTRE RIGHT) © THE FRICK COLLECTION, N.Y., (BOTTOM LEFT) COLLECTION, THE MUSEUM OF MODERN ART, N.Y., PURCHASE FUND, (BOTTOM RIGHT) THE METROPOLITAN MUSEUM OF ART; BEQUEST OF MARY STILLMAN HARKNESS, 1950

found both pure and oil-glazed tempera techniques, such as those of the early Italian and Flemish masters, very appropriate for their own work, particularly for blond effects. A strong revival of tempera followed, elevating it to a significant position among standard 20th-century techniques. *See* also PAINTING.

BIBLIOGRAPHY.—Cennino Cennini, *Il libro dell' Arte,* a 15th-century manuscript, 3 vol.: vol. i, Italian text (1932–33), vol. ii, Eng. trans. by D. V. Thompson, *The Craftsman's Handbook* (1932–33, 1954), vol. iii, D. V. Thompson (ed.), *The Practice of Tempera Painting* (1936); Ralph Mayer, *The Painter's Craft* (1948), and *The Artist's Handbook of Materials and Techniques,* rev. ed. with extensive bibliography (1957). (RH. M.)

TEMPERAMENT. In psychology, temperament refers to that part of personality concerned with emotional dispositions and reactions, their speed and intensity, and the term is often used to refer to the prevailing mood of a person. Historically, the concept originated with the Hippocratic school of medicine and was based upon a theory of the varying proportions of four bodily humours or excretions in the body (Lat. *tempero,* "to mingle," "compound in definite proportions"): these four fluids—blood, phlegm, black bile and yellow bile—produced, according to their degree of predominance, the sanguine, phlegmatic, melancholic or choleric temperament. This theory constitutes the first of many attempts to group human beings into types, and its stress on the physiological determination of temperament is echoed in modern theories emphasizing the action of the endocrine glands; *e.g.,* an increase in the secretion of adrenaline produces heightened emotional reactivity. However, modern psychology attributes more importance to the activity of the autonomic nervous system (*see* NERVOUS SYSTEM), and particularly its sympathetic branch, in accounting for emotional reactivity: autonomic over-responsiveness is intimately linked with neurotic dispositions, but such responses can easily be conditioned (*see* CONDITIONING), and consequently individual differences in ease of conditioning (also probably innate) also play their part in determining temperament.

External physical characteristics have often been supposed to make possible the identification of different temperaments, but empirical studies have almost universally failed to substantiate such hypotheses. Very thin, lean people tend to be introverted and to develop symptoms of anxiety and depression when suffering from neurotic disorders; very broad, fat people tend to be extraverted and to develop symptoms of hysteria and psychopathy when suffering from neurotic disorders. But these relationships are slight and of little help in prognosis or treatment.

Modern investigations have left little of the original theory of Hippocrates standing, and modern methods involve two great steps: (1) the experimental measurement and determination of a person's habitual emotional reactions when subjected to standardized stressful situations; and (2) the detailed statistical analysis, by correlational and factor-analytic methods, of the relationships observed. The resulting picture is highly complex, revealing, among other things, a tendency for individuals to respond consistently to stress with only one of the many parts of the body controlled by the autonomic system, and it is this part which is most vulnerable in neurotic breakdown. *See* also PERSONALITY. (H. J. EY.)

TEMPERAMENT, MUSICAL, a term applied principally to systems for tuning instruments of fixed pitch, such as the keyboard instruments. These systems were designed to produce, by means of tempering (*i.e.,* a slight mistuning), a maximum number of intervals (especially major thirds) both melodically and harmonically acceptable to the ear.

Just Intonation.—When two notes are sounded together, they are pleasing to the ear when their wavelengths or frequencies bear a simple numerical ratio to each other (or very nearly so). Unisons have a frequency ratio of 1:1 and octaves 2:1. These intervals are so fundamental that they are accepted in all systems and need not be tempered. The next most harmonious interval is the perfect fifth, with the ratio 3:2 (its complement being the perfect fourth, with the ratio 4:3).

To add intervals together, the frequency ratios have to be multiplied together. Now; if fifths are added together, a new note is formed each time. For instance, in just intonation (*i.e.,* in mathematically exact intervals), starting with C and G, one then obtains D such that D and C have the ratio 9:4 or, bringing the D down an octave, the ratio 9:8. This interval is a major tone and is still reasonably concordant. If the process is repeated, one obtains an E giving with C an interval of two major tones with the ratio of 81:64, which is unbearably discordant. This may explain why in early medieval music, dominated by the theory of the pure fifth, the tone, fourth, fifth and seventh were regarded as concordant but the major third as discordant. If, however, a natural major third is sung, the frequency ratio will be 5:4, *i.e.;* 80:64.

The difference between the pure major third and the one obtained from four pure fifths is an interval with the ratio of 81:80 and is known as the comma of Dydimus. The reason for this is that just as $3 \times 3 \times 3 \times 3 = 81$ is a different number from $5 \times 16 = 80$, so the two E's are actually different notes. The problem lies in the fact that musical notation recognizes only one E and musicians similarly want only one E. The question then arises as to what extent this duplication applies to other notes or to, say, a whole major scale.

At this stage it becomes difficult to visualize ratios of large numbers and an additive numerical measure for intervals becomes necessary. This is easily obtained by taking logarithms of the frequency ratios; by convention, the octave is divided into 1,200 cents. Thus:

$$\text{No. of cents} = \frac{1{,}200}{\log 2} \times \log \text{ of frequency ratio.}$$

The ratios and cents of the two independent basic intervals are: perfect fifth = 3:2, 701.955; and major third = 5:4, 386.314. As all musical intervals are derived from ratios of products of the prime numbers 2, 3 and 5 only, cents of all the other intervals can be obtained by addition and subtraction of these two intervals. (Trumpets and horns produce "harmonics" that include other notes based on the numbers 7, 11, 13 and 17, but the player is expected to modify these in performance.) Thus the following intervals are obtained: perfect fourth (4:3, 1,200 − 701.955 = 498.045); major tone (9:8, 701.955 − 498.045 = 203.910); diatonic semitone or minor second (16:15, 498.045 − 386.314 = 111.731); minor sixth (8:5, 1,200 − 386.314 = 813.686); minor third (6:5, 701.955 − 386.314 = 315.641); major sixth (5:3, 498.045 + 386.314 = 884.359); chromatic semitone or augmented unison (25:24, 386.314 − 315.641 = 70.673); and the difference between a diatonic semitone and a chromatic semitone = 41.058 cents. Also, the comma of Dydimus = $4 \times 701.955 - 2 \times 1{,}200$ − 386.314 = 21.506. This interval is consequently more than one-fifth of an equal-tempered semitone.

In the C major scale the relative frequencies and cents to the nearest whole number from the first note are: C (24, 0); D (27, 204); E (30, 386); F (32, 498); G (36, 702); A (40, 884); B (45, 1,088); C (48, 1,200). All the possible intervals in the octave can be checked by subtracting the cents. They all come out right except D − A = 680 is not a fifth and D − F = 294 is not a minor third. As they are both too narrow by 22 cents (the comma of Dydimus), the solution might appear to be to alter D to 182 cents. This makes D − A and D − F right but then D − G and D − B would be too wide by a comma. Further, the note B♭ would have to be 996 to be a fourth above F or 1,018 to be a minor third above G, again differing by a comma.

If adjacent notes in the scale are subtracted the result is the following intervals: 204, 182, 112, 204, 182, 204, 112. In other words, an octave consists of two diatonic semitones of 112 cents, three major tones of 204 cents and two minor tones of 182.404 cents (ratio 10:9). As music recognizes only one type of tone, one requirement for a practical scale is to standardize on some compromise size of tone.

Tempered Scales.—*Mean Tone Temperament.*—It has been shown that if the fifths are kept pure by making all tones of 204 cents, the major thirds are unbearably sharp by a whole comma. If all the tones are minor, neither fifths nor thirds are right. If, however, one takes an average, or mean tone, of 193.157 cents, the major thirds will be pure and the fifths will be only a quarter

of a comma flat. The semitones will have to be equal to half the difference between the octave and five mean tones, or (1,200 − 5 × 193.157) ÷ 2 = 117.107; *i.e.*, one-quarter comma sharp. The chromatic semitone will then be tempered to 193.157 − 117.107 = 76.050, also a quarter comma sharp. The fifth will be 3 × 193.157 + 117.107 = 696.578. In the mean-tone system the difference between a minor second and an augmented unison, *e.g.*, C♯ to D♭, is still 41 cents, or just over two-fifths of an equal-tempered semitone. Consequently, on keyboard instruments, if there are only five accidental keys within the octave it is necessary to decide which are the most useful five. The usual choice is C♯, E♭, F♯, G♯ and B♭. If an A♭ is needed (*e.g.*, to use with E♭) and G♯ is used instead, the interval is 772.628 − 310.264 = 462.364, which is 35.681 cents out for a pure fourth and produces a howling sound known as the "wolf."

Equal Temperaments.—There is a second criterion for tempering a practical scale: fretted instruments, such as guitars, lutes and viols, that have strings tuned to any interval other than unisons or octaves need a cyclic division of the octave so that pure octaves can be obtained from all appropriate combinations of two strings and two frets. The only way to do this is to divide the octave into a certain number of exactly equal intervals. As a scale consists of five tones and two semitones, the octave must be divided into $5x + 2y$ parts where $x:y$ is approximately 193:112, or 1.7:1.

(1) The 12-note system. The simplest case is when $x = 2$ and $y = 1$. The octave is then divided exactly into 12 and this gives the normal equal temperament. All semitones and minor seconds are tempered to 100 cents each and the tone becomes 200 cents, etc. The same note has to serve both for C♯ and for D♭ (although music does require two different notes) and F has to be used for E♯, and so on. The major third is 14 cents too sharp and is just on the limit of pleasantness. The overriding merit of this system is its simplicity.

(2) The 19- and 31-note systems. If $x = 3$ and $y = 2$, the octave is divided into 19 equal intervals of 63 cents. The notes are in the order: C, C♯, D♭, D, etc., with a note between E and F for both E♯ and F♭. If $x = 5$ and $y = 3$, the octave is divided into 31 equal intervals of 38.7 cents. The order of notes is then C, D𝄫, C♯, D♭, C𝄪, D, . . . E, F♭, E♯, F, ... This system is remarkably close to mean-tone tuning (and to just intonation), as can be seen in the table of comparisons below. A harpsichord built in 1606 by Vito Trasuntino has 31 keys to the octave; it could be tuned exactly according to this system.

	Just	Mean Tone	12-note system	19-note system	31-note system
Chromatic semitone . .	71	76	100	63	77
Diatonic semitone . .	112	117	100	126	116
Tone . . .	182 or 204	193	200	190	194
Minor third . . .	316	310	300	316	310
Major third . . .	386	386	400	379	387
Fifth	702	697	700	695	697

History.—During the late medieval and Renaissance periods the 12-note equal temperament must have been familiar, as is shown by (1) the extensive use of fretted instruments such as the lute and viol; (2) organists' instructions to tune by fifths, tempering them just sufficiently to make the thirds usable; and (3) a few keyboard and viol fantasias that modulate enharmonically through all the possible keys. The equal-tempered major third, especially on keyboard instruments, is, however, offensively sharp to sensitive ears and toward the end of the Renaissance theorists were experimenting with other systems of tuning that produced better thirds.

The baroque era brought the decline of the lute and viol and the rise of wind instruments whose fingering favoured purer thirds, and this encouraged the introduction of mean-tone tuning. This system, however, allows little modulation and no enharmonic modulation.

Accordingly, as the technique of composition developed in the 18th century, despite attempts at duplication of accidentals on organs, the change to equal temperament was advocated by J. S. Bach and others. It should be remembered that an infinite range of intermediate tunings between mean tone and equal temperament is possible; it is probable that such intermediate tunings were much used in previous centuries.

(G. F. Ol.)

TEMPERANCE. The word temperance, which strictly means moderation, acquired a particular meaning in connection with intoxicating liquor and is here used in that sense.

History.—Ever since man in some distant age first discovered the process of fermentation by which sugar is converted into alcohol and carbonic acid, and experienced the intoxicating effects of the liquor so produced, there has been a temperance problem. The records of the ancient oriental civilization contain many references to it and from very remote times efforts were made by priests, sages and lawgivers in India, Persia, China, Palestine, Egypt, Greece and Carthage to combat the injury of drunkenness to the individual and society. But the evil appears never to have been so great or the object of so much attention in the ancient world as it came to be in western countries, especially in modern times.

The intoxicating drinks consumed by ancients were wines obtained from grapes or other fruits and beers made from various kinds of grain. These products were not confined to the east, but were also known to the ancient civilizations of Mexico and Peru and to other peoples who used the sugar-containing juices and other substances indigenous to their countries. In the time of the Romans the people of northern Europe used fermented liquors made from honey (mead), barley (beer) and apples (cider) in place of grape wine. All such drinks produce intoxication if taken in sufficient quantity, but their action is much slower and less violent than that of distilled spirits. This difference is a result of the greater concentration of alcohol in distilled spirits, which leads to a more rapid absorption of the alcohol into the blood stream and to a higher concentration of alcohol in the blood.

Potable distilled spirits were apparently not known until the 13th century. The distilled essence of wine or aqua vitae (brandy) is mentioned then as a new discovery by Arnaldus de Villa Nova, a chemist and physician, who regarded it, from the chemical or medical point of view, as a divine product. It probably came into use gradually, but once the art of distillation (*q.v.*) had been mastered it extended to other alcoholic products in countries where wine was not produced. (*See* also Alcoholic Beverages, Distilled.)

The concentrated form of alcohol thus evolved long carried with it the prestige of a divine essence given to it in the middle ages. It had potent properties and was held to possess great virtue, a view embodied in the name "water of life," and at one time almost universally held. Ardent spirits seemed particularly desirable to the people of the cold and damp regions of northern Europe, where people took to it with avidity when it became cheap and easily accessible. That happened in England in the early part of the 18th century, and from the ensuing bad results there eventually arose the modern temperance movement. The legislature had been busy with the liquor traffic for more than two centuries previously, but its task had been the repression of disorder: drunkenness was a nuisance and had to be checked in the interests of public order. It is significant that though drunkenness had been prevalent from the earliest times, necessity for attempts at legislative control did not arise until after the introduction of distilled spirits.

Intemperance was one of many problems which arose during the latter half of the 18th century to become the subject matter of social reform in the 19th century. A breach had been made in the unthinking traditional belief in the virtue of alcoholic liquor, and medical thought, as soon as it began to be concerned with health as distinguished from the treatment of disease, took up the matter. In 1804 Thomas Trotter of Edinburgh, Scot., published a book on the subject, an expansion of his academic thesis written in 1788. Benjamin Rush of Philadelphia, who had studied in Edinburgh and London, wrote a striking paper on the subject in the same year; and soon after this the organized temperance movement got under way in the United States.

Temperance and Prohibition Organizations.—In 1808 a temperance society was founded at Saratoga, N.Y., and 1813 the Massachusetts Society for the Suppression of Intemperance made its appearance. These seem to have been the earliest organizations, though the device of a pledge of abstinence had been introduced in 1800. The movement made rapid progress, mainly under the influence of the churches. In 1826 the American Society for the Promotion of Temperance was founded in Boston, Mass., and by 1833 there were 6,000 local societies in several states with more than 1,000,000 members. The campaign was directed primarily against the use of distilled spirits only, and the proposal to include all alcoholic drinks in the pledge of total abstinence, though adopted by a few societies, was rejected in 1833 by the American society, but accepted in 1836 and retained thereafter.

In Europe the earliest organizations were formed in Ireland. A temperance club is said to have been started at Skibbereen in 1818, and others followed; but it was in 1829 that the organized movement began to make effective progress with the formation of the Ulster Temperance society. By the end of that year there were 25 societies in Ireland and 2 or 3 in Scotland. In 1830 the movement spread to Yorkshire and Lancashire, and supported a newspaper called the *Temperance Societies' Record,* according to which there were then 127 societies with 23,000 paying members and 60,000 associated abstainers. In 1831 the British and Foreign Temperance society was founded in London with the bishop of London (Charles Blomfield) for president and Archbishop John Sumner for one of the vice-presidents. This important society, of which Queen Victoria became patron on her accession in 1837, came to an end in 1850, when the cause was under an eclipse.

The most remarkable episode in the temperance campaign at this period was the mission of the Rev. Theobald Mathew of Cork, commonly known as Father Mathew, greatest of all temperance missionaries. He traveled through Ireland in the years 1838–42 and everywhere excited intense enthusiasm. People flocked to hear him and took the pledge in crowds. In 1841 the number of abstainers in Ireland was estimated to be 4,647,000; in three years the consumption of spirits fell from 10,815,000 to 5,290,000 gal. This was not all due to Father Mathew, because great depression and distress prevailed at the same time, but he exercised an extraordinary influence. In 1843 he went to England, where he had less (though still great) success, and in 1850 to America. He died in 1856, by which time the cause had fallen into a depressed state in both countries. In the United States a flash of enthusiasm of a similar but smaller character, known as the Washingtonian movement, had appeared. In 1845 a law prohibiting the public sale of liquor was passed in New York state but repealed in 1847; in 1851 Maine adopted prohibition.

The Church of England Temperance society, much the largest at mid-20th century, was founded in 1862 and reconstituted in 1873. It was incorporated in 1907, reconstituted again in 1911 and 1921, and recognized by the church assembly in 1923. It carries on an extensive publication department and educational courses, police court and prison gate missions to seamen, traveling vans and inebriate homes.

The United Kingdom alliance, founded in 1853, became the chief political organization. Its object was prohibition of the drink trade, but it adopted the policy of local prohibition by means of local option. It began publication of the "national drink bill," a calculation of the national expenditure on alcoholic liquors. Other legislative organizations in Britain are the National Temperance league, the British Temperance league, the National United Temperance council, the National Temperance federation, the Temperance Legislation league and others in Scotland, Ireland and Wales. There is also a Royal Army Temperance association and a Royal Naval Temperance society. International societies enumerated are the World League Against Alcoholism, the World Prohibitive federation, the International Orders of Good Templars and the World's Woman's Christian Temperance union.

Thus temperance and total abstinence became the objects of the united efforts of education and legislation in many regions. This union was probably promoted in part by the fact that once a person has become an alcoholic recovery usually requires total abstinence. In later years the organized movement in many countries embraced both voluntary and compulsory abstinence and combined the inculcation of individual abstinence with the promotion of legislation for the reduction or suppression of the alcoholic liquor traffic. On the whole the latter predominated, particularly in the United States, where organized agitation made temperance or prohibition partly a political question and produced various experiments in legislation. (*See* PROHIBITION.) Besides combining the moral and the political elements the modern movements were characterized by the following features: (1) international organization; (2) organized co-operation of women; (3) juvenile temperance; (4) teaching of temperance in schools and elsewhere; (5) scientific study of alcohol and inebriety.

International Organization.—The first international temperance organization appears to have been the Order of Good Templars, formed in 1851 at Utica, N.Y. It spread over the United States and Canada, and in 1868 was introduced into Great Britain. Several years later it was extended to Scandinavia, where it became very strong. Temperance societies had previously existed in Norway from 1836 and in Sweden from 1837; these seem to be the earliest examples on the continent of Europe. The Good Templars organization also spread to several other European countries, to Australasia, India, south and west Africa and South America. Several other international societies were founded, and international congresses were held, the first in 1885 at Antwerp, Belg. A World's Prohibition conference was held in London in 1909, attended by about 300 delegates from temperance societies in nearly all parts of the world. This conference resulted in the foundation of an International Prohibition federation, embracing every country in Europe with three or four minor exceptions, the United States, Mexico, Argentina, much of the British commonwealth, China, Japan, Israel, Tunisia and Hawaii.

Alcoholics Anonymous.—This is an organization of former alcoholics who have become total abstainers. It is a national organization, started in the United States but gradually extended to other countries. Its primary object is to aid, by education and personal service, other alcoholics to recover from their addiction.

W.C.T.U.—The organization of women, which has become international, dates from 1874, when the National Women's Christian Temperance union was founded at Cleveland, O. It employs educational and social as well as political means in promoting legislation. About 1883 Frances Willard, who had been the moving spirit of the union, carried this organization of women into other lands and formed the World's Woman's Christian Temperance union, which by the 1960s had branches in about 50 countries.

Juvenile Groups.—The inclusion of children in temperance organization goes back to 1847, when a society of juvenile abstainers who had taken the pledge was formed at Leeds; it took the name of Band of Hope. The practice spread, and in 1851 a Band of Hope union was formed. Within the next century a number of such unions were established in Scotland, Ireland and separate counties in England; the Bands of Hope were said to number about 15,000 in all.

School Programs.—The teaching of temperance in schools, a feature of social education and moral propaganda, was begun by private effort in 1852 when John Hope inaugurated a regular weekly visitation of day schools in Edinburgh. In 1875, at the invitation of the National Temperance league, Sir Benjamin Richardson wrote his *Temperance Lesson Book,* which was adopted by many schools as a primer. In 1889 school teaching by traveling lecturers was taken up by the U.K. Band of Hope union, and the example was followed by many other societies which spent large sums on itinerant lecturers. The Church of England Temperance society carried on similar work. In 1906 the board of education in Ireland made hygiene and temperance a compulsory subject in the public schools; in 1909 the English board of education issued a syllabus of temperance teaching; in Scotland courses in hygiene and temperance were adopted by many local educational authorities.

The question of temperance teaching in the United States was first taken up by the Woman's Christian Temperance union in

1879; it was believed that by teaching the physiological effects of alcohol to all children the problem of intemperance would be effectually solved, and a systematic political campaign was planned and carried out for the purpose of obtaining compulsory legislation to give effect to this idea. The campaign was successful in New York in 1884, in Pennsylvania in 1885 and subsequently in other states. Laws were passed in every state making antialcohol teaching part of the curriculum in the public schools. Teaching of temperance became compulsory in Canada except in Quebec and Prince Edward Island; in France in 1902; in Sweden in 1892; and in Ireland. It became optional in Australia, South Africa, some parts of India, Belgium, Finland, Denmark, Norway, Germany and Switzerland. *See* also PROHIBITION; LOCAL OPTION; LIQUOR LAW. (A. SL.; A. J. CN.)

TEMPERATURE: *see* CALORIMETRY; HEAT; INVERSION OF TEMPERATURE; METEOROLOGY; SEASONS; THERMOMETRY.

TEMPERATURE, BODY: *see* ANIMAL HEAT; FEVER.

TEMPERING: *see* HEAT-TREATMENT.

TEMPEST, DAME MARIE (MARY SUSAN ETHERINGTON) (1866–1942), English actress, was born in London on July 15, 1866. Educated in Belgium, she studied piano in Paris and won three medals as a student of voice at the Royal Academy of Music in London. For nearly a decade at the end of the 19th century she starred in light opera and made her first American tour with the Duff Comic Opera company. In 1895 a London musical comedy producer had to buy out her American bookings in order to get her back to London to star in *An Artist's Model*. This was followed by roles in *Geisha, Greek Slave* and *San Toy*. In 1900 she abandoned light opera for straight comedy roles and made an auspicious start by creating the role of Nell Gwyn in *English Nell*. Returning from a second tour of America in 1911, Dame Tempest went into management and in 1914 began an eight-year world tour, returning to London in 1922. In 1925 she appeared in London in *Hay Fever*, a play written for her by Noel Coward and in 1937 she was created dame commander of the Order of the British Empire. During the bombings of London in the early part of World War II, her home was demolished and most of her possessions lost. Her last appearance on the stage was in *The First Mrs. Fraser*. Illness prevented her from taking part in a new show she was rehearsing shortly before her death which occurred in London, on Oct. 14, 1942.

TEMPLARS. The Knights Templars, or Poor Knights of Christ and of the Temple of Solomon (*pauperes commilitones Christi templique Salomonici*), formed one of the three great military orders, founded in the 12th century. Unlike the Hospitallers and the Teutonic Knights it was a military order from its very origin. Its founders were a Burgundian knight named Hugues de Payns and Godeffroi de St. Omer, a knight from northern France, who in 1119 undertook the pious task of protecting the pilgrims who, after the first crusade, flocked to Jerusalem and the other sacred spots in the Holy Land. They were quickly joined by six other knights and soon afterwards organized themselves as a religious community, taking an oath to the patriarch of Jerusalem to guard the public roads, to forsake worldly chivalry, "of which human favour and not Jesus Christ was the cause," and, living in chastity, obedience and poverty, according to the rule of St. Benedict, "to fight with a pure mind for the supreme and true King."

To this nascent order of warrior monks Baldwin II., king of Jerusalem, handed over a part of his royal palace lying next to the former mosque of al-Aksa, the so-called "Temple of Solomon," whence they took their name. They had at first no distinctive habit, wearing any old clothes that might be given to them. Nor was their community exclusive. Their primitive rule seems to have enjoined them especially to seek out excommunicated knights, and to admit them, after absolution by the bishop, to their order, and they thus served a useful purpose in at once disciplining and converting the unruly rabble of "rogues and impious men, robbers and committers of sacrilege, murderers, perjurers and adulterers"[1] who streamed to the Holy Land in hope of plunder and salvation. It was this rule which led later to the most important privilege of the order, the immunity from sentences of excommunication pronounced by bishops and parish priests.

This practice might have brought them at once under the suspicion of the Church, and it soon became expedient to obtain the highest sanction for the new order and its rules. In the autumn of 1127 accordingly Hugues de Payns, with certain companions, appeared in Europe, where he was fortunate enough to secure the enthusiastic support of the all-powerful abbot of Clairvaux. Grateful pilgrims had already begun to enrich the order; the *De laude novae militiae*, a glowing panegyric of this new and holy conception of knighthood, addressed by Bernard to Hugues de Payns by name, insured the success of his mission. In 1128 the council of Troyes discussed and sanctioned the rule of the order.

Rule of the Temple.—No ms. of the original French Rule of the Temple (*Règle du Temple*) exists; but in essentials the later copies preserve the matter and spirit of the primitive Rule, and they prove that to the end the order was, in principle at least, submitted to the same strict discipline as at the beginning. Of a secret Rule, in spite of the most diligent research, no trace has ever been found. It is now generally held that none ever existed. The legend of its existence, so fatal to the order, is probably traceable to the fact that the complete Rule was jealously guarded by the chief office-bearers of the order.

The *Règle du Temple* in its final form as we now possess it contains the rules for the constitution and administration of the order; the duties and privileges of the various classes of its personnel; the monastic rules, regulations as to costume and as to religious services; rules for the holding of chapters, and a summary of offences and their punishment; the procedure at the election of a grand master and at receptions into the order; a definition of the relations of the order to the pope, and to other religious orders.

At the head of the order was the master of the Temple at Jerusalem (in Cyprus after the fall of the Latin Kingdom), known as the grand master. His authority was very great—except in certain reserved cases his word was law—but he was not absolute. Thus in matters of special importance—alienation of the estates of the order, attack on a fortress, declaration of war, conclusion of an armistice, reception of a new brother—he had to consult the chapter, and was bound by the vote of the majority; nor could he modify or abrogate a decree of the council of the order without their consent. He had to obtain the consent of the chapter also to the nomination of the grand commanders of the provinces of the order; the lesser offices were absolutely in his gift. He was elected by a complicated process, a chapter summoned *ad hoc* electing a "commander of the election" and one other brother who, after vigil and prayer, co-opted two more, these four choosing another two, and so on till the number of the twelve apostles had been reached. A chaplain, representing Jesus Christ, was then added to complete the electoral college. (*See* Curzon, *Règle du Temple*, p. xxxv.).

Of peculiar importance were the *chaplains* (*fratres capellani*). These did not originally form part of the order, which was served by priests from outside. The bull *Omne datum optimum* of 1163 imposed on clerics attaching themselves to the order an oath of life-long obedience to the grand master; by the middle of the 13th century the chaplains took the same oath as the other brothers and were distinguished from them only by their orders and the privileges these implied (*e.g.*, they were spared the more humiliating punishments, shaved the face, and had a separate cup to drink out of). The order thus had its own clergy, exempt from the jurisdiction of diocesan bishops and parish priests, owing obedience to the grand master and the pope alone.

It remains to be said that the brethren were admitted either for life or for a term of years. Married men were also received, but on condition of bequeathing one half of their property to the order.

History.—Long before St. Bernard's death (1153) the new order was established in almost every kingdom of Latin Christendom, and every establishment was the recipient of rich endowments. Henry I., for example, granted them lands in Normandy; and Louis VII. in the latter years of his reign gave them a

[1] Bernard of Clairvaux, *De laude novae militiae*, cap. v. (in Migne, *Patrol lat.* 182, p. 928).

piece of land outside Paris, which in later times became known as the Temple, and was the headquarters of the order in Europe, and Stephen of England granted them the manors of Cressing and Witham in Essex, and his wife Matilda that of Cowley, near Oxford. Spiritual privileges were granted to them by the popes as lavishly as temporal possessions by the princes and people. Pope Adrian IV. allowed them to have their own churches; Eugenius III. added to these the right to have churchyards. They were, moreover, as defenders of the Church exempted from the payment of tithes. Finally, they were exempted from the action even of general censures and decrees of the popes, unless mentioned in them by name. Very soon the order refused to submit in any way to the ordinary jurisdiction of the diocesan bishops and formed in effect a separate ecclesiastical organization under the pope as supreme bishop. The result was that, scarce twenty-five years after its foundation, the order was at open feud with bishops and parish priests, and the popes found it necessary to issue decree after decree to protect it from violence and spoliation.

So long, however, as the attention of the papacy and of Christendom was fixed on the problem of recovering and safeguarding the Holy Land, the position of the Templars was unassailable and all efforts to curb the growth of their power vain. The later history of the Templars is therefore the history of the CRUSADES (*q.v.*).

Fall of Latin Kingdom.—There is something Homeric in the story of the fall of the Latin kingdom as related by the historians of the next century. A French knight, Gerard de Riderfort, coming to the East in quest of fortune, attached himself to the service of the ambitious Raymond, count of Tripoli; looking for the hand of some wealthy widow in reward. Failing ignominiously in this, Gerard enrolled himself among the Templars, biding his time for revenge, and was elected grand master in 1184. Baldwin IV., king of Jerusalem, died in 1185, leaving the throne to his young nephew Baldwin V., the son of his sister Sibylla, under the guardianship of Raymond, whose term of office was brief, as the little king died in September 1186. This was Gerard's opportunity. The Templars carried the body of their dead sovereign to Jerusalem for burial; and then, unknown to the barons of the realm, Gerard and the patriarch crowned Sibylla and her husband Guy de Lusignan, Raymond's rival. This was the overthrow of Raymond's ambition; and both Latin and Arabic writers are agreed that the Christian count and the Mohammedan sultan now entered into an alliance. To break this friendship and so save the kingdom, Gerard and the grand master of the Hospitallers were sent north to make terms with Raymond. But the rash valour of the Templars provoked a hopeless contest with 7,000 Saracens. The grand master of the Hospitallers was slain; but Gerard made his escape with three knights to Nazareth (1st May 1187). In this emergency Raymond became reconciled with Guy; and Gerard placed the Temple treasures of Henry II. at his king's disposal. Once more it was the Templars' rashness that led to the disastrous battle of Hittin (4th July). Gerard and the king fell into the hands of Saladin, but were released about a year later; Raymond of Tripoli made his escape through treachery or fortune; and 230 Templars fell in or after the battle, for the fight was scarcely over before Saladin ordered all the Templars and Hospitallers to be murdered in cold blood. One after another the Christian fortresses of Palestine fell into the hands of Saladin. Jerusalem surrendered on 2nd–3rd October 1187, and the treasures of the Temple coffers were used to purchase the redemption of the poorer Christians, part of whom the Templar warriors guarded on their sad march from the Holy City to Tripoli. Part of their wealth was expended by Conrad of Montferrat in the defence of Tyre; but, when this prince refused to admit Guy to his city, both the Templars and the Hospitallers from the neighbouring parts flocked to the banner of their released king and accompanied him to the siege of Acre (22nd August 1189). In his company they bore their part in the two years' siege and the terrible famine of 1190–91; and their grand master died in the great battle of 4th October 1189, refusing to survive the slaughter of his brethren.

On the capture of Acre Philip Augustus established himself in the palace of the Templars, who are, however, stated to have sympathized with Richard. The English king sold them the island of Cyprus for 100,000 besants; but, unable to pay the purchase money, they transferred the debt and the principality to Guy of Lusignan. Richard consulted them before deciding on any great military movement; and in June 1192 they advocated the bold plan of an advance on Egypt rather than on Jerusalem. It was in the disguise of a Templar and in a Templar galley that Richard left the Holy Land, after the failure of his own military schemes. When Acre was recovered, the Templars, like the Hospitallers, received their own quarters in the town, which from this time became the centre of the order. On the death of Henry of Champagne (1197) they vetoed the election of Raoul de Tabarie; after the death of his successor Amalric they refused to renew the truce with Saladin's brother, Saif al-Dīn, and led an expedition against the Saracens before the arrival of the new king, John de Brienne, at whose coronation in 1210 William de Chartres, the grand master, was present. Seven years later, with the aid of Walter de Avennis and of the Teutonic Knights, they commenced the building of their fortress of Castle Pilgrim, near Acre, on a rocky promontory washed by the Mediterranean on every side except the east.

It was from this castle that in May 1218 the fifth crusade started for the expedition against Egypt. The Templars were the heroes of the siege of Damietta, at which William de Chartres was slain. "First to attack and last to retreat," they saved the Christian army from annihilation on 29th August 1219; and when the city surrendered (5th November) the only one of its twenty-eight towers that had begun to give way had been shaken by their engines. On the other hand, it was largely owing to their objections that John de Brienne refused the sultan's offer to restore Jerusalem and Palestine.

From the very first the Templars seem to have been opposed to Frederick II., and when he landed at Acre (7th September 1228) they refused to march under the banners of an excommunicated man, and would only accompany his host from Acre to Joppa in a separate body. They were accused of notifying Frederick's intended pilgrimage to the Jordan to the sultan, and they were certainly opposed to Frederick's ten years' peace with Al-Kāmil, the sultan of Egypt, and refused to be present at his coronation in Jerusalem. On neither side was the treaty fully observed; and preparations were made in Europe for a fresh crusade. In the meantime open dissension broke out between the Templars and the Hospitallers, the former advocating negotiations with the sultans of Damascus and Kerak, the latter with the sultan of Egypt; and when Richard of Cornwall arrived (11th October) he had to decide between the two rival orders and their opposing policies. After some hesitation he concluded a treaty with the sultan of Egypt, much to the annoyance of the Templars, who openly mocked his efforts. On his departure the three orders came to open discord: the Templars laid siege to the Hospitallers in Acre and drove out the Teutonic Knights "in contumeliam imperatoris." They were successful on all sides. The negotiations with Damascus and Kerak were reopened, and in 1244 Hermann of Perigord wrote to the princes of Europe that after a "silence of fifty-six years the divine mysteries would once more be celebrated in the Holy City."

It was in this moment of danger that the sultan of Babylon called in the barbarous Kharizmians, whom the Mongol invasions had driven from their native lands. These savages, entering from the north, flowed like a tide past the newly built and impregnable Templar fortress of Safed, swept down on Jerusalem, and annihilated the Christian army near Gaza on October 18th, 1244. From this blow the Latin kingdom of the East never recovered; 600 knights took part in the battle; the whole force of the Templars, 300 in number, was present, but only 18 survived, and of 200 Hospitallers only 16. The masters of both orders were slain or taken prisoners, and Jerusalem was lost to Christendom once more. The havoc caused by the Kharizmians was alleged by pope Innocent IV. as the reason for again summoning Christendom to the rescue of the Holy Land. Recognizing the fact that the true

way to Jerusalem lay through Egypt, Louis IX. led his host to the banks of the Nile, being accompanied by the Templars. Their master, William de Sonnac, attempted in vain to restrain the rash advance of the count of Artois at the battle of Mansura (8th February 1250), which only three Templars survived. St. Louis, when captured a few weeks later, owed his speedy release to the generosity with which the order advanced his ransom-money.

A new enemy was now threatening Mohammedan and Christian alike. For a time the Mongol advance may have been welcomed by the Christian cities, as one after another the Mohammedan principalities of the north fell before the new invaders. But this new danger stimulated the energies of Egypt, which under the Mameluke Bibars encroached year after year on the scanty remains of the Latin kingdom. The great Frankish lords, fearing that all was lost, made haste to sell their lands to the Templars and Hospitallers before quitting Palestine for ever. But they lost their power of resistance, and became so enfeebled as to welcome the treaty which secured them the plain of Acre and a free road to Nazareth as the result of the English crusade of 1272. While thus weak against external foes, the Templars were strong enough for internal warfare. In 1277 they espoused the quarrel of the bishop of Tripoli, formerly a member of the order, against his nephew Bohemond, prince of Antioch and Tripoli, and began a war which lasted three years. In 1276 their conduct drove Hugh III., king of Cyprus and Jerusalem, from Acre to Tyre. In the ensuing year, when Mary of Antioch had sold her claim to the crown to Charles of Anjou, they welcomed this prince's lieutenant to Acre and succeeded for the moment in forcing the knights of that city to do homage to the new king. Thirteen years later (26th April 1290) Tripoli fell, and next year Acre, after a siege of six weeks, at the close of which (16th May) William de Beaujeu, the grand master, was slain. The few surviving Templars elected a new master, and sailed for Cyprus, which now became the headquarters of the order.

Power and Influence of the Order.—For more than a hundred years the Templars had been one of the wealthiest and most influential factors in European politics. If we confine our attention to the East, we realize but a small part of their enormous power. Two Templars were appointed guardians of the disputed castles on the betrothal of Prince Henry of England and the French princess in 1161. Other Templars were almoners of Henry III. of England and of Philip IV. of France. One grand master was godfather to a daughter of Louis IX.; another, despite the prohibition of the order, is said to have been godfather to a child of Philip IV. They were summoned to the great councils of the Church, such as the Lateran of 1215 and the Lyons council of 1274. Frederick II.'s persecution of their order was one of the main causes of his excommunication in 1239; and his last will enjoined the restoration of their estates. Their property was scattered over every country of Christendom, from Denmark to Spain, from Ireland to Cyprus. But the wealth of the Templars was due not so much to their territorial possessions as to the fact that they were the great international financiers and bankers of the age. The Paris Temple was the centre of the world's money market. In it popes and kings deposited their revenues, and these vast sums were not hoarded but issued as loans on adequate security. Above all, it was the Templars who made the exchange of money with the East possible. It is easy, indeed, to see how they were the ideal bankers of the age; their strongholds were scattered from Armenia to Ireland, their military power and strict discipline ensured the safe transmission of treasure, while their reputation as monks guaranteed their integrity. Thus they became the predecessors, and later the rivals, of the great Italian banking companies. To take interest (usury) was of course unlawful. The method of circumventing this seems to have been that the mortgagees paid to the mortgagors a *nominal* rent which was used towards the reduction of the debt. The difference between this and the *real* rent represented the interest. A document throwing a vivid light on the banking methods of the Templars and Hospitallers is a charter of Margaret, queen of the English, A.D. 1186, from the abbey of Fontevrault, printed in *Calendar of Documents*, France (1899), vol. i., ed. J. H. Round, No. 1084.

Suppression of the Order.—Never had the order of the Temple been to all appearance more powerful than immediately before its ruin. Sovereign power, in the sense of that of the Teutonic Knights in Prussia or the Knights of St. John in Rhodes and later in Malta, it had never possessed; but its privileges and immunities constituted it a church within the church and—in France at least—a state within the state. Philip IV., indeed, in pursuance of his policy of centralizing power in the crown, had from 1287 onwards made tentative efforts to curtail the power and wealth of the order; but soon his necessities forced him to a temporary change of policy. In January 1293 the privileges of the order in and about Paris were confirmed and extended, and in 1297 Philip borrowed 5,200 *livres tournoises* from the Paris Temple. Then came the great quarrel with Pope Boniface VIII., and on the 10th of August 1303 the king signed with Hugues de Peraud, the general visitor of the French Templars, a formal treaty of alliance against the pope. On the 6th of February 1304 Boniface's successor, Benedict XI., once more confirmed all the Templars' privileges; while Philip, for his part, appointed Hugues de Peraud receiver of the royal revenues and, under pressure of the disastrous campaign in Flanders, in June granted a charter exempting the order from all hindrances to the acquisition of property. Two years later the king took refuge in the Temple from the violence of the Paris mob, and so late as the spring of 1307 was present at the reception of a new Templar.

Yet for some two years past the king had been plotting a treacherous attack on the order. His motives are clear: he had used every expedient to raise money, had robbed and expelled the Jews and the Lombard bankers, had debased the coinage; the suppression of the Templars would at once rescue him from their unwelcome tutelage and replenish his coffers. He cherished also another ambition. The question of an amalgamation of the great military orders had often been mooted; the project had been approved by successive popes in the interests of the Holy Land; it had been formally proposed at the Lyons council of 1274, only to be rejected by the opposition of the Templars and Hospitallers themselves. To Philip this scheme commended itself as an opportunity for bringing the orders under the control of the French crown; there was to be but one order, that of the "Knights of Jerusalem," of which the grand master was always to be a prince of the royal house of France. Clearly, it only needed an excuse and a favourable opportunity to make him attack the Templars; and, once having attacked them, nothing short of their entire destruction would have been consistent with his safety. The excuse was found in the denunciation of the order for heresy and unspeakable immoralities by a venal informer; the opportunity was the election of a pope, Clement V., wholly devoted to the interests of the king of France.

For perhaps half a century there had been strange stories circulating as to the secret rites practised by the order at its midnight meetings, stories which probably had their origin in the extreme precautions taken by the Templars, originally perhaps for military reasons, to secure the secrecy of their proceedings, which excited popular curiosity and suspicion.

In the spring of 1304 or 1305 a certain Esquiu de Floyran of Béziers pretended to betray the "secret of the Templars" to James II. of Aragon. The pious king, who had every reason to think well of the order, did not affect to be convinced; but the prospect of spoils was alluring, and he seems to have promised the informer a share of the booty if he could make good his charges. Esquiu now turned to Philip of France, with more immediate success. For the purpose of collecting additional evidence the king caused twelve spies to find admission to the order, and in the meantime sought to win over the pope to his views. Clement V. owed the tiara to the diplomacy of Philip's agents, perhaps to their gold; but though a weak man, and moreover a martyr to ill health, he was not so immediately accommodating as the king might have wished, expressing his disbelief in the charges against the order, and, though promising an inquiry, doing his best to procrastinate. Philip determined to force his hand. All France

was at this time under the jurisdiction of the Inquisition, and the Inquisition could act without consulting the pope. The grand inquisitor of France, William of Paris, was Philip's confessor and creature. The way was thus open for the king to carry out his plan by a perfectly legal method. His informers denounced the Templars to the Inquisition, and the grand inquisitor—as was the customary procedure in the case of persons accused of heresy—demanded their arrest by the civil power. On the 14th of September 1307, accordingly, Philip issued writs to his *baillis* and seneschals throughout the kingdom, directing them to make preparations to arrest the members of the order.

The Templars had for some time past been aware of the charges against them. On the 6th of June 1306 Pope Clement had summoned Jacques de Molay, the grand master, from Cyprus to France, in order to consult him on the projected crusade. He had obeyed the call, and, in an interview with the pope, had taken the opportunity to demand a full inquiry. They had, however, taken no measures to defend themselves; the sudden action of the king took them wholly by surprise; and on the night of Friday, the 13th of October 1307, their arrest was effected without difficulty, Jacques de Molay himself with sixty of his brethren being seized in Paris.

The Templars were caught in toils from which there was no escape. To force them to confess, they were first tortured by the royal officials, before being handed over to the inquisitors to be, if need were, tortured again. In Paris alone thirty-six died under the process. The result was, at the outset, all that the king could desire. Of 138 Templars examined in Paris between the 19th of October and 24th of November, some of them old men who had been in the order the greater part of their lives, 123 confessed to spitting on (or "near") the crucifix at their reception. Many of the prisoners confessed to all the charges, however grotesque. But the most damning confession was that of the grand master himself, publicly made with tears and protestations of contrition and embodied in a letter (October 25) sent to all the Templars in France. He had been guilty, he said, of denying Christ and spitting on the cross.

To the pope, meanwhile, the proceedings in France were to the highest degree unwelcome. He had, indeed, become convinced, if not of the general guilt of the order, at least of the guilt of some of its members. But the affair was one which he desired to reserve for his own judgment; Philip's action he interpreted, rightly, as an encroachment of the civil power on the privileges and property of the Church, and his fears were increased when the French king, without consulting him, sent letters to King James of Aragon, Edward II. of England, the German king Albert and other princes, calling upon them to imitate his example. On the 27th of October Clement issued letters suspending the powers of the Inquisition in France. What followed is not clear, for the documentary evidence for these months is very defective; but on the 22nd of November the pope issued a bull calling on all kings and princes to arrest the Templars everywhere, his motive probably being to forestall the probable action of the secular powers and keep the affair in his own hands. All scruples and hesitations now vanished. In England the Templars were arrested on the 10th of January 1308, in Sicily on the 24th of the same month, in Cyprus on the 27th of May; in Aragon and Castile the process was less easy, for the knights, forewarned, had put their fortresses into a state of defence, notably their strong castle of Monzon, which was only taken after a long siege on the 17th of May, while the last of the Templars' strongholds, Castellat, did not fall until Nov. 2nd.

Meanwhile, on the 26th of May, Philip had made his solemn entry into Poitiers, where the pope and cardinals had already assembled for the purpose of conferring with the king on the matter. After stormy debate, an arrangement was made. The king agreed to hand over to the papal commissioners the property and persons of the Templars; Clement, for his part, withdrew the sentence of suspension against the grand inquisitor of France and ordered an inquisition into the charges against individual Templars by the diocesan bishops with assessors nominated by himself. The examination of the grand master, of the grand visitor of France, and of the grand preceptors of Cyprus, Normandy and Aquitaine he reserved to himself. Inquisition was to be made into the conduct of the order in each country by special papal commissions; and the fate of the order as a whole was to be decided by a general council, summoned at Vienne for the 1st of October 1311, when the question of the guilt of the order might be considered. Meanwhile the pope and cardinals had elaborated the organization of the new inquisition. There was much confusion and delay, however, and the actual public trial did not begin till the 11th of April, 1310. Many Templars, trusting in the assurance implied in their citation, had volunteered to defend the order and withdrew their previous confessions. They were soon undeceived; the commission was packed with creatures of the crown. The evidence given in Paris for or against the order was, it was soon found, used against the individual Templars on their return to the provinces; the retraction of a confession, under the rules set up for the diocesan inquisition, was punished with death by fire. Sixty-seven Templars perished in this way during May 1310. Meanwhile Clement and Philip had come to terms. The pope condemned the Templars. The council of Vienne met in October 1311. A discussion arose as to whether the Templars should be heard in their own defence. Clement, it is said, broke up the session to avoid compliance; and when seven Templars offered themselves as deputies for the defence he had them cast into prison. Towards the beginning of March Philip came to Vienne, and he was seated at the pope's right hand when that pontiff delivered his sermon against the Templars (3rd April 1312), whose order had just been abolished, not at the general council, but in private consistory (22nd March). On 2nd May 1312 he published the bull *Ad Providam*, transferring the goods of the society, except for the kingdoms of Castile, Aragon, Portugal and Majorca, to the Knights of St. John.

The final act of the stupendous tragedy came early in 1314. Jacques de Molay, the grand master, had not hitherto risen to the height of his great position; the fear of torture alone had been enough to make him confess, and this confession had been used to extract avowals from his brethren, subject as they were to unspeakable sufferings and accustomed to yield to the military chief. Humiliation on humiliation had been heaped on the wretched man, public recantations, reiterated confessions. Before the papal commission he had flamed into anger, protested, equivocated—only in the end to repeat his confession once more. The same had happened before the commission of cardinals at Chinon; the audience with the pope, which he demanded, he had never obtained. On the 6th of May 1312 Pope Clement issued his final decision as to the fate of the Templars in general; that of the five great offices of the order he reserved in his own hand. With this a silence falls over the history of the Templars; the fate of the order had been decided, that of the individuals still under trial was of little interest to contemporary chroniclers. Then the veil is suddenly lifted. Jacques de Molay has found his wonted courage at last, and with him Gaufrid de Charney, the preceptor of Normandy; on the 18th or 19th of March 1314 they were brought out on to a scaffold erected in front of Notre Dame, there in the presence of the papal legates and of the people to repeat their confessions and to receive their sentence of perpetual imprisonment. Instead, they seized the opportunity to withdraw their confessions and to protest to the assembled thousands the innocence of the order. King Philip the Fair did not wait to consult the Church as to what he should do; he had them burnt.

A word must be added as to the significance of the work of the Templars and of the manner of their fall in the history of the world. Two great things the order had done for European civilization: in the East and in Spain it had successfully checked the advance of Islam; it had deepened and given a religious sanction to the idea of the chivalrous man, the *homo legalis*, and so opened up, to a class of people who for centuries to come were to exercise enormous influence, spheres of activity the beneficent effects of which are still recognizable in the world. On the other hand, the destruction of the Templars had three consequences fateful for Christian civilization. (1) It facilitated the conquests of the Turks by preventing the Templars from playing in Cyprus

the part which the Knights of St. John played in Malta. (2) It partly set a precedent for, partly confirmed, the cruel criminal procedure of France, which lasted to the Revolution. (3) It set the seal of the highest authority on the popular belief in witchcraft and personal intercourse with the devil, sanctioned the expedient of wringing confessions of such intercourse from the accused by unspeakable tortures, and so made possible the hideous witch-persecutions which darkened the later middle ages and, even in Protestant countries, long survived the period of Reformation.

On the question of the guilt or innocence of the Templars in respect of the specific charges on which the order was condemned opinion has long been divided. Their innocence was maintained by the greatest of all their contemporaries, Dante (Purg. xx. 92), and by the historian Villani and others. In more recent times a certain heat was introduced into the discussion of the question owing to its having been for centuries brought into the arena of party controversy, between Protestants and Catholics, Gallicans and Ultramontanes, Freemasons and the Church. Thus in 1654 Pierre Du Puy, librarian of the Bibliothèque Royale, published his work on the Templars to confute those who sought to establish their innocence in order to discredit a king of France. On the other hand, Nicolas Gurtler published his *Historia Templariorum* (Amsterdam, 1691, 2nd ed. 1703) to show, as a good Protestant, that the Templars had the usual vices of Roman Catholics, while, according to Loiseleur, the later editors of Du Puy were Freemasons who, under false names, garbled the old material and inserted new in the interests of the supposed origin of their own order in that of the Templars. Several Roman Catholic champions of the order now entered the field, notably the Premonstratensian canon R. P. M. Jeune, prior of Etival, who in 1789 published at Paris his *Histoire critique et apologétique de l'ordre des chevaliers . . . dits Templiers*, a valuable work directed specifically against Gurtler and Du Puy. In the 19th century a fresh impetus was given to the discussion by the publication in 1813 of Raynouard's brilliant defence of the order. The challenge was taken up, among others, by the orientalist F. von Hammer-Purgstall who, in a paper entitled "Mysterium Baphometis" (*Fundgraben des Orients*, vol. vi., Vienna 1818), attempted to prove that the Templars followed the doctrines and rites of the Gnostic "Ophites," and gave reproductions of obscene representations of supposed Gnostic ceremonies and of mystic symbols said to have been found in Templars' buildings. W. F. Wilcke (*Geschichte des Tempelherrenordens*, 1826, 2nd. ed. 1860), while rejecting Hammer's main conclusions as unproved, argued for the existence of a secret doctrine based not on Gnosticism but on the monotheism of Islam, of which Baphomet (Mahomet) was a symbol. On the other hand, Wilhelm Havermann (*Geschichte des Ausganges des Tempelherrenordens*, 1846) decided in favour of the innocence of the order. This view was also taken by a succession of German scholars; in England, by C. G. Addison; and in France by a whole series of conspicuous writers: Mignet, Guizot, Renan, and others. On the other hand, Hans Prutz, in a series of brilliant contributions, of which the most well-considered is his *Entwicklung und Untergang des Tempelherrenordens* (1888), maintained that the custom of denying Christ and spitting on the cross was often, and in some provinces universally, practised at the reception of the brethren, "as a coarse test of obedience, of which the original sense had partly been forgotten, partly heretically interpreted under the influence of later heresies." Prutz points out that the failure of the Crusades had weakened men's absolute belief in Christianity, at least as represented by the mediaeval Church (*Kulturgeschichte der Kreuzzüge*, p. 268 ff.).

H. C. Lea, in his *History of the Inquisition* (1888, vol. iii.), had already come independently to the conclusion that the Templars were innocent. Lastly appeared the fascinatingly interesting and closely reasoned book of Professor H. Finke (1907) which, based partly on a mass of new material drawn from the Aragonese archives, had for its object to establish the innocence of the order on an incontrovertible basis (*Papsthum und Untergang des Templenordens*, 1907).

Bibliography.—The original sources given by Du Puy, though often valuable, were selected and edited with a purpose, but Michelet's *Procès des Templiers* (2 vols., 1851, 1861), gives the original minutes of the trial preserved at the Bibliothèque Nationale. This is analysed by Gmelin (*op. cit.*). Of documents published elsewhere, the most important are in Finke (*op. cit.*, from the Aragonese archives) and K. Schottmüller (*Untergang des Templerordens*, 1887, from the Vatican archives). The Rule of the Temple has several times been published; the most convenient edition is that of H. de Curzon, *La Règle du Temple* (1886); *see* also Maillard du Chambure, *Règle et statutes secrets des Templiers* (1840). In addition to the works already mentioned, the following may be named: M. Lavocat, *Procès des frères et de l'ordre du Temple* (Paris, 1888); G. Schnürer, *Die ursprüngliche Templerregel* (1903); C. G. Addison, *The Knights Templars* (3rd ed. 1854), which contains a valuable account of the suppression of the order in England; Marquis d'Albon, *Cartulaire général de l'Ordre du Temple* (1922); G. Lizerand (ed.), *Le Dossier de l'affaire des Templiers* (1923). (W. A. P.; X.)

TEMPLE, FREDERICK (1821–1902), English divine, archbishop of Canterbury, was born in Santa Maura, one of the Ionian Islands, the son of Major Octavius Temple. He was educated at Blundell's School, Tiverton, and at Balliol College, Oxford. The "Tractarian Movement" had set in five years before he went up to Oxford, but the memorable tract, No. 90, had not yet been written. After much discussion and reflection he drew closer to the camp of "the Oxford Liberal Movement." In 1842 he took a "double-first" and was elected fellow of Balliol, and lecturer in mathematics and logic. Four years later he took orders, and with the aim of helping forward the education of the very poor, he accepted the headship of Kneller Hall, which served at that time for the training of masters of workhouse and penal schools. But the experiment was not altogether successful, and Temple himself advised its abandonment in 1855. He then accepted a school-inspectorship, which he held until he became headmaster of Rugby in 1858. In the meantime he had attracted the admiration of the prince consort, and in 1856 he was appointed chaplain-in-ordinary to the queen. In 1857 he was select preacher at his university.

At Rugby Temple showed great energy and bold initiative. Whilst making the school a strong one on the classical side, he instituted scholarships in natural science, built a laboratory, and gave importance to that side of the school work. He had the courage also to reform the games, in spite of all the traditions of the playing fields. His school sermons were deeply impressive: they rooted religion in the loyalties of the heart and the conscience, and taught that faith might dwell secure amid all the bewilderments of the intellect, if only the life remained rooted in pure affections and a loyalty to the sense of duty. Two years after he had taken up his work at Rugby *Essays and Reviews* appeared. The first essay in the book, "The Education of the World," was by Dr. Temple. Temple refused, so long as the storm lasted, to comply with the request that he would repudiate his associates, and it was only at a much later date (1870) that he saw fit quietly to withdraw his essay. In the meantime, however, he printed a volume of his Rugby sermons, to show definitely what his own religious positions were. His appointment by Gladstone as bishop of Exeter in 1869 raised a fresh storm.

G. A. Denison, archdeacon of Taunton, Lord Shaftesbury, and others formed a strong committee of protest, whilst Pusey declared that "the choice was the most frightful enormity ever perpetrated by a prime minister." At the confirmation of his election counsel was instructed to object to it, and in the voting the chapter was divided. But Gladstone stood firm, and Temple was duly consecrated on Dec. 21, 1869. On the death of Dr. John Jackson in 1885, he was translated to London, the appointment gave general satisfaction. In 1884 he was Bampton Lecturer, taking for his subject "The Relations between Religion and Science." In 1885 he was elected honorary fellow of Exeter College, Oxford.

Temple led a strenuous life as bishop of London. His normal working day at this time was one of fourteen or fifteen hours, and he was felt by many of his clergy and by candidates for ordination to enforce almost impossible standards of diligence and efficiency. The working classes instinctively recognized him as their friend. When, in view of his growing blindness, he offered to resign the bishopric, he was induced to reconsider his proposal, and on the sudden death of Archbishop Benson in 1896, though now seventy-six years of age, he accepted the see of Canterbury.

As archbishop he presided in 1897 over the decennial Lambeth Conference. In the same year Temple and his brother archbishop issued an able reply to an encyclical of the pope which denied the validity of Anglican orders. In 1900 the archbishops again acted together, when an appeal was addressed to them by the united episcopate, to decide the vexed questions of the use of incense in divine service and of the reservation of the elements. After full hearing of arguments they gave their decision against both the practices in question. During his archbishopric Dr. Temple was deeply distressed by the divisions which were weakening the Church of England, and many of his most memorable sermons were calls for unity. His first charge as primate on "Disputes in the Church" was felt to be a most powerful plea for a more catholic and a more charitable temper, and again and again during

the closing years of his life he came back to this same theme. While speaking in the house of lords on Dec. 2, 1902 on the Education bill of that year, he was seized with sudden illness, and, though he revived sufficiently to finish his speech, he never fully recovered, and died on Dec. 23, 1902. He was interred in Canterbury cathedral four days later.

TEMPLE, RICHARD GRENVILLE-TEMPLE, 1ST EARL (1711–1779), English statesman, eldest son of Richard Grenville (d. 1727) of Wootton, Buckinghamshire, was born on Sept. 26, 1711. His mother was Hester (*c.* 1690–1752), daughter, and ultimately heiress, of Sir Richard Temple, Bart. (1634–1697), of Stowe, Buckinghamshire, and sister of Richard Temple, Viscount Cobham, whose title she inherited under a special remainder in 1749; in the same year, her husband having been long dead, she was created Countess Temple. Her son, Richard Grenville, was educated at Eton, and in 1734 was returned to parliament as member for the borough of Buckingham. In 1752, on the death of his mother, he inherited her titles together with the estates of Stowe and Wootton; and took the name of Temple in addition to his own surname of Grenville. By the marriage of his sister Hester with William Pitt, afterward earl of Chatham, Temple's career was linked with that of his brother-in-law.

In Nov. 1756 Temple became first lord of the admiralty in the ministry of Devonshire and Pitt. He was disliked by George II, who dismissed both him and Pitt from office in April 1757. But when the coalition cabinet of Newcastle and Pitt was formed in June of the same year, Temple received the office of privy seal. He alone in the cabinet supported Pitt's proposal to declare war with Spain in 1761, and they resigned together on Oct. 5. From this time Temple became one of the most violent and factious of politicians, and it is difficult to account for the influence which he exerted over his brother-in-law. He was at variance with his younger brother, George Grenville, when the latter became first lord of the treasury in April 1763, and he had no place in that ministry; but the brothers were reconciled before 1765, when Temple refused to join the government, and persuaded Pitt to refuse likewise.

By 1765, however, the old friendship between the brothers-in-law was dissolving; and when at last in July 1766 Pitt consented to form a government, Temple refused to join, being bitterly offended because, although offered the head of the treasury, he was not to be allowed an equal share with Pitt in nominating to other offices. Temple forthwith began to inspire the most virulent libels against Pitt; and in conjunction with his brother George he concentrated the whole Grenville connection in hostility to the government. After George Grenville's death in 1770 Lord Temple retired almost completely from public life. He died on Sept. 12, 1779.

See *The Grenville Papers* (1852), a considerable portion of which consists of Earl Temple's correspondence; Horace Walpole, *Memoirs of the Reign of George II,* 3 vol. (1847); *Memoirs of the Reign of George III,* 4 vol. (1845 and 1894); Earl Waldegrave, *Memoirs 1754–58* (1821); Sir N. W. Wraxall, *Historical Memoirs,* ed. H. B. Wheatley, 5 vol. (London, 1884); *Correspondence of Chatham,* ed. by W. S. Taylor and J. H. Pringle, 4 vol. (1838–40); W. E. H. Lecky, *History of England in the Eighteenth Century,* vol. ii and iii (7 vol., 1892).

TEMPLE, SIR WILLIAM, BART. (1628–1699), English statesman, diplomatist and writer, was born on April 6, 1628, at Blackfriars, London, the eldest son of Sir John Temple, later master of the rolls in Ireland. Educated at a school at Bishop's Stortford, Hertfordshire, and at Emmanuel college, Cambridge, he avoided taking part in the controversies which were then convulsing England by setting out in 1648 to travel on the continent, visiting France, Germany and the Netherlands, and acquiring that knowledge of foreign languages and conditions abroad that was to be of such value to him. While passing through the Isle of Wight he made the acquaintance of Dorothy Osborne, daughter of Sir Peter Osborne, governor of Guernsey, with whom he was to conduct a long and famous correspondence, and whom he eventually married on Christmas day, 1654.

For eight years after his marriage Temple resided mainly in Ireland, and it was as a member for Carlow borough in the Irish convention which assembled early in 1660, and for Carlow county in the parliament which took its place a year later, that he first entered public life. After the prorogation of the Irish parliament in May 1663, however, he moved to England, settled at Sheen, near Richmond, Surrey, and, attaching himself to Lord Arlington, then recently appointed secretary of state, embarked under his auspices on a diplomatic career. His first mission, to the bishop of Münster in the summer of 1665, was unsuccessful, Temple proving no match for the shifty and evasive bishop; but later in the same year he was appointed to the position he particularly coveted, that of minister resident to the Spanish court at Brussels, and on Jan. 31, 1666, to give him the increase of dignity which this important post demanded, he was created a baronet.

Scarcely had he established himself in Brussels when the outbreak of the War of Devolution and the French invasion of the Spanish Netherlands confronted him with the twin problems of bringing to an end the existing conflict between England and the Dutch Republic, and finding means of checking French aggression in an area where English interests were vitally concerned. The first problem was solved by the treaty of Breda of July 21, 1667, in the negotiation of which he took little part, the second problem by the Triple alliance of Jan. 13, 1668, between England, the United Netherlands and Sweden, which was acclaimed at the time, and has been recognized ever since, as his most brilliant achievement. Unfortunately Charles II and Arlington remained faithful to the policy of the Triple alliance for little more than a year, concluding on May 22, 1670, the infamous secret treaty of Dover, by which England became pledged to wage war on the United Netherlands in alliance with France. Thus Temple received no more substantial reward for his services than appointment as ambassador at The Hague, where he made the acquaintance of the young Prince of Orange, afterward King William III; and on his return to England in Sept. 1670 he met with a cold welcome and practical relegation to obscurity.

With the failure of the Anglo-French attack on the Dutch, however, he came again into favour, the more so as the minister upon whom Charles II was beginning to rely, Sir Thomas Osborne, earl of Danby, was a cousin of his wife, and an old acquaintance. Having first repeated his earlier successes by negotiating the treaty of Westminster with the Dutch on Feb. 9, 1674, he was reappointed ambassador at The Hague in May, dispatched to the general peace conference at Nijmegen in July 1676, allowed to co-operate with Danby in bringing about the marriage of William of Orange and the Princess Mary in Nov. 1677, sent back to The Hague to arrange a fresh treaty with the Dutch in July 1678, and required to pay a last visit to Nijmegen, where a general peace was signed in Jan. 1679. Danby, indeed, who knew little of foreign affairs, was most anxious to secure his services as secretary of state, but repeated pressure from king and minister alike failed to induce him to accept. In Feb. 1679, in the midst of the panic produced by the revelations of Titus Oates, he was hurriedly summoned home and again urged to take office, but once more refused, on the plea of poverty, ill-health and inability to secure a seat in the house of commons.

It was thus, rather strangely, only after Danby's fall that Temple undertook to play any part in the domestic affairs of his own country. On April 21, 1679, the privy council was remodeled in a way which it was believed might satisfy all parties. Its numbers, exclusive of certain supernumeraries, were reduced to 30, of whom half were to be great officials and half noblemen or gentlemen of wealth and distinction without office. Temple was allowed to claim the credit of having suggested the change, and accepted membership of the new body. For a time, indeed, he shared with the marquis of Halifax, the earl of Sunderland and the earl of Essex a place of considerable influence within it, and, to strengthen his position, secured return as member for Cambridge university to the second Exclusion parliament on Aug. 22, 1679.

Before the year was out, however, Temple found that the new council was little more than a screen behind which the king was pursuing a policy of which he could not approve, and thereafter his attendance became decidedly irregular. On Jan. 24, 1681, he was struck off the council, and although only in his 53rd year went into

permanent retirement. He died on Jan. 27, 1699.

BIBLIOGRAPHY.—Jonathan Swift (ed.), *The Works of Sir William Temple*, 2 vol. (1720, 1731); 4 vol. (1754, 1770, 1814); G. C. Moore Smith (ed.), *The Letters of Dorothy Osborne to William Temple* (1928); *Early Essays and Romances of Sir William Temple* (1930); Abel Boyer, *Memoirs of the Life and Negotiations of Sir William Temple* (1714); T. P. Courtenay, *Memoirs of the Life, Works and Correspondence of Sir William Temple*, 2 vol. (1836); Clara Marburg, *Sir William Temple* (1932); H. E. Woodbridge, *Sir William Temple, The Man and His Work* (1940). (A. Bc.)

TEMPLE, WILLIAM (1881–1944), English clergyman, second son of Frederick Temple (*q.v.*), archbishop of Canterbury, was born at Exeter on Oct. 15, 1881, and educated at Rugby and at Balliol college, Oxford, where he was president of the Oxford union. From 1904–10 he was fellow and lecturer in philosophy at Queen's college, Oxford. In 1914 he resigned the headmastership of Repton school, which he had held from 1910, to take charge of St. James's, Piccadilly. He was editor of *The Challenge* (1915–18). He resigned from St. James's in 1917, was made a canon of Westminster (1919) and in 1921 bishop of Manchester. He was president of the Workers' Educational association (1908–24). In 1929 he became archbishop of York and in 1942 archbishop of Canterbury. Temple died on Oct. 26, 1944.

TEMPLE, a city of Bell county, Tex., U.S., 67 mi. N.E. of Austin, was founded by the Santa Fe railroad in 1881 and incorporated in 1882. It was named for B. M. Temple, an engineer with the railroad. Temple is a trade centre for an agricultural, stock raising and industrial region with fertile land and a mild climate. Chief industrial products are school equipment, rock wool, furniture, tile, shoes, cottonseed products and feedstuffs. It is a division point of the Santa Fe, with railroad shops located there. A local medical centre, the city is served by several public and private hospitals; the public education system includes a junior college. Temple has a council-manager form of government, in effect since 1922. For comparative population figures *see* table in TEXAS: *Population.* (B. H. PR.)

TEMPLE ARCHITECTURE. The word "temple," from the Latin word *templum*, originally signified any holy enclosure or building that is screened from the eyes of the profane. It is the dwelling place of the god; it serves the purpose of housing his image; and it is the architectural frame of the liturgy and cult of the god. The architectural forms of temples are manifold; they are primarily conditioned by the requirements of the cults, which in their turn reflect different religious beliefs. In later times "temple" came to signify, in a broader sense, any shrine of any denomination or even of fraternal orders. (*See* also RELIGIOUS ARCHITECTURE.)

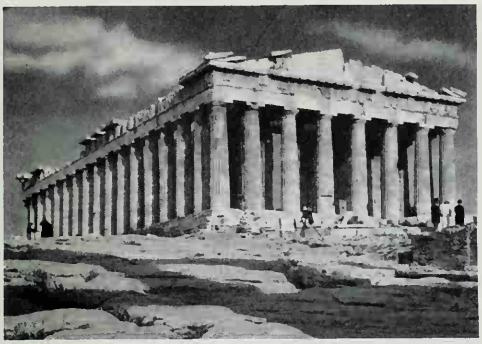
Hypaethral Temples.—This form of temple, often called a temenos, is generally dedicated to a nature deity thought to promote the fertility of the soil, of animals and of men. This being a common idea in primitive religious thought, the geographical and chronological spread of open-air temples is very wide. Their architectural form is vague and insignificant. A wall or hedge fenced in the holy place; an altar was built for the offerings, and a grove of holy trees or a spring was the seat of the deity. During the Bronze Age in Crete and Greece this form of temple was a common feature, to judge from the many representations of related cult scenes found on ring stones and seals. The best known examples are in Cyprus: Idalium (4th–3rd century B.C.), Tamassus (6th–5th century B.C.) and Ajia Irini (10th–7th century B.C.). In Gaul and Great Britain the druid temples were of similar type, and Nordic sagas testify to their existence in the Scandinavian countries in pre-Christian times. Mountain peaks and hilltops were often chosen for hypaethral temples. They are found in Greece .(Mount Ida [Psiloriti], Parnassus, Olympus) and are particularly common in Asia Minor. The Hittite temples of Yazilikaya, with famous rock reliefs, and Qareh Dagh serve as early examples; while Phrygian sites, like Midas City south of modern Eskisehir, belong to a later period, as does Nimrud Dagh in eastern Anatolia. Such temples seem to have been dedicated mostly to rain-producing weather gods or astral deities, divine incarnations of the celestial bodies.

Cave Temples.—Natural or artificial caves were also used as temples in different regions. In the Mediterranean area such temples are found on the island of Malta where they date back to the middle of the 2nd millennium B.C. A typical example is the hypogeum of Hal Saflieni, an elaborate system of communicating caves of a very articulate architectural form, with finely cut door jambs and lintels, carved altar stones and wall decorations in relief or painting. The Kamares cave and the Zeus grotto, both on the slopes of Mount Ida in Crete, are basically of the same type but of less elaborate shape. Much more developed in form is the cave temple at Abu Simbel in upper Egypt, dating from the 13th century B.C. and the great temple caves in India and China. The latter should be considered as translations into rock carving of real temple architecture; they have lost the character of caves.

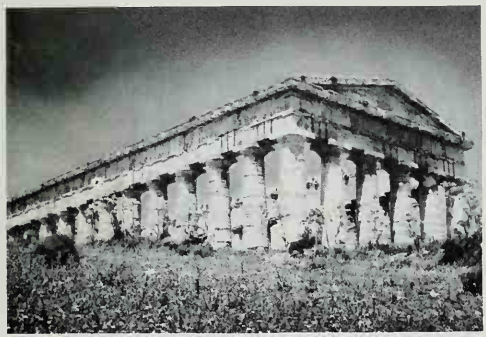
Egypt.—The Egyptians erected monumental temples in stone, and their generic form, later to be enlarged and further developed, emerged as early as the Old Kingdom (the 3rd millennium B.C.) in the shape of Pharaonic mortuary temples connected with the pyramids at Saqqara (Sakkara) and Giza (*see* PYRAMID). The mortuary temples were dedicated to the divine Pharaoh, and it can be deduced that they in essential parts duplicated the salient features of the royal palace. Through a long entrance corridor a large forecourt was reached and from there an inner court, facing the temple proper. To reach the holy of holies, poorly lit corridors and a series of small rooms were passed until the central cella of the temple was entered. Access to that part was reserved for the priest, just as audiences to the living Pharaoh were granted only to his highest officials. What in his lifetime served the divine Pharaoh as a ceremonial court residence fulfilled a similar purpose for his cult after his death. The mortuary temple was the only type of temple existing prior to the New Kingdom (18th–20th dynasties, 1580–1100 B.C.), and the earliest preserved regular temple, dedicated to a god, was that built by Queen Hatshepsut at Deir el Bahri close to Thebes (*c.* 1500 B.C.). What in the mortuary temples was grouped according to a more concentrated and entangled plan was stretched out along a common central axis, but the architectural elements were the same: entrance ramp, forecourt, gate, colonnaded inner court and the temple cella proper, surrounded by smaller rooms used for storing votive offerings and the like. The actual sanctuary was the innermost part of the whole complex and was strikingly small in comparison with the preceding courtyards and halls. This was true also for the fully developed temples of the 19th dynasty (1350–1205 B.C.), such as the great temples at Luxor and Karnak which may well stand as typical examples of Egyptian temple architecture. The entire holy area was walled in and a gate, flanked by two upward tapering towers (the outer pylon), led into the first courtyard. A second pylon led to a colonnaded room (the hypostyle hall), lit by small clerestory windows; from here the inner courtyard was reached through two other pylons and a series of rooms. At the far end of the courtyard was the temple itself, dwarfed by all the preceding structures. The layout was monumental and displayed along a common axis which marked the ceremonial way of the holy processions, typical to Egyptian liturgy. This type of temple building, once firmly established in the Ramesside period, continued essentially unchanged through the entire period of Egyptian ancient history. One of the best preserved examples is the Ptolemaic temple of Horus at Edfu (3rd century B.C.). (*See* EGYPTIAN ARCHITECTURE.)

Western Asia.—In comparison with Egyptian temple architecture the Sumerian, Babylonian and Assyrian temples followed a less rigid pattern. The plans lacked axiality and the enclosed area contained, in addition to various rooms for the priests and storerooms for temple treasures and archives, a relatively small temple building, generally without a cult image. The principal feature of the temple complex was the ziggurat (*q.v.*) or stepped pyramid, with stairs or an inclined plane leading to the terraces and the summit. The basic religious ideas governing the temple form were the same as those expressed in the hypaethral temples on mountain peaks. The deities worshipped were conceived as astral gods, incarnated in the heavenly bodies. In some cases, at least, each step of the ziggurat was dedicated to a separate divinity, as in the famous example at Babylon, in which each of the seven stages was dedicated to a separate planet.

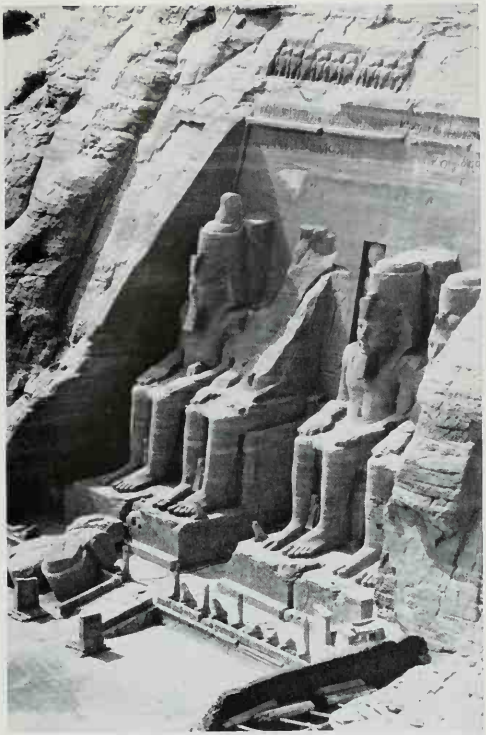
The Hittite temples at Bogazköy in Asia Minor (*c.* 1400 B.C.)



Parthenon, a 5th century B.C. Doric temple. Athens, Greece



Greek Doric temple, 5th century B.C., at Segesta, Sicily



Rock-cut temple of Rameses II at Abu Simbel, Egypt. Nineteenth dynasty (about 1270 B.C.)

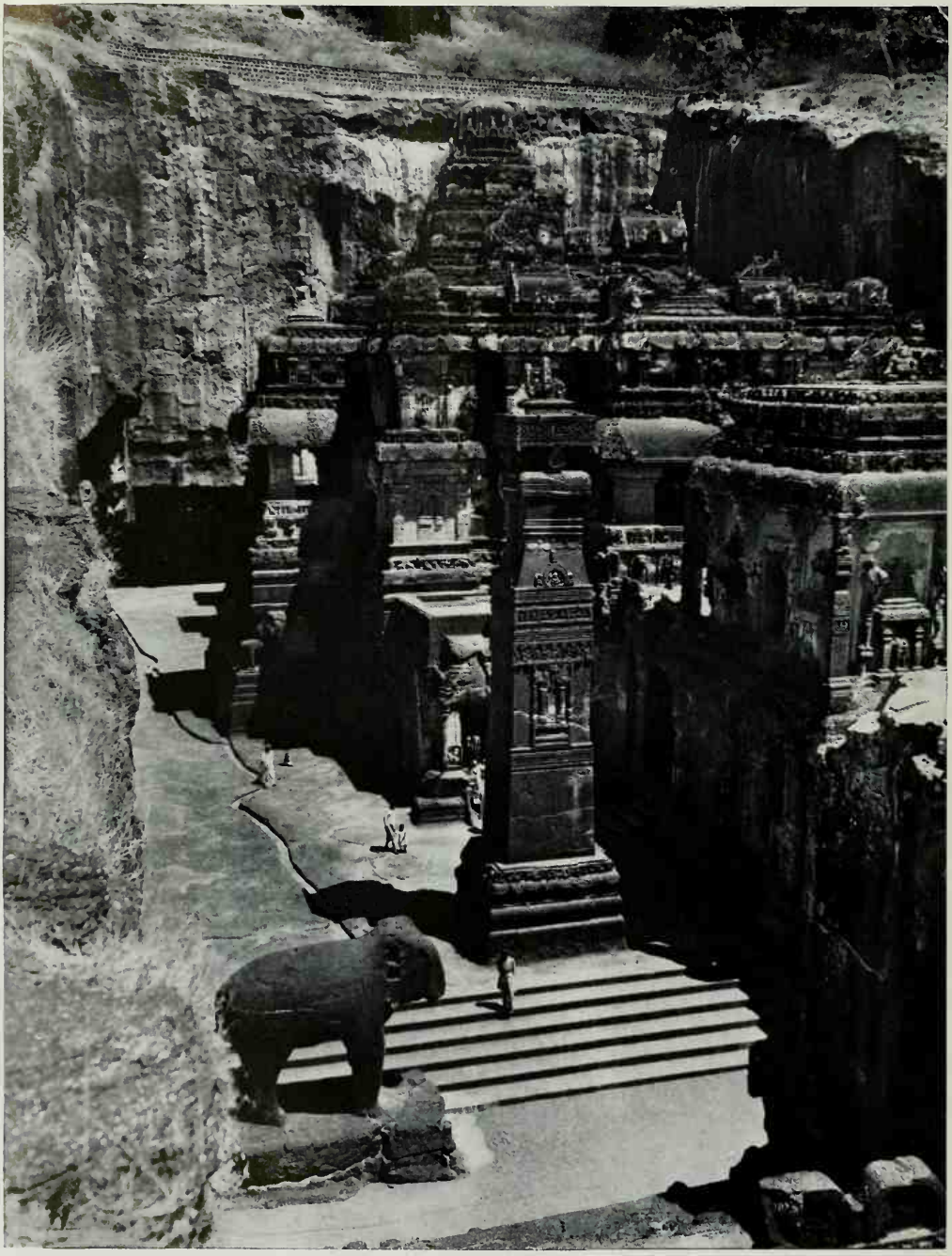
Interior, showing part of the dome of the Pantheon; Roman, A.D. 110–125

Temple of Fortuna Virilis; Roman, 1st century B.C.

EGYPTIAN, GREEK AND ROMAN TEMPLES

BY COURTESY OF (TOP RIGHT) UNITED ARAB REPUBLIC TOURIST OFFICE, (CENTRE LEFT) E.N.I.T., ROME; PHOTOGRAPHS, (TOP LEFT) ALISON FRANTZ, (BOTTOM LEFT) FOTOTECA UNIONE, ROME, (BOTTOM RIGHT) ALINARI

TEMPLE ARCHITECTURE

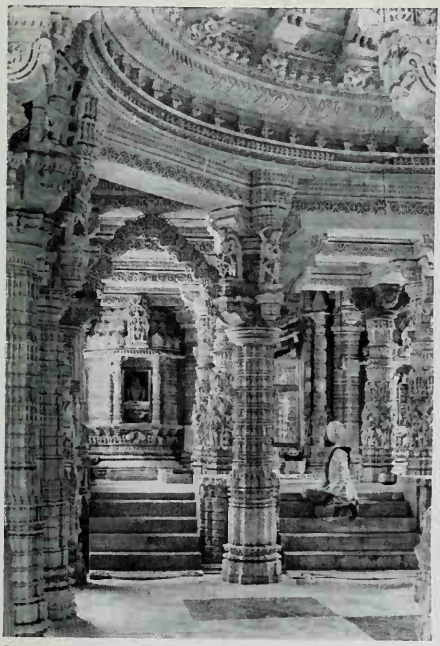


INDIAN ROCK-CUT CAVE TEMPLE

Kailasanatha, Ellora, India, an unusual rock-cut monolith with all sides exposed. Dravidian style, 8th to 10th centuries

PHOTOGRAPH, ELIOT ELISOFON, © TIME, INC.

Yashamon gate of the tomb shrine of Iyemitsu, Nikko, Japan. Tokugawa style, middle of 17th century



Dilwara temple, Mt. Abu, India. Jaina style, 11th century

Hall of Annual Prayers, Temple of Heaven complex, Peking, China; Ming dynasty, reconstructed in the 19th century

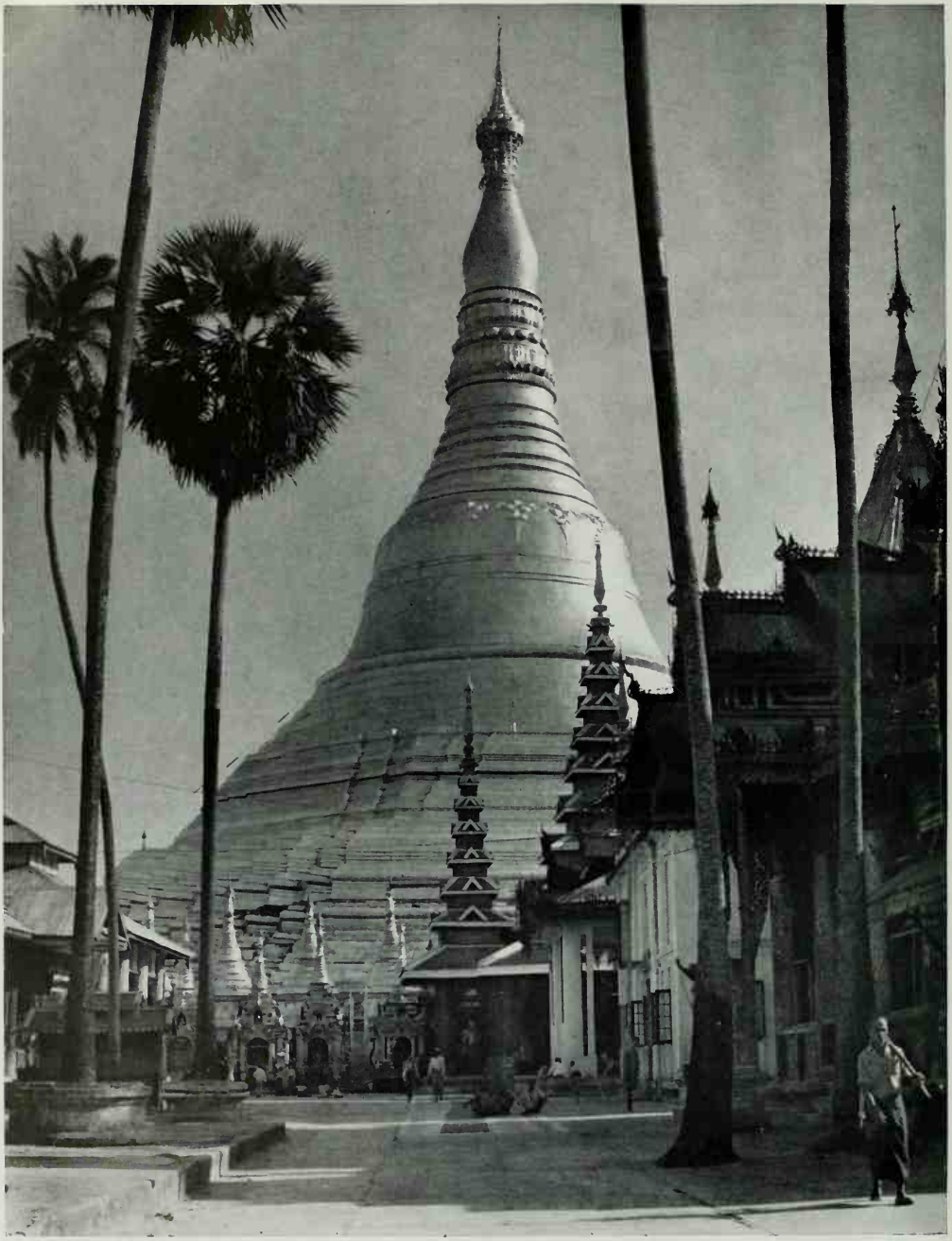
Lama temple, Peking, characteristic of northern Chinese tradition; 18th century

Subramanya shrine in the temple of the great pagoda at Tanjore, India. Dravidian style, Chola period, c. 907–1053

EAST ASIAN TEMPLES

PHOTOGRAPHS, (TOP LEFT) OE COU FROM EWING GALLOWAY, (TOP RIGHT, CENTRE RIGHT, BOTTOM RIGHT) EWING GALLOWAY, (BOTTOM LEFT) HERBERT G. PONTING

TEMPLE ARCHITECTURE



SHWE DAGON, RANGOON, BURMA

Rising to a height of more than 350 ft., the pagoda consists of a central stupa surrounded by many smaller ones; gilded brick construction. Its present size dates from the reign of Queen Shin Sawbu (1453–72)

PHOTOGRAPH, W. SUSCHITZKY

displayed a different plan: a central rectangular courtyard, on all sides surrounded by a series of rooms, one of which was the shrine proper with the cult image and the altar. It was so placed that it seems to have been the architect's intention rather to hide it from the worshipper than to make it the focus of his attention.

The Jewish Temple.—The temple used by the Jews during their migration from Egypt to the Promised Land is known as the Tabernacle. It is described in the Bible (Ex. xxv–xxvii; xxxvi–xxxvii) and was, in accordance with the nomadic way of life of the Jews of the period, a tent shrine. The tent, made of brightly colored linen and woolen material, contained a small inner room, the holy of holies, and a larger outer room, both surrounded by a verandalike structure. The total size was about 30 by 60 ft., and it was placed in a closed court measuring 150 by 75 ft. In the holy of holies stood the ark of the covenant, and the outer room contained an altar, the table for the show bread and the seven-branched candlestick.

This simple transportable tent shrine became the model for the great Temple of Solomon, probably erected in the years 959–952 B.C. and destroyed by Nebuchadrezzar in 586 B.C. It is described in all its splendour in the Bible (I Kings vi–vii), but as nothing of the building exists and the description is far from clear, it remains problematic. Two courtyards lay in front of the shrine after Egyptian fashion, and the interior of the temple was lighted by latticed clerestory windows as in Karnak. These seem to be the only egyptianizing features of the temple, and more significant stylistic parallels must be sought at Phoenicia whose King Hiram of Tyre assisted Solomon in planning the building.

When the Babylonian captivity came to an end in 539 B.C. the re-erection of the temple began. It was preceded by the prophetic vision of Ezekiel (Ezek. xl–xliii) which may have influenced the architect of this so-called Second Temple and certainly meant much for the planning of Herod's Temple. The former was an attempt to reproduce the fabulous Solomonic building in a time of poverty and distress, while the latter, built between 20–9 B.C. must have been one of the more important monuments of the early Roman empire in the near east. The outer courtyard, the Court of the Gentiles, with a colonnaded perimeter of about 4140 ft., rested on mighty retaining walls, at points reaching the height of 140 ft. The Temple itself rose on an elevated platform and was reached by passing two more courtyards. The decorative details of the exterior reflected the late Hellenistic and early imperial taste, while the temple itself was erected in traditional conformity with the ritual and religious requirements of the cult. It was destroyed at the capture of Jerusalem by the emperor Titus in A.D. 70. (*See* PRE-HELLENIC ARCHITECTURE.)

The Aegean World and Greece.—In Minoan Crete the separate, monumental temple building was unknown. The common form of their cult places was the hypaethral shrine on a mountaintop or in a holy grove. There was also the house chapel, incorporated in the great palaces of Crete. One such chapel was located off the central court of the palace of Knossos, and was dedicated to the protecting deity of the palace. A similar house chapel, but simpler in its architectural form, has been found in Greece proper, at Asine (14th century B.C.). Since Knossos from the 15th century B.C. was under Achaean rule and the Achaeans in Greece spoke an early form of the Greek language, it is surprising that the most typical product of Greek architectural creativity, the classical temple, had no forerunners during the Bronze Age. As a generic form it had features in common with the main room of the Mycenaean palaces, the megaron, but the megaron was not a separate building, as was the later Greek temple, nor was it a cult room, and it is highly debatable whether it had the typical gable roof so characteristic of the Greek temple. Moreover, no Mycenaean megaron survived the 12th century B.C., and no Greek temple can be traced further back than the 8th century B.C. The great majority of Greek temples were oriented in approximate east-west direction with the entrance facing east.

After some hesitant beginnings exemplified by some early structures at Thermon in Aetolia, the canonical form of the Greek temple began to emerge during the 7th century B.C. Its core was an oblong rectangular room (naos or cella) housing the cult image, with a porch in front (pronaos) created by the projecting side walls (antae). Columns were added between the antae to support the architrave spanning the porch. Balancing the frontal structure was a similar porch (opisthodome) at the rear, usually without an entrance. A peristyle or colonnade, found at an early stage in most of the important Greek temples, surrounded the building and added to the impression of quadrilateral completeness, so characteristic of the classical Greek peripteral temple. The entire building rested on a low stepped platform (stylobate) which lifted it above the surrounding ground level and at the same time made the shaded colonnade accessible from all sides. The roof was a slightly sloping gable roof with projecting eaves. The altar stood outside the temple, usually in front of the entrance at the east end. Altar, temple and surrounding area were enclosed and formed the holy precinct (hieron) often with additional smaller buildings. On the more important temple sites, such as Olympia and Delphi, the different Greek city states built their own buildings, called treasuries, serving as repositories for their votive gifts and as meeting points during the festivals. All over the hieron votive statuary was erected. Other votive gifts were placed in the naos itself and sometimes, as in the Parthenon at Athens, the rear part of the temple served as a special treasury. The exterior of the temple (the peristyle, the pronaos and the triangular gable fields of the roof) gave ample opportunity for applying sculptural decorations, in form of friezes and pedimental groups. Many of these decorations were brightly painted in red, blue, purple and brown. The interior was windowless and lit only from the door. In the dim twilight of the naos the great cult image dominated the room.

The temples of classical Greece were built in stone or marble and the masonry was laid with admirable precision. The earliest temples were of simpler construction and the materials were sun-dried brick on stone foundations and timber for columns and superstructure (entablature). Many of the decorative elements of the classical temple reveal their origin from wooden or terracotta forms, as the Doric frieze and the different types of column capitals.

Greek temples were spread over the entire Greek world and the great majority of them were of the so-called Doric order. Early examples are the temple of Apollo at Corinth, the temple of Hera at Olympia and many well-preserved temples in Sicily and south Italy, such as those in Syracuse, Selinus, Segesta, Agrigentum (Agrigento) and Paestum. The outstanding examples of Doric temples of the 5th century B.C. are the temple of Hephaestus (449–445 B.C., formerly known as the Theseum) and the Parthenon (447–432 B.C.), both in Athens.

In the east Greek provinces of Asia Minor, the Ionian order was commonly used and the temples displayed not only different decorative principles but also a more complicated plan, as the temple of Apollo at Didyma, south of Miletus. Characteristic of many Ionian temples is their great size, the most famous ones being the temples of Artemis in Ephesus and Sardis.

A specific form, rare in Greece, is the round temple with a circular naos surrounded by an annular peristyle. Such a temple is called a tholos and examples exist in Delphi (5th century B.C.) and in Epidaurus and Olympia (4th century B.C.). (*See* GREEK ARCHITECTURE; ORDER.)

Central Italy and Rome.—Material remains of early Etrusco-Italic temples are few and insufficiently known. Most of them were built of perishable material, such as sun-dried brick and wood, and what remains is mainly fragments of the painted terracotta revetments that adorned them. From these it can be concluded that Greece and the east Greek provinces of Ionia were the main sources of inspiration for the decorative details. In plan these temples differed in some essential respects from the normal Greek temple. The peristyle was missing and so was the stylobate. Thereby the frontal and axial approach to the building was stressed in a new way. The Roman architect Vitruvius gave an elaborate description of a specific type of Etruscan temple with three cellas. Actually, the three-cella temple was rare, and its use was limited to instances when three deities (a triad) were worshipped in one

temple, as in the temple on the Capitoline hill in Rome, which was built during the period of Etruscan influence and dedicated to Jupiter, Juno and Minerva. Etruscan usage had a direct impact on the formation of the later Roman temple. The deep pronaos and the wide intercolumniations are two other features of the Roman temple derived from Etruscan forms. The frontality and axiality inherent in the Etruscan plan was in Roman republican times further emphasized by placing the temple on a high podium, accessible by a flight of steps at the front side. The ritual orientation of the Greek temple and its free placement in an open hieron did not occur on Italic soil. Roman republican temples of the 3rd and 2nd centuries B.C. often had a forecourt, symmetrically laid out in front of the temple. The temple formed therefore the monumental backdrop of the forecourt, with a distinct front façade. (*See* ETRUSCANS.)

In decorative details the Roman temples followed Greek and Hellenistic patterns and developed them further. The Corinthian order was favoured over the Ionic, and the Doric was rarely used. Both the outside and the inside of the cella walls were richly decorated with marble incrustations, pilasters and half columns; and the buildings became more and more luxurious. A typical Roman form was the so-called pseudoperipteral temple with half columns decorating the exterior cella walls so as to give the false effect to a peripteral temple. Outstanding examples of this kind are the temple of Fortuna Virilis in Rome (first half of 1st century B.C.) and the so-called Maison Carrée at Nîmes (1st century A.D.).

In imperial times the court in front of the temple grew out to an entire colonnaded forum dominated by the temple placed close to its rear end. The forum of Caesar (47 B.C.) was thus dominated by the temple of Venus Genetrix, that of Augustus by the temple of Mars Ultor, and that of Domitian and Nerva by the temple of Minerva. The great temple of Jupiter at Heliopolis, Syria (now Baalbek, Lebanon) illustrates the end of this development (begun during the reign of Hadrian and completed early in the 3rd century A.D.). The giant temple towers over a vast courtyard which is preceded by a hexagonal forecourt with a triple entrance gate. The temple itself has a colonnade on all sides, a feature which came into fashion again in the times of the emperor Hadrian. Thus the temple of Venus and Roma, built under his reign, and the so-called Hadrianeum, both in Rome, were peripteral temples.

The round temple with annular peristyle is common: the temple of Vesta in Forum Romanum (originating in the 6th century B.C.); and those temples at Tivoli (1st century B.C.); in Forum Boarium at Rome (1st century A.D.) and at Baalbek (3rd century A.D.). The most magnificent of them all is the Pantheon (*q.v.*), a vast domed building erected under the reign of Hadrian. *See* ROMAN ARCHITECTURE.

(EK. S.)

India.—Indian temples show clearly the development from sacred cave to free-standing structure. The earliest existing monumental temples are all rock cut and developed with caves. The earlier temples, such as those at Bihar, and at Karli, near Bombay, show a great cave interior, divided by piers or columns into nave and aisles, with a small stupa, or solid masonry curve-topped structure enclosing a relic of Buddha, and a façade cut in the face of the rock. A second common type of rock-cut temple consists of those in which, in addition to the shrine proper, there are great courtyards surrounded by cells for monks. The most remarkable groups of cave temples are those at Ajanta and Ellora. The dates vary from the 2nd century B.C. to the 6th and 7th centuries A.D. The Dravidian- and Jain-constructed temples both show a power and heaviness of general mass with an amazing intricacy of carved detail that is more typical of rock-cut than free-standing structures, and the characteristic pyramidal towers, with their strongly marked horizontal mouldings and vertical projections have almost the character of artificial mountains and cliffs. The part played by figure sculpture in its lavishness, its multitude of figures and their bold relief, is unlike that played in any other structural style, and without doubt owes much, also, to rock-cut prototypes. The plan type shows usually a vast rectangular enclosure entered through several monumental gateways, crowned with enormous, solid towers. Near the centre is the shrine proper, containing the cult statue, sometimes approached by vast, colonnaded halls, and every exposed surface of the great towers and the interior columns and walls is covered with a rich network of sculpture and ornament. (*See* INDIAN ARCHITECTURE.)

East Indies.—Indian influence is marked in the temple ruins found in scattered places through Cambodia, Burma, Java and nearby islands. All of the great temples are elaborations of the stupa idea and consist in essence of one or more stupas on raised terraces. Thus in the Shwe Dagon pagoda at Rangoon are a multitude of small stupas crowned with high, spirelike finials, surrounding a central stupa that rises to a height of over 350 ft., and the whole surface is heavily gilded. At Borobudur, Java, the great Buddhist temple (*c.* 7th or 8th century A.D.) consists of a series of terraces around a hill 150 ft. high, at the top of which is a large stupa surrounded by 16 smaller ones. All of the stone lining of the terrace walls is covered with exquisite relief sculpture, not only illustrating the life of Buddha, but also containing many pictures of ordinary, everyday events. The most remarkable of these eastern Buddhist temples is Angkor Wat at Angkor in Cambodia (attributed to the first half of the 12th century and apparently originally intended as a Brahman temple, being converted to Buddhist uses later). This consists of three great symmetrical terraces with elaborate stairs, corner towers, porticos and colonnaded halls, with a great stupalike tower as the climax in the centre. The sculpture and ornament, although typically Khmer, show distinct traces of Hindu influence. The extremely classic character of some of the pier and column capitals is noteworthy.

China.—Farther east an entirely contradictory tradition governed temple design, that of the wooden-columned prayer halls of China. As all Chinese monumental buildings, the Chinese temple structure had a wooden framework of horizontal beams and vertical columnar supports which was raised on a stone platform and crowned by a large ornamental roof with overhanging eaves. Since the entire weight of the roof was directed upon the columnar supports through a system of complicated bracket capitals, the Chinese architect was free to use curtain walls and light screen windows throughout the building for decorative purposes. The planning of a temple complex, whether Confucian, Buddhist or Taoist, also followed that of an imperial palace or a great residential home, with three identical buildings grouped on the three sides of an open court, which was entered from its fourth side, directly opposite the central structure, through a gatehouse. Additional buildings or gateway structures could be laid along the main axis in a similar manner, with or without connecting corridors or verandas between them. In Buddhist temples, often there was on either side of the main axis in the outer court a small tower, one containing a drum and the other a bell. Aesthetically, the architectural character varied with the locality. Those in the north were full of monumental grandeur, as in the temple of Confucius or such monasteries in the western hills as that of Wo-fu-szu (both rebuilt in the Ming and the Ch'ing dynasties). Those in the Yangtze valley were more picturesque, with exaggeratedly curved roofs and slimmer supports, as in the 18th century Sheng-yin-szu in the West Lake in Hangchow. (*See* CHINESE ARCHITECTURE.)

Japan.—In Japan the elaborate formality of the typical Chinese temple is absent. In fact, the native Shintō temples were based upon very primitive house ideas; they were simple, rectangular buildings with heavily thatched roofs. Architectural effect was gained, however, through the monumental torii or gateways which crossed the main entrance paths. Many votive stone lanterns also decorated the grounds. In the Buddhist temples the halls themselves were based entirely on Chinese prototypes, but the detail was typically Japanese in its composition and its mannered realism. Outstanding examples include the temples of the magnificent group at Nikko, largely of the 17th century, and the Honganji temple at Kyōto. The genius for placing temples is well seen in the island temples that dot the Inland sea, especially those of the sacred island of Itsuku-shima, near Hiroshima, where the chief temple was originally founded in 587. *See* JAPANESE ARCHITECTURE.

(WE. F.)

TEMPLER, SIR GERALD WALTER ROBERT (1898–), British field marshal, best-known for his successful operations against rebel forces and terrorists in Malaya, was born at

Colchester, Essex, on Sept. 11, 1898, and was educated at Wellington college, Crowthorne, Berkshire, and Sandhurst. He was commissioned in 1916 and served in France toward the end of World War I. In 1936 he was awarded the distinguished service order for gallantry during special operations in Palestine. In 1939–40 he served at Lord Gort's headquarters in France and later with an emergency formation which was attempting to check the German advance to the French coast. In 1943 he commanded the 1st and 56th infantry divisions and the 6th armoured division in the Allies' successful campaign in the Mediterranean. In 1944 he commanded the 6th armoured division in Italy, and in 1945 he became director of civil affairs in the military government of occupied Germany. A year later he became director of military intelligence at the war office and in 1948 vice-chief of the imperial general staff. In 1950 he was appointed general officer commanding eastern command with the rank of full general. He had been knighted in 1949. Templer is, however, best known for the part he played as high commissioner and director of operations in Malaya between 1952 and 1954, during which time the Communist bands in that country were ruthlessly sought out and finally reduced to impotence as a result of his skillful and energetic leadership. He was chief of the imperial general staff from 1955 to 1958.

Field Marshal Templer was created knight grand cross of the Order of St. Michael and St. George in 1953 and of the Order of the Bath in 1955. (E. B. Bn.)

TEMPO: *see* Time.

TEMUCO, capital of Cautín province, Chile, situated in the heart of Araucania (*q.v.*). Pop. (1952) 51,597, (1960) 111,641 (mun.). It was founded in 1881 as a frontier outpost on a strategic crossing point of the Cautín river. The land was ceded in a Chilean-Araucanian compact ending the Indian wars, drawn up on adjacent Cerro Ñielol. This forested hill, preserved as a natural park, affords vistas of the central valley and of the volcano-studded Andean cordillera, appropriate to this lake-country gateway. Temuco is an agricultural, livestock and forest products centre. Of special interest is the Araucanían Indian market. (J. T.)

TENANT: *see* Landlord and Tenant.

TENANT-RIGHT, in law, a term expressing the right which a tenant has, either by custom or by law, against his landlord for compensation for improvements at the determination of his tenancy. In England it is governed for the most part by statute (*see* Landlord and Tenant; Small Holdings).

TENASSERIM, a division of Lower Burma, bordering on Thailand. Area, 37,614 sq.mi. Pop. (1941) 2,110,420. It was divided into six districts: Toungoo, Salween, Thatôn, Amherst, Tavoy and Mergui. It formed the tract south of Pegu conquered from Burma in 1826, which was for many years known as the Tenasserim province. The southern extremity of the division approaches the insular region of Malaya, and it is fringed along its entire western coast by a number of islands, forming in the north the Moscos and in the south the Mergui Archipelago. The eastern frontier is formed by a mountain range 5,000 ft. high, the waterparting between Tenasserim and Siam.

TENBY, a municipal borough in the Pembrokeshire county parliamentary division of Wales, 10 mi. E. of Pembroke by road. Pop. (1961) 4,752. Area 1.7 sq.mi. Standing on a promontory on the south coast of Pembrokeshire overlooking Carmarthen bay to the east, the walled town is built of the gray limestone of the peninsula. It has a harbour and two sandy beaches. The alternation of cliffs of gray limestone and Old Red Sandstone along this coast is very beautiful, and Tenby, with its mild climate, is visited all the year round. Tenby is first mentioned as a Danish fishing station, and the old Welsh name, Dinbych-y-Pysgod, means "the precipice of fishes." Flemish clothworkers went there in the early 12th century, and a 16th-century Flemish house still exists. Ruins of the castle (12th–13th century) stand on a hill northeast of the town; St. Mary's church is mostly 15th century. The early charters were confirmed in 1580 and the 15th-century town walls repaired in 1588 when a Spanish invasion was feared. During the 15th and 16th centuries Tenby traded by sea with Bristol and had wealthy mercantile families. It declined after the Restoration until the coming of the railway.

Caldey, 2½ mi. S. of Tenby, is an island privately owned by a priory of Cistercian monks of the Strict Observance (Trappists) who went there in 1929 from Chimay in Belgium.

TENCH (*Tinca tinca*). A cyprinid fish of Europe, Asia Minor and Siberia, greenish, with small scales and rounded fins. It thrives in weedy ponds. Specimens exceeding 8 lb. are rare. The flesh is more palatable than that of most cyprinids.

TENCIN, ALEXANDRINE CLAUDE DE GUÉRIN DE (1682–1749), prominent French social figure, patroness of writers and herself a writer, was born at Grenoble, France, on April 27, 1682. She became a nun early in life, but soon abandoned her vows in obscure circumstances around which many legends grew. After the death of Louis XIV she sought fortune at court and became the mistress of Cardinal Dubois, of the regent and of many other personalities. To one of her lovers, the Chevalier Destouches, she gave a son who became the philosopher D'Alembert. Imprisoned in the Bastille in 1726 on a false though plausible charge of murder, she was released after an intervention by her brother Pierre, then an archbishop, later a cardinal. Thereafter she sought less scandalous distinction, as a hostess. Bernard de Fontenelle, Montesquieu (whom she twice assisted in the publication of his works), Pierre de Marivaux and Jean Marmontel frequented her salon. She was esteemed by Pope Benedict XIV. Her best-known work is an autobiographically inspired novel, *Memoires du comte de Comminge* (1735). She died on Dec. 4, 1749.

See P. M. Masson, *Madame de Tencin* (1909). (Rt. S.)

TEN COMMANDMENTS: *see* Decalogue.

TENDINITIS is an inflammation of the sheaths of the tendons. These sheaths are composed of thin, filmy tissue that permits the sliding motion of tendons within them. The cause of inflammation is irritation of the sheaths by prolonged or abnormal use of the tendons. Less often it may follow invasion of the tendon sheaths by bacteria with subsequent infection. It is in many instances an occupational hazard (wrist tendons in the typist; tendons about the shoulder in baseball pitchers; or tendons about the elbow in pneumatic drill operators).

Inflammation in the sheath of the tendon produces swelling, redness and pain along the course of the involved tendon, and motion of the tendon produces severe pain. Swelling of the sheath narrows the space through which the tendon may slide, causing stiffness in the involved area. A grating sensation may be felt over the involved tendon sheath as the tendon moves.

Treatment of the condition depends upon placing the involved tendon and its sheath at complete rest by means of splints, cast or bandages. With rest the inflammation subsides and motion may gradually be resumed. Local injections of hydrocortisone may hasten the abatement of the inflammation. Recurrences following improvement are not uncommon. Repeated episodes of tendinitis may produce permanent thickening of the sheath, and the thickened, irregular sheath limits motion of the tendon within it.

(W. F. E.)

TENDON OF ACHILLES: *see* Achilles, Tendon of.

TENEBRAE, an office sung in Roman Catholic churches on the afternoon or evening of Wednesday, Thursday and Friday of Holy Week. Durandus (vi, 72, n. 2) derives the name from the fact that "the church on these days cultivates darkness (*tenebras colit*)" in sorrow for the Lord's Passion, celebrating His exequies on three days since for three days He was dead, and it symbolizes the darkness that fell on the earth while the Sun of Justice was crucified.

TENERIFE (Teneriffe), the largest of the Canary Islands; area, 795 sq.mi. Of irregular shape, 60 mi. long, 30 mi. wide, it is the main portion of the Spanish province of Santa Cruz de Tenerife (pop. 1950, 321,949). The celebrated peak locally known as the Pico de Teyde (or Teide), with its supports and spurs, occupies nearly two thirds of the whole island. It has a double top; the highest point, El Piton, is 12,172 ft. above the sea; the other, Chahorra, connected with the first by a short narrow ridge, has a height of 10,289 ft. They are both orifices in the same grand dome of trachyte, where snow lingers four months.

For more than one half of its circumference the base of the true peak rises from an elevated but comparatively level tract.

On the southeast, south and southwest there is a high curved ridge overlooking the Pumice Stone plains and presenting a very steep face to the peak. Between the ridge and the sea the slope is more gradual, and there are intervening tablelands. Peaks rise from the ridge, one of which (Guajara) attains the height of 8,900 ft. Both El Piton and Chahorra have craters on their summits, from which issue steam and a little sulphurous vapour. The crater on El Piton is partly surrounded by a wall of lava, which has been made white by the action of sulphurous vapours. The crater is about 300 ft. across, with a depth of 70 ft. The crater on Chahorra has a diameter of 4,000 ft.; its depth is scarcely 150 ft. The entrance to the Llano at a sort of natural gateway (called *Portillo*) between two basaltic hills, is about 7,000 ft. above the sea. Between two and three hours are consumed in crossing the Llano to the base of the cone, the lower part of which is ascended to a point 9,750 ft. above the sea.

To the northwest of the grand cone, several thousands of feet below Chahorra, there are many small cones of eruption, showing that the intensity of volcanic action was greatest on that side. Eastward from the ridge bounding the Pumice Stone plains extends a chain of mountains to the northeastern extremity of the island. The highest peaks are Izaña (7,831 ft.), Perejil (6,027) and Cuchillo (5,633).

There is no record of eruptions from either crater of Pico de Teyde. In 1795 a great quantity of lava was poured out from three vents on the eastern side; and in the same year lava streams issued from a crater near Guimar, halfway between Santa Cruz and the peak. In the year 1706 a vent on the northwestern side of the peak discharged a copious stream, which flowed down to the sea and nearly filled up the harbour of Garachico. For three months in 1798 much lava and other volcanic matter were ejected.

Santa Cruz, the capital of Tenerife and of the Canaries, and Laguna, the former capital, are described in separate articles. A good road connects Santa Cruz and Orotava, a town on the north coast 25 mi. W.N.W. Date palms form a striking feature in the landscapes. The town of Orotava (pop. 1950, 20,487) is 1,040 ft. above the sea. The houses are solidly built, but the town has a deserted aspect. A stream of water is conducted through every street.

Puerto de la Cruz (mun. pop. 1950, 12,476) 3 mi. N. of the town, is a clean place. The streets are broad and the houses well built. (*See* also CANARY ISLANDS.)

TENGGERESE. In the isolated mountain districts of East Java, near Tosari, live a primitive people known as the Tenggerese. Because of the rugged character of their homeland they have managed to preserve their identity despite the fact that they number only about 10,000 in the densely populated island of Java. Although their religious beliefs indicate some influence from the Hindu-Javanese cultures of the lowlands they have remained pagans. Little is known about their material culture other than that they live in small mountainside villages, make clearings in the jungle and raise dry-land rice. A. A. J. Nieuwenhuis who has studied their physical type says they are mixed Veddoid and Proto-Malayan.

See A. A. J. Nieuwenhuis, *Enn Anthropologische studie van Tenggerezen en Slamet-Javanen* (1948). (F.-C. CE.)

TENIERS, the name of a family of Flemish artists who flourished at Antwerp and Brussels during the 17th century.

DAVID TENIERS, the elder (1582–1649), was born at Antwerp, the son of Julian Teniers, a mercer who settled at Antwerp in 1558. He received his first training in the painter's art from his brother Juliaen, then studied under Rubens in Antwerp, and subsequently under Elsheimer in Rome; he became a member of the Antwerp gild of painters in 1606. Though he tried his skill in large religious, historical and mythological compositions, his fame depends chiefly on his landscapes and paintings of peasants carousing, of kermesse scenes and the like, which are marked by a healthy sense of humour, and which are not infrequently confused with the early works of his son David. A large painting by the elder Teniers at St. Paul's church in Antwerp, represents the "Works of Charity." At the Vienna Gallery are four landscapes painted by Teniers under the influence of Elsheimer, and four small mythological subjects, among them "Vertumnus and Pomona," and "Juno, Jupiter and Io." The National Gallery has a characteristic scene of village life, "Playing at Bowls," a "Conversation," and a "Rocky Landscape." Other examples of his work are to be found at the galleries of Leningrad, Madrid, Munich, Dresden and Berlin ("The Temptation of St. Anthony"). He died at Antwerp on July 29, 1649.

DAVID TENIERS, the younger (1610–1690), his more celebrated son, was born in Antwerp and was baptized on Dec. 15, 1610. Through his father, he was indirectly influenced by Elsheimer and by Rubens. We can also trace the influence of Adrian Brouwer at the outset of his career. In 1637 Teniers married the ward of Rubens, Anne Breughel, the daughter of John (Velvet) Breughel. He became a "master" in the gild of St. Luke in 1633. The Berlin Museum possesses a group of ladies and gentlemen dated 1634. Some first-rate works—the "Prodigal Son" and a group of "Topers" in the Munich Gallery, and a party of gentlemen and ladies at dinner, termed the "Five Senses," in the Brussels Museum—are instances of the perfection attained by the artist when he was probably scarcely twenty. His touch is of the rarest delicacy, his colour at once gay and harmonious. He was little over thirty when the Antwerp gild of St. George enabled him to paint the marvellous picture now in the Hermitage Gallery in Leningrad —the "Meeting of the Civic Guards." Correct to the minutest detail, yet striking in effect, the scene, under the rays of glorious sunshine, displays acquired knowledge and natural good taste. Another work of the same year (1643), now in the National Gallery, London, is "The Village Fête."

Teniers was chosen by the common council of Antwerp to preside over the gild of painters in 1644. The archduke Leopold William, who had assumed the government of the Spanish Netherlands, employed Teniers not only as a painter but as keeper of the collection of pictures he was then forming. With the rank and title of "ayuda de camara," Teniers took up his abode in Brussels in 1651. Immense sums were spent in the acquisition of paintings for the archduke. A number of valuable works of the Italian masters, now in the Vienna Museum, came from Leopold's gallery after having belonged to Charles I. and the duke of Buckingham. De Bie (1661) states that Teniers was some time in London, collecting pictures for the duke of Fuensaldaña, then acting as Leopold's lieutenant in the Netherlands. Paintings in Madrid, Munich, Vienna and Brussels show what the imperial residence was at the time of Leopold, who is represented as conducted by Teniers and admiring some recent acquisition. No picture in the gallery is omitted, every one being inscribed with a number and the name of its author, so that the *ensemble* of these paintings might serve as an illustrated inventory of the collection. Still more interesting is a canvas, now in the Munich Gallery, showing Teniers at work in the palace, with an old peasant as a model and several onlookers. When Leopold returned to Vienna, the pictures travelled to Austria. Teniers remained in high favour with the new governor-general, Don Juan, a natural son of Philip IV.

Shortly after the death of his wife in 1656 Teniers married Isabella de Fren, daughter of the secretary of the council of Brabant, and strove to prove his right to armorial bearings. In 1663 Teniers founded the Academy at Brussels.

Teniers died in Brussels on April 25, 1690. David, his eldest son, a painter of talent and reputation, died in 1685. One of this third Teniers's pictures—"St. Dominic Kneeling before the Blessed Virgin," dated 1666—is in the church at Perck.

See T. Smith, *Catalogue Raisonné III.* (1831); John Vermoelen, *Notice historique sur David Teniers et sa famille* (Antwerp, 1871); Alph. Wauters, *Histoire des environs de Bruxelles* (1851); and *Les tapisseries bruxelloises* (1878); Max Rooses, *Geschichte der Malerschule Antwerpens* (Munich 1880); A. v. Wurzbach, *D. Teniers* (1877); J. van den Branden, *Gesch. d. Antw. Schilderschool* (1883); A. Rosenberg, *David Teniers* (Leipzig, 1901).

TENISON, THOMAS (1636–1715), English archbishop, was born at Cottenham, Cambs., on Sept. 29, 1636. He was educated at Corpus Christi college, Cambridge, of which he became a fellow in 1659. After holding cures at Cambridge, where he gave devoted care to sufferers from the plague, and at Norwich, he was presented in 1680 by Charles II to the vicarage of St.

Martin's-in-the-Fields, London, where he endowed schools, established a public library, and was a champion of Protestantism during the reign of James II. Monmouth sent for him before his execution. Under William III, Tenison was in 1689 named a member of the ecclesiastical commission appointed to prepare a reconciliation of the Dissenters, the revision of the liturgy being specially entrusted to him. He preached a funeral sermon on Nell Gwyn (d. 1687) in which he represented her as truly penitent—a charitable judgment which did not meet with universal approval. He was made bishop of Lincoln in 1691 and archbishop of Canterbury in 1695. He attended Queen Mary during her last illness and preached her funeral sermon in Westminster abbey. During William's absence in 1695 Tenison was appointed one of the seven lords justices to whom his authority was delegated. With Burnet he attended the king on his deathbed. He enjoyed little favour with Queen Anne, but was a commissioner for the union with Scotland in 1706. A strong supporter of the Hanoverian succession, he was one of the three officers of state to whom on the death of Anne was entrusted the duty of appointing a regent till the arrival of George I, whom he crowned on Oct. 31, 1714. Tenison died in London on Dec. 14, 1715.

Tenison's works include *The Creed of Mr. Hobbes Examined* (1670) and *Baconia, or Certain Genuine Remains of Lord Bacon* (1679). He was one of the founders of the Society for the Propagation of the Gospel.

TENNENT, GILBERT (1703–1764), American Presbyterian clergyman and evangelist, the son of William Tennent (*q.v.*), was born Feb. 5, 1703, in County Armagh, Ire., and educated by his father, with perhaps a year at Yale college. He entered the Presbyterian ministry in Delaware in 1725, and the year after settled at New Brunswick, N.J. Under the influence of his father and of a Dutch pastor, T. J. Frelinghuysen, he soon became a rousing preacher and evangelist. Very quickly he emerged as the natural leader of the "Log college men" who had gathered about his father. He was also the dominant figure in the Great Awakening revival movement centred in the Log college or New Side group in the Presbyterian Church. When George Whitefield (*q.v.*), the English evangelist, was persuaded by William Tennent to aid the New Side, he found in Gilbert a like spirit. He persuaded Gilbert to follow him up in his evangelistic tour of New England. As a "Sun of Thunder" Tennent did so, and in so doing became a most controversial figure throughout the colonies. He is regarded as second only to Whitefield as a popular evangelist. When the Presbyterian Church broke up (1741) into New Side and Old Side, Tennent and Jonathan Dickinson became the two leaders of the New Side. In 1743 Tennent took over Whitefield's tabernacle in Philadelphia and made of it Second Presbyterian church. Ten years later he and Samuel Davies raised over £4,000 on a trip through England and Scotland for the building of Nassau hall (later Princeton university). He was a leader in the move that reunited the New Side and the Old Side in 1758. Tennent died July 23, 1764.

BIBLIOGRAPHY.—L. J. Trinterud, *Forming of an American Tradition: a Re-examination of Colonial Presbyterianism,* pp. 53–57 (1949); C. H. Maxson, *Great Awakening in the Middle Colonies,* ch. 3 (1920); E. S. Gaustad, *Great Awakening in New England,* pp. 32–37 (1957). (L. J. T.)

TENNENT, WILLIAM (*c.* 1673–1745), American colonial Presbyterian clergyman and educator, was born perhaps in the north of Ireland, and graduated from the University of Edinburgh in 1695. He began as a Presbyterian, but in 1704 and 1706 took orders in the Anglican Church in Ireland. On Sept. 6, 1718, he landed at Philadelphia, and ten days later joined the Presbyterian Church. For two years he served in East Chester, N.Y., and for eight years at Bedford, N.Y., settling later along Neshaminy creek just north of Philadelphia. There he served as pastor and began in 1736 a small school, called derisively the "log college." During his career Tennent educated about 21 of the most useful and gifted ministers of the colonial Presbyterian Church, among them his sons Gilbert Tennent (*q.v.*) and William Tennent II; Samuel Blair; Charles Beatty; John Rodgers; Samuel Finley; and John Redman. They became known as the New Side, opposed to the more rigidly orthodox Old Side of the Scotch-Irish majority. Tennent welcomed the English evangelist George Whitefield (*q.v.*) to Philadelphia and persuaded him to assist the New Side in its revival effort, known as the Great Awakening. Tennent and the New Side enlisted also the active support of such men as Jonathan Dickinson and Ebenezer Pemberton, thus assuring that colonial Presbyterianism would be predominantly New School. Indirectly, Tennent's Log college was a forerunner of Princeton university, inasmuch as his example inspired the New Side to found Nassau hall, which later became Princeton university. Tennent died May 6, 1745.

BIBLIOGRAPHY.—L. J. Trinterud, *Forming of an American Tradition: a Re-examination of Colonial Presbyterianism,* ch. 3–5 (1949); T. C. Pears, Jr., and Guy S. Klett, "William Tennent and the Log College," *Journal of the Presbyterian Historical Society,* 28:37–64, 105–128, 167–204 (1950); W. B. Sprague, *Annals of the American Pulpit,* vol. iii, pp. 23–27 (1858). (L. J. T.)

TENNESSEE, popularly known as the "Volunteer state," is a south-central state of the United States. It is bounded on the north by Kentucky and Virginia; on the east by North Carolina along the line of the crest of the Unaka mountains to within 26 mi. of Georgia, where the boundary turns due south; on the south by Georgia, Alabama and Mississippi along the 35th parallel of north latitude; on the west by the Mississippi river, which separates it from Arkansas and Missouri. The extreme length of the state from east to west is 432 mi., and the extreme breadth is 112 mi., its area being 42,244 sq.mi. of which 482 are water surface. The name "Volunteer state" was given to Tennessee because of its remarkable record in furnishing volunteers in the War of 1812 and in the Mexican War. It is the 34th state of the union in size. The state capital is Nashville and the state entered the union June 1, 1796, as the 16th state. The iris is the state's official flower, the tulip poplar its official tree, and the mockingbird its official bird. The state flag consists of a red field upon the centre of which is depicted a blue circle edged with white and which contains three white five-pointed stars. At the fly end of the flag is a blue stripe separated from the red field by a narrow white stripe. The name "Tennessee," is of Indian origin and is generally believed to be derived from the name of an ancient Cherokee capital.

PHYSICAL GEOGRAPHY

Physical Features.—The state lies between the extremes of approximately 35° and 36° 41′ N. and 81° 40′ and 90° 18′ W. It is popularly divided into three large divisions known as east, middle and west Tennessee. The first extends from the heights of the Unaka ridges along the North Carolina border, across the valley of the Tennessee river to the heights of the Cumberland plateau. The middle section includes a part of the Cumberland plateau, all the Highland rim plateau and the Central basin, and extends westward to the Tennessee river. The western division includes the plateau region from the Tennessee river to the steep slope which overlooks the Mississippi flood plain; also a narrow strip of lowland which extends to the Mississippi river. From a maximum elevation of 6,642 ft. at Clingmans Dome near the North Carolina border, in Sevier county, the surface descends to 182 ft. on the Mississippi river in Shelby county. The Unaka mountains, which occupy a belt 8 mi. wide along the state's eastern border, are a series of somewhat irregular ridges developed on complexly folded and faulted crystalline rocks. Sixteen peaks exceed 6,000 ft. in height and maximum elevation reaches 6,285 ft. That part of the Great Appalachian valley which traverses Tennessee is known as the valley of east Tennessee. It consists of parallel ridges and valleys developed by erosion on folded sandstones, shales and limestones, the valley quality predominating because the weak limestones were of great thickness. The valley areas vary in height from 600 ft. in the southwest to 1,600 ft. in the northeast. The northeast ridges are more numerous and higher than the southwest.

Along the southwest border of the valley a steep escarpment, known as Walden's ridge, rises to the Cumberland plateau. This plateau has a mean elevation of about 1,800 ft. and a rolling topography, the northern portion sloping toward the northwest.

The western edge of the plateau is much broken by deep indentations of stream valleys, and drops suddenly downward about 1,000 ft. to the Highland rim plateau, so named from the scarp

which forms an encircling rim around the Central basin. It is fairly level except where it is cut by river valleys. The Central basin lies for the most part 400 to 600 ft. below the rim; a few hills or ridges rise to the level of the rim. The basin is elliptical in form, extending nearly across the state from northeast to southwest with an extreme width of about 60 mi.; near its centre is the city of Murfreesboro, and Nashville lies in the northwest.

Between the Central basin and the lower Tennessee river is the western Highland rim. Westward from the lower Tennessee river the surface rises rapidly to the summit of a broken cuesta or ridge and then descends gently and terminates abruptly in a bluff overlooking the Mississippi flood plain. This is called the plateau slope of west Tennessee, which is part of the east Gulf coastal plain. The bluff, 150 to 200 ft. in height, traverses the state in a rather straight course; between it and the meandering Mississippi, except at a few points where that river touches it, lie low bottom lands varying in width according to the bends of the river and containing numerous swamps and ponds. In the northern portion, principally in Lake county, is Reelfoot lake, which occupies a depression formed by an earthquake in 1811. It is 18 mi. long and has a maximum width of 3 mi.

The whole of the Appalachian province of Tennessee and the southern portions of the Cumberland plateau, the Highland rim, and the Central basin are drained southward and westward by the Tennessee river and its tributaries. The valley of the lower Tennessee is drained northward by the same river. The northern portions of the Cumberland plateau, Highland rim and Central basin are drained northward and westward by the Cumberland river and its tributaries. The western slope of the east Gulf plains is drained directly into the Mississippi by several small streams.

Climate.—Tennessee has a delightful climate. The mean summer temperature varies according to elevation from 62° F. on the Unaka mountains to 72° on the Cumberland plateau, or 75° in the valley of east Tennessee and on the Highland rim, 77° in the Central basin and about 78° on the east Gulf plains of west Tennessee. The mean winter temperature for each of these divisions varies little from 38°, and the mean annual temperature ranges only from 57° in east Tennessee to 58° in middle Tennessee and to 60° in west Tennessee. Killing frosts are rare between the third week in April and the middle of October, especially in the southern and western parts of the state. An average annual precipitation of about 50 in. is quite equally distributed over the state, and a little more than one-half of it is well distributed through the spring and summer months. The average annual snowfall is about 8 in. The warm, moisture-bearing winds blow low from the south or southwest with a free sweep across the state in a direction nearly parallel with the trend of the mountains. The commingling of these winds with upper cold currents from the north gives rise frequently to westerly and occasionally to easterly winds.

Soil.—The Central basin, the less elevated parts of the valley of east Tennessee, and parts of the outer portion of the Highland rim have a fertile soil of limestone origin. There are narrow strips of rich alluvium along many of the rivers. The soils on the mountains, on the ridges of the valley of east Tennessee and on the eastern slope of the east Gulf plains vary greatly, according to the rocks from which they are derived. In the Cumberland plateau, in the inner portion of the Highland rim and in the higher parts of the western slope of the east Gulf plains, the soil for the most part is sandy and thin.

Plants and Animals.—At one time the greater part of the state was covered with forests, but less than 10% of the virgin forest remains. There are some great stands of oak, elm and beech that are original growth. Common throughout the state are sycamore, basswood, cherry, walnut, hickory, locust, poplar and maple trees. Hemlock, pine, spruce and balsam are found in abundance in eastern Tennessee. The swampy sections of western Tennessee are similar to the deep south in their vegetation. Cypress, pecan, tupelo, cotton gum and chinquapin grow there. Altogether, more than 150 types of trees are found within the state's borders. Redbud, dogwood, wild grape, passionflower and trailing arbutus are the most widely seen flowering shrubs and vines. Types of wildflowers include violets, daisies, black-eyed Susans, goldenrod, asters, gentians and Jerusalem artichokes. In the mountain regions, the commonest flower is the rhododendron; others frequently seen are many types of trillium, azalea, wild honeysuckle and mountain laurel. The state flower, the iris, is most commonly found in middle Tennessee, as is yellow jasmine. In western Tennessee lilies and orchids abound in the west woodlands.

During pioneer days, bison, elk, Virginia deer, black bears, wolves, panthers and lynxes were common in Tennessee. Virginia deer are still fairly numerous and there are bears and lynxes in the mountainous and wooded parts of the state. Fox, skunks and opossums are probably the best known of the fur-bearing animals found in the state. Others include raccoons, minks, muskrats, rabbits, squirrels, weasels, woodchucks and chipmunks. The Russian wild boar is infrequently seen in the wooded parts of eastern Tennessee. Many birds, once common to the region have become extinct there. These include the passenger pigeon, the Carolina parakeet, the ivory-billed woodpecker and the prairie chicken. The bald eagle is found in the wilds of the Cumberland plateau and a rare golden eagle is sometimes seen in the Great Smokies. The great blue heron, the green heron and nearly all of the other types of wading and swimming birds are numerous along the state's waterways. Common birds found include the mockingbird, robin, bluebird, killdeer, meadow lark, cardinal, towhee, screech owl, hawk, chickadee, whippoorwill, flicker, woodpecker and wren. The largest game bird is the wild turkey. Other game birds are grouse, quail, ring-necked pheasant, bobwhite and dove. All of the common North American fishes are found in Tennessee's streams and lakes. Particularly numerous are catfish, buffalo fish, drum, bass, pike, jack salmon, crappie, perch and bream. Of the reptiles found in the state, the rattlesnake, copperhead and cottonmouth moccasin are poisonous.

State and National Parks, Historic Sites.—The 20 state parks of Tennessee comprise more than 250,000 ac. of land. Practically all of the parks have lakes and streams and all of them are wildlife refuges. The largest, Natchez Trace State park, is in western Tennessee and contains 42,000 ac. It is named for the famous early road that connected Nashville and Natchez, Miss. Fall Creek Falls State park (15,777 ac.) contains Fall Creek falls, whose height of 256 ft. make it the highest falls in the eastern United States. David Crockett State park (1,000 ac.) is located immediately west of Lawrenceburg where Crockett lived. Other state parks are Big Ridge (3,600 ac.); Booker T. Washington (350 ac.); Cedars of Lebanon (8,300 ac.); Chickasaw (11,215 ac.); Cove Lake (850 ac.); Cumberland Mountain (1,425 ac.); Harrison Bay (1,500 ac.); Montgomery Bell (5,000 ac.); Norris Dam (4,000 ac.); Paris Landing (1,200 ac.); Pickett (11,752 ac.); Reelfoot Lake (50 ac.); Shelby Forest (12,512 ac.); Standing Stone (10,000 ac.); T. O. Fuller (1,000 ac.); Warriors' Path (1,000 ac.); and Chucalissa (30 ac.). The western half of the Great Smoky Mountains National park is situated on the eastern boundary of Tennessee—the other portion of the park is in North Carolina. The park comprises about 800 sq.mi., with approximately 241,000 ac. in Tennessee. In the park are mountain peaks more than 6,000 ft. in height. Part of the Appalachian trail runs through the park area. Considered one of the most beautiful of the national parks because of the unique, deep blue haze that overhangs the mountains, the park was dedicated in 1940.

Meriwether Lewis National monument is in western Tennessee about 35 mi. W. of Columbia, Tenn. Places of historic interest within the monument area include a section of the long-abandoned Natchez trace, and the grave of Meriwether Lewis (*q.v.*). Fort Donelson National Military park is situated on the site of one of the earliest and most decisive battles of the American Civil War. There Gen. U. S. Grant defeated a Confederate force and captured the fort in 1862. Chickamauga and Chattanooga National Military park is partly in Tennessee and partly in Georgia. The Tennessee portion of the park contains the battlefield of Lookout mountain, and Orchard Knob, Missionary Ridge and Signal Point on Signal mountain. The park commemorates two of the most famous military engagements of the Civil War. At Lookout mountain in Nov. 1863 Union forces captured the approaches to Chattanooga and forced a general retreat of the Confederate army

into Georgia. Stones River National battlefield near Murfreesboro is located on the site of one of the hardest fought battles of the Civil War. At Stones River the Union army of the Cumberland and the Confederate army of Tennessee fought a three-day battle in Dec. 1862–Jan. 1863. Though both armies suffered heavy casualties, the Union army occupied Murfreesboro and the Confederate army was forced to retreat. Shiloh National Military park commemorates the first major battle of the western campaign during the Civil War. The battle, considered a Confederate defeat, took place in 1862.

HISTORY

Exploration and Settlement.—Eastern Tennessee was recognized as a common hunting ground by the Cherokee (*q.v.*), Creeks and other Indian tribes. The Iroquois of New York claimed much of the central portion by right of conquest, and the western section was the home of the Chickasaws (*see* MUSKOGEAN INDIANS).

The first white man known to have visited Tennessee was the Spanish explorer, Hernando de Soto, who in 1540 entered briefly the southeastern portion of the state. Some historians believe De Soto traversed the entire southern portion of the state, crossing the Mississippi river from a point near the present site of Memphis; the generally accepted theory, however, is that he followed a more southerly course through northern Alabama and Mississippi.

Pere Marquette, the French missionary and explorer, in his voyage down the Mississippi in 1673, camped upon the western border of the state and eight years later Robert Cavelier, sieur de La Salle, and his companions left Canada to complete the exploration of the river. La Salle built Ft. Prud'homme in 1682 upon the Chickasaw bluff, near the present site of Memphis, but it was soon abandoned. Later the French built Ft. Assumption, where Memphis now stands, and kept a garrison there, but made no attempt at colonization. The territory was a part of the English grant to Sir Walter Raleigh in 1584, and of the later Stuart grants, including that in 1663 to the proprietors of Carolina. James Adair of South Carolina, a fur trader and explorer, is supposed to have been the first to go from the English colonies into Tennessee. A party of Virginians led by Thomas Walker (1715–94) in 1750 reached the Cumberland river and Cumberland mountains and named them in honour of the royal duke. In 1756–57 Ft. Loudoun, named in honour of John Campbell, earl of Loudoun, was built on the Little Tennessee river, about 30 mi. S. of the present site of Knoxville, as an outpost against the French who were then active in the whole Mississippi valley, and was garrisoned by royal troops. The fort was captured by the Cherokee Indians in 1760, and both the garrison and the neighbouring settlers were massacred.

The first permanent white settler in Tennessee was William Bean, who in 1768 built a cabin along the Watauga river in the northeastern portion of the state. Within the next few years settlements were established near Bean's cabin and also along the Holston and Nolichucky rivers. Hundreds of additional settlers arrived after the defeat of the Regulator insurrection, a popular uprising against excessive taxes, in North Carolina in 1771. These settlers negotiated a ten-year lease with the Cherokees for the land they occupied, and then drew up a compact of government called the Watauga association, mutually binding themselves to observe a body of law. A general committee of 13 was elected to exercise legislative powers and this committee elected 5 of its number to exercise executive and judicial functions. In 1775 the land was purchased from the Cherokees as a part of the negotiations by which Richard Henderson made his famous purchase (*see* also KENTUCKY: *History*).

Part of North Carolina.—With the approach of the American Revolution, the dream of becoming a colony with a royal governor was abandoned, and on petition of the inhabitants the territory was annexed to North Carolina in 1776 as the Washington district, which in 1777 became Washington county, with the Mississippi river as the western boundary. The population increased rapidly and soon several new counties were created.

During the American Revolution the Watauga and Holston settlers joined to defeat decisively the British forces under Maj. Patrick Ferguson. Though warned by Ferguson to cease aiding the Colonial forces, the mountaineers under John Sevier, Isaac Shelby and William Campbell, nevertheless crossed the mountains to attack Ferguson on Oct. 7, 1780, caught up with him at King's mountain, S.C., and aided by some North Carolina troops, killed or captured his entire force. In 1779 James Robertson (1742–1814) led an overland expedition that planted a settlement on the Cumberland river in middle Tennessee. The women and children of Robertson's party arrived at the settlement in the spring of 1780 after a harrowing four-month voyage down the Tennessee river and then up the Ohio and Cumberland rivers, under the command of John Donelson. The infant settlement of Nashborough, named after Gen. Francis Nash of North Carolina, was self-governed after the fashion of its sister settlement in east Tennessee. The Cumberland compact, which contained the first provision for recall of elected officials in the history of the United States, remained in force until 1783, when North Carolina made the renamed town of Nashville the county seat of the newly created Davidson county.

The State of Franklin.—After the American Revolution the legislature of North Carolina offered in 1784 to cede its western territory to the general government, provided the cession be accepted within two years. The Watauga settlers, indignant at this transfer without their consent, and fearing to be left without any form of government whatever, assembled in convention at Jonesboro on Aug. 23, 1784, and chose delegates to a later convention to form a new state. Meanwhile North Carolina repealed the act of cession and created the western counties into a new judicial district. A second convention of the Watauga settlers broke up in confusion without accomplishing anything; but a third adopted a constitution, which was submitted to the people, and ordered the election of a legislature. This body met early in 1785, elected Sevier governor of the new state of Franklin, filled a number of offices and passed several laws in anticipation of an autonomous existence.

In May 1785 William Cocke presented to the federal congress a memorial requesting recognition of Franklin as a state. Congress refused to accede to this request, however, and subsequently North Carolina renewed its efforts to effect a return of the "Franklinites" to the allegiance of the mother state. For a time two sets of officials claimed authority in Franklin; but North Carolina finally prevailed after remitting taxes unpaid since 1784. In 1788 Sevier's term as governor of Franklin expired and no successor being elected the state of Franklin came to an end. Sevier was arrested on a charge of treason, but was allowed to escape, and soon afterward was again appointed brigadier general of militia.

The State of Tennessee.—On Dec. 12, 1789, North Carolina again ceded the territory to the United States government, stipulating that all the general provisions of the ordinance of 1787 should apply except that forbidding slavery. Congress accepted the cession and, on May 26, 1790, passed an act for the government of the "Territory south of the River Ohio." William Blount was appointed the first governor, and in 1792 Knoxville became the seat of government. The chief events of Blount's administration were the contests with the Indians, the purchase of their lands and the struggle against Spanish influence.

A census ordered by the territorial legislature in 1795 showed more than 60,000 free inhabitants (the number prescribed before the territory could become a state), and accordingly a convention to draft a state constitution met in Knoxville on Jan. 11, 1796. The instrument, which closely followed the constitution of North Carolina, was proclaimed without submission to popular vote. Sevier was elected governor, and Blount and Cocke United States senators. In spite of the opposition of the Federalist party the state of Tennessee was admitted to the union on June 1, 1796.

With the subsequent rapid increase of population, the dread of Indians and Spaniards declined. Churches and schools were built, and soon many of the comforts and some of the luxuries of life made their appearance. The question of a circulating medium of exchange was acute during the first half of the 19th century, and state banks were organized, but they suspended specie payments in times of financial stringency. The bank of Tennessee, organized

in 1838, had behind it the credit of the state, and it was hoped that money for education and for internal improvements might be secured from its profits. The management became a question of party politics, and later, during the American Civil War, its funds were used to advance the Confederate cause.

The development of west Tennessee along the Mississippi was rapid after the beginning of the 19th century. This section was purchased from the Chickasaw Indians in 1818, and in the following year Tennessee's most populous county, Shelby, was created with its county seat at Chickasaw Bluffs (now Memphis).

The peculiar topographical conditions made the three sections of the state almost separate commonwealths, and demand for better means of communication was insistent. The policy of state aid to internal improvements found advocates very early, but a definite program was not laid out until 1829 when commissioners for internal improvements were appointed and an expenditure of $150,000 was authorized. In 1835 the state agreed to subscribe one-third to the capital stock of companies organized to lay out turnpikes, railways, etc., and four years later the proportion became one-half. The general policy was continued, and in 1861 more than $17,000,000 of the state debt was due to these subscriptions, from which there was little return.

Andrew Jackson was the dominant figure in Tennessee politics during the second quarter of the 19th century. The state's first representative in congress and one of its first superior court judges, Jackson's career had gone into virtual eclipse until his well-earned fame in the War of 1812 revived it. His nomination for president in 1824 ushered in a period during which he literally dictated his party's policies, but his attempt to dictate the election of Martin Van Buren as his successor in 1836 brought about a revolt against him in his own state. Sen. Hugh Lawson White received Tennessee's electoral votes that year and his followers became the leaders of the newly formed Whig party. During the next 20 years that party carried every presidential election in Tennessee, including even that of 1844 when a native son, James K. Polk, was the Democratic candidate for president.

The American Civil War.—When the Whig party split nationally on the slavery issue Tennessee became nominally Democratic. Proslavery sentiment was strong in middle and west Tennessee, but less strong in east Tennessee; loyalty to the Union, however, was strong in all sections of the state. The lack of a strong secessionist sentiment among the people of the state helped prevent the adoption of extreme measures when the representatives of nine southern states met in Nashville in 1850 to consider the sectional issues confronting the country. In 1860 Tennesseans again demonstrated their loyalty by voting for the Constitutional Union party's candidates, John Bell and Edward Everett, in the presidential elections. As late as Feb. 9, 1861, voters defeated a proposal to hold a convention for the consideration of seceding from the Union.

This pro-Union sentiment changed almost overnight when Pres. Abraham Lincoln called on Tennessee to provide troops to help put down the southern rebellion. On May 7, 1861, the state entered into a "military league" with the Confederacy, and on June 8 the people of middle and west Tennessee voted overwhelmingly to sever their ties with the Union; by an equally overwhelming majority east Tennessee voted to remain in the Union. Thus, when Gov. Isham G. Harris proclaimed that Tennessee had declared its independence of the union, Andrew Johnson of east Tennessee refused to resign his seat in the United States senate and was upheld in that action by the people of his section. East Tennessee remained Unionist throughout the Civil War, and has been strongly Republican in its politics ever since.

Next to Virginia, the state was the chief battleground during the Civil War. In Feb. 1862 Gen. U. S. Grant and Commodore A. H. Foote captured Ft. Henry on the Tennessee river and Ft. Donelson on the Cumberland. The Confederate line of defense was broken and Gen. D. C. Buell occupied Nashville. Grant next ascended the Tennessee river to Pittsburg Landing, with the intention of capturing the Memphis and Charleston railway, and on April 6 and 7 defeated the Confederates in the battle of Shiloh. The capture of Island No. 10 in the Mississippi on April 7 opened the river as far south as Memphis, which was captured in June. On Dec. 31 and Jan. 2 Gen. William S. Rosecrans (Federal) fought with Gen. Braxton Bragg (Confederate) the bloody but indecisive battle of Stones river (Murfreesboro). In June 1863 Rosecrans forced Bragg to evacuate Chattanooga. Bragg, however, turned upon his pursuer, and on Sept. 19 and 20 in one of the bloodiest battles of the war, defeated Rosecrans at Chickamauga. General Grant then assumed command, and on Nov. 24–25 defeated Bragg at Chattanooga, thus opening the way into east Tennessee. There Confederate Gen. A. E. Burnside at first met with success, but was shut up in Knoxville by Gen. James Longstreet, who was not able, however, to capture the city, and on the approach of Gen. W. T. Sherman retired into Virginia. Almost the whole state was now held by Federal troops, and no considerable military movement occurred until after the fall of Atlanta in Sept. 1864. Then Gen. J. B. Hood moved into Tennessee, expecting Sherman to follow him. Sherman, however, sent reinforcements to Thomas and continued his march to the sea. Hood fought with Gen. John M. Schofield at Franklin, and on Dec. 15–16 was utterly defeated by Gen. George H. Thomas at Nashville, the Federals thus securing virtually undisputed control of the state. (*See* AMERICAN CIVIL WAR).

Reconstruction.—After the occupation of the state by the Federal armies in 1862, Andrew Johnson was appointed military governor by the president (confirmed March 3, 1862), and held the office until he was inaugurated vice-president on March 4, 1865. Republican electors attempted to cast the vote of the state in 1864, but were not recognized by congress.

Tennessee was the first of the Confederate states to be readmitted to the union (July 24, 1866) after ratifying the 13th and 14th amendments to the United States constitution. Tennessee freed its slaves by an amendment to the state constitution ratified by a vote of the people on Feb. 22, 1865, but suffrage was not conferred upon the Negro until two years later (Feb. 25, 1867). The state escaped "carpet bag" government, but the native whites in control, under the leadership of William G. Brownlow, exhibited almost every phase of the Reconstruction policy. All persons who had either directly or indirectly taken part in the war against the union or had given aid to the Confederacy were denied the right of suffrage. In the election of 1869, the acting governor, D. W. C. Senter, ordered the election commissioners to issue to all actual citizens of the state permits to vote. The Democrats united with the conservative Republicans and Senter was easily elected. At the same time a Democratic and conservative legislature was elected, thus placing the state government again in the control of officers elected by the majority of the people. The Ku Klux Klan, originating in 1865 at Pulaski, Tenn., spread over the state and the entire south, and in 1869 nine counties in the middle and western section were placed under martial law because of the Klan's activities against the Loyal (or Union) league, an organization supporting the union, and the Negroes.

A constitutional convention, which met in Jan. 1870, revised the old constitution, and the revision became the state constitution when ratified by the people the following May. In 1873 a school law was passed which provided for state and county superintendents and separate schools for white and Negro children and levied a state tax to aid in paying the expenses of these schools. A progressive step was the final compounding of the old state debt at 50 cents on the dollar by an act of the legislature in 1883.

In 1904 the Adams law, which prohibited the sale of intoxicating liquors in towns of 5,000 inhabitants or less, was passed. Successive laws concerning prohibition were passed in 1910, 1913, 1915 and 1917; nevertheless, Tennessee voted for repeal in 1933. In 1915 a law was passed providing for mothers' pensions. In 1917 the letting of prison labour to private contractors was prohibited, and an act was passed forbidding the limiting of the output of coal in order to increase the price. In the same year a state budget commission was created. On Aug. 18, 1920, the state ratified the proposed amendment to the federal constitution, providing for woman suffrage. As the 36th state to ratify, Tennessee brought the number to the requisite three-fourths.

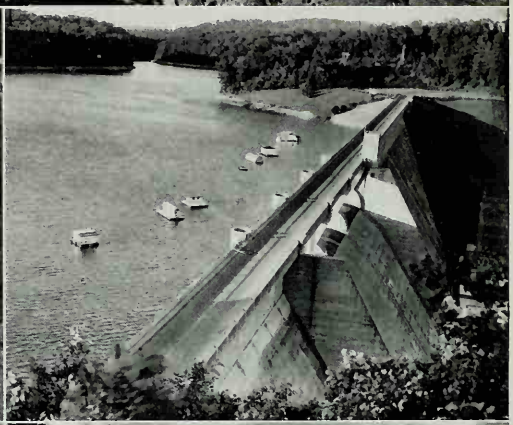
After World War I Tennessee became a major industrial state,

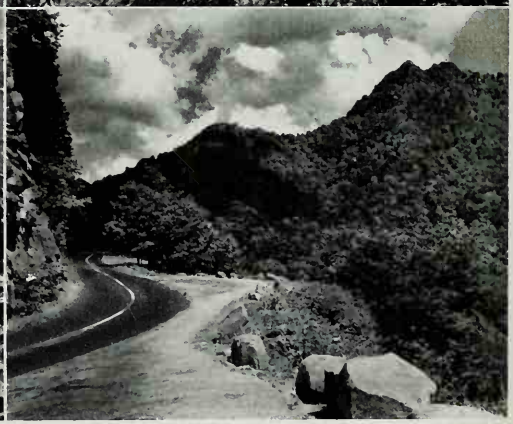
BY COURTESY OF (TOP LEFT, TOP RIGHT, CENTRE RIGHT, BOTTOM RIGHT) TENNESSEE CONSERVATION DEPARTMENT PHOTOGRAPH, (BOTTOM LEFT) WILLARD SMITH

VIEWS OF CENTRAL AND EAST TENNESSEE

Top left: Hardwood forest in White county, central Tennessee
Top right: View of the Tennessee river and Chattanooga from atop Lookout mountain, site of the Civil War "battle above the clouds," Nov. 24–25, 1863. Lookout mountain is part of the Chickamauga and Chattanooga National Military park
Centre right: Horse and cattle farm in the central part of the state
Bottom left: Merner-Pfeiffer library of Tennessee Wesleyan college, Athens. In 1857 the school was organized as Athens Female college
Bottom right: Burley tobacco field in a valley of eastern Tennessee

TENNESSEE





BY COURTESY OF (TOP, CENTRE LEFT, BOTTOM LEFT, BOTTOM RIGHT) TENNESSEE CONSERVATION DEPARTMENT; PHOTOGRAPH, (CENTRE RIGHT) JOHN E. THIERMAN

SCENES IN TENNESSEE

Top: Aerial view of Ducktown. Sulfur fumes, released in smelting copper, have denuded the terrain

Centre left: The Hermitage, residence of Andrew Jackson (1767–1845) northeast of Nashville. It was completed in 1835

Centre right: Houseboats on the Clinch river at Norris dam, 25 mi. northwest of Knoxville. In 1936 the dam was completed as part of the TVA and was named in honour of Sen. George W. Norris (1861–1944)

Bottom left: Aerial view of Memphis and the Mississippi river. Bridge (background) connects the southwest section of the state with Arkansas

Bottom right: "The Chimneys" in Great Smoky Mountains National park

largely as a result of the Tennessee Valley authority (TVA). This giant network of navigation-flood control-hydroelectric power dams on the Tennessee river and its tributaries began with the construction of Wilson dam at Muscle Shoals, Ala., in 1916. This dam, intended originally for the production of nitrates and left unfinished at the end of the war, became the focus of the controversy in the 1920s between private power interests and those who advocated governmental development of the hydroelectric potential of the Tennessee river. The dedication of Sen. George W. Norris of Nebraska to the development of public power facilities which could be used as a "yardstick" against which to measure the rates of private power companies reached fruition in 1933 when the TVA was created as a part of Pres. Franklin D. Roosevelt's New Deal program. The subsequent building of a network of dams which provided cheap electric power to residents and industries in the area, and the later development of giant steam plants which provided additional power both for industry and for atomic power installations in Tennessee and Kentucky did much to improve Tennessee's economy and attract additional industry to the area.

Modern Politics and Issues.—Except for the years 1881–83, 1911–15 and 1921–23, the Democratic party has controlled the executive office since 1870. In 1920 and again in 1928 the state gave its electoral votes to the Republican presidential candidate. Tennessee returned to the Democratic fold in 1932, but in 1952, 1956 (despite the fact that in that year Tennessean Estes Kefauver was the Democratic nominee for vice-president) and 1960 the Republican party won the state's electoral votes. Tennessee remained firmly Democratic in state elections and in the election of senators, and of representatives in all but the two east Tennessee congressional districts.

GOVERNMENT

Tennessee's constitution has been amended only once (in 1953) since its adoption in 1870. The constitution's cumbersome amending process, though somewhat liberalized at that time, still requires that a proposed amendment pass both houses of two consecutive general assemblies, the first time by a majority of the total membership of each house and the second time by a two-thirds vote of the total membership, after which it must be ratified by a majority of all citizens of the state voting in the gubernatorial election. Amendments may also be proposed by a constitutional convention and ratified in like manner by the people, but the legislature must first propose the holding of such a convention and this proposal must be approved by a majority of those voters expressing themselves on the proposal at a governor's election. Other changes made included the lengthening of the governor's term of office from two to four years with the provision that he could not succeed himself; the granting to the governor of the power of veto over individual items in appropriations bills; the consolidation of certain city and county functions; the raising of legislator's salaries; the granting of a certain amount of home rule to localities; and the repeal of the poll tax, which, in fact, had not been regularly collected.

State elections are held in even-numbered years on the first Tuesday after the first Monday in November, but local and county elections are held the first Thursday of August, which, in even-numbered years, is also the date for statewide primary elections.

Legislative.—The governor is the state's only popularly elected executive officer. He is elected for a term of four years and is not eligible to succeed himself. The speaker of the state senate is, by virtue of that office, lieutenant governor and succeeds the governor in the event of his death or removal from office. The treasurer, comptroller of the treasury and secretary of state are elected by joint ballot of both houses of the general assembly. The secretary of state's term is four years; the others are two-year terms. The attorney general is appointed by the members of the supreme court for an eight-year term. The heads of the major departments of the state government are appointed by and serve at the pleasure of the governor. These departments are agriculture, conservation, corrections, education, employment security, finance and taxation, highways and public works, insurance and banking, labour, mental health, public health, public welfare and safety.

Following the historic U.S. supreme court decision in *Baker* v. *Carr* on March 26, 1962, that Tennessee's legislative apportionment was a proper subject for review by federal courts, a special session of the legislature adopted reapportionment legislation for the first time since 1901; this gave greater representation to urban areas but was found by a federal court still to be inadequate. The court allowed time for the 1963 session to remedy the inequity. State senators and representatives are elected for two-year terms. The number of representatives is limited by the constitution to 99; there are 33 senators. The legislature meets biennially, in odd-numbered years for a maximum of 75 days in regular session and 20 days in an extraordinary session called by the governor. Bills may originate in either house, but no bill can become law unless it receives a majority vote of the full membership of each house. A veto by the governor can be overridden by the same majority vote. No minister, priest or rabbi is eligible for membership in the legislature.

Judiciary.—The administration of justice is vested in a supreme court, a court of appeals, chancery courts, circuit courts, criminal courts, courts of general sessions, county courts, justice of the peace courts and recorders' courts. The supreme court consists of five judges elected by the state at large for a term of eight years, one for each of the three grand divisions (east, middle and west) and two for the state at large. The judges designate one of their number to preside as chief justice. The court has appellate jurisdiction only. The court of appeals is composed of nine judges, three from each grand division of the state, elected for a term of eight years. This court has jurisdiction of appeals in all equity cases, with a few exceptions, and it may transfer any case to the supreme court. The state is divided into 14 chancery divisions, in each of which a chancellor is elected for a term of eight years, and a court of chancery is held in every county seat in each division. Two of the divisions each have two chancellors.

The state is divided into 21 circuits in each of which a circuit judge is elected for a term of eight years, and a circuit is held in every county seat. Some circuits have more than one judge. There are also 12 criminal judges, principally in the most heavily populated areas, who are elected for eight-year terms.

Each county is divided into civil districts, varying in number according to population or other factors, and each district elects at least two justices of the peace for a term of six years; each county town or incorporated town also elects one justice of the peace. A recorder has concurrent jurisdiction with a justice of the peace. In several of the counties the county court is composed of a county judge, elected for a term of eight years, together with the justices of the peace. In other counties the court is composed of the justices of the peace alone who elect one of their number as chairman. The government of each county is vested principally in the county court. A county trustee, whose duty it is to collect state and county taxes, and a sheriff are elected by the people of the county for a term of two years; a clerk of the county court and a registrar are also elected by the voters of the county for a term of four years.

Finance.—Educational appropriations consume the largest portion of the total expenditure of state funds, with highways second and public welfare third. Chief sources of revenue are general sales tax, motor vehicle licences, general property tax, gasoline tax, tax on insurance companies, sales tax on tobacco and alcoholic beverages, an excise tax and an inheritance tax. Various other licences, fines and forfeitures help to augment the state's revenue.

Total state revenues in the second half of the 20th century were nearly $400,000,000 annually; total expenditures were about $375,000,000 of which about 29% was for education, nearly 24% was for highways, nearly 14% for welfare and more than 10% for cities and local governments. The gross debt at that time was about $115,000,000. Though personal income was up considerably since World War II, the state's per capita income remained among the lowest in the nation.

POPULATION

The population of Tennessee in 1800, four years after it obtained statehood, was 105,602. Its population ranked 15th among the 22 states and territories that then composed the union, and was classified as 100% rural. The period from the first census in 1790 to 1800 was the one of the most rapid growth in the state's history—the population increased by nearly 196% during that time. Immediately before the Civil War, in 1860, the state had a population of 1,258,520 of whom about 275,000 were Negro slaves.

*Tennessee: Places of 5,000 or More Population (1960 Census)**

Place	Population				
	1960	1950	1940	1920	1900
Total state	3,567,089	3,291,718	2,915,841	2,337,885	2,020,616
Alcoa	6,395	6,355	5,131	3,358	—
Athens	12,103	8,618	6,930	2,580	1,849
Bristol	17,582	16,771	14,004	8,047	5,271
Brownsville	5,424	4,711	4,012	3,062	2,645
Chattanooga	130,009	131,041	128,163	57,895	30,154
Clarksville	22,021	16,246	11,831	8,110	9,431
Cleveland	16,196	12,605	11,351	6,522	3,858
Columbia	17,624	10,911	10,579	5,526	6,052
Cookeville	7,805	6,924	4,364	2,395	—
Covington	5,298	4,379	3,513	3,410	2,787
Dickson	5,028	3,348	3,504	2,263	1,363
Donelson	17,195	1,765	—	—	—
Dyersburg	12,499	10,885	10,034	6,444	3,647
Eagleton Village	5,068	—	—	—	—
East Ridge	19,570	9,645	2,939	—	—
Elizabethton	10,896	10,754	8,516	2,749	—
Fayetteville	6,804	5,447	4,684	3,629	2,708
Fountain City	10,365	—	—	—	—
Franklin	6,997	5,475	4,120	3,123	2,180
Gallatin	7,901	5,107	4,829	2,757	2,409
Greeneville	11,759	8,721	6,784	3,775	1,817
Harriman	5,931	6,389	5,620	4,019	3,442
Humboldt	8,482	7,426	5,160	3,913	2,866
Inglewood	26,527	—	—	—	—
Jackson	33,849	30,207	24,332	18,860	14,511
Johnson City	29,892	27,864	25,332	12,442	4,645
Kingsport	26,314	19,571	14,404	5,692	—
Knoxville	111,827	124,769	111,580	77,818	32,637
La Follette	6,204	5,797	4,010	3,056	366
Lawrenceburg	8,042	5,442	3,807	2,461	823
Lebanon	10,512	7,913	5,950	4,084	1,956
Lewisburg	6,338	5,164	3,582	2,711	1,421
Lynn Gardens	5,261	—	—	—	—
McMinnville	9,013	7,577	4,649	2,814	1,980
Madison	13,583	—	—	—	—
Maryville	10,348	7,742	5,609	3,739	—
Memphis	497,524	396,000	292,942	162,351	102,320
Milan	5,208	4,938	3,035	2,057	1,682
Millington	6,059	4,696	730	657	—
Morristown	21,267	13,019	8,050	5,875	2,973
Murfreesboro	18,991	13,052	9,495	5,367	3,999
Nashville	170,874	174,307	167,402	118,342	80,865
Newport	6,448	3,892	3,575	2,753	1,630
Oak Ridge	27,169	30,229	—	—	—
Paris	9,325	8,826	6,395	4,730	2,018
Pulaski	6,616	5,762	5,314	2,780	2,838
Red Bank-White Oak	10,777	—	—	—	—
Rockwood	5,345	4,272	3,981	4,652	2,899
Shelbyville	10,466	9,456	6,537	2,912	2,236
Springfield	9,221	6,506	6,668	3,860	1,732
Tullahoma	12,242	7,562	4,549	3,479	2,684
Union City	8,837	7,665	7,256	4,412	3,407
Whitehaven	13,894	1,311	—	—	—
Woodbine-Radnor-Glencliff	14,485	—	—	—	—
Woodmont-Green Hills-Glendale	23,161	—	—	—	—

*Populations are reported as constituted at date of each census.
Note: Dash indicates place did not exist during reported census, or data not available.

It had advanced to the 8th most populous state, and was about 96% rural in character. At the beginning of the 20th century, the state had slipped to 14th in rank of population with a total of 2,020,616 persons by the 1900 census. It was still largely rural in character; about 84% of the people were classified as nonurban dwellers. The decade between 1910 and 1920 was the period of the slowest growth in the state's history. In 1920 the population was only 7% greater than it had been in 1910. Since 1920 the Negro population has been steadily rising, but its proportion of the total population has been declining since 1880. The 1950 population showed a total of 3,291,718 persons living within the state, making Tennessee the 16th in size of population among the states. The rural population had declined to 55.9% of the total at that time. There were four standard metropolitan areas in 1950. They were Chattanooga (Hamilton county in Tennessee and Walker county in Georgia); Knoxville (Anderson, Blount and Knox counties); Memphis (Shelby county); and Nashville (Davidson county). These areas contained 41.0% of the total population. By 1960 these metropolitan areas housed 45.8% of the total population.

In 1960 Tennessee had a population of 3,567,089, an increase of 275,371 or 8.4% over 1950. The state ranked 17th in size of population. The population per square mile in 1960 was 84.4, as compared with 49.6 for the U.S. as a whole. The 1960 urban population was 1,864,828 or 52.3% of the total, continuing the trend toward increasing urban residence. Distribution by race and nativity, according to the 1950 census was as follows: 83.4% native white; 0.5% foreign-born white and 16.1% nonwhite, practically all Negro. In line with the rest of the U.S. the percentage of persons 65 years old or over was increasing, being 7.2% in 1950. The percentage of the population 14 years old and over that was in the labour force (50.9 in 1950) was steadily decreasing, the result of prolonged schooling for that age group.

EDUCATION

The first steps toward the creation of a common school system for Tennessee were taken in 1823, when an act was passed setting aside for school purposes the revenue and taxes accruing from public lands. From that beginning until the adoption of a uniform school code in 1925, more than 1,000 different acts were passed pertaining to the educational system.

Under the new code the administration is vested in a state commissioner of education; a state board of education, consisting of the commissioner, the governor and nine other members; a county superintendent and a county board in each county; and in cities, city boards of education.

The commissioner and the other members of the state board are appointed by the governor. The county superintendent is elected biennially by either the county court or by popular ballot. The county board consists of seven members (unless otherwise provided), elected by the county court or by popular ballot, for a term of seven years, one retiring annually.

The state board of education prescribes the rules and regulations for the operation of the school program, and prescribes standards and requirements for curriculum, teacher certification, school buildings, grading and instructional materials. It also maintains and controls the five state colleges and Tennessee Agricultural and Industrial state university.

The quarterly county court has power to issue bonds, levy taxes and provide funds for buildings for the county schools. The court is under obligation to maintain at least one four-year high school within the county.

Public schools are free to all persons above six years of age. Attendance is compulsory between the ages of 7 and 16 unless high school standing has been attained. Separate schools are maintained for whites and Negroes, though a few schools have been integrated since the U.S. supreme court decision in 1954 declaring segregated schools unconstitutional. Expenditures per pupil in average daily attendance were in excess of $170 during the second half of the 20th century.

Universities and Colleges.—The largest institution of higher learning provided by the state is the University of Tennessee, at Knoxville. Chartered by the legislature of the Southwest Territory in 1794 under the name of Blount college, it became East Tennessee college in 1807 and East Tennessee university in 1840. It was incorporated as Agricultural and Mechanical college in 1869 and took its present name in 1879. The main campus remains at Knoxville, but the graduate school; school of biological sciences; colleges of medicine, dentistry and pharmacy; school of nursing; and division of public health are at Memphis. The agricultural extension branch office and the school of social work are at Nashville; substations of the agricultural experiment station are at Jackson, Columbia, Greeneville, Crossville, Springfield and Lewisburg; and there is a branch of the university at Martin. Other state-controlled schools include Memphis State university (1909); East Tennessee State college at Johnson City (1911); Middle Tennessee State college at Murfreesboro (1909); Tennessee Polytechnic institute at Cookeville (1915); Austin Peay State college at Clarksville (1927); and Tennessee Agricultural and Industrial State university at Nashville (1912).

Privately controlled institutions of higher learning include Belmont college, Nashville (Baptist, 1951); Bethel college, McKenzie (Presbyterian, 1842); Carson-Newman college, Jefferson City

(Baptist, 1851); Christian Brothers college, Memphis (Roman Catholic, 1871); David Lipscomb college, Nashville (Church of Christ, 1891); Fisk university, Nashville (American Missionary association affiliate, 1867), originally established by the Freedmen's bureau as a school for Negroes; George Peabody College for Teachers, Nashville (nonsectarian, 1785); King college, Bristol (Presbyterian, 1867); Knoxville college, Knoxville (United Presbyterian, 1875) for Negroes; Lambuth college, Jackson (Methodist, 1923); Lane college, Jackson (Methodist Episcopal, 1882) for Negroes; Le Moyne college, Memphis (Congregational Christian, 1871) for Negroes; Lincoln Memorial university, Harrogate (nonsectarian, 1897); Maryville college, Maryville (United Presbyterian, 1819); Scarritt College for Christian Workers, Nashville (Methodist, 1892); Siena college, Memphis (Roman Catholic, 1922); Southern Missionary college, Collegedale (Seventh-day Adventist, 1893); Southwestern at Memphis, Memphis (Presbyterian, 1848); Tennessee Wesleyan college, Athens (Methodist, 1857); Tusculum, Greeneville (United Presbyterian, 1794); Union university, Jackson (Baptist, 1825); University of Chattanooga, Chattanooga (nonsectarian, 1886); University of the South, Sewanee (Protestant Episcopal, 1857); and Vanderbilt university, Nashville (nonsectarian, 1873), one of the best-known schools in the south and endowed by Cornelius Vanderbilt.

HEALTH, WELFARE AND CORRECTIONS

The state department of public health, headed by a commissioner of cabinet rank, is made up of 11 divisions which have jurisdiction over central administration, local health service, crippled children's service, dental health, laboratories, preventable diseases, sanitary engineering, tuberculosis control, vital statistics, hospital survey and construction, medical care and industrial hygiene. A department of mental health provides psychological, nursing, statistical, fiscal and personnel service, as well as directing work in forensic psychiatry, mental health education, psychiatric social work, nutrition and the engineering and planning of mental health installations. There are state mental hospitals at Knoxville, Nashville and Bolivar. There is a psychiatric hospital in Memphis, a home for the mentally defective in Donelson and 13 community mental health centres in various parts of the state.

The department of correction has its main penitentiary at Nashville. A branch known as Brushy Mountain prison, is located at Petros in east Tennessee. There the prisoners work principally in the state coal mines and on a farm. Fort Pillow state farm in west Tennessee is primarily for first offenders.

The department of public welfare administers programs of public assistance to the aged, the blind, the disabled and to dependent children. It also aids in rehabilitation and training of sightless persons; regulates public and private child care agencies; oversees adoptions and placing of dependent children in foster homes.

THE ECONOMY

Tennessee was primarily an agricultural state until the second quarter of the 20th century, when manufacturing became the major source of the state's wealth. Pre-Civil War Tennessee, especially the middle and western portions, belonged to that complex of southern states in which cotton was "king"; however, the state's economy was never as completely geared to cotton as that of the states of the deep south. Tobacco has almost always been the second most valuable crop in the state, with corn, hay and wheat also ranking high in value of production. Since World War II Tennessee's already diversified farming has been affected by two significant trends: a greater emphasis on livestock production and the mechanization of farm operations. The breeding of show cattle, particularly of the Jersey, Angus and Hereford breeds, has become important in the state's rapidly growing livestock and dairy industry. Tennessee walking horses are bred almost solely for show. The once comfortable and utilitarian plantation horse with its "rocking chair ride" became less a riding horse, though its exaggerated gaits are much admired in some show circles.

Agriculture.—The number of acres of farmland remained fairly constant in the second half of the 20th century, but with more intensive cultivation there had been a constant decline in acres of cropland harvested. In the mid-1950s, 4,860,793 ac. were classed as cropland or nearly 1,400,000 ac. less than in 1925. Much former cropland had been turned into pasture needed for the growing livestock industry in Tennessee. During this same period the size of the average farm increased from 70.8 to 86.9 ac. and the average value of land and buildings from about $3,000 to about $7,500. The number of farm operators declined from more than 250,000 to slightly more than 200,000, of whom approximately 120,000 were full owners, 30,000 part owners and 50,000 tenants; there were about 400 managers. Of the farm operators about 10% were nonwhites, most of them Negroes. Tennessee farmers averaged total cash receipts from farming of more than $475,000,000 annually, about evenly divided between livestock and livestock products and crops. Leading income producers in the livestock category were dairy products, cattle and hogs. Leading crops were corn, hay, tobacco, wheat and cotton.

Forests.—In the 1960s Tennessee's forest lands totaled 12,408,000 ac. and contributed a large portion of the state's agricultural and industrial income through lumber and wood products, furniture and pulp paper industries.

Minerals.—Tennessee ranks about 27th among the states in the value of its mineral products. The principal minerals are coal, cement, stone, phosphate rock, zinc and copper. Other mineral products of economic value were clays, lime, manganese, sand and gravel, lead, gold, silver, barite, petroleum, natural gas and celestite. Tennessee produces most of the copper mined in the southern states; this metal, with the gold and silver by-products of the smelting process, is largely confined to Polk county in the southeastern portion of the state. The coal-producing area is in the Cumberland plateau; building marble, for which Tennessee is famous, comes from the eastern part of the state; and phosphate deposits are localized in middle Tennessee.

Manufacturing.—Tennessee experienced a rapid industrial development in the second half of the 20th century. In 1933 there were 1,561 manufacturing establishments employing 94,909 wage earners; 25 years later the number of establishments had increased to 4,100 and the number of employees to 295,000. Chief industries were chemicals, food and kindred products, textiles, apparel, lumber and wood products, pulp and paper, furniture and fixtures, printing and publishing, primary metals, fabricated metals, leather and leather products. Among the biggest industrial installations were the giant aluminum plant at Alcoa, chemical plants at Kingsport and Old Hickory and newsprint mills at Calhoun.

Transportation.—Railway mileage in Tennessee, 4,078 mi. in 1920, had decreased to less than 3,500 mi. by 1960. Paralleling the decline of railway mileage was a steady growth of the highway system which totaled more than 70,000 by 1960, of which about 8,000 mi. were state roads and the remainder county. Much freight formerly handled by railroads had been taken over by highway trucking firms; much additional freight was being carried by boats on the Tennessee and Cumberland rivers, where huge dams and locks constructed by the Tennessee Valley authority and by the U.S. army corps of engineers had made possible massive barge tows and extremely low shipping charges.

See also references under "Tennessee," in the Index volume.

BIBLIOGRAPHY.—William T. Alderson and Robert H. White, *A Guide to the Study and Reading of Tennessee History* (1959). *History–General:* Cartter Patten, *A Tennessee Chronicle* (1953); Philip M. Hamer (ed.), *Tennessee; A History*, 4 vol. (1933); Joseph H. Parks and Stanley J. Folmsbee, *The Story of Tennessee* (1954); Mary U. Rothrock, *Discovering Tennessee* (1936); Robert H. White, *Tennessee: Its Growth and Progress* (1936); Tennessee Historical Commission, *Tennessee Old and New*, 2 vol. (1946).

History–Various phases: Donald Davidson, *The Tennessee*, 2 vol. (1946–48); Robert L. Kincaid, *The Wilderness Road* (1947); Laura Thornborough, *The Great Smoky Mountains* (1937); Tennessee Secretary of State, *Tennessee Blue Book* (1890–); Stanley F. Horn, *The Army of Tennessee* (1953), *The Decisive Battle of Nashville* (1956); T. M. N. Lewis and Madeline Kneberg, *Tribes That Slumber* (1958); Samuel C. Williams, *Dawn of Tennessee Valley and Tennessee History* (1937), *Early Travels in the Tennessee Country, 1540–1800* (1928), *History of The Lost State of Franklin* (1933), *Tennessee During the Revolutionary War* (1944), *Beginnings of West Tennessee, in the Land of the Chickasaws* (1930); Thomas P. Abernethy, *From Frontier to Plantation in Tennessee* (1932); Lyman C. Draper, *King's Mountain and Its Heroes* (1954); Blanche Henry Clark, *The Tennessee Yeoman, 1840–1860*

(1942); Claude A. Campbell, *The Development of Banking in Tennessee* (1932); Henry T. Malone, *Cherokees of the Old South* (1956); Robert S. Cotterill, *The Southern Indians* (1954); Stanley J. Folmsbee, *Sectionalism and Internal Improvements in Tennessee, 1796–1845* (1939); Robert H. White, *Development of the Tennessee State Educational Organization* (1929); Chase C. Mooney, *Slavery in Tennessee* (1957); Robert S. Henry, *First With the Most Forrest* (1944); James W. Patton, *Unionism and Reconstruction in Tennessee, 1860–1869* (1934); Thomas B. Alexander, *Political Reconstruction in Tennessee* (1950); Daniel M. Robison, *Bob Taylor and the Agrarian Revolt in Tennessee* (1935); Harriette Arnow, *Seedtime on the Cumberland* (1960).

See also Robert H. White, *Messages of the Governors of Tennessee*, 5 vol. (1952–59).

Current statistics on production, employment, industry, etc., may be obtained from the pertinent state departments; the principal figures, together with the current history, are summarized annually in the *Britannica Book of the Year*, American edition. (W. T. A.)

TENNESSEE RIVER, the largest tributary of the Ohio river, U.S. It is formed by the confluence of the Holston and the French Broad rivers 4.5 mi. above Knoxville, Tenn. The upstream tributaries of the Tennessee drain the southern portion of the Appalachian mountains, and the highest part, known as the Great Smokies, receives the largest total rainfall in the eastern half of the U.S. The Holston, the Clinch and their tributaries rise on the mountain slopes that flank the valleys in western Virginia; the French Broad and the Little Tennessee and their tributaries rise in the high mountains of western North Carolina; the Hiwassee and its tributaries rise in the mountains of western North Carolina and northern Georgia.

The Tennessee flows south-southwest to Chattanooga, there turns west through the Cumberland plateau and into the northeast corner of Alabama, continues west across the northern part of Alabama, turns north on the boundary between Alabama and Mississippi and, continuing north across Tennessee and Kentucky, unites with the Ohio at Paducah, Ky. The drainage basin includes parts of seven states.

The Tennessee itself is 652 mi. long, and with the Holston and the North fork of the Holston forms a channel about 900 mi. long. Its drainage basin covers about 38,520 sq.mi.

History.—The name of the river, like that of the state, may have come from a village of the Cherokee Indians, located on the Little Tennessee near its mouth, and named "Tanase" or "Tennassee."

Detailed exploration of the Tennessee river took place during the period of rivalry between the French and the English for the territory west of the Appalachians. A few small forts and posts had been established by each rival. Previous to this, other early explorers and fur traders had entered the lower river from the Ohio.

The lands along the upper (Holston and Watauga) valley began to receive American settlers by the 1770s and 1780s. These had crossed the Appalachians from North Carolina or had arrived by way of the Appalachian valley routes (such as the Shenandoah) from Virginia and Pennsylvania. The first house in Knoxville, at the head of the Tennessee, was built in 1786; the townsite was soon surveyed. By this time it became possible for the river itself to serve as a route for further movement southwestward. Many settlers continued downstream on flatboats to the lands between Knoxville and Chattanooga, or into what is now northern Alabama. However, more westward settlements resulted from travel on overland routes through the Cumberland gap (60 mi. northeast of Knoxville) or on roads westward across the Cumberlands. The role of the Tennessee as a westward passage was negligible as compared to the Ohio.

Originally the Tennessee was navigated with difficulty, and only by shallow-draft boats. The upper course was shallow where mountain ridges were broached, filled with short rapids which hindered navigation except at high water and the river was subject to seasonal fluctuations. The rapid run-off from mountainous upstream tributaries, followed by periods of lower flow, resulted in both downstream floods and periods of insufficient water for safe navigation. The middle course of the Tennessee, through the Cumberlands, was tortuous, contained whirlpools and was interrupted by the famous Muscle Shoals (*q.v.*). Only the lower course, north-flowing in western Tennessee, was easily navigated.

The Tennessee furnished an early but circuitous route to eastern markets for flour, bacon, whisky and other products of the pioneers. Flatboats made the downstream journey to the Ohio, and thence to the Mississippi and to New Orleans. From this port shipments went by sea to Atlantic coast markets. Upstream traffic, via keelboats, was not important. Steamboat operation on the river dates mainly from the mid-1820s. Although it continued in modest form through the century, the advent of railroads in the valley, after the 1840s, kept river traffic from assuming the significance it had on other western, more easily navigated rivers.

Only two cities of considerable size are on the Tennessee, Knoxville at the head of navigation, and Chattanooga at an important river crossing and route centre. Traffic westward from Chattanooga focuses through the "gorge" of the river, cut through the Cumberland plateau. In early settlement days the site of Chattanooga was an important "landing." Cities of moderate size on the river (all in Alabama) are Guntersville, Decatur, Florence and Tuscumbia; these date from early settlement. Sheffield, also on the river, was not founded until the 1880s. Savannah, in western Tennessee, is the largest town on the lower course. Many of the early "landings" and "ferries" were by the 1950s only hamlets or were in ruins.

The northward-flowing lower course of the Tennessee was of utmost importance during the Civil War. It and its valley offered an invasion route into the western Confederacy. The very downstream part of the course is paralleled by the Cumberland river (*q.v.*). Confederate Ft. Henry was on the Tennessee; Ft. Donelson was 12 mi. away on the Cumberland. The army of General Grant, accompanied by federal gunboats, struck southward in the valley in February 1862. The Confederate armies fell back to Corinth, Miss., and the Union armies moved almost to the southern boundary of Tennessee. Here was fought the battle called either Pittsburgh Landing or Shiloh. With the eventual capture of the middle and western valley by Union forces, and the fall of Chattanooga, the entire Tennessee valley fell to Union forces. The upper valley in mountainous east Tennessee saw little fighting except at Knoxville; in fact, many upper east Tennesseans remained loyal to the Union.

River traffic on the Tennessee remained of only moderate importance into the early 1900s. Interestingly, for a time, marble from the Knoxville and other quarries was actually shipped to eastern building projects via the circuitous river route. The development of the river system to a higher place among the inland waterways of the United States came during the 20th century (*see* TENNESSEE VALLEY AUTHORITY).

The Tennessee is today a series of reservoirs, called lakes, each impounded by a dam. The nine mainstream dams are Kentucky, in Kentucky; Pickwick Landing, in Tennessee; Wilson, Wheeler and Guntersville, in Alabama; and Hales Bar, Chickamauga, Watts Bar, Ft. Loudoun, Knoxville and confluence of Holston and French Broad, in Tennessee. The lakes formed by the dams are extensively used for recreation. (L. Dd.)

TENNESSEE VALLEY AUTHORITY, a U.S. government agency established in 1933 to engage in improvement activities along the Tennessee river and its tributaries. The jurisdiction of the Tennessee Valley authority (TVA) is generally limited to the drainage basin of the Tennessee river, but some of its activities extend beyond the region.

The river basin, an area of almost 41,000 sq.mi., covers parts of seven states: Alabama, Georgia, Kentucky, Mississippi, North Carolina, Tennessee and Virginia. Its 1965 population was approximately 3,500,000.

The Tennessee, fifth largest among U.S. rivers in volume of stream flow, provided a means of westward travel by American frontiersmen as early as 1800. Travel on the river was interrupted, however, at Muscle Shoals, Ala., where the river dropped approximately 100 ft. in the course of 20 mi. Early efforts by state and private interests to cope with the shoals met with failure or with only limited success. It was not until World War I that national attention was drawn to the Tennessee. At that time the U.S. government determined to erect two nitrate manufacturing plants at Muscle Shoals for the manufacture of munitions;

these plants were to free the nation from dependence on Germany and Chile for nitrates. The government also determined to construct a dam on the river to supply hydroelectric power to run the nitrate plants, along with a steam-electric plant to provide additional power. None of these undertakings had been finished by the end of the war, when it was decided to suspend construction work, though Wilson dam was later completed. The government's investment in the Muscle Shoals properties up to 1918 was somewhat more than $100,000,000.

In the next 15 years there were as many as half a dozen serious proposals for liquidating the government's investment at Muscle Shoals. Some failed to reach the action stage and congress refused to approve others. Two bills providing for government development of the properties were approved by congress, but were vetoed by the president. In 1933, two unusual circumstances combined to produce a decision on the question: (1) in that year the country was in the depths of a great depression; (2) Pres. Franklin D. Roosevelt strongly favoured a course regarding the Muscle Shoals properties that would put idle men to work and stimulate the economy of the region. These circumstances, placed against the background of a long history of frustration and delay, were sufficient to cause congress to pass a bill introduced jointly by Sen. George W. Norris of Nebraska and Rep. Lister Hill of Alabama to establish the Tennessee Valley authority. The bill was signed by President Roosevelt on May 18, 1933.

The act established the TVA as a regional agency of the U.S. government. In legal form the authority was a public corporation; though it was a governmental agency it had many of the attributes of a private corporation. The TVA established its administrative headquarters at Knoxville, Tenn. It was governed by a board of three directors appointed by the president by and with the advice and consent of the senate. The TVA normally employed somewhat more than 15,000 people. The greatest concentrations of TVA personnel were in Knoxville and Chattanooga, Tenn., and at Muscle Shoals, Ala.

The TVA Program.—Those who drafted the TVA act visualized that agency as something approaching a federal department of natural resources for the Tennessee valley. The program outlined for the new agency did not contemplate for it any new or startling powers; for most of the things it was to do were already being done, in the Tennessee valley and elsewhere, by a host of governmental agencies. What was new about the TVA was that its activities were to be focused on a river drainage basin, and that they were to be brought together in the hands of a single agency. The TVA, then, represented a new approach to the administration of natural resources in the United States.

The 1933 act, recognizing water as a basic natural resource, gave the TVA three broad, basic powers over the Tennessee river: (1) it was directed to take such action as might prove necessary to control floods on the river; (2) it was required to develop navigation on the river; (3) it was instructed to produce and market electric power "so far as [may prove] consistent with such purposes." The constitutionality of the TVA was immediately challenged but was upheld by the supreme court in the case of *Ashwander* v. *TVA* (1936) and in later decisions.

The TVA eventually operated 32 major dams, 9 of them on the main stream of the Tennessee and 23 on its tributaries. Of this total the TVA built 21; it purchased a few from a private power company; and a few, originally built and owned by a private industry, were operated by the authority as part of its river control system. (The TVA also marketed electric power from five dams constructed by the U.S. army corps of engineers on the Cumberland river, but since these are in another drainage basin they are not considered here.) All dams in the system were managed as a unit, with emphasis first of all on flood control. It was, of course, not possible to avoid all floods, for that would have required complete control of every minor tributary stream, which was not economically practicable. Large floods, however, were prevented; no major flood damage occurred on the Tennessee river after the system of dams was completed.

Navigation, the second major responsibility of the TVA, increased markedly after 1933. In that year the river carried little commercial traffic: mainly short-haul bulk items such as sand and gravel. The TVA built navigation locks into each of its nine main dams, deepened and otherwise improved the channel, and encouraged both local governments and private businesses to develop port facilities along the river. As a result, river traffic increased from 33,000,000 ton-miles in 1933 to 600,000,000 in 1950 and more than 2,000,000,000 in the 1960s.

A third major activity named in the 1933 act, the production and sale of electric power, was a highly controversial issue and was resisted by privately owned power companies. At first it was confined to the output of the turbines installed in the big dams. In carrying out this function, the TVA entered into contracts with municipalities and co-operatives to supply wholesale power for distribution and joined with them in purchasing the facilities of privately-owned electric utility companies in the region. These purchases established an integrated power service area in which the TVA was the sole supplier of power. To the end of World War II, and for several years thereafter, hydroelectric plants supplied most of the power. The continuing growth of power requirements for homes, farms and industry, together with a vast expansion in power requirements for defense plants such as those of the Atomic Energy commission, made necessary the extensive construction of steam plants. In 1950, 90% of the power produced by the TVA system came from the big dams, but by 1962 this had shrunk to about 25%. The decrease in the percentage of hydroelectric power was accounted for by the big new steam-electric plants built after 1950. One of these plants was the largest of its kind in the United States and another was the second largest. They consumed a total of about 18,000,000 tons of coal a year.

In 1962 the TVA power system, which included 48 dams and 10 steam plants, possessed a generating capacity of nearly 12,000,000 kw. Power generation during that year was about 65,000,000,000 kw.hr. TVA sold its power in bulk, almost half of it to federal agencies, principally the Atomic Energy commission; of the remaining half, about one-third went to large industries, which were served directly, and about two-thirds to 155 locally-owned municipal and co-operative distribution systems. Electric rates in the Tennessee valley, reflecting the economies resulting from the authority's emphasis on mass markets, were among the lowest in the nation. Critics of TVA asserted that its rates were lower than those of privately owned utility companies because the TVA was not subject to taxation and was not obligated to make a profit on its operations. The TVA power system was financed by appropriations from congress, by corporate earnings, and (after 1960) by electric power bonds. A schedule of payments was eventually worked out by which over a period of years the TVA would refund to the national treasury all federal electric power investment in the valley.

Another major provision of the 1933 act made the TVA responsible for the operation of the nitrate plants at Muscle Shoals. As nitrates can be used in the manufacture of both munitions and fertilizers, the TVA developed the Muscle Shoals plants into a vast laboratory for the development and production of experimental fertilizers. The TVA did not produce fertilizer for commercial sale.

The 1933 act also charged the TVA with responsibility for improving forestry practices in the region. In pursuance of this directive the authority conducted studies of forest conditions and forest practice, produced and distributed hundreds of millions of pine seedlings, and worked closely with governmental forest agencies and with forest industries throughout the basin.

Two auxiliary but nevertheless important program activities are worthy of notice. First, construction of the "Great Lakes of the South" resulted in a vast increase in recreational opportunities in the valley. The TVA played an active part in developing these opportunities and in encouraging the practice of outdoor recreation. As a consequence, such activities as boating, fishing, and camping, known only in rudimentary form in 1933, increased sharply. Recreation, indeed, assumed the proportions of a major industry. A second important auxiliary activity was found in the authority's health program. By co-operating with state and local

agencies and conducting a vigorous mosquito control program on its reservoirs, the TVA participated in the eradication of one of the area's most dread diseases. In a region where malaria had been endemic, no case of malaria traceable to the river was reported after 1948.

Methods of Work.—Two outstanding features characterized TVA's approach to the regional administration of natural resources. First was application of the principle of multiple use throughout the watershed. It is possible to develop a river for the single purpose of preventing flooding, or of promoting navigation, or of producing electric power; but it is also possible to develop it for all three purposes simultaneously, and for other subsidiary purposes as well. In 1933 there were very few multiple-use water installations in the United States, and there was no river basin to which the multiple-use principle was applied in its entirety. In this respect the TVA was a unique institution in U.S. river basin development.

TVA's operating methods were unique in yet a second important respect. Instead of building the agency into a regional supergovernment, the TVA board of directors decided to work, to the maximum degree possible, through the valley's existing governments—the states, counties and municipalities. This made for an intricate network of agreements and contracts among the governments of the basin covering a wide variety of activities, such as health, forestry, agriculture, parks and other recreational facilities, navigation and electric power. The TVA thus became one of the foremost practitioners of co-operative administration to be found in the United States.

The TVA is held in high esteem in other countries. Some 3,000 foreign visitors, many of them high government officials, visit the valley each year to see for themselves how the TVA operates and what it has achieved. It has been the inspiration for similar projects in other lands—in India, for example, with its Damodar valley development, and in Colombia with its Cauca river project. The TVA inspired no similar river valley developments in the United States, though a number of principles that it developed were put into operation in various river basins.

BIBLIOGRAPHY.—Preston J. Hubbard, *Origins of the TVA* (1961); C. Herman Pritchett, *The Tennessee Valley Authority* (1943); David E. Lilienthal, *TVA: Democracy on the March* (1953); Gordon R. Clapp, *The TVA: an Approach to the Development of a Region* (1955); Tennessee Valley Authority, *TVA: the First Twenty Years,* ed. by Roscoe C. Martin (1956). (R. C. MA.)

TENNGLER, ULRICH, German jurist, was born in the middle of the 15th century in Heidenheim near Nördlingen. He served as chief clerk of Nördlingen, 1479–83, and afterward, until his death, was governor of Höchstädt. He died sometime in the first four months of 1511. He is noted chiefly as the author of the *Layenspiegal* (1509 and later editions; after 1516 the work was frequently issued in editions containing also Sebastian Brant's *Klagspiegal*).

The *Layenspiegal* was a systematic encyclopaedia of popular jurisprudence, which had wide influence and for more than half a century was the authority in Germany for the settlement of legal questions.

TENNIEL, SIR JOHN (1820–1914), English illustrator and satirical artist, especially known for his work in *Punch* and his illustrations for *Alice in Wonderland,* was born in London on Feb. 28, 1820. Although he became a probationer, and then a student of the Royal Academy, he soon left the schools, where at that time there was little teaching. In 1836 he sent his first picture to the exhibition of the Society of British Artists and in 1845 contributed a 16-ft. cartoon to the competition of designs for mural decoration of the new Palace of Westminster. For this he received a prize of £100 and a commission to paint a fresco in the Upper Waiting hall (or "Hall of Poets") in the house of lords. He was already known and appreciated as a humorist and his early companionship with Charles Keene fostered and developed his talent for caricature.

In 1850 he was invited by Mark Lemon to fill the position of joint cartoonist (with John Leech) on *Punch,* in succession to Richard Doyle. Gradually he took over altogether the weekly drawing of the political "big cut," which John Leech was happy to resign into his hands in order to restrict himself to his pictures of life and character. Leech's work consisted for the most part of farce; Tenniel's was high comedy and not infrequently tragedy; and the freedom of the humorist heightened the severe beauties of the satirist. When Leech died his friend continued his work alone.

About 2,300 cartoons, innumerable minor drawings, double-page cartoons for *Punch's Almanac* and other special numbers and 250 designs for *Punch's Pocket-books* comprise the sum of Sir John Tenniel's work for the periodical in the service of which he spent the greater portion of his life. His most famous cartoon was probably "Dropping the Pilot" (1890), on the subject of Bismarck's resignation.

In his drawings for *Punch* he raised the political cartoon into a classic composition from which a sense of nobility is never absent. It was in recognition not only of his ability as an artist in black and white but of his service in infusing good humour into one phase of political life that a knighthood was conferred upon him in 1893. He retired in 1901. He illustrated many books and his drawings for *Alice's Adventures in Wonderland* (1866) and *Through the Looking-Glass* (1870) gained him international fame and immortality.

See *Cartoons by Sir John Tenniel* (1901); and M. H. Spielmann, *History of Punch* (1895). (H. Es.)

TENNIS. Two games bearing the name tennis are closely connected by origins but are widely different. One, "lawn tennis," may be regarded as the descendant of the other, which is sometimes termed "royal tennis," "real tennis" or simply "tennis" in England but is called "court tennis" in the U.S. Lawn tennis, generally called tennis in the U.S., is the more popular and will be considered first.

LAWN TENNIS

As an independent member of the family of ball games of which the parent root was probably buried in Egypt or Persia, 500 years before the Christian era, lawn tennis is relatively modern and it had its beginning in England. It is played by men and women, at night as well as day, indoors as well as outdoors, on various surfaces—dirt, grass, clay, concrete, asphalt, wood and composition.

In 1874 an Englishman, Maj. Walter C. Wingfield, devised and applied for a patent for a "new and improved portable court for playing the ancient game of tennis." Wingfield's court was wider at the base lines than at the net and so often referred to as the hourglass court, and the game was introduced as Sphairistiké, shortened to Stické, or as its enemies liked to call it, "Sticky." The court was set up outdoors, on the grass, and under the new name tennis-on-the-lawn or lawn tennis its popularity quickly grew. About this time the All England Croquet club, located in Wimbledon, a suburb of London, discovered that its exchequer was empty, or almost so, and added the term "Lawn Tennis" to its title and several grass courts to its facilities. The first championship was held on these Wimbledon courts in June 1877, on a rectangular surface measuring 78 ft. long and 27 ft. wide, with a net 3 ft. 3 in. high at the centre. There were 22 competitors, of whom the majority were either court tennis or rackets players. The winner, Spencer W. Gore, was a former rackets star and a native of London.

It is interesting to note that the dimensions of the lawn tennis singles court remained unchanged after that time. In 1878 the height of the net was set at 4 ft. 9 in. at the posts, and 3 ft. at the centre. In 1880 the net was lowered to 4 ft. at the posts, and in 1884 to 3 ft. 6 in.

From the start until 1907, the championships at Wimbledon were won by natives of the British Isles. The early Wimbledon, however, with its bleacherless courts and general air of calm, bore little resemblance to its successors. Yet from the start the All England club had a decisive influence on the game. The Lawn Tennis association (English) was founded in 1886, the International Lawn Tennis federation not until 1912. In the interim, the game had spread to other continents.

Lawn tennis was said to have been introduced into the United

States by Miss Mary E. Outerbridge, who spent the winter of 1874 in Bermuda and there saw the game played by British officers of the garrison. She returned with a lawn tennis net, some rackets and balls, obtained from the regimental stores through the courtesy of the colonel. One of her brothers, A. Emilius Outerbridge, helped her lay out a court on the grounds of the Staten Island Cricket and Baseball club. James Dwight and F. R. Sears, Jr., of Boston, Mass., obtained a tennis set from England and laid out a court at Nahant, a seaside resort in Boston harbour, a little later. The first tournament in the United States, a handicap event, was played at Nahant in Aug. 1876. The first official championships of the United States, playing under English rules and with English balls, was held at the Casino, Newport, R.I., on Aug. 31, 1881. There were 25 competitors entered in the singles and 13 pairs in the doubles. The singles was won by Richard D. Sears of Boston, who held the title every year until 1888. In 1881, the United States National Lawn Tennis association, with Gen. R. S. Oliver of the Albany Tennis club as president, was formed.

This period might be called the age of innocence of lawn tennis. In both England and the United States, the players were amateurs and sportsmen, the tournaments few in number, and watched by small galleries. In the United States, from the time of Sears until 1900, most U.S. champions were volleyers, Oliver S. Campbell and Robert D. Wrenn, who held between them seven singles titles in the 1880s and 1890s, especially so. During this period the game was played chiefly along the eastern seaboard. In 1900 a cup was donated by Dwight F. Davis, then an undergraduate at Harvard, for international competition. (From the start, all Davis Cup competitions consisted of four singles and one doubles match.) The first contest was played in Longwood, near Boston, on Aug. 8, 9 and 10, 1900, and was won by the United States, three matches to none. (One was not played, and one left unfinished.)

In 1907, Norman E. Brookes of Australia became the first overseas player to win the Wimbledon men's title. Meanwhile in the United States there was a swing away from the net game. Both William A. Larned, who held the U.S. title from 1907 to 1912, and William T. Tilden, who dominated world tennis from 1920 to 1925, were able to play the net and play it well. Yet both won the majority of their titles from the base line. In 1912 a new influence was felt in the game. For the first time in U.S. tennis the champion did not come from the eastern seaboard. That year the title was won by Maurice E. McLoughlin, a Californian trained on courts of asphalt. He introduced the cannon ball service, a weapon used by every U.S. star since then.

By this time tennis may be said to have come of age. In 1915 the national championships of the United States were removed from the tranquillity of the Casino at Newport, R.I., to the urban setting of the West Side Tennis club at Forest Hills, Long Island. Attracted by the spectacular playing of McLoughlin, crowds turned out for the matches. At the same period, interest grew in the Davis Cup matches. In 1903, when it had been captured by the Doherty brothers, R. F. and H. L., for Great Britain at the Longwood Cricket club, in Brookline, Mass., only two nations were represented, and hardly 2,000 spectators saw the contests each afternoon. In 1914, seven nations from three continents competed, and the challenge round played at Forest Hills just after World War I broke out was witnessed by 10,000 people. Australia with the left-handed veteran Brookes, and Capt. Anthony Wilding (killed in action in Flanders in 1915) met McLoughlin and R. Norris Williams, an undergraduate at Harvard. There have perhaps been greater matches in tennis history; but few so dramatic, so grim or so spectacular. The speed of shots on both sides was tremendous. The rallies were few. It was a different style from the tennis played by the Dohertys which was based on accuracy rather than sheer speed. Tennis had changed as the world was changing.

Tilden won the title at Wimbledon in 1920, the first but not the last male American to do so. He repeated the next year, and in 1923 the first Californian, William M. Johnston, won the men's singles. In 1920 Tilden and Johnston beat the Australians at Auckland, N.Z., taking the Davis Cup back to the United States. With Williams and Vincent Richards they formed the best balanced team ever to appear on a court in an international match, and through their efforts Australian, Japanese and French challengers for the trophy were regularly defeated until 1927. The U.S. influence on sport at this period was visible in many ways. Money was now an important factor in competitive tennis. Huge crowds attended national championships and Davis Cup matches in every land. As an example, the old grounds seating several thousand spectators under wooden roofs at the Worple road, Wimbledon, were abandoned in 1921 for a new, modern, concrete tennis stadium, one mile away, seating 15,000. This stadium which was opened by King George V on June 22, 1922, attracted players from Spain, France, the United States, Italy, Ireland, Scotland, Sweden, Holland, South Africa, Belgium, Rumania, Denmark, Greece, New Zealand, Australia and England that summer. It cost £140,000, and was probably paid for in a few years. Likewise in 1923 the wooden stands of the West Side Tennis club at Forest Hills were replaced by a concrete stadium seating more than 13,000. The first Wightman Cup matches inaugurated the stadium.

By this time young Frenchmen were advancing in the technique of the game, and a French team composed of J. R. Lacoste, J. Borotra, H. Cochet and J. Brugnon (the "Four Musketeers") reached the challenge round of the Davis Cup for the first time in 1925. Two years later they were the survivors of 25 nations, and crossing the ocean found themselves for the third time challenging the holders, the United States, at Philadelphia, Pa. This time they discovered that age had stemmed the power of the two great U.S. stars, Tilden and Johnston. They carried the cup back to France. This was the first of six years of French successes in tennis.

In 1933, a total of 32 different countries challenged France for the Davis Cup. The trophy was finally won by Great Britain, who defeated the French in the challenge round at Paris. For the first time in 20 years this international sporting cup was returned to the land which gave the game birth.

With its young stars Fred Perry and "Bunny" Austin, Great Britain was in the driver's seat until the arrival of another star from California, J. Donald Budge. Perry had won three successive championships at Wimbledon (1934–36) and three times at Forest Hills (1933, 1934 and 1936).

Budge not only succeeded Perry as Wimbledon and U.S. champion but set a new record when in one year (1938) he won the four major singles championships of the world—Australia, France, Great Britain and the United States. One of Budge's most noted victories was his defeat of Baron Gottfried von Cramm, German star, in the deciding match of the Davis Cup Interzone final at Wimbledon in 1937. Tilden declared afterward that it was one of the finest matches he had even seen. Von Cramm won the first two sets and Budge evened the score. Then the German took a commanding 4–1 lead in the final set, but Budge again pulled even and finally, after having had five match points in the 14th game, he won it for the set and match. Budge again defeated Von Cramm in the 1937 final at Forest Hills to win his first U.S. championship.

The loss of Budge to the professional game in 1938 after he had repeated his Wimbledon, Forest Hills and Davis Cup victories proved to be a severe blow to U.S. Davis Cup play. In 1939 Robert L. Riggs, who had won at Wimbledon and was to become U.S. champion about ten days later, and Frank Parker, who had two years of challenge round experience, let the cup slip from their grasp after they had led Australia 2–0 in the challenge round tie at Merion Cricket club, Haverford, Pa.

So again, as in 1914, Australia captured the Davis Cup as World War II halted international competition, this time for six years. Bombs damaged the Wimbledon grounds and other clubs throughout Europe. In the United States, the U.S. Lawn Tennis association helped keep tennis alive during the war. The development of junior tennis, a major activity of the USLTA, was emphasized and a nationwide program provided opportunities for thousands of boys and girls under 18 years of age to learn the game.

The USLTA championships were held during the war period. Most of the leading players were in the armed forces but a few on furlough or stationed near Forest Hills competed with the younger players. The wartime champions included: (in U.S. navy) Lt.

Donald McNeill; Seaman Robert L. Riggs; Lt. Frederick R. Schroeder, Jr.; and Lt. Joseph R. Hunt (killed in a plane accident); and Sgt. Frank A. Parker (U.S. army).

After the war interest in the game revived quickly throughout the world. In 1946, 20 nations challenged for the Davis Cup which had been held since 1939 without competition by Australia. The Americans won 5–0 in Melbourne and retained the cup for the next three years at Forest Hills. John A. ("Jack") Kramer, Frederick R. ("Ted") Schroeder, Frank A. Parker and Richard A. ("Pancho") Gonzalez were the outstanding Americans during these four years and none of them lost a singles match in these challenge rounds.

The public, so long starved of the glamour of top tennis entertainment, flocked to the major championships. At Wimbledon the first winner was the tall Frenchman Yvon Petra, but the U.S. provided the next five winners, all of whom were players of great character and individuality of style. Kramer, later to amass the greatest fortune ever made from the game as a professional player and promoter, was the "power-player" par excellence; the tall Robert Falkenburg relied on service speed and volleying the return; Schroeder based his game on a superb net attack and an invulnerable backhand, coupled with remarkable stamina; J. Edward ("Budge") Patty possessed an immaculate touch, while Richard Savitt put his trust in service power and heavily hit ground strokes.

Another great American of this era was Gonzalez who won his national title at Forest Hills in 1948 and 1949. He turned professional at the then unusually young age of 21 and maturing in his new status became undoubtedly the world's greatest player of the 1950s, and, in the view of many, the greatest player yet seen.

In 1950 the balance of power swung back to Australia. The fine athletic physique and speed play of Frank A. Sedgman and his doubles partner, Kenneth B. McGregor, won back the Davis Cup for Australia and this partnership retained the trophy in 1951 and 1952, in which latter year Sedgman had already won at Wimbledon. In 1951 this partnership won all four of the major doubles championships—an achievement without precedent.

Despite the loss of two such fine players to the professional ranks in Jan. 1953, the Australians were able to replace them by the 18-year-olds Lewis A. Hoad and Kenneth R. Rosewall from Sydney. Much of the credit for Australia's extraordinary feat of continuing to hold the Davis Cup in these circumstances went to the shrewd captaincy of Harry Hopman.

The more mature and experienced Americans E. Victor Seixas and M. Anthony Trabert, who had won Wimbledon and the U.S. title respectively in 1953, recaptured the Davis Cup in 1954 at Sydney in front of a world record crowd of 25,600 on the opening day. But only eight months later the Australians, tuned to concert pitch by Hopman, beat the same American team at Forest Hills by 5–0. Since then the Australian domination was consistent except for the successful American raid at Brisbane in Dec. 1958 when Alejandro Olmedo, a Peruvian by birth, succeeded in beating the Australians Ashley J. Cooper and Malcolm J. Anderson, and, in partnership with Hamilton F. Richardson, Anderson and Neale A. Fraser in the doubles.

The U.S.–Australian monopoly of the top honours was first breached during the 1950s by Jaroslav Drobny, exiled from Czechoslovakia, whose amazing repertoire of strokes, including his remarkable left-handed flexibility overhead, won him the French title in 1951 and 1952 and the Wimbledon crown in 1954 in a memorable final of a record number of 58 games against Rosewall.

After Drobny followed other Europeans in Sven Davidson (Sweden), Kurt Neilsen (Denmark), Nicola Pietrangeli (Italy) and Manuel Santana (Spain), all of whom appeared as winners or finalists in the major championship finals.

In 1955 Trabert won three of the major championships and then turned professional and the following year Hoad repeated this feat but was defeated in the U.S. final by his compatriot, Rosewall. After their successful Davis Cup defense in 1956 Rosewall turned professional and Hoad followed him into the paid ranks immediately after becoming the first postwar player to retain a Wimbledon men's singles title.

This left another young Australian, Ashley Cooper, as the world's leading amateur player and in 1958 he won the Australian, Wimbledon and U.S. championships, leaving only the French title to his elder and left-handed compatriot, Mervyn Rose.

For the next two seasons major honours were shared by Pietrangeli, Olmedo and the left-handed Neale Fraser. Two new stars emerged in 1960 from Queensland, Rodney G. Laver and Roy Emerson. Laver proved the superior of the two because his game was less brittle than Emerson's. In 1962 Laver achieved the elusive "grand slam" of all four major championships which had previously been captured only by Donald Budge in 1938, as already noted.

The inevitable procedure followed and Laver turned professional after Australia's successful defense of the Davis Cup in 1962 to pit his maturing skill against Rosewall who had inherited the mantle of Gonzalez as the world's leading exponent. In 1963 Charles R. (Chuck) McKinley captured the Wimbledon title for the United States for the first time in eight years and, teamed with Dennis Ralston, went on to take the Davis Cup from the Australians, 3–2, at Adelaide in Dec. 1963. Australia regained the cup in Sept. 1964, at Cleveland, O., with Emerson and Fred Stolle over McKinley and Ralston, 3–2.

A feature of the end of the 1950s and the early 1960s was the advance of the Latin American countries. The Peruvian Alex Olmedo was the first player from his continent to win a major singles championship when he won the Australian title in 1959 and followed this by capturing the Wimbledon crown. He was followed in 1963 by Rafael Osuna of Mexico, who won the U.S. championship and had been a member of his country's Davis Cup team, which upset all tradition by beating the United States in 1962, and reached the challenge round against Australia.

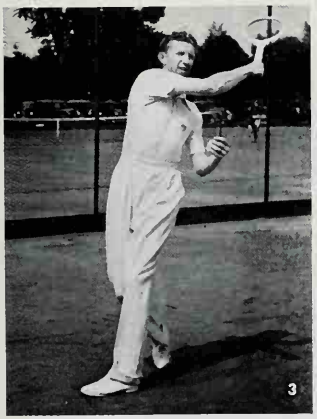
Interest in international competition continued high, with more than 40 nations annually challenging for the Davis Cup.

Women Players.—Tennis was played by women from the start; but their entry into competitive tournaments was delayed, and it was not until 1884, seven years after the championships began, that Maud Watson became the first woman's champion at Wimbledon by winning from an entry of 13 players. In the United States a tournament for women was held at the Philadelphia Cricket club, and was won by Ellen Hansell in 1887. This tournament was held again in 1888 and it became the first official championship for women in 1889. It was won by Bertha Townsend. Elisabeth Moore, later singles champion, described the costumes of the period in this way: "The girls were wearing lawn dresses with leg of mutton sleeves, sailor hats, and ornamental tennis shoes. My dresses, which were of white lawn, swept just clear of the grass on ordinary shots; but brushed it when I stooped."

In those days the game was played chiefly from the base line. In England the standard of play was and remained for some time much higher than in the United States, yet women still served underhand. In 1904 a new star appeared, May Sutton, a girl of 16 and the first of a long line of Californian champions who revolutionized U.S. and world tennis. She was gifted with a strong constitution and a heavily topped forehand drive which enabled her to win the American title that year, and also to defeat Miss D. K. Douglass (afterward Mrs. Lambert Chambers) at Wimbledon in 1905, thus becoming the first American ever to capture an English title. She again defeated her rival to win the championships in 1907, and 22 years later she returned to Wimbledon and reached the quarter finals of the singles in a strong field.

An important influence on the game for women was Suzanne Lenglen, a French girl who first appeared at Wimbledon in 1919. Born at Compiègne, France, in 1899, she won her first championship at St. Cloud in 1914. Between 1919 and 1926 she dominated the game as Tilden dominated play for men, winning at Wimbledon six times.

The next player who pointed the path for women was Helen Wills, later Helen Wills Moody Roark. She played Mlle. Lenglen only once, at Cannes in 1926, when she was a young girl, losing 6–3, 8–6. Like her rival she had every stroke, supported by a more powerful forehand drive, the weapon of every champion.



Three photographs above show Don Budge, U.S., winner of four major singles championships in 1938, playing a forehand drive. (1) Gathering speed in forward swing, racket's axis almost parallels the court; (2) it meets ball in front of the player's right hip, and (3) then follows through, imparting a top spin to the ball

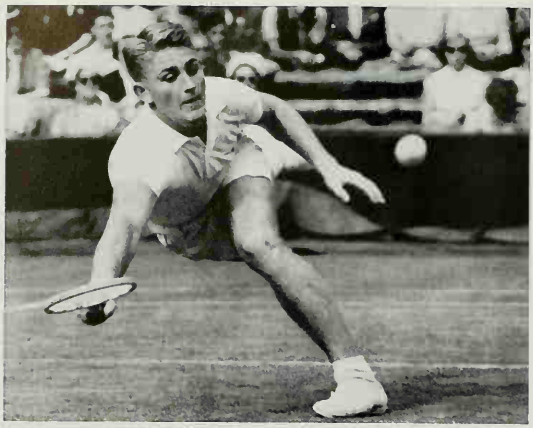
William Tilden, U.S., who dominated world tennis from 1920 to 1925, in play at the 1930 U.S. singles matches at Forest Hills

Helen Wills Moody Roark, U.S., one of the more noteworthy women players in tennis history, smashes the ball as she defeats Helen Jacobs, U.S., in 1938 for the All-England women's singles title at Wimbledon

TENNIS STROKES AND OUTSTANDING PLAYERS

PHOTOGRAPHS, (TOP LEFT, CENTRE LEFT, BOTTOM LEFT, BOTTOM RIGHT) UPI COMPIX, (TOP RIGHT) WIDE WORLD

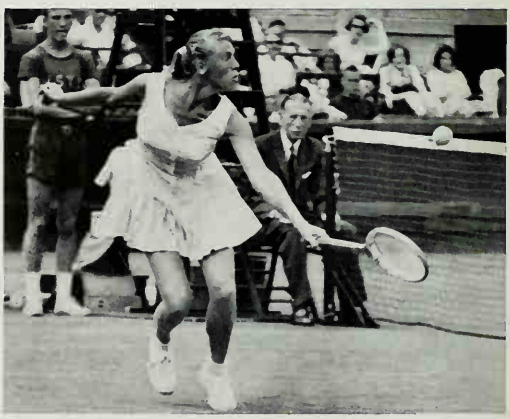
TENNIS



Lewis Hoad, Australia, goes low to make a forehand return during play in the 1953 men's singles at Wimbledon

Jack Kramer, U.S., serves, his racket meeting the ball high above his head

Tony Trabert, U.S., jumps to make an overhead smash as his partner, Victor Seixas, ducks out of the way in the 1954 men's doubles at Wimbledon



Ann Haydon, England, playing backhand, strokes the ball just over the net in the 1961 women's singles at Forest Hills

TENNIS STROKES AND OUTSTANDING PLAYERS

PHOTOGRAPHS, (TOP LEFT, BOTTOM LEFT) KEYSTONE, (TOP RIGHT) "LIFE," © 1949, TIME, INC., (BOTTOM RIGHT) WIDE WORLD

She had a matchless record in competitive tennis, winning the first of her seven American titles in 1923 and the first of her eight Wimbledon titles four years later.

Another great woman player, also an American and a Californian, was Alice Marble. Despite illnesses which would have discouraged most players, she persevered and ended as Wimbledon and U.S. titleholder in 1939. She was the first feminine champion to develop a really hard service like that of a man, and possessed an all-round game to which she added unusual severity at the net and overhead.

Miss Marble won her fourth and last U.S. championship in 1940, defeating for the third time the resolute Helen Jacobs, who had succeeded Mrs. Moody as U.S. titleholder in 1932 and as Wimbledon champion four years later. The next ten years failed to produce any great woman champion. During the war years Mrs. Sarah Palfrey Cooke, an easterner, and Pauline Betz, a Californian, led the field, Mrs. Cooke winning her second championship in 1945 and Miss Betz her fourth in 1946, after which they turned professional and Margaret Osborne duPont, Louise Brough, Doris Hart and Shirley Fry, sometimes referred to as "the big four," dominated the game at Forest Hills, Wimbledon and in other countries wherever they competed, including France, England, South Africa and Australia. Then in 1951 Maureen ("Little Mo") Connolly from California, like Miss Sutton in 1904, took complete command, winning the first of three successive U.S. championships at Forest Hills before she had reached her 17th birthday. Miss Connolly was the second youngest player to capture the title. Miss Sutton was only a few days younger when she won the championship in 1904 and Miss Wills was almost a year older when she captured her first U.S. title in 1923.

On her first attempt at Wimbledon in 1952 Miss Connolly won the crown and successfully defended it in 1953 and 1954. In 1953 she also won the Australian, French and United States titles and so this mistress of ground stroke play became the first woman to hold the four major championships simultaneously. She sustained a leg injury while horseback riding which prevented her from defending her United States title in 1954. Later Miss Connolly, perhaps the only player of the postwar era to rank with Suzanne Lenglen and Helen Wills Moody Roark, became a professional.

Doris Hart won the U.S. championship in 1954 and retained it in the following year when Louise Brough had recaptured the Wimbledon title after an interval of five years. In 1956 Althea Gibson of New York became the first Negro player to win a major title when she won the premier hard court title in the world—the French championship—at Paris. She was the dominating player for the next two seasons, notably by reason of her service power, winning both the Wimbledon and U.S. championships. The U.S. strength in depth weakened at this time and the British trio of Angela Mortimer (1955), Shirley Bloomer (1957) and Christine Truman (1959) won the French championship, as did the Hungarian Suzy Kormoczy in 1958.

The first player to break the 13-year-old postwar U.S. monopoly of the Wimbledon and U.S. championship titles was Maria Esther Bueno, a superb stylist from Brazil. She won both these championships in 1959 with artistic ease and retained her Wimbledon title in 1960. She, however, lost her American title to her doubles partner, Darlene R. Hard of California. Miss Hard had already won the French championship that year, and went on to retain her U.S. title in 1961 following the severe illness of Miss Bueno.

In the season of 1961 two British players won major championships: Ann Haydon took the French title and Angela Mortimer became the first British singles winner at Wimbledon for 24 years.

While the star of Australia's men players waned, the prowess of their women players waxed and Margaret Smith, a player of rare physique and power, won the Australian championship in 1960 and retained it for the next three years. In 1962 Miss Smith took the French and United States championships but was thwarted at Wimbledon in the first round, the title eventually going to Karen Susman (U.S.). In 1963 Miss Smith lost her French title to her compatriot, Lesley Turner; became the first Australian woman ever to win Wimbledon; but lost her U.S. championship to Miss Bueno, who had recovered from the after effects of her severe attack of jaundice.

Wightman Cup.—In 1923 a cup was donated by Mrs. George Wightman, the former Hazel Hotchkiss of California and herself national singles champion of the United States in 1909, 1910 and 1911 and winner of some 47 other U.S. titles, for competition between teams of women players from the United States and England. The Wightman Cup was first contested at Forest Hills, Long Island, on Aug. 11 and 13, 1923. The annual contest is held alternately in the two countries, with five singles and two doubles matches. The first contest was won by the United States. By 1930 each side had won on four occasions: but then the United States' team won for the next nine years until the competition was abandoned in 1940 owing to the war. After the war the United States' dominance was even more pronounced, producing a string of 12 more consecutive victories, and never losing more than two of the seven rubbers. The British girls, with Christine Truman contributing to three wins, succeeded for the first time in 28 years at Wimbledon in 1958 by the minimum margin of 4–3 and repeated this performance in 1960. In 1961 the United States recaptured the Cup and held it through the early 1960s.

The Federation Cup.—A new international competition for the women's team tennis championship of the world was inaugurated with the Federation Cup by the ILTF in honour of its 50th anniversary in 1962. The first competition for the cup, held at the Queen's club, London, in 1963 was won by the U.S. team when Billie Jean Moffitt and Darlene Hard beat the top-seeded Australian team of Margaret Smith and Lesley Turner in the doubles finals. (J. R. Tu.; E. S. Br.; R. McW.)

STANDARD GRIPS AND STROKES

Many outstanding players and coaches have written books and manuals on how to play tennis, a few of which are listed in the *Bibliography*, below. The following descriptions of the standard grips and strokes originally were prepared by Tilden, one of the all-time great players, for the United States Lawn Tennis association's *Official Tennis Guide*, 1942 (*Official Tennis Guide with the Official Rules, 1942*, copyright 1942 by A. S. Barnes and Co.). Tilden's instructions have been revised only to comply with subsequent changes in the rules (see *Rules of Lawn Tennis*, below). (X.)

Forehand Ground Stroke Grip.—Hold the racket as if it were standing on the edge of the frame, the handle pointing toward you, the short strings perpendicular to the ground, the long strings parallel to it. Then "shake hands" with it—literally. The handle settles comfortably into your palm. Close your fingers tightly around the handle. The racket is then, or should be, part of the straight line of your arm. There should be no angle at your wrist. The hand rests along the outside, or right side, of the handle, the thumb and fingers wrapped around the handle, naturally and comfortably. The racket thus becomes an extension of the arm that can be moved as easily as if it belonged there.

Backhand Ground Stroke Grip.—Obviously, the forehand grip, as described in the foregoing paragraph, is impossible for a backhand shot, since it would be swung against the back of the hand itself. Therefore, a change is necessary. Move the hand one-quarter of a circle on the handle, backward, so that the hand rests directly on top of the handle, the knuckles of the hand pointing to the sky. Grip the handle tightly. If it is more comfortable to run the thumb up the handle, do so. If not, wrap it around as in the forehand grip. The shot off the backhand travels directly across the wrist with this grip. Swing freely from the shoulder with both these grips and keep the wrist stiff and locked on all drives. The grip is about the same for a volley.

Service Grip.—The grip for service is just midway between the forehand and the backhand grip. Take your forehand grip and then turn the hand half way back to the backhand position.

On all grips, hold the racket tight at the moment of impact with the ball. Loosen your grip between shots so as not to cramp the hand from too great a strain.

Forehand Drive.—The forehand drive is one continuous swing, which is considered in three sections:

The speed of the shot is decided by the length of the swing behind the body. The farther back you start your swing the faster the racket will move forward, and the speedier your shot will be.

The direction of your stroke, whether it is hit straight or cross court, is determined by that portion of the swing in front; that is, alongside your body. In other words, as you face off court, it is the space within the lines from your hips. If you meet the ball back near your right hip, the shot goes toward the right; if you meet it directly in front of the belt buckle, it goes on a straight line; and if you meet it up in front of your left hip, the shot goes cross court to your left.

When awaiting the ball, face your opponent across the net. As soon as you see the ball coming toward you, decide which side, fore or backhand, you will play it. Then change your body position accordingly so that you face off the side line, while your shoulders are along the line of the shot you are going to make. As the ball reaches you, drop your weight to your back foot and commence your swing well behind your body, meeting the ball with the racket in its forward flight just in front of your body and follow through to the end of your swing.

This brings us to the third portion of the swing, the portion beyond or nearer the net than the body, which is called the follow-through. This determines the spin or twist on the shot and is the factor that holds the ball in court. For the drive the racket meets the ball flat and passes up and over the top of the ball, thus spinning the ball away from you and downward. It is the same spin as a pitcher's drop in baseball. The racket passing under the ball, making it float or drag, is the "chop" or "slice" stroke (*see* below).

The swing, as described, is in three parts, but is made in one free movement that carries the whole body along with it. It is as essential in tennis to follow through from the beginning of your swing to the extreme end of your reach as it is in golf or baseball. It gives direction and control of length.

There is a certain height at which to hit a tennis ball correctly. There are other heights at which one should never hit it unless caught out of position. The drive should be hit as a height never lower than your knees or higher than your shoulders. If the ball must be met at either of these heights, do not drive. Merely defend it and put it back and wait for another chance to drive. The height at which a tennis ball may be driven hardest is about waist high and at a distance about three feet from your body.

In making the drive, first be certain your form is correct. See to it that your feet are correctly placed so that you are sideways to the net. Next make sure your racket face meets the ball flat and that your racket passes up, over and slightly outside the ball, thus pulling it toward you. Do not try to hit hard until you have learned to hit correctly. Speed will come later, but correct form can be learned only before you have formed bad habits in stroking that are almost impossible to break.

Backhand Drive.—The laws of footwork are exactly the reverse of those applying to the forehand drive. The back foot is now the left, while the front foot is the right. The swing is from well behind the body, as in the forehand shot. The grip, as previously explained, is changed one-fourth on the handle, so that the hand is on top of the handle, while the shot travels across the wrist, which is locked and stiff. The racket head should be slightly in advance of the hand when it meets the ball, and the face of the racket is flat. The ball travels across the short strings. There is less top spin to this stroke than the forehand, because the follow-through is less. Use your left hand as a balance, well behind your body, and be sure to keep your body well away from the ball, so as not to cramp your swing. Bend to meet the ball with the racket head slightly on the side away from you and pull it toward you cross court.

Service.—Service is the "opening gun" in tennis. It is putting the ball in play, and every man should learn a service that will be a distinct advantage. Service has made many players famous. Maurice E. McLoughlin, the "California Comet," was world renowned for his service. Robert Lindley Murray, Elia Fottrell, Harold Throckmorton, S. Howard Voshell and many other players were noted for their services. All these men hit their deliveries hard, but they never sacrificed accuracy to mere speed. They were always careful to put the ball in court. They served very few double faults.

Foot faults should be avoided. The easiest way to serve legally is to remember that one foot must be on the ground behind the line and the other foot must not touch inside the court until your racket has struck the ball. Stand at least six inches behind the base line and you will not foot fault.

Slice Service.—This is the most common form of service. It requires the least effort and has the greatest chance of going in court.

To hit this service, stand behind the base line with your feet at an angle of about 45° to the line. Drop your weight back on your right foot. Toss the ball well above your head and slightly to the right in advance of it. Swing your racket up and forward with one rhythmic movement, at the same time shifting the weight forward toward the left foot. Meet the ball as high above your head as you can comfortably reach. The racket face is outside, or on the right side of the ball, and slightly over it. It travels from right to left. Swing directly through this, imparting a twist to the ball to spin it from left to right. This tends to hold the ball in court.

American Twist Service.—The stand and grip for this service are the same as for the slice. The ball is thrown slightly to the left of the head and almost directly above it. If anything, it may be slightly behind. The body is bent backward and the ball hit at a point more to the left than in the slice. The racket travels up and over the ball which is below and inside the strings. The ball is really between your face, turned upward to watch it, and the face of your racket which is travelling from left to right away from your body. This service carries extreme twist and a high bound, but with not so much speed as a slice or flat service. There is more wrist turn to this service than in the slice. The body movement causes a distinct pull on the stomach muscles and those along the right side. The slice service is hit with the shoulder and arm muscles.

Reverse Twist.—This delivery is a totally different style of service and is not one for regular use. The server, to deliver this service, should stand facing the net with both toes on a line with the base line. The ball is thrown about shoulder high. The ball is caught and carried on the strings for a short space. The racket passes up and over the ball, which is outside the racket face from the server. This service carries extreme twist but little speed or control. The hitting plane is about shoulder high instead of over the head. The weight control is faulty. In fact, it is not a service for a beginner to learn.

Volley: Forehand and Backhand.—The volley is the shot made by taking the ball in the air before it has struck the ground. It is "hitting it on the fly," as the common expression goes. The laws of footwork and body position are the same for volleys as for ground strokes, the body always sideways to the net and weight moving forward.

The ground stroke or drive is made from the back of the court. The volley is made from the front of the court or net position. Volleys should not be made from behind the service line if it is possible to drive.

The drive has a distinct follow-through and a long swing. The volley has little or no follow-through and a very short swing. All volleys should be blocked or undercut and not stroked or topped.

The shot is made with a short swing from behind the body. The racket face meets the ball flat or, if the shot must be made from below the top of the net and lifted, slightly under the ball. The wrist is stiff and locked. The angle of the shot is determined by the angle made by the flight of the ball coming to you and the racket face. The racket moves forward, meets the ball and stops completely. It does not follow through. It blocks the ball, which rebounds by its own speed plus the body weight behind the racket.

Smash.—The overhead smash is the stroke killing the ball in the air when it is lobbed over your head. This shot should be modelled as nearly after your service as you can do it, except that it is hit with less twist and more speed. (W. T. T.)

RULES OF LAWN TENNIS

Tennis is one of the few games that are played around the world under the same rules. The official rules are promulgated by the ILTF and are widely available in the publications of the various affiliated national tennis associations and other organizations interested in promoting the sport.

The Singles Game

The Court, Net and Fixtures.—The size, shape and markings on the court as determined by the rules are shown in fig. 1. The net is 3 ft. high at the centre, where it is held down taut by a strap not more than 2 in. wide, and the cord or wire (maximum diameter one-third of an inch) from which it is suspended is attached to or passes over the tops of the posts, which are 3 ft. 6 in. high.

A band covers the cord or wire for not less than 2 in. nor more than $2\frac{1}{2}$ in. in depth on each side. The spaces behind the base lines and at the sides indicated on the diagram are required in Davis Cup and other official ILTF championships. The centre line is 2 in. wide; all other lines are not less than 1 in. nor more than 2 in. wide, except the base line, which may be 4 in. wide. The centre mark is 4 in. long and 2 in. wide.

The permanent fixtures of the court include the posts and the net with its cord or wire, strap and band, and also any back or side stops, the umpire, linesmen and spectators and their chairs or stands, and all other objects around and above the court.

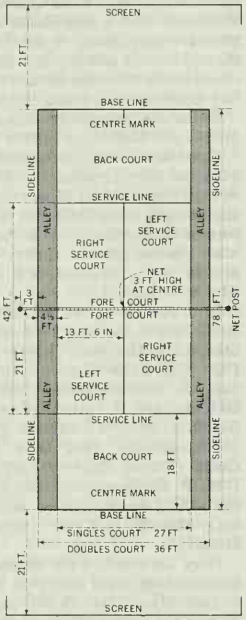


FIG. 1.—PLAYING AREA FOR LAWN TENNIS. THE ALLEYS ARE USED ONLY IN DOUBLES PLAY

The Ball.—The tennis ball has a uniform outer surface, any seams being stitchless. It must be more than $2\frac{1}{2}$ in. and less than $2\frac{5}{8}$ in. in diameter, and more than 2 oz. and less than $2\frac{1}{16}$ oz. in weight.

Play.—The choice of sides and the right to serve or receive in the first game of a series is decided by toss (spinning or throwing up a racket). The winner may choose, or require his opponent to choose: (1) the right to be server (serve first) or receiver, in which case the other player chooses the side; or (2) the side, in which case the other player chooses the right to be server or receiver.

The service is delivered alternately from behind the right and left courts, beginning with the right. Immediately before commencing to serve, the server stands with both feet at rest behind the base line, and within imaginary continuations of the centre mark and the sideline. The receiver may stand wherever he pleases on his own side of the net and the server must not serve until the receiver is ready. The server then tosses the ball into the air and before it hits the ground strikes it with his racket. The delivery is completed (the ball is in play) at the moment of impact. There is no restriction regarding the kind of service that may be used.

The server throughout the delivery may not materially change his position (location) by walking or running, and must not touch inside the court with either foot until his racket has struck the ball (see *Standard Grips and Strokes: Service*, above). The ball served must pass over the net and hit the ground within or upon any line bounding the diagonally opposite service court. In the absence of a linesman and umpire it is customary for the receiver to determine whether the service is good. The server continues to serve until he wins or loses one game, when the service passes to his opponent. The players change sides at the end of the first, third and every subsequent alternate game.

If the service lands outside the proper service court, it is a fault. It also is a fault if the player breaks any of the rules described above, if he misses the ball when he swings at it, or if the ball touches a permanent fixture (other than the net, strap or band) before it lands in the proper service court.

After one fault, another ball is served from behind the same half of the court. If this, too, is a fault, the server is charged with a double fault and loses the point.

A let is called when a ball breaks or is obstructed as specified in the rules, or to provide for an interruption of play. When a let is called for a service, that one service is replayed; when called under any other circumstance, the point is replayed. The service is a let if the ball touches the net, strap or band but either falls within the proper court or touches the receiver or anything he wears or carries, or if the receiver is not ready.

The receiver must return a good service by hitting the ball on the first bounce so that it goes over the net to fall anywhere within his opponent's side of the court, after which the players alternately hit the ball back and forth across the net until one of them fails to make a good return and a point is scored. After the service either player may volley the ball, that is, hit it before it touches the ground.

When the ball is in play, a player fails to make a good return and loses a point if the ball bounces on the ground twice before he returns it; if his return lands outside any of the lines bounding his opponent's court; if he touches the ball with his racket more than once in making a return or if the ball touches him or anything he wears or carries, except his racket; if he or his racket or anything he wears or carries touches the net, posts, cord or wire, strap or band, or the ground within his opponent's court; or if he volleys the ball before it has passed the net.

Scoring.—The receiver scores a point if the server delivers a double fault. After the service, either player or side scores a point if the opposition fails to make a good return of the ball. A game is won by one side scoring at least four points; the first point is 15, the second is 30, the third is 40 and the fourth is game point.

A game must be won by at least two points, so that after a tie score at three points, or 40–40 (deuce), one side must score two consecutive points to win. The first point scored after deuce is called vantage, advantage or simply "ad." If the other side wins the next point the score returns to deuce. A score of zero is called love; *e.g.*, 30–love is 30–0.

A set consists of one side winning at least six games. A tie score at five games is a deuce set, after which one side must win two consecutive games to win the set.

In major tournament play (*i.e.*, championships and Davis Cup tournaments), a match consists of three out of five sets for men; two out of three for women.

The Doubles Game

The Court.—As shown in fig. 1, the doubles court is 36 ft. wide, $4\frac{1}{2}$ ft. wider on each side than the singles court. Those portions of the singles side lines that lie between the service lines are called the service side lines (those portions of the singles side lines between service lines and the base lines may be omitted if desired). The server may stand anywhere behind the base line between the centre mark and the doubles side line.

Order of Serving and Receiving.—The pair who will serve in the first game of each set decide which partner will do so, and the opposing pair decide similarly for the second game. The partner of the player who served in the first game serves in the third; and the partner of the player who served in the second game serves in the fourth. The players continue throughout the set to serve in that order. Similarly, each pair decide which shall receive the first service by the opponents, the other receiving the second, and they continue in that order.

Play.—The rules applying to the play of the singles game ap-

ply also to doubles. A fault is incurred as in singles, and also if the ball served touches the server's partner or anything he wears or carries; but, if the service is not a let, and touches the receiver's partner or anything he wears or carries before it hits the ground, the server wins the point.

After the service, either partner of the opposing pairs may return any shot, but if his partner should touch the ball with his racket their opponents win the point. (E. S. Br.)

ROYAL OR COURT TENNIS

History.—Court tennis may well be called a royal game, having been popular with various kings of England and France. In the ball games of the Greeks and Romans may be seen the rudiments

THE BETTMANN ARCHIVE INC.

FIG. 2.—COPPER ENGRAVING OF EARLY COURT TENNIS, 1612. AT THIS DATE THE RACQUET WAS IN USE BUT THE NET HAD NOT YET REPLACED THE FRINGED OR TASSELLED ROPE STRETCHED ACROSS THE COURT. NOTE LE PETIT TROU AND LA LUNE, "WINNING OPENINGS" ON THE FAR END WALL

of the French *jeu de paume*, which is the ancestor of modern tennis in a direct line. The origin of the name is quite obscure. The most probable derivation is from the word *Tenez!* (Take it! Play!), especially in view of the large number of French terms that adhere to the game; *e.g.*, grille, tambour (drum, from the sound on the board that formed the face of that buttress) and *dedans*. Further, a poem dealing with the game, written in Latin elegiacs by R. Frissart, makes the striker cry *"Excipe!"* (Take it!) before each stroke; this seems to correspond with the custom which enjoins the rackets marker to call "play" whenever a legitimate stroke has been made.

In the *Alexiad* of Anna Comnena (about A.D. 1120) is a reference to a game played on horseback in which a staff, curved at the end and strung with strings of plaited gut was used. This game was played in a court called "a court for goff" (*sic;* according to the Lexicon of *Alexandrine Greek*) and some similar game, corrupted through *tchangan* into *chicane*, was played in France. Good authorities also find a more ancient derivation of the game in Egypt, in Persia and among the Arabs before Charlemagne. In A.D. 1300 the game was also known as *La Boude*. Throughout the century indeed it was played in France and by the highest in the land. Louis X died from a chill contracted after playing, and Charles V was devoted to the game. In England the game, or some form of it, was known, Chaucer alluding to it in the words "But canstow playen racket to and fro."

The 16th century was the golden age of court tennis in France. Francis I built a magnificent court at the Louvre, where his son Henry II, one of the best players of his day, built a second and larger court. Charles IX was a keen player and granted a royal charter to the Corporation of Tennis professionals. Henry IV was equally keen and in 1596 there were said to be 250 courts in Paris alone. An English visitor in 1598 wrote "There be more tennis players in France than ale-drinkers or malt-wormes (as they call them) with us." Louis XIII and Louis XIV both played tennis but by 1657 the number of courts in Paris had declined to 114. The Royal tennis court at Versailles was the scene of the famous tennis court oath of June 20, 1789, which marked the start of the French Revolution.

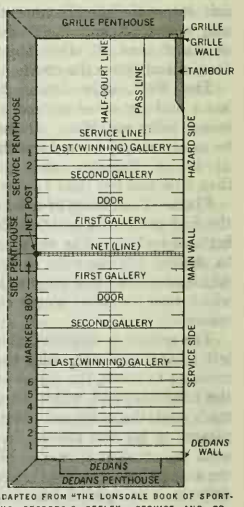
In England, too, court tennis was in its heyday in the 16th century. Henry VII was a good player and played on courts in Woodstock, Wycombe, Westminster, Sheen and Windsor. At Windsor in 1506 he entertained the king of Castile who played a tennis match against the marquis of Dorset. Henry VIII was an outstanding tennis player and in 1530 built the court at Hampton Court palace which is still in use. He also built four courts at Whitehall and one at St. James's palace. Elizabeth I was a keen spectator. Charles I and Charles II were both good tennis players. Charles II built a magnificent new court at Whitehall and Pepys made several allusions to the king playing tennis. A print of 1641 shows James II as a boy in a tennis court. Tennis in England declined in the 18th century and in the second half of the 20th century there were 15 courts in play.

In the late 19th century courts were built in Australia at Hobart (1854) and Melbourne (1882) which are still in play.

Court tennis was introduced to the U.S. in the 1870s. The game was played regularly at the Racquet Court club in New York in 1876 where Robert Moore was the professional instructor. The club merged about 1890 with the Racquet and Tennis club, which became the foremost club in America sponsoring court tennis and other racquet games. Courts were built in Boston, Mass. (the first in 1876), Newport, R.I. (1880), Chicago, Ill. (1893), and later at Tuxedo, N.Y.

It is possible to trace a world's open championship of court tennis back to the early 19th century and holders of this title have been: 1816, Marchisio (Italy); 1819, Philip Cox (Great Britain); 1829, J. E. Barre (France); 1862, E. Tompkins (Great Britain); 1871, G. Lambert (Great Britain); 1885, Tom Pettitt (United States); 1896, P. Latham (Great Britain); 1905, C. Fairs (Great Britain); 1907, P. Latham (Great Britain); 1908, C. Fairs (Great Britain); 1912, G. F. Covey (Great Britain); 1914, J. Gould (United States); 1922, G. F. Covey (Great Britain); 1928, P. Etchebaster (France); 1955, J. Dear (Great Britain); 1957, A. Johnson (Great Britain); 1959, N. R. Knox (United States).

Peter Latham had the distinction of being world champion of rackets also. Jay Gould and N. R. Knox are the only amateurs ever to have won the world championship. Pierre Etchebaster, a French Basque and professional at the New York Racquet and Tennis club, held the title for 27 years until he retired at the age of 60 undefeated, and was undoubtedly the outstanding player of the modern era.



ADAPTED FROM "THE LONSDALE BOOK OF SPORTING RECORDS," SEELEY, SERVICE AND CO., LTD., LONDON

FIG. 3.—PLAYING AREA FOR COURT TENNIS

General Description of Game.—Court tennis is played on a four-walled indoor court somewhat larger than a lawn tennis court (*see* fig. 3). Although there is no standard size, the following dimensions may be regarded as suitable: the space enclosed by the walls, 110 ft. by 38 ft. 8 in.; the floor however, measuring but 96 ft. by 31 ft. 8 in., the difference being the width of a roofed corridor (the "penthouse") which runs along the two end walls and one of the side walls. Across the middle of the court a net is stretched and the first object of the game is to strike the ball over this with a racket. The net is 5 ft. high at the ends, 3 ft. at the middle, and divides the floor into two equal parts—the "service" side and the "hazard" side. The cement floor and walls should be smooth but not polished.

The height of the court to the tie beam is 30 ft. The height of the play line, above which the ball must not go, is 18 ft. at the sides and 23 ft. at the ends. The roof of the penthouse slopes downward toward the court, the lower edge being 7 ft. 1½ in. from the floor, the upper edge 10 ft. 7 in., the width 7 ft. The *dedans* is an open gallery in the wall at the service end of the court, under the penthouse. It is 21 ft. 8 in. in width; the upper edge is 6 ft. 10 in. from the floor, the lower edge 3 ft. 3 in. Looking from the *dedans*, the right-hand or main wall has one peculiarity, the tambour, a sloping buttress toward the far end, to form which the wall is built inward, reducing the breadth of that part of the court to 30 ft. 2 in. In the right-hand corner of the hazard side end wall (as viewed from the *dedans*) is the grille, an opening 3 ft. 1 in. square. The left-hand wall, along which runs the penthouse, is not continuous, being broken by a long opening between the floor and the penthouse similar to the *dedans*, and at the same height from the ground. The low walls under this opening and the *dedans* are called the batteries. There is no wall in front of the marker's box, through which the court is entered on either side of the net post. This long opening in the left-hand wall is divided into galleries and doors, the latter situated where the entrances to the court used to be in early courts. The galleries are marked by posts, which also serve to support the penthouse. The galleries, *dedans* and grille are known as the openings; three of these—the grille, *dedans* and winning gallery—are winning openings, for if a ball in play is struck into one of these, the striker scores a point. In the earlier French courts there were other winning openings such as: *Pais* (the board), an upright board 9 ft. by 1 ft. in the left-hand corner of the *dedans* wall; *le petit trou* or *le trou*, a hole 16 in. square at the bottom of the other side of the wall; and *la lune*, a round opening high up by the play line, one at each end of the court. The court has certain lines painted on the floor which are also continued perpendicularly on the walls. On the hazard side is the half-court line, the pass line and the service line. The first is required in doubles or in certain forms of handicap game. The pass line is drawn 7 ft. 8 in. from the main wall, the service line 21 ft. 1 in. from the grille wall. The rectangle contained by the pass and service lines forms the service court. The other lines, both on the hazard side and service side, mark the chases, which will be explained below.

Equipment.—The balls vary in the three countries where the game is chiefly played—England, United States and France. In England balls weigh approximately 2½ oz. and are 2½ in. in diameter. They are made of strips of cloth wound tightly together, tied with twine and covered with white melton cloth. The French ball is harder and faster than the English; and the American ball lies somewhere between the two. The racket is usually about 27 in. long and weighs about 16 oz. The head is about 9 in. long and 6 in. broad, but there are no restrictions as to size or weight. The head is somewhat pear-shaped, but its centre line does not correspond with the centre line of the handle, since the head is curved upward. Some of the early rackets were strung diagonally (*i.e.*, in diamonds); later, vertical-horizontal stringing was universally adopted; then followed knotting at the points of intersection, a practice which later disappeared.

Scoring.—A match in England and the U.S. consists ordinarily of the best of three or the best of five sets, a set being scored by the player who gets six games first. Sets of "eight games," *i.e.*, the set won by the player who gets eight games first, are generally played in France.

Occasionally vantage sets are played; *i.e.*, when each player has scored five games or seven games as the case may be, the set continues until one player gets an advantage of two games. In these conditions sets sometimes continue for a very long time, and sets where over 60 games have been played are recorded. In matches for the world's championship the best of 13 sets of six games are generally played on three different days with an interval between them. Scoring is as in lawn tennis. The score of the player who won the last stroke or made the last chase is called first.

The Game.—The players decide who shall serve by spinning a racket on its head. One spins and the other calls "rough" or "smooth," the rough side of the head of the racket showing the knots of some of the lower strings. The winner takes the service side, service being an advantage. In four-handed games the winners of the spin often take the hazard side, since by doing so they can decide which of their opponent's service each one will elect to take. He serves from any part of the court and in any way he thinks best, and the ball must go over the net, strike the side penthouse, and fall into the service court. His opponent ("striker-out") tries to return the ball over the net before it has touched the ground a second time; he may volley or half-volley (strike the ball as soon as it touches the court) his return. For a stroke to be good, it must be made before the second bounce of the ball, and the ball must go over the net (even if it touches it), and must not strike the wall above the play line, nor touch the roof or rafters.

The Chase.—The Chase is an ancient and somewhat complicated device which gives some advantage to skill over brute force. In most games a player has to return the ball either on the volley or after it has bounced once; a ball that bounces twice is called "not up." In court tennis this is not so and a ball that is "not up" makes a chase at the point where it strikes the ground on its second bounce. Neither player scores a point when a chase is made; the point goes into cold storage until it is played off, either when one player has reached 40 or when there are two chases in cold storage in the same game. When this happens the players change ends and play off the chase or chases. The nearer the chase is to the back wall, the better it is and the more difficult it is to beat. This is where skill is at a premium. The player who allowed the ball to bounce a second time and make a chase now has to make a better chase to win the point. If any stroke he makes in the course of the rally is going to make a worse chase, his opponent can leave it and thereby win the point.

The marker standing by the net is responsible for calling out the score and for naming the chases. To enable him to mark the chases the floor of the court is ruled off in transverse lines. On the service side these start at one-yard intervals from the back wall and are numbered from one to six. From chase six to the net further lines are ruled, corresponding to the centre of the galleries which line the side wall and called by the name of the gallery to which they refer, *i.e.*, last gallery, second gallery, door, first gallery and line. A ball entering directly into one of these galleries makes a chase in exactly the same way as when the ball bounces twice. On the hazard side there are no chase lines at the back of the court and a ball "not up" in this area loses a point as in other games. The galleries are the same as on the service side except that the equivalent of the last gallery is called the winning gallery and a ball entering it rings a bell and wins a point outright. Under the winning gallery there are two chase lines marked one and two. (AB.)

BIBLIOGRAPHY.—A. Wallis Myers, *Fifty Years of Wimbledon, 1877 to 1926* (1926); U.S. Lawn Tennis Association, *Fifty Years of Lawn Tennis in the United States* (1931); *Spalding's Tennis Annual* (1912 *et seq.*); S. Wallis Merrihew, *The Quest of the Davis Cup* (1928); J. R. Lacoste, *Lacoste on Tennis* (1928); Helen Jacobs, *Tennis* (1941); William T. Tilden, *How to Play Better Tennis* (1957); L. Budge, *Tennis Made Easy* (1945); Maureen Connolly, *Power Tennis* (1954); Maurice Brady, *Encyclopaedia of Lawn Tennis* (1958); Lord Aberdare, *The Story of Tennis* (1959); *Sports Illustrated Book of Tennis* (1961); *International Lawn Tennis Almanac* (annual); U.S. Lawn Tennis Association, *Official Handbook* (annual); W. M. Fischer Lawn Tennis Library, New York, N.Y. (E. S. BR.; AB.)

TENNYSON, ALFRED TENNYSON, 1ST BARON (1809–1892), English poet, was born at Somersby, Lincolnshire, on Aug. 6, 1809, fourth of 12 children of the Rev. George Clayton Tennyson. At a very early age he began to write in prose and verse. At Christmas 1815 he was sent to the grammar school at Louth, where he remained for five years, and then returned to Somersby to be trained by his father. In the rectory the boys had the run of an excellent library, and there the young poet based his wide knowledge of the English classics. Tennyson was already writing copiously—he had constructed "an epic of 6,000 lines" at the age of twelve, composed a drama in blank verse when he was 14, and so on. In 1827 Frederick Tennyson (1807–1898), the eldest surviving brother, uniting with his younger brothers Charles and Alfred, published at Louth an anonymous collection of *Poems*

by Two Brothers. The "two" were Charles and Alfred (whose contributions predominated), and who shared the surprising profits, £20. On Feb. 20, 1828, Charles and Alfred matriculated at Trinity college, Cambridge, where Frederick was already a student. The brothers were shy, and at first made few friends; but they gradually gathered selected associates around them, and Alfred grew to be looked up to in Cambridge "as to a great poet and an elder brother" by a group which included Richard Chenevix Trench, Monckton Milnes (Lord Houghton), James Spedding, W. H. Thompson, Edward FitzGerald, W. H. Brookfield, and, above all, A. H. Hallam (1811–1833). Charles Tennyson (1808–1879) afterwards took the additional name of Turner. He published four volumes of sonnets which have been highly praised. In June 1829 Alfred Tennyson won the Chancellor's prize medal for his poem called "Timbuctoo." With great imperfections, this study in Miltonic blank verse displays the genius of a poet, in spite of obscurity both of thought and style. Here are already both richness and power. But by this time Tennyson was writing lyrics of still higher promise, and, as Arthur Hallam early perceived, with an extraordinary earnestness in the worship of beauty. The results of this enthusiasm and this labour of the artist appeared in the volume of *Poems, chiefly Lyrical*, published in 1830. This book would have been astonishing as the production of a youth of 21, even if, since the death of Byron six years before, there had not been a singular dearth of good poetry in England. Here at least, in the slender volume of 1830, was a new writer revealed, and in "Mariana," "The Poet," "Love and Death," and "Oriana," a singer of wonderful though still unchastened melody. Through these, and through less perfect examples, was exhibited an amazing magnificence of fancy, at present insufficiently under control, and a voluptuous imagery, tending to an oversweetness.

In the summer of 1830 Tennyson and Hallam volunteered in the army of the Spanish insurgent Torrijos, and marched about a little in the Pyrenees, without meeting with an enemy. Tennyson came back to find his father ailing, and in Feb. 1831 he left Cambridge for Somersby, where a few days later Dr. George Tennyson died. The new incumbent was willing that the Tennysons should continue to live in the rectory, which they did not leave until six years later. Arthur Hallam was now betrothed to Emily Tennyson (afterwards Mrs. Jesse, 1811–1889), and stayed frequently at Somersby. This was a very happy time, and one of great physical development on Alfred's part. He took his share in all kinds of athletic exercises, and it was now that Brookfield said, "It is not fair that you should be Hercules as well as Apollo." This high physical zest in life seems to have declined after 1831, when his eyes began to trouble him, and he became liable to depression. The poetical work of these three years, mainly spent at Somersby, was given to the world in the volume of *Poems* which (dated 1833) appeared at the end of 1832.

This was certainly one of the most astonishing revelations of finished genius ever produced by a young man of less than four-and-twenty. Here were "The Lady of Shalott," "The Dream of Fair Woman," "Oenone," "The Lotus-Eaters," "The Palace of Art," and "The Miller's Daughter," with a score of other lyrics, delicious and divine. The advance in craftsmanship and command over the *matériel* of verse shown since the volume of 1830 is astonishing. It was well that its publication was completed before the blow fell upon Tennyson which took for a while all the light out of him. In Aug. 1833 Arthur Hallam started with his father, the historian, for Tirol. On the way young Hallam died suddenly (Sept. 15) of a broken blood-vessel at Vienna. His body was brought back to England, and buried on Jan. 3, 1834.

Hallam's death affected Tennyson extremely. He grew less than ever willing to come forward and face the world; his health became "variable and his spirits indifferent." The earliest effect of Hallam's death upon his friend's art was seen, in the summer of 1834, in *The Two Voices;* and to the same period belong the beginnings of the *Idylls of the King* and of *In Memoriam*, over both of which he meditated long. In 1835 he visited the Lakes, and saw much of Hartley Coleridge, but would not "obtrude on the great man at Rydal," although "Wordsworth was hospitably disposed." Careless alike of fame and of influence, Tennyson spent these years mainly at Somersby, in a uniform devotion of his whole soul to the art of poetry. In 1837, to their great distress, the Tennysons were turned out of the Lincolnshire rectory where they had lived so long. They moved to High Beech, in Epping Forest, which was their home until 1840. The poet was already engaged, or "quasi-betrothed," to Emily Sellwood, but ten years passed before they could afford to marry. At Torquay, in 1838, he wrote *Audley Court* on one of his rare excursions, for he had no money for touring, nor did he wish for change; he wrote at this time, "I require quiet, and myself to myself, more than any man when I write." In 1840 the Tennysons moved to Tunbridge Wells, and a year later to Boxley, near Maidstone, to be close to Edmund Lushington, who had now married Cecilia Tennyson. Alfred was from this time frequently a visitor in London.

In 1842 the two-volume edition of his *Poems* broke the ten years' silence which he had enforced himself to keep. Most of those studies of home-life in England, which formed so highly popular a section of Tennyson's work—such as "The Gardener's Daughter," "Walking to the Mail," and "The Lord of Burleigh"—were now first issued, and, in what we have grown to consider a much higher order, "Locksley Hall," "Ulysses," and "Sir Galahad." To the older and more luxurious lyrics, as reprinted in 1842, Tennyson did not spare the curbing and pruning hand, and in some cases went too far in restraining the wanton spirit of beauty in its youthful impulse. It is from 1842 that Tennyson took his place as the leading poet of his age in England. Among the friends whom he now made, or for the first time cultivated, were Carlyle, Rogers, Dickens, and Elizabeth Barrett.

Material difficulties now, however, for the first time intruded on Tennyson's path. He became the victim of a certain "earnest-frothy" speculator, who induced him to sell his little Lincolnshire estate at Grasby, and to invest the proceeds, with all his other money, and part of that of his brothers and sisters, in a "Patent Decorative Carving Company"; in a few months the whole scheme collapsed, and Tennyson was left penniless. He was attacked by so overwhelming a hypochondria that his life was despaired of, and he was placed for some time under the charge of a hydropathic physician at Cheltenham, where absolute rest and isolation gradually brought him round to health again. The state of utter indigence to which Tennyson was reduced greatly exercised his friends, and in Sept. 1845, at the suggestion of Henry Hallam, Sir Robert Peel was induced to bestow on the poet a pension of £200 a year. Tennyson's health slowly became restored, and in 1846 he was hard at work on *The Princess;* in the autumn of this year he took a tour of Switzerland, and saw great mountains and such "stateliest bits of landskip" for the first time. In 1847 nervous prostration again obliged him to undergo treatment.

The Princess was now published, in a form afterwards considerably modified and added to. Carlyle and FitzGerald "gave up all hopes of him after *The Princess,*" or pretended that they did. It was true that the bent of his genius was slightly altered, in a direction which seemed less purely and austerely that of the highest art; but his concessions to public taste vastly added to the width of the circle he now addressed. The home of the Tennysons was now at Cheltenham; on his occasional visits to London he was in the habit of seeing Thackeray, Coventry Patmore, Browning and Macready, as well as older friends, but he avoided "society."

In his absent-minded way Tennyson was very apt to mislay objects; in earlier life he had lost the ms. of *Poems, chiefly Lyrical*, and had been obliged to restore the whole from scraps and memory. Now a worse thing befell him, for in February 1850, having collected into one "long ledger-like book" all the elegies (*In Memoriam*) on Arthur Hallam which he had been composing at intervals since 1833, he left this only ms. in the cupboard of some lodgings in Mornington Place, Hampstead Road. By extraordinary good chance it had been overlooked by the landlady, and Coventry Patmore was able to recover it. *In Memoriam* was published, in its original anonymous form, in May 1850. The public was at first greatly mystified by the nature and object of this poem, which was not merely a chronicle of Tennyson's emotions under bereavement, nor even a statement of

his philosophical and religious beliefs, but, as he long afterwards explained, a sort of *Divina Commedia*, ending with happiness in the marriage of his youngest sister, Cecilia Lushington. In fact, the great blemishes of *In Memoriam*, its redundancy and the dislocation of its parts, were largely due to the desultory manner of its composition. The poet wrote the sections as they occurred to him, and did not think of weaving them together into a single poem until it was too late to give them real coherency.

On June 13, 1850, Tennyson was married at Shiplake to Emily Sarah Sellwood (1813–1896). Of this union no more need be said than was recorded long afterwards by the poet himself, "The peace of God came into my life before the altar when I wedded her." On Nov. 19, 1850, Queen Victoria appointed Tennyson poet laureate, in succession to Wordsworth. The salary connected with the post was very small, but it had a secondary value in stimulating the sale of his books. The young couple took a house at Warninglid, in Sussex, which did not suit them, and then one in Montpelier Row, Twickenham, which did better. In April 1851 their first child was born dead. At this time Tennyson was brooding much upon the ancient world, and reading little but Milton, Homer and Virgil. In the summer of 1851 he made a tour in Italy, of which *The Daisy* is the immortal record. Of 1852 the principal events were the birth of his eldest son Hallam, the second Lord Tennyson, in August, and in November the publication of the *Ode on the Death of the Duke of Wellington*.

In the winter of 1853 Tennyson took a little house and farm called Farringford, near Freshwater, in the Isle of Wight, which he leased at first, and afterwards bought; this beautiful place, ringed round with ilexes and cedars, entered into his life and coloured it with its delicate enchantment. In 1854 he published *The Charge of the Light Brigade*, and was busy composing *Maud* and its accompanying lyrics; and this volume was published in July 1855, just after he was made D.C.L. at Oxford: he was received on this occasion, which may be considered his first public appearance, with a "tremendous ovation." The reception of *Maud* from the critics, however, was the worst trial to his equanimity which Tennyson had ever had to endure, nor had the future anything like it in store for him. He had risen in *Maud* far above his ordinary serenity of style, to ecstasies of passion and audacities of expression which were scarcely intelligible to his readers.

It is odd that this irregular poem, with its copious and varied music, its splendid sweep of emotion, its unfailing richness of texture—this poem in which Tennyson rises to heights of human sympathy and intuition which he reached nowhere else, should have been received with bitter hostility, have been styled "the dead level of prose run mad," and have been reproved more absurdly still for its "rampant and rabid bloodthirstiness of soul." There came a reaction of taste and sense, but the delicate spirit of Tennyson had been wounded. For some years the world heard nothing from him; he was at Farringford, busying himself with the Arthurian traditions. He had now become an object of boundless personal curiosity, being already difficult to find, and the centre of amusing legends. It was in 1857 that Bayard Taylor saw him, and carried away the impression of a man "tall and broad-shouldered as a son of Anak, with hair, beard and eyes of southern darkness." This period of retirement embraced a tour in Wales in 1857, a visit to Norway in 1858, and a journey through Portugal in 1859.

In 1857 two Arthurian poems had been tentatively and privately printed, as *Enid and Nimue, or the True and the False*, to see how the idyllic form would be liked by the inner circle of Tennyson's friends. In the summer of 1859 the first series of *Idylls of the King* was at length published, and achieved a popular success far beyond anything experienced before by any English poets, save perhaps Byron and Scott. Within a month of publication, 10,000 copies had been sold. The idylls were four in number, "Enid," "Vivien" (no longer called "Nimue"), "Elaine" and "Guinevere." These were fragments of the epic of the fall of King Arthur and the Table Round which Tennyson was so long preparing, and which he can hardly be said to have ever completed, although nearly thirty years later he closed it. The public and the critics alike were entranced with the "sweetness" and the "purity" of the treatment. A few, like Ruskin, were doubtful about "that increased quietness of style"; one or two already suspected that the "sweetness" was obtained at some sacrifice of force, and that the "purity" involved a concession to Victorian conventionality.

Urged by the duke of Argyll, Tennyson now turned his attention to the theme of the Holy Grail, though he progressed with it but fitfully and slowly. In 1861 he travelled in Auvergne and the Pyrenees, with Clough, who was to die a few months later; to this year belong "Helen's Tower" and the "Dedication" of the *Idylls* to the prince consort, "These to his Memory." The latter led to Tennyson's presentation in April 1862 to the queen, who "stood pale and statue-like before him, in a kind of stately innocence," which greatly moved his admiring homage. From this time forth the poet enjoyed the constant favour of the sovereign, though he could never be moulded into a conventional courtier. He now put the Arthurian legends aside for a time, and devoted himself to the composition, in 1862, of "Enoch Arden," which, however, did not appear until 1864, and then in a volume which also contained "Sea Dreams," "Aylmer's Field" and, above all, "The Northern Farmer," the first and finest of Tennyson's remarkable studies in dialect. Tennyson's popularity in Great Britain was growing all the time to an extent unparalleled in the whole annals of English poetry. This universality of fame led to considerable practical discomfort; he was besieged by sightseers, and his nervous trepidation led him perhaps to exaggerate the intensity of the infliction. In 1867 he determined to make for himself a haven of refuge against the invading Philistine, and bought some land on Blackdown, above Haslemere, then a secluded corner of England; here Mr. (afterwards Sir) James Knowles began to build him a house, ultimately named Aldworth. This is the time of two of his rare, privately printed pamphlets, *The Window; or, the Loves of the Wrens* (1867), and *The Victim* (1868).

The noble poem *Lucretius*, one of the greatest of Tennyson's versified monographs, appeared in May 1868, and in this year *The Holy Grail* was at last finished; it was published in 1869, together with three other idylls belonging to the Arthurian epic, and various miscellaneous lyrics, besides *Lucretius*. The reception of this volume was cordial, but not so universally respectful as that which Tennyson had grown to expect from his adoring public. The fact was that the heightened reputation of Browning, and still more the sudden vogue of Swinburne, Morris and Rossetti (1866–70), considerably disturbed the minds of Tennyson's most ardent readers, and exposed him to a severer criticism than he had lately been accustomed to endure. His next volume (1872), *Gareth and Lynette* and *The Last Tournament*, continued, and, as he then supposed, concluded *The Idylls of the King*, and for the time being he dismissed it from his mind.

Believing that his work with the romantic Arthurian epics was concluded, Tennyson now turned his attention to the drama. He put before him a scheme of illustrating "the making of England" by a series of great historical tragedies. His *Queen Mary*, the first of these chronicle-plays was published in 1875, and played by Sir Henry Irving at the Lyceum in 1876. Although it was full of admirably dramatic writing, it failed on the stage. Extremely pertinacious in this respect, the poet went on attempting to storm the theatre, with assault upon assault, all practically failures until the seventh and last, which was unfortunately posthumous. Meanwhile *Harold*, a tragedy of doom, was published in 1876; but, though perhaps the finest of its author's dramas, it was not staged. During these years Tennyson's thoughts were largely occupied with the building of Aldworth. His few lyrics were spirited ballads of adventure, inspired by an exalted patriotism—"The Revenge" (1878), "The Defence of Lucknow" (1879)—but he reprinted and finally published his old suppressed poem, *The Lover's Tale*, and a little play of his, *The Falcon*, versified out of Boccaccio, was produced by the Kendals in 1879.

In 1880, when he was over 70, he published the earliest of six important collections of lyrics, this being entitled *Ballads and other Poems*, and containing the sombre and magnificent "Rizpah." In 1881 *The Cup* and in 1882 *The Promise of May*, two little plays, were produced without substantial success in London theatres: the second of these is perhaps the least successful of all the poet's longer writings, but its failure annoyed him unreasonably. In September 1883 Tennyson and Gladstone set out on a voyage round the north of Scotland, to Orkney, and across the ocean to Norway and Denmark. During the voyage Gladstone offered Tennyson a peerage, which he accepted. On March 11, 1884, he took his seat in the House of Lords as Baron Tennyson of Aldworth and Farringford. In the autumn of this year his tragedy of *Becket* was published, but the poet at last despaired of the stage, and disclaimed any hope of "meeting the exigencies of our modern theatre." Curiously enough, after his death *Becket* was the one of all his plays which enjoyed a great success on the boards.

In 1885 was published another interesting miscellany, *Tiresias and other Poems,* with a posthumous dedication to Edward FitzGerald. In this volume, it should be noted, *The Idylls of the King* was completed at last by the publication of "Balin and Balan"; it contained also the superb address "To Virgil." In April 1886 Tennyson suffered the loss of his second son, Lionel, who died in the Red Sea on his return from India. The untiring old poet was steadily writing on, and by 1886 he had another collection of lyrics ready, *Locksley Hall Sixty Years After,* etc.; his eyes troubled him, but his memory and his intellectual curiosity were as vivid as ever. He was past 80 when he published the collection of new verses entitled *Demeter and other Poems* (1889), which appeared almost simultaneously with the death of Browning, an event which left Tennyson a solitary figure indeed in poetic literature. His latest drama, *The Foresters,* was produced in New York in March 1892, with Miss Ada Rehan as Maid Marian. During this year Tennyson was steadily engaged on "Akbar's Dream," "Kapiolani" and other contents of the posthumous volume called *The Death of Oenone,* 1892.

Soon after entering his eighty-fourth year, however, symptoms of weakness set in, and early in September his condition began to give alarm. He retained his intellectual lucidity and an absolute command of his faculties to the last, reading Shakespeare with obvious appreciation until within a few hours of his death. With the splendour of the full moon falling upon him, his hand clasping his Shakespeare, and looking, as we are told, almost unearthly in the majestic beauty of his old age, Tennyson passed away at Aldworth on the night of Oct. 6, 1892. On Oct. 12 he received a public funeral at Westminster Abbey.

No living poet has ever held England—no poet but Victor Hugo has probably ever held any country—quite so long under his unbroken sway as Tennyson did. As he recedes from us, however, we begin to see that he has a much closer relation to the great Georgian writers than we used to be willing to admit. He is more the last of that great school than the first of any new one. The qualities in which he seems to surpass his immediate predecessors are exactly those which should be the gift of one who sums up the labours of a mighty line of artists. He is remarkable among them for the breadth, the richness, the substantial accomplishment of his touch; he has something of all these his elders, and goes farther along the road of technical perfection than any of them. We still look to the earlier masters for supreme excellence in particular directions: to Wordsworth for sublime philosophy, to Coleridge for ethereal magic, to Byron for passion, to Shelley for lyric intensity, to Keats for richness.

Tennyson does not excel each of these in his own special field, but he is often nearer to the particular man in his particular mastery than any one else can be said to be, and he has in addition his own field of supremacy. This consists, perhaps, in the beauty of the atmosphere which Tennyson contrives to cast around his work, moulding it in the blue mystery of twilight, in the opaline haze of sunset: this atmosphere produces an almost unfailing illusion or mirage of loveliness, so that, even where (as must sometimes be the case with every poet) the thought and the imagery have little value in themselves, the fictive aura of beauty broods over the otherwise undistinguished verse. In his luminous subtlety and his broad undulating sweetness, his relationship with Virgil has long been manifest; he was himself aware of it. But he was also conscious that his exquisite devotion to mere lucidity and beauty might be a snare to him, and a happy instinct was always driving him to a study of mankind as well as of inanimate nature. (E. G.; X.)

The standard biography is *Alfred, Lord Tennyson: A Memoir,* by Hallam, 2nd Baron Tennyson, 2 vol. (1897), who also edited Tennyson's *Works* with a memoir in 1913 (9 vol.); A. E. Baker published a *Concordance to the poetical and dramatical works of Tennyson* in 1914. For bibliography *see* R. H. Shepherd, *Tennysoniana* (1866, new ed. 1879). For biography and criticism consult A. C. Bradley, *A Commentary on Tennyson's In Memoriam* (1901); A. Lang, *Alfred Tennyson* (1901); A. C. Benson, *The Life of Lord Tennyson* (1904); T. R. Lownsbury, *Life and Times of Lord Tennyson* (1916); H. I. 'A. Fausset, *Tennyson; A modern portrait* (1923); H. Nicolson, *Tennyson* (1925).

TENONITIS: *see* TENDINITIS.

TENOR, the highest adult male voice. Its range is approximately two octaves, an octave below and an octave above middle C, but this varies with individuals. The word "tenor" is derived from the Latin *tenere,* "to hold" and refers to the fact that in early polyphonic music the highest of the men's voices held the leading part or *cantus firmus,* to which other parts were added.

Tenor voices are described according to quality—a light voice being designated lyric tenor or *tenore di grazia,* sometimes *tenore leggiero;* and the heavier voice being known as dramatic or heroic tenor, in German *Heldentenor.* The male voice that lies above the natural tenor voice is a countertenor, which is a natural voice, not a falsetto like the male alto, and extremely rare. Alto tenor is a term used to characterize the changing voice of the adolescent boy with a range limited to the lower tones of the alto and the upper tones of the real tenor. *See* SINGING. (A. E. PE.)

TENREC (TENDRAC) (*Tenrec ecaudatus*), an insect-eating mammal, one of the largest representatives of the order Insectivora, the length being from 12 to 16 in. It is restricted to Madagascar. The coat consists chiefly of bristles and hairs, with some flexible spines. The general hue is brown tinged with yellow. From 12 to 16 or even 20 young are produced at birth.

In habits the tenrec is nocturnal; its home is in the brush in the mountain regions, and in the cool season it hibernates in deep burrows. The long flexible snout is used to root up worms and grubs, and ground insects form part of its nourishment. When fat before hibernation, tenrecs are much valued as food by natives. *See* INSECTIVORA.

TENSOR ALGEBRA. Tensors were invented to express geometric entities arising in the study of manifolds (*see* MANIFOLDS; TENSOR ANALYSIS). Since manifolds were originally considered only in the small in terms of co-ordinate systems, tensors were expressed in terms of functions involving many indices which obey certain transformation laws when the co-ordinate system is changed. Subsequently, a global study of manifolds resulted in an invariant treatment of tensors and their algebraic analysis. This is the subject of tensor or multilinear algebra.

Tensors and Vector Spaces.—Technically, a tensor of type (s,r) is an element of a vector space $V_{s,r}$ associated with a given vector space V of dimension d (*see* VECTOR SPACES). Any basis of V induces a basis of $V_{s,r}$ in terms of which every tensor T in $V_{s,r}$ can be represented by its components

$$T^{j_1 j_2 \cdots j_s}_{i_1 i_2 \cdots i_r}, \quad i_1, \ldots, i_r, j_1, \ldots, j_s = 1, \ldots, d$$

The vector space $V_{s,r}$ can be constructed from V in the following way: Let V^* denote the dual space of V; that is, the vector space of linear functions of V. Then $V_{s,r}$ is the collection of scalar-valued functions of $r+s$ variables, r of them in V^* and s of them in V, which are linear in each variable when the others are left fixed. Thus each T in $V_{s,r}$ gives the scalar value $T(v^1, \ldots, v^s, \alpha_1, \ldots, \alpha_r)$ for every $v^1, \ldots, v^s$ in V and $\alpha_1, \ldots, \alpha_r$ in V^*. If $e^1, \ldots, e^d$ is a basis of V, let $\phi_1, \ldots, \phi_d$ be the basis of V^* which is dual to $e^1, \ldots, e^d$ so that $\phi_i(e^j) = 0$ if $i \neq j$ and $\phi_i(e^i) = 1$. Because T is linear in each variable separately, T is completely determined by its values

$$T(e^{i_1}, \ldots, e^{i_s}, \phi_{i_1}, \ldots, \phi_{i_r}) = T^{i_1 \cdots i_s}_{i_1 \cdots i_r}$$

If $\phi_{i_1} \otimes \ldots \otimes \phi_{i_s} \otimes e^{i_1} \otimes \ldots \otimes e^{i_r}$ is the tensor which gives the value 1 on $(e^{i_1}, \ldots, e^{i_s}, \phi_{i_1}, \ldots, \phi_{i_r})$ and 0 for any other choice of basis vectors, then

$$T = \sum_{i,j=1}^{d} T^{i_1 \cdots i_s}_{i_1 \cdots i_r} \phi_{i_1} \otimes \ldots \otimes \phi_{i_s} \otimes e^{i_1} \otimes \ldots \otimes e^{i_r}$$

Suppose a different basis $f^1, \ldots, f^d$ of V had been chosen and suppose $f^i = \sum_{k=1}^{d} a^i_k e^k$. Then the dual basis of $f^1, \ldots, f^d$ is the set of linear functions $\Psi_1, \ldots, \Psi_d$, where $\Psi_i = \sum_{l=1}^{d} b^i_l \phi_l$ and (b^i_l) is the inverse of the nonsingular matrix (a^i_k). In terms of the new basis, T is represented by its components

$$T(f^{j_1}, \ldots, f^{j_s}, \Psi_{i_1}, \ldots, \Psi_{i_r}) =$$

$$\sum_{k,l=1}^{d} a^{j_1}_{k_1} a^{j_2}_{k_2} \ldots a^{j_s}_{k_s} b^{l_1}_{i_1} b^{l_2}_{i_2} \ldots b^{l_r}_{i_r} T^{k_1, \ldots, k_s}_{l_1, \ldots, l_r}$$

Thus a tensor of type (s,r) can be described as a function of $r+s$ indices depending upon a basis which obeys the transformation law in the last formula when the basis is changed. The advantage of the invariant definition of $V_{s,r}$ (*i.e.*, independent of a basis) is that it is applicable to the case where the scalars form only a ring and also to infinite-dimensional vector spaces. It is also aesthetically more pleasing. The description of T in terms of its components $T^{i_1, \ldots, i_s}_{i_1, \ldots, i_r}$ is useful in explicit computations.

Construction and Properties of Tensor Algebra.—Two tensors can be multiplied to give another tensor. If T is a tensor of type (s,r) and $\overline{T}$ is a tensor of type $(\bar{s},\bar{r})$, then $T\overline{T}$ is a tensor of type $(s+\bar{s}, r+\bar{r})$, where

$$T\overline{T}(v^1, \ldots, v^s, v^{s+1}, \ldots, v^{s+\bar{s}}, a_1, \ldots, a_r, a_{r+1}, \ldots, a_{r+\bar{r}}) =$$

$$T(v^1, \ldots, v^s, a_1, \ldots, a_r) \cdot \overline{T}(v^{s+1}, \ldots, v^{s+\bar{s}}, a_{r+1}, \ldots, a_{r+\bar{r}})$$

In terms of a basis the components of $T\overline{T}$ are

$$T^{i_1, \ldots, i_s\ i_{s+1}, \ldots, i_{s+\bar{s}}}_{i_1, \ldots, i_r\ i_{r+1}, \ldots, i_{r+\bar{r}}}$$

From this multiplication an associative algebra called the tensor algebra $\mathcal{T}$ can be constructed over the vector space V. The elements of $\mathcal{T}$ are the scalars together with the sum of all the spaces $V_{0,r}$ for $r = 1,2,3, \ldots$. Two tensors of type $(0,r)$ and $(0,\bar{r})$ can be multiplied by the rule above to give a tensor of type $(0, r+\bar{r})$. This rule together with the distributive law determines the multiplication in $\mathcal{T}$. The tensor algebra $\mathcal{T}$ has these properties, which characterize it completely:

1. It contains $V = (V_{0,1})$ and a unity element, the scalar 1. The smallest algebra in $\mathcal{T}$ containing V and 1 is $\mathcal{T}$.

2. If $\mathcal{R}$ is any associative algebra with unity, any linear mapping of V into $\mathcal{R}$ can be extended to a map of $\mathcal{T}$ into $\mathcal{R}$ which preserves multiplication.

Tensor Product of Vector Spaces.—The multiplication of two tensors is a special case of the tensor product of vector spaces. If $V_1, \ldots, V_k$ are k vector spaces, their tensor product $V \otimes \ldots \otimes V_k$ is the vector space of scalar-valued functions in k variables in $V_1^*, \ldots, V_k^*$ respectively which are linear in each variable separately. If v^i lie in V_i, $i = 1, \ldots, k$, then their tensor product $v^1 \otimes \ldots \otimes v^k$ has the value $v^1(\alpha_1)v^2(\alpha_2) \ldots v^k(\alpha_k)$ on $(\alpha_1, \ldots, \alpha_k)$ in $V_1^*, \ldots, V_k^*$ respectively. The space $V_{s,r}$ is the tensor product of V^* taken s times and V taken r times.

Tensor Fields.—The tensors which arise in the study of manifolds are obtained by attaching to each point of a smooth d-dimensional manifold a d-dimensional vector space, its tangent space or space of directional derivatives. The tensors considered are those associated with each tangent space. A tensor field of type (s,r) on a manifold is a function which assigns to each point of the manifold a tensor of type (s,r) associated with the tangent space at that point. If a co-ordinate system with co-ordinate functions $x_1, \ldots, x_d$ is chosen about some point m of the manifold, then the directional derivatives $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d}$ form a basis of the tangent spaces of points near m. In terms of this basis, a tensor field of type (s,r) has components $T^{i_1, \ldots, i_s}_{i_1, \ldots, i_r}$ which are functions of the co-ordinates $x_1, \ldots, x_d$. On the whole, tensor analysis studies the analytical properties of tensor fields, while tensor algebra deals only with the algebraic properties of tensors associated with a fixed vector space.

Associated Algebras.—Closely related to the tensor algebra $\mathcal{T}$ over a vector space V are several other algebras, the most important for geometric and topological applications being the Grassman or exterior algebra over V. In this algebra, the elements to be multiplied are skew-symmetric tensors. A tensor T of type $(0,r)$ is skew-symmetric if $T(\alpha_{\pi(1)}, \alpha_{\pi(2)}, \ldots, \alpha_{\pi(r)}) = \epsilon(\pi)T(\alpha_1, \alpha_2, \ldots, \alpha_r)$; $\epsilon(\pi)$ is $+1$ or -1 depending upon whether the permutation π of the integers $1,2, \ldots, r$ is even or odd. Any tensor T of type $(0,r)$ can be made skew-symmetric by alternating it; *i.e.*, $\mathrm{ALT}_r(T)$ is skew-symmetric, where $\mathrm{ALT}_r(T)(\alpha_1, \ldots, \alpha_r) = \frac{1}{r!}\sum \epsilon(\pi)T(\alpha_{\pi(1)}, \ldots, \alpha_{\pi(r)})$, summed over all permutations π of the integers $1, \ldots, r$. The exterior algebra $\mathcal{G}$ over V is the space of skew-symmetric tensors in $\mathcal{T}$ where the multiplication is the product of two tensors made skew-symmetric. Thus if T and $\overline{T}$ are skew-symmetric tensors of type $(0,r)$ and $(0,\bar{r})$ respectively, $T \wedge \overline{T} = \mathrm{ALT}_{r+\bar{r}}(T\overline{T})$ is a skew-symmetric tensor of type $(0, r+\bar{r})$. Under this multiplication $\mathcal{G}$ becomes an associative algebra with unity of dimension 2^d which contains $V = (V_{0,1})$ and in which $v \wedge v = 0$ for each v in V.

An algebra with the symmetric tensors can be formed in a similar manner. A tensor T of type $(0,r)$ is symmetric if $T(\alpha_{\pi(1)}, \ldots, \alpha_{\pi(r)}) = T(\alpha_1, \ldots, \alpha_r)$ for every permutation π of the integers $1, \ldots, r$. Any tensor of type $(0,r)$ can be made symmetric, *i.e.*, $\mathrm{SYM}_r(T)$ is a symmetric tensor of type $(0,r)$ where $\mathrm{SYM}_r(T)(\alpha_1, \ldots, \alpha_r) = \frac{1}{r!}\sum T(\alpha_{\pi(1)}, \ldots, \alpha_{\pi(r)})$, the sum taken over all permutations π of the integers $1, \ldots, r$. The polynomial algebra over V is the space of symmetric tensors in $\mathcal{T}$ where the multiplication is the product of two tensors made symmetric. Thus if T and $\overline{T}$ are symmetric tensors of type $(0,r)$ and $(0,\bar{r})$ respectively, then $T \times \overline{T} = \mathrm{SYM}_{r+\bar{r}}(T\overline{T})$.

A generalization of Grassman algebras called Clifford algebras is important for the study of certain representations of orthogonal groups. If B is a symmetric tensor of type (2,0) over V, the orthogonal group of B is the group of nonsingular linear transformations l on V for which $B(l(v), l(v)) = B(v,v)$ for all v in V. The Clifford algebra of B, which contains V and in which $v \cdot v = B(v,v)1$, is useful for studying the spin representations of this group. The Clifford algebra of B is constructed from $\mathcal{T}$ in the following way: Let $\mathcal{J}_B$ denote the ideal in $\mathcal{T}$ generated by all elements of the form $v^2 - B(v,v)1$ for v in V; *i.e.*, $\mathcal{J}_B$ is the set of all linear combinations of elements of $\mathcal{T}$ of the form $T(v^2 - B(v,v)1)$ or $(v^2 - B(v,v)1)T$. The Clifford algebra of B is the algebra $\mathcal{T}/\mathcal{J}_B$; *i.e.*, $\mathcal{T}$ with the elements $\mathcal{J}_B$ set equal to zero. If the field of scalars is not of characteristic two, and if $e^1, \ldots, e^d$ is a basis of V, then $e^{i_1} \cdot e^{i_2} \cdot e^{i_3} \cdot \cdots e^{i_k}$ $(i_1 < i_2 < \ldots < i_k)$ is a basis of the Clifford algebra, so that the Clifford algebra has dimension 2^d. If B is identically zero, the Clifford algebra is the previously constructed Grassman algebra. *See* also DIFFERENTIAL GEOMETRY.

See C. Chevalley, *The Construction and Study of Certain Importan Algebras* (1955); N. Bourbaki, *Algebre Multilineaire* (1948). (I. M. S.)

TENSOR ANALYSIS. The concept of tensors and the knowledge of some of their properties can be traced back to Karl Gauss, Georg Riemann and Elwin Christoffel, but their algebra and analysis have been shaped into a systematic method only recently, 1900, by G. Ricci and T. Levi-Civita, who coined for this powerful branch of mathematics the name of *absolute differential calculus.*

Its chief aim is to construct and discuss relations or laws *generally covariant;* such, that is, as remain valid in passing from one to any other system of co-ordinates. This makes the subject of prime importance for differential geometry. It has especially become the object of a very widespread interest after the advent of *generalized relativity* (1916), whose principal requirement is precisely such unrestricted covariance of physical laws. There are interesting applications of tensor analysis—both finite and infinite dimensional—to classical mechanics, quantum mechanics, fluid mechanics, elasticity and aeronautics.

Definitions, Algebra of Tensors.—Consider a continuous n-dimensional manifold or "space" S_n (*see* MANIFOLDS), whose element or point $P(x_\iota)$, is determined by assigning the values of n real independent variables or co-ordinates x_ι. Let $Q(x_\iota + dx_\iota)$ be another point of S_n. Then the ordered point-pair P, Q or the set of differentials dx_ι is called a *vector.* (To begin with, the idea of "size" or "length" is foreign to this concept, S_n being thus far a nonmetrical manifold.) Let the x_ι be transformed into any other system of n co-ordinates x'_ι, the former being continuous functions of the latter with continuous derivatives $\partial x_\iota / \partial x'_\kappa$ and non-vanishing, finite Jacobian $J = \left|\frac{\partial x_\iota}{\partial x'_\kappa}\right|$; then dx_ι are transformed into

$$dx'_\iota = \frac{\partial x'_\iota}{\partial x_\alpha} dx_\alpha$$

to be summed over $\alpha=1$ to n, and vice versa, $dx_\iota=(\partial x_\iota/\partial x'_\alpha)dx'_\alpha$. (The convention will be adopted that every term in which an index occurs *twice*, is to be summed over all its values.) Any set of n magnitudes A^ι, functions of the x, which are transformed by this rule, *i.e.*, into

$$A'_\iota=\frac{\partial x'_\iota}{\partial x_\alpha}A^\alpha,$$

form a *contravariant tensor* of rank one, the A^ι being its n *components*. For contravariant tensors upper indices are used, an exception being made for dx_ι, the prototype of all such tensors. Next, any n magnitudes A_ι which are transformed as the differentiators $\partial/\partial x_\iota$, *i.e.*, into

$$A'_\iota=\frac{\partial x_\alpha}{\partial x'_\iota}A_\alpha,$$

form a *covariant* tensor of rank 1, lower indices being used for such tensors. These two kinds of tensors, of rank 1, are also termed *vectors; e.g.*, three-vectors, four-vectors (such as the relativistic four-velocity or four-potential, in space-time, S_4), etc., according as $n=3$, 4, etc. Similarly, any n^2 magnitudes $A_{\iota\kappa}$ transformable into

$$A'_{\iota\kappa}=\frac{\partial x_\alpha}{\partial x'_\iota}\frac{\partial x_\beta}{\partial x'_\kappa}A_{\alpha\beta}$$

form a covariant, and any $A^{\iota\kappa}$ transformable into

$$A'^{\iota\kappa}=\frac{\partial x'_\iota}{\partial x_\alpha}\frac{\partial x'_\kappa}{\partial x_\beta}A^{\alpha\beta}$$

form a contravariant tensor of *rank two*. Again, n^2 magnitudes $A_\iota{}^\kappa$ transformable into

$$A'_\iota{}^\kappa=\frac{\partial x_\alpha}{\partial x'_\iota}\frac{\partial x'_\kappa}{\partial x_\beta}A^\beta{}_\alpha$$

are said to form a *mixed* tensor of rank 2, covariant in ι and contravariant in κ. The extension to any rank is obvious. Any $n^{r_1+r_2}$ magnitudes $A^{ik\cdots}_{\iota\kappa\cdots}$, with r_1 lower and r_2 upper indices, which are transformed according to the rule

$$\left(A^{ik\cdots}_{\iota\kappa\cdots}\right)'=\frac{\partial x_\alpha}{\partial x'_\iota}\frac{\partial x_\beta}{\partial x'_\kappa}\cdots\frac{\partial x'_\iota}{\partial x_a}\frac{\partial x'_\kappa}{\partial x_b}\cdots A^{ab\cdots}_{\alpha\beta\cdots} \qquad (1)$$

form a mixed tensor of rank $r=r_1+r_2$. This is the most general concept of a tensor.

A tensor of rank zero, called also a *scalar*, is a single function of the x, *invariant* with respect to any transformations of co-ordinates, $f'=f$. $A_{\iota\kappa}$ is *symmetrical* if $A_{\iota\kappa}=A_{\kappa\iota}$, and *antisymmetrical* or a *skew* tensor if $A_{\iota\kappa}=-A_{\kappa\iota}$, implying $A_{\iota\iota}=0$. Similarly for $A^{\iota\kappa}$. Analogous definitions hold for mixed tensors, and for higher ranks. Symmetry and antisymmetry are invariant properties.

The transformed tensor components being linear homogeneous functions of the original ones, the sums of corresponding components of tensors of same rank and kind form again a tensor. Thus, A_ι, B_ι being covariant vectors, so is $A_\iota+B_\iota=C_\iota$. Similarly, $A_{\iota\kappa}+B_{\iota\kappa}=C_{\iota\kappa}$, etc. The addends, functions of x_ι, must be taken at the same point of S_n. If a tensor vanishes in one, it will vanish also in any other co-ordinate system. Consequently, any tensor equation, if valid in one, holds also in any other co-ordinate system. This is the chief reason of the importance of tensors in pure geometry and physics.

The *outer product* of two tensors of ranks r and s, *i.e.*, the array of n^{r+s} products of their components is again a tensor, of rank $r+s$. Thus $A_\iota B_\kappa=C_{\iota\kappa}$, $A_{\iota\kappa}B_\lambda=C_{\iota\kappa\lambda}$, $A_{\iota\kappa}B^\lambda=C^\lambda{}_{\iota\kappa}$. The *contraction*, an operation of almost magical efficiency, applicable to any mixed tensor, consists in equating one of its r_1 lower to one of its r_2 upper indices and summing over it from 1 to n. The result is again a tensor, with r_1-1 covariant and r_2-1 contravariant indices; *e.g.*, a contraction of $A^\lambda_{\iota\kappa}$ gives $A^\lambda{}_{\iota\lambda}=A_\iota$, a vector; $B^{\lambda\mu}_{\iota\kappa}$ contracted once becomes $B^{\lambda\kappa}_{\iota\kappa}=B^\lambda{}_\iota$, and this yields $B^\iota{}_\iota=B$, a scalar.

The *inner product* is the outer product (supposed mixed) contracted once or more. Thus, the inner product of A_ι, B^κ is $A_\lambda B^\kappa$, a scalar; that of $A_{\iota\kappa}$, $B^{\lambda\mu}$ gives $A_{\iota\kappa}B^{\lambda\kappa}=C^\lambda{}_\iota$, a mixed tensor, and after a second contraction, $A_{\iota\kappa}B^{\iota\kappa}=C^\iota{}_\iota=C$, a scalar. (Unlike $A_\lambda B^\kappa$, $A_\kappa B_\kappa$ is *not* invariant; it has no tensor character, and likewise for $A^\kappa B^\kappa$. Conversely, if $A_\kappa B^\kappa$ is a scalar for *any* covariant vector A_κ, B^κ is a contravariant vector. Similarly, if $A_{\iota\kappa}B^{\iota\kappa}$ is a scalar for any covariant $A_{\iota\kappa}$, $B^{\iota\kappa}$ is contravariant. This is an efficient method of establishing the tensor character of a set of n, n^2, etc., magnitudes.

Differentiation.—The differentiation of tensors with respect to the co-ordinates yields, in certain circumstances, further tensors. If f is a scalar function of the x, or scalar field, $\partial f/\partial x_\iota=f_\iota$ is a covariant vector, the *gradient* of f (but $\partial^2 f/\partial x_\iota\partial x_\kappa$ is not a tensor; again, if du be an invariant, dx_ι/du is, but d^2x_ι/du^2 is not a vector). Further, if A_ι be a vector, $B_{\iota\kappa}=\partial A_\iota/\partial x_\kappa-\partial A_\kappa/\partial x_\iota$ is a skew tensor, *the rotation* of A_ι. Finally, if $A_{\iota\kappa}$ be antisymmetric, $B_{\iota\kappa\lambda}=\partial A_{\iota\kappa}/\partial x_\lambda+\partial A_{\kappa\lambda}/\partial x_\iota+\partial A_{\lambda\iota}/\partial x_\kappa$ is again such a tensor, *the expansion* of $A_{\iota\kappa}$.

In what precedes only such properties of tensors were treated as are independent of any metrical considerations, the space S_n being thus far a nonmetrical, amorphous manifold. Let now its metrics be fixed by laying down the *line-element*, a quadratic differential form with coefficients $g_{\iota\kappa}=g_{\kappa\iota}$, prescribed functions of the x,

$$ds^2=g_{\iota\kappa}dx_\iota dx_\kappa, \qquad (2)$$

to be considered as *invariant* and to serve as the measure of the (squared) size or length of the vector dx_ι. Then, dx_ι being contravariant, $g_{\iota\kappa}$ will be covariant. Equivalently we may say that a certain symmetrical tensor $g_{\iota\kappa}$ is being impressed upon S_n as the *fundamental* or *metrical tensor*, converting it into a metrical manifold, a *Riemannian space*.

The minors $g^{\iota\kappa}$ of $g=|g_{\iota\kappa}|$, divided by g, form again a symmetrical tensor, the contravariant metrical tensor, to be used along with $g_{\iota\kappa}$. The outer product and contraction can be used in raising and lowering tensor indices. Thus $g^{\iota\kappa}A_\kappa=A^\iota$ and $g_{\iota\kappa}B^\kappa=B_\iota$.

The angle θ between two (copunctal) vectors A^ι and B^ι is defined by the invariant $\cos\theta=g_{\iota\kappa}A^\iota B^\kappa/AB$; A, B being their sizes. The unit vector $\dot{x}_\iota=dx_\iota/ds$ determines, locally, a *direction* in S_n. The angle between two directions $\dot{x}_\iota$, $\dot{y}_\iota$ is given by $\cos\theta=g_{\iota\kappa}\dot{x}_\iota\dot{y}_\kappa$.

Integrals.—The integral $\int dx_1, \cdots, dx_n$, briefly $\int dx$, extended over a region of S_n, is transformed into $\int Jdx'$, and the determinant g into $g'=J^2g$. Consequently, $\int\sqrt{g}dx$ is an invariant metrically impressed upon that region, its size or *volume* (area, if $n=2$). This concept is readily extended to any submanifold of S_n characterized by x_ι as functions of $m<n$ parameters p_a, the rôle of metrical tensor being taken over by

$$h_{ab}=g_{\iota\kappa}(\partial x_\iota/\partial p_a)(\partial x_\kappa/\partial p_b).$$

If, by a proper choice of the co-ordinate system, all components $g_{\iota\kappa}$ are reducible to constants, the metrical space S_n is Euclidean or homaloidal (flat). This is but a very special case of a Riemannian space. In general such a reduction is not possible, and S_n is non-Euclidean (*v. infra*).

Differentiation of tensors, aided by metrics, yields an unlimited number of new tensors. The oldest of such is Riemann's set of four-index symbols, of 1861. The simplest metrically differential tensor, however, was discovered in 1869 by Christoffel. This is *the covariant derivative* of a vector A_ι,

$$A_{\iota,\kappa}=\frac{\partial A_\iota}{\partial x_\kappa}-\{\iota\kappa,\lambda\}A_\lambda, \qquad (3)$$

where $\{\iota\kappa,\lambda\}=g^{\lambda\mu}[\iota\kappa,\mu]=\{\kappa\iota,\lambda\}$ and

$$[\iota\kappa,\mu]=\tfrac{1}{2}(\partial g_{\mu\iota}/\partial x_\kappa+\partial g_{\kappa\mu}/\partial x_\iota-\partial g_{\iota\kappa}/\partial x_\mu)$$

are Christoffel's symbols (*not* forming a tensor). Again,

$$B^\lambda{}_{,\kappa}=\partial B^\iota/\partial x_\kappa+\{\alpha\kappa,\iota\}B^\alpha$$

is a mixed tensor.

$$C,^{\iota\kappa}{}_\lambda=\partial C^{\iota\kappa}/\partial x\lambda+\{\lambda\alpha,\iota\}C^{\alpha\kappa}+\{\lambda\alpha,\kappa\}C^{\iota\alpha}, \qquad (4)$$

$$C_{\iota\kappa},\lambda=\partial C_{\iota\kappa}/\partial x^\lambda-\{\iota\lambda,\alpha\}C_{\alpha\kappa}-\{\kappa\lambda,\alpha\}C_{\iota\alpha} \qquad (5)$$

are the covariant derivatives of $C^{\iota\kappa}$ and $C_{\iota\kappa}$. The covariant

derivative of $g_{\iota\kappa}$ itself vanishes identically. Noteworthy for their applications are

$$A^{\iota}=\mathrm{Div}(A^{\iota\kappa})\equiv\frac{1}{\sqrt{g}}\frac{\partial}{\partial x_\kappa}(\sqrt{g}A^{\iota\kappa}) \text{ and } \mathrm{div}(A^{\iota})\equiv\frac{1}{\sqrt{g}}\frac{\partial}{\partial x_\iota}(\sqrt{g}A^{\iota}),$$

the vector divergence of a skew $A^{\iota\kappa}$ and *the scalar divergence of* A^{ι}; *e.g.*, one group of Maxwell's equations is represented by $\mathrm{Div}(F^{\iota\kappa})=C^{\iota}$, where C^{ι} is the four-current and $F^{\iota\kappa}$ the electromagnetic skew tensor (six vector); the second group being expressed by equating to zero the "expansion" (*v. supra*) of $F_{\iota\kappa}=g_{\iota\alpha}g_{\kappa\beta}F^{\alpha\beta}$.

Geodesics, Null lines.—The most characteristic curves of the metrical S_n are its *null* or *minimal lines*, $ds=0$, and "shortest" curves or *geodesics* defined by $\delta\int ds=0$, with fixed terminals. The equations of a geodesic are

$$\ddot{x}_\iota+\{\alpha\beta,\iota\}\dot{x}_\alpha\dot{x}_\beta=0, \tag{6}$$

where $\dot{x}=dx/ds$. They determine, *e.g.*, in the case of space time, the motion of a free particle in the metrical, and gravitational, field $g_{\iota\kappa}$. The null lines are imaginary or real according as (2) is a definite or nondefinite form. The former is the case of spaces proper as contemplated by the pure geometer, and the latter that of space time with one positive and three negative $g_{\kappa\kappa}$. The null lines of space time represent light propagation. The definiteness or nondefiniteness of the quadratic form and its *index of inertia* (number of negative $g_{\kappa\kappa}$'s) are invariant properties.

Parallel Displacement of Vectors.—A fruitful contribution to tensor analysis, due to Levi-Civita (1917) is the concept of parallel displacement of a vector with respect to a curve. The metrical S_n can in general be imagined as imbedded locally in a Euclidean space E of $\frac{1}{2}n(n+1)$ dimensions. Levi-Civita defined parallel displacement in S_n as a consequence of a parallel displacement in E. In spite of the ultra-spacial construction, parallelism is intrinsic, expressible by the S_n-metrics alone.

If $x_\iota=f_\iota(s)$ are the parametric equations of any chosen curve C, then the differential equations along C

$$\frac{dA^{\iota}}{ds}+\{\alpha\beta,\iota\}A^{\alpha}\dot{x}^{\beta}=0 \tag{7}$$

with the initial vector $A^{\iota}(s_0)=A^{\iota}_0$ determine uniquely a contravariant vector A^{ι} at each point of the curve. The vector A^{ι} at the point corresponding to $s=s$ is said to be obtained by a parallel displacement of the vector A^{ι}_0 with respect to the given curve C, and is said to be parallel to A^{ι}_0. Any two vectors generated as above, one at each point, are parallel with respect to C.

Let it be required to construct a curve whose tangents follow from each other by parallel displacement along the curve itself. This gives the equations (6), so that the curve is a geodesic. The latter is thus throughout "parallel to itself"—an interesting generalization of the property of Euclidean straights. Accordingly, a vector transferred along a geodesic remains equally inclined to it.

If a vector X^{ι} be carried by parallel displacement along two different paths a, b from o to o^1, the two vectors at o^1, equal in size, generally differ in direction. One is parallel to X^{ι} via a, the other via b. In fine, parallelism depends on the route of transfer or is nonintegrable (unless S_n is essentially Euclidean). Thus also, if the vector be carried around a circuit, it returns at O with its direction changed.

The covariant derivative (*v–supra*) of the vector is the tensor $X_{\iota,\alpha}$. Its covariant derivative $X_{\iota\ \alpha,\beta}$ is not symmetric in α and β. A straightforward calculation shows that

$$X_{\iota,\beta,\alpha}-X_{\iota,\alpha,\beta}=R^{\kappa}_{\iota\alpha\beta}X_\kappa, \tag{8}$$

where

$$R^{\kappa}_{\iota\alpha\beta}=\frac{\partial}{\partial x_\alpha}\{\iota\beta,\kappa\}-\frac{\partial}{\partial x_\beta}\{\iota\alpha,\kappa\}+\{\iota\alpha\ \gamma\ \kappa\}\{\iota\beta,\gamma\}-\{\beta\gamma,\kappa\}\{\iota\alpha,\gamma\}. \tag{9}$$

This is the Riemann-Christoffel or curvature tensor whose vanishing is the necessary and sufficient condition for the reducibility of $g_{\iota\kappa}$ to a constant tensor, or the criterion of a homaloidal (Euclidean) space.

Riemann's Four-index Symbols and Curvature.—Riemann's own set of *four-index symbols* is the covariant tensor $(\iota\mu,\ \lambda\kappa)\equiv R_{\iota\mu\lambda\kappa}=g_{\mu\alpha}R^{\alpha}_{\iota\lambda\kappa}$. Conversely, $R^{\alpha}_{\iota\lambda\kappa}=g^{\mu\alpha}(\iota\mu,\ \lambda\kappa)$. Like (9), the symbols are antisymmetric in κ, λ. Three more linear relations hold between them. This reduces the number of independent symbols to $\frac{1}{12}n^2(n^2-1)$; *e.g.*, six for a three-space, twenty for an S_4, and but *one* for a surface, say (12, 12). This symbol divided by g is a scalar of the surface, its *Gaussian curvature* (K.) (*See* DIFFERENTIAL GEOMETRY.) If a vector is parallelly displaced around a circuit on a surface, it returns with its direction changed by an amount equal to $\int Kd\sigma$, the Gaussian curvature of the surface integrated over the surface area enclosed by the circuit. As a consequence, *the excess* (over π) of the angle sum in any geodesic triangle is $\int Kd\sigma$, a famous theorem due to Gauss.

In the case of three or more dimensions the curvature properties can no longer be expressed by a single magnitude, but require for their description the knowledge of the whole curvature tensor or the associated Riemann symbols. The concept of Gaussian curvature is now replaced by that of *Riemannian curvature*. This is, at any point $O(x)$ of S_n, the set of Gaussian curvatures K of geodesic surfaces of all possible orientations laid through O. If $h_{\alpha\beta}$ be the metrical tensor of such a surface as submanifold, then $K=\frac{1}{h}(12,\ 12)_k$, the symbol to be calculated with $h_{\alpha\beta}$. This and the determinant h can be expressed in terms of the tensor $g_{\iota\kappa}$ of the manifold and the vector pair, ξ^{ι}, η^{ι}, fixing the orientation. The result is

$$K=\frac{(\iota\kappa,\ \lambda\mu)\ \xi^{\iota}\eta^{\kappa}\xi^{\lambda}\eta^{\mu}}{(g_{\iota\lambda}g_{\mu\kappa}-g_{\iota\mu}g_{\kappa\lambda})\ \xi^{\iota}\eta^{\kappa}\xi^{\lambda}\eta^{\mu}}, \tag{10}$$

In general, K will depend on position and on orientation. In other words, with regard to curvature, S_n may be nonhomogeneous as well as anisotropic (*e.g.*, space time within or around matter); but if K is everywhere isotropic, it is also constant throughout S_n. This is Schur's theorem. By (10) the necessary and sufficient condition for isotropy of Riemannian curvature becomes $(\iota\lambda,\ \kappa\mu)=K(g_{\iota\kappa}g_{\lambda\mu}-g_{\mu\iota}g_{\lambda\mu})$ or $R^{\alpha}_{\iota\kappa\lambda}=K(\delta^{\alpha}_{\kappa}g_{\iota\lambda}-\delta^{\alpha}_{\lambda}g_{\iota\kappa})$ with constant K.

The tensor (9), being mixed, yields the contracted curvature tensor $R^{\alpha}_{\iota\kappa\alpha}=R_{\iota\kappa}$ which turns out to be symmetrical. Its scalar, $R=g^{\iota\kappa}R_{\iota\kappa}$, is the *curvature invariant* of the manifold. For an isotropic n-fold, $R=-(n-1)nK$ and $R_{\iota\kappa}=g_{\iota\kappa}R/n$. A capital use of $R_{\iota\kappa}$ was made by Albert Einstein for constructing the gravitational field equations, $R_{\iota\alpha}-\frac{1}{2}Rg_{\iota\alpha}=-\kappa T_{\iota\alpha}$, where $\kappa\infty$ gravitation constant, and $T_{\iota\alpha}$ is the tensor of matter, embodying energy, momentum, and stress. The covariant derivative of $R_{\iota\kappa}$ is connected with grad R by the n relations $g^{\lambda\lambda}R_{\iota\kappa,\lambda}=\frac{1}{2}\partial R/\partial x_\iota$, which were again utilized by Einstein and which follow from the identical relations $(\iota\kappa,\ \lambda\mu),_{\nu}+(\iota\underline{x},\ \underline{\mu\nu}),_{\lambda}+(\kappa\iota,\ \nu\lambda),_{\mu}=0$. These remarkable identities, discovered by Luigi Bianchi, give also a very simple proof of Schur's theorem.

Non-Riemannian Tensor Analysis.—The year 1918 saw the beginning of epoch-making generalizations of the Riemannian tensor analysis. It was first observed that many of the results of the Riemannian tensor calculus were independent of the metric structure of the Christoffel symbols $\{\iota\kappa,\lambda\}$ in terms of the metrical tensor $g_{\iota\kappa}$, but were merely dependent on the law of transformation of the Christoffel symbols under a transformation of co-ordinates. There thus arose the fruitful conceptions of non-Riemannian geometries and tensor calculi in which the fundamental geometric object is not a metrical tensor $g_{\iota\kappa}$ but a linear connection with components $\Gamma^{\lambda}_{\iota\kappa}$, which undergo the non-tensorial transformation

$$\bar{\Gamma}^{\gamma}_{\alpha\beta}=\Gamma^{\lambda}_{\iota\kappa}\frac{\partial x_\iota}{\partial\bar{x}_\alpha}\frac{\partial x_\kappa}{\partial\bar{x}_\beta}\frac{\partial\bar{x}_\gamma}{\partial x_\lambda}+\frac{\partial^2 x_\lambda}{\partial\bar{x}_\alpha\partial x_\beta}\frac{\partial\bar{x}_\gamma}{\partial x_\lambda} \tag{11}$$

under a transformation of co-ordinates x_ι to new co-ordinates $\bar{x}_\iota$. This transformation law is the same as the one for the Christoffel symbols $\{\iota\kappa,\lambda\}$ in the Riemannian tensor analysis.

Both the symmetric case $\Gamma^{\lambda}_{\iota\kappa}=\Gamma^{\lambda}_{\kappa\iota}$ and the asymmetric case

have been studied in many papers and books. The numerous attempts at unified field theories in general relativity theory have often shown the need for further researches in non-Riemannian tensor analysis. A historical survey by E. Schrödinger in *Nature* (1944) is illuminating.

One of the most interesting applications of non-Riemannian tensor analysis was made by E. Cartan (1927) to the differential geometry of an n-dimensional continuous group manifold. There exist two fundamental asymmetric (nonsymmetric) connections which in turn define uniquely the symmetric linear connection with the property that the covariant derivative of the curvature tensor is zero. In the special case of a semisimple group, the group manifold reduces to a Riemannian space which is not necessarily of constant Riemannian curvature.

General Tensor Analysis.—The differential calculus and theory of functions used in most of the topics discussed so far are classical (prior to 1906). The phenomenal rise of general function theory and topology in the decade 1927–37 was accompanied by a searching investigation of the foundations of differential geometry and tensor analysis in n-dimensional spaces. This necessitated the introduction of the theory of topological spaces into the treatment of co-ordinates and co-ordinate transformations.

Above all, the influence of general function theory and topology was marked in general differential geometry and its accompanying general tensor analysis. The co-ordinate spaces of such a general tensor analysis must be capable of supporting an appropriate differential calculus. It is for this reason that the permissible co-ordinate spaces are linear topological spaces and especially normed linear spaces, *i.e.*, in the place of the set $x = (x_1, x_2, \ldots, x_n)$ of the n numerical co-ordinates x_i of the classical cases we now have an element x of these more general mathematical spaces—which in some instances may be function spaces or other infinitely dimensional spaces. There are numerous applications of general tensor analysis to the vibrations of elastic media and to the mechanics of continuous media in general.

BIBLIOGRAPHY.—G. Ricci and T. Levi-Civita, *Méthodes de Calcul différentiel absolu et leurs applications, Math. Annalen* (1901); A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik* (1916); G. F. B. Riemann, *Ueber die Hypothesen, welche der Geometrie zu Grunde liegen* (2nd ed., 1921); H. Weyl, *Raum-Zeit-Materie* (5th ed., 1923), *Mathematische Analysis des Raumproblems* (1923); A. S. Eddington, *The Mathematical Theory of Relativity* (2nd ed., 1924); T. Levi-Civita, *Lezioni di Calcolo differenziale assoluto* (1925), Eng. trans., M. Long (1927); L. P. Eisenhart, *Riemannian Geometry* (1926), *Non-Riemannian Geometry* (1927), *Continuous Groups of Transformations* (1933); O. Veblen, *Invariants of Quadratic Differential Forms* (1927); E. Cartan, *Lecons Sur La Géométrie Des Espaces De Riemann* (1928), *La Théorie des Groupes Finis et Continus et l'Analysis Situs* (1930); A. D. Michal, "Affinely Connected Function Space Manifolds," *Amer. Journ. Math.* (1928), "General Differential Geometries and Related Topics," *Bull. Amer. Math. Soc.* (1939), *Matrix and Tensor Calculus with Applications to Mechanics, Elasticity and Aeronautics* (1946), "Studies on Geodesics in Vibrations of Elastic Beams," *Proc. Nat. Acad. Sci.* (1945); A. J. McConnell, *Applications of the Absolute Differential Calculus* (1931); H. Hopf and W. Rinow, "Ueber den Begriff der vollständigen differentialgeometrischen Fläche," *Commentarii Mathematici Helvetici* (1931); O. Veblen and J. H. C. Whitehead, *The Foundations of Differential Geometry* (1932); T. Y. Thomas, *The Differential Invariants of Generalized Spaces* (1934), "Recent trends in Geometry," *Amer. Math. Soc. Semicentennial Publ.*, vol. ii (1938), "Imbedding Theorems in Differential Geometry," *Bull. Amer. Math. Soc.* (1939), "Qualitative Analysis of the Flow of Fluids in Pipes," *Amer. Journ. Math.* (1942); J. A. Schouten and D. J. Struik, *Einführung in die Neueren Methoden der Differential geometrie*, vol. i (1935), vol. ii (1938); L. Brillouin, *Les Tenseurs En Mécanique Et En Élasticité* (1938); E. Schrödinger, "The Union of the three fundamental fields (gravitation, meson and electromagnetism)," *Proc. Roy. Irish Acad.* (1944), "On Distant Affine Connection," *ibid.* (1945); G. C. Hall, *Matrices and Tensors* (1963). (L. SN.; A. D. MI.)

TENTERDEN, CHARLES ABBOTT, 1st Baron (1762–1832), lord chief justice of England, was born at Canterbury on Oct. 7, 1762, the son of a hairdresser. He was educated at King's school, Canterbury, and Corpus Christi college, Oxford, and entered at the Middle Temple in 1787. For several years he practised as a special pleader under the bar, and was finally called at the Inner Temple in 1796. He joined the Oxford circuit and made rapid headway. In 1801 he was appointed recorder of Oxford. In 1802 his *Law Relative to Merchant Ships and Seamen* appeared. In 1808 he refused a seat on the bench; this, however, he accepted in 1816, being made a judge of the court of common pleas. On the resignation of Lord Ellenborough in 1818 he was promoted to the chief justiceship of the king's bench, and presided over several important state trials, notably that of Arthur Thistlewood (*q.v.*) and the Cato street conspirators (1820). He was raised to the peerage in 1827 as Baron Tenterden of Hendon. Never a great lawyer, Tenterden made his way by steady hard work. He was an uncompromising Tory and had no sympathy with the reform of the criminal law carried out by Sir Samuel Romilly. He strongly opposed the Catholic Relief bill and the Reform bill. Tenterden died on Nov. 4, 1832.

See E. Foss, *The Judges of England* (1870).

TENTERDEN, a municipal borough in the Ashford parliamentary division of Kent, Eng., 19 mi. S.S.E. of Maidstone. Pop. (1961) 4,935. An agricultural centre, it lies on a hill above the Newmill channel, a tributary of the Rother, whose fertile valley (Rother levels) was once an estuary; in the 18th century the sea was within two miles of Tenterden, which was a limb of the Cinque Port of Rye. St. Mildred's church is mainly Early English, and its Perpendicular tower is associated with the legend of Goodwin sands, which tells how the abbot of St. Augustine's, Canterbury, diverted funds for repairing the sea wall protecting Earl Godwine's island in order to build the steeple; in 1099 an inundation took place, and "Tenterden steeple was the cause of the Goodwin Sands." The nearby church of High Halden has an octagonal wooden tower with a belfry of wooden tiles (shingles). In 1449 Tenterden was incorporated with Rye and granted a charter which did not function until 1463. In 1600 it was reincorporated under the "Mayor, Jurats and Commons" as the town and hundred of Tenterden, the members of the corporation ranking as barons of the Cinque Ports. In this capacity they had the privilege of attending the court of Sheppey to elect the lord wardens.

Smallhythe place, 2 mi. S., is maintained by the National trust as the Ellen Terry memorial museum, having been her residence for 25 years.

TENURE, JUDICIAL. Essential to the maintenance of an independent judiciary is a tenure of office adequate to ensure freedom from political and economic pressure. The truth of this proposition was recognized by the authors of the constitution of the United States, and all judges of federal courts created pursuant to art. iii hold office during "good behaviour"; *i.e.*, for life. Most other federal judges are granted similar tenure by statute. Thus, though a federal judge is eligible to retire at full salary after a given number of years of service, he may not be compelled to surrender his office except for misbehaviour. Similar constitutional and statutory protection is afforded judicial compensation, which may not be reduced during the tenure of an incumbent. Both the federal and English judicial systems have demonstrated the value of and necessity for such extended tenure of office, even at the occasional price of retaining a superannuated or less competent judge.

The state court systems in the U.S. have lagged far behind, with the result that political influence is rife in the administration of justice in the states. Only two states, Massachusetts and New Hampshire, have absolute life tenure, but Rhode Island has it in fact if not in law. The terms of judges in other state courts vary from 2 years in the Vermont trial and appellate courts to 21 years in the supreme court of Pennsylvania. Life tenure is usually associated with a judiciary appointed by the executive, as in the federal and English systems. Where the judiciary is elected, terms of office have been severely restricted as part of the Jacksonian notion of democratic control of the courts. Most reform movements do not, however, recommend life tenure. They seek a system whereby an incumbent is required to run for reelection on the basis of his record, giving the electorate a choice of retention or rejection. This method of tenure, generally known as the Missouri plan, has the support of the American Bar association, among others, although there is no evidence to support the proposition that the Missouri plan has any advantages over the life tenure system in effect in the federal courts. (P. B. K.)

England.—Until the end of the 17th century most judges were appointed *durante bene placito* ("during the royal pleasure"); and there were many instances of judges being dismissed. When the Act of Settlement, 1701, s. 3, came into force in 1714, judges were appointed *quamdiu se bene gesserint* ("during good behaviour"), and they were removable (and probably removable only)

upon the address of both houses of parliament. The Judicature act, 1925, s. 12, preserves this position, which applies to all judges of the high court and court of appeal; and by ss. 13–15 their salaries and pensions were established and charged on the consolidated fund, so that they are not subject to the annual vote of parliament. However, these provisions are subject to the omnipotence of parliament, and statute has reduced and, more recently, increased the salaries and pensions. All judges appointed since the Judicial Pensions act, 1959, must retire when they become 75 years old. A county court judge is removable by the lord chancellor for "inability or misbehaviour," and he must retire when he is 72 years old, subject to a possible extension to 75 (County Courts act, 1959, s. 8). The lord chancellor similarly has power to remove any justice of the peace, and many other of the lower judicial officers. The independence of English judges is thus secured less by formal provisions of law than by the public opposition likely to be aroused by any improper attempt to coerce them. (R. E. My.)

TENURE, LAND: *see* Land Tenure: Economic and Agrarian Aspects; Land Tenure, Primitive.

TENZING NORGAY (1914–), a Sherpa mountaineer, the first man, with Sir Edmund Hillary, to set foot on the summit of Everest (29,028 ft.), was born in 1914 (he believes), in Solo Khumbu, the district south of Everest which is inhabited by the Sherpas. He was named Namgyal Wangdi. At the age of 18 he migrated to Darjeeling, where he lived by carrying loads, and in 1935 and 1936 went as porter on Everest expeditions. In 1938 he carried to Camp VI (27,200 ft.). After World War II he set up as sirdar, or organizer of porters, and accompanied a number of expeditions. In 1952 the Swiss made two attempts on the southern route up Everest, on both of which Tenzing was sirdar. In 1953 he went as sirdar of the British expedition, and formed the second summit pair with Hillary. From a tent at 27,900 ft. on the southeastern ridge they reached the summit at 11:30 A.M. on May 29.

The qualities which distinguished Tenzing from other Sherpas were his mountaineering skill and his ambition to reach the summit. After Everest he received great honours and is regarded with religious awe by many Nepalese and Indians. He has not, however, lost his native simplicity and sincerity of character.

See Tenzing Norgay, *Man of Everest*, as told to James Ramsey Ullman (1955). (C. W. F. N.)

TEOSINTE (*Euchlaena mexicana*), a tall, stout, annual grass, native to Mexico, very closely allied botanically to maize (*see* Corn). Usually branching at the base, it grows in large clumps, with tassels (staminate flowers) like those of maize. The bundles of fruiting spikes, enclosed in husks, with the silk (long styles) hanging from the upper ends, are somewhat similar to the ears of maize. Teosinte is grown in Mexico as a fodder plant and sparingly in southern United States as a soiling crop.

In Mexico, hybrids of teosinte and maize are found, giving rise to the view that the two plants have developed from a common ancestral form.

TEPEXPAN MAN is one of the most important skeletal remains of early American Indians because of the good data concerning its geological position and contemporary faunal remains. It was found in 1947 by Helmut de Terra, near the village of Tepexpan in the Valley of Mexico. The skeleton is that of an adult male between 55 and 65 years of age. The skull is large, with an endocranial capacity of 1,540 c.c. The cranial index is 79.89 (breadth/length × 100).

Most other crania of early American Indians, such as those found at Lagoa Santa, Braz., and Punín, Ecuador, have been narrower, with an index below 75. There is always some variation within a group, and this relatively broader-headed individual could easily have been a member of a group whose mean cranial index was much less. A stature of 170 cm. (5 ft. 7 in.) has been estimated from the long bones. This stature is greater than that of most pre-Columbian Indians, who are closer to 160 cm.

The skeleton lay face down and flexed at a depth of 14 in. in silt deposits of the El Riso sand, a lake facies of the Upper Becerra formation. This silt suggests a moist period and was overlain by *caliche* deposits suggesting a dry climatic period. Geologically these deposits permitted an estimate of 11,000 to 12,000 years ago for the silt. A radiocarbon date of 11,003 ± 500 years ago has been derived from a peat specimen belonging to the same level as that of the skeleton.

A number of fossil elephant remains and an obsidian flake were found in related deposits. These, and bones of horse, bison, pocket gopher and deer indicate a habitat with more luxuriant vegetation than in modern times and indicate that this man may have been a big-game hunter.

See Helmut de Terra, Javier Romero and T. D. Stewart, *Tepexpan Man* (1949); H. M. Wormington, *Ancient Man in North America* (1957). (W. S. Ln.)

TEPHILLIN: *see* Phylactery.

TEPLICE-SANOV (Teplitz-Schönau), a town in North Bohemia, Czech., in the valley of the Bilina. The saline-alkaline waters of the thermal mineral springs were used in the treatment of gout, rheumatism and some skin complaints and as an "after-cure" to Karlovy Vary (Carlsbad), etc.

Although traditionally believed to have been known and used since the 8th century, the springs are not historically mentioned until the 16th century, and the function of the town as a watering place has tended to disappear in modern times in the face of the advance of industry. The rich lignite field of Brüx-Komotau-Teplitz led to the establishment of numerous manufactures. Hardware, machinery, glass, pottery, chemicals, cotton and woolen goods are among the principal products and in all of these Teplice has a brisk trade. Industrial life grew mainly in the suburbs. Pop. (1961) 44,136.

TEQUILA, *see* Alcoholic Beverages, Distilled: *Classification of Potable Spirits.*

TERAMO, an episcopal see of Abruzzi e Molise, Italy, the capital of the province of Teramo, 16 mi. W.S.W. by rail of Giulianova, a junction on the Ancona-Brindisi railway. Pop. (1957 est.) 42,070 (commune). The town stands on the left bank of the Tordino, where it is joined by the Vezzola, at an altitude of 876 ft. above sea level.

The cathedral has a Romanesque Gothic portal of 1332 by a Roman marble worker named Deodatus; it contains a splendid silver antependium by Nicolo da Guardiagrele (1433–48). The tower is fine. The church of S. Antonio is also in the Romanesque Gothic style. Under the church of S. Anna dei Pompetti remains of Roman houses and of the original cathedral have been discovered. Two miles away is an astronomical observatory, presented to the state by Vincenzo Cerulli in 1917. The ancient Interamna Praetuttiorum was the chief town of the tribe of the Praetuttii.

TERAPHIM, a Hebrew word, found only in the plural, of uncertain etymology. The name appears to be applied to some form of idol (*cf.* Gen. xxxi, 19 and 30), but details as to its precise configuration, etc., are lacking. From I Sam. xix, 13, 16 it would seem that in the early monarchical period a regular place in every household was still reserved for the teraphim; while in the 8th century, Hosea (iii, 4) speaks of "ephod and teraphim" as essential elements in the national worship. Later the teraphim were banned by the prophets. The meaning of the story in Gen. xxxv, 2–4 clearly is that the employment of teraphim was given up by Israel in order that they might serve Jehovah alone at Bethel. In Judges and Hosea the teraphim are closely associated with the ephod; both are mentioned in connection with divination (*cf.* II Kings xxiii, 24; Ezek. xxi, 21 and 26; Zech. x, 2). In view of Ezek. xxi, 21 and Hosea iii, 4, it is difficult to suppose that the teraphim were purely household idols.

TERATOLOGY, the study of abnormalities and monstrosities in plants and animals. *See* Monster.

TERBIUM (symbol Tb, atomic number 65, atomic weight 158.93, stable isotope Tb^{159} [100%]) is a very rare metallic element of the rare-earth group. It was discovered in 1843 by C. G. Mosander, who originally named the oxide erbia, but since 1877 it has been known as terbia. It was first obtained fairly pure by G. Urbain in 1905. It is best separated from other members of the group by the fractional crystallization of the bromates. The ignition of the oxalate or sulfate in air gives a dark brown oxide,

Tb_4O_7, in which part of the terbium is believed to have a valency of 4; this oxide slowly decomposes when treated with dilute acids giving the trivalent ion in solution. The dark brown oxide can be reduced to Tb_2O_3 by heating in a stream of hydrogen. The trivalent compounds are very pale pink but are characterized by having a rather faint absorption band in the blue region of the spectrum. The salts are paramagnetic. Since 1953 kilogram quantities in high purity have been separated by ion-exchange methods. The metal can be produced by reducing the anhydrous fluoride with calcium, and is a silvery metal which melts in the neighbourhood of 1,360° C. and boils in the neighbourhood of 2,800° C. It crystallizes in the hexagonal close-packed system a = 3.6010 Å, c = 5.6936 Å, and has a calculated density of 8.253 g. per cubic centimetre. *See* RARE EARTHS. (F. H. SP.)

TERBORCH (TER BORCH), **GERARD** (1617–1681), Dutch painter, who developed his own distinctive type of interior genre, in which he depicted with grace and fidelity the atmosphere of well-to-do middle-class life, was born at Zwolle. His father had been an artist and had visited Rome, but from 1621 was employed as a tax collector. Surviving drawings made by the young Terborch in 1625 and 1626 are proudly inscribed and dated by his father. In 1632 Gerard was in Amsterdam and in 1634 he was a pupil of Pieter Molyn (Molijn) in Haarlem. He visited England in 1635, Rome in 1640, and from 1646 spent two or three years in Münster, Westphalia, where the peace congress was in session. The masterpiece of this period is the "Peace of Münster" (National Gallery, London), of 1648, which portrays the delegates of Holland and of Spain assembled to sign the peace treaty. After a stay in Madrid he had finally returned to his own country by the end of 1650. In 1654 he was married in Deventer, where he settled for the remainder of his life, dying there on Dec. 8, 1681.

Terborch's works consist almost equally of portraits and genre pieces. His characteristically delicate technique can be appreciated in the portraits, which are painted on a small, almost miniature scale, though many of them are full-lengths. In colour they tend to be subdued, due largely to the sober costume of the times, but by subtlety of tonal gradations and mastery in rendering diverse surface textures he was able to achieve an extraordinary richness of effect. His superb colour sense appears to greater advantage in genre subjects, though always employed with masterly restraint. In his earlier years he painted many guardroom subjects in the manner of Pieter Codde and Willem Duyster, but later, from about the time when he finally settled in Holland, he painted calm, exquisitely drawn groups, posed easily and naturally against shadowy backgrounds, and imbued with an almost aristocratic elegance that is unique among Dutch painters of his time. Among many fine examples of Terborch's art may be mentioned "The Letter" (Buckingham palace, London), "The Concert" (Berlin), "Fatherly Advice" (Amsterdam) and "Woman Peeling Apples" (Vienna).

See E. Plietzsch, *Gerard Ter Borch* (1944). (R. E. W. J.)

TERCEIRA, an island in the Atlantic ocean, belonging to Portugal, and forming part of the Azores archipelago. Pop. (1950) 60,372; area, 153 sq.mi. Terceira, *i.e.*, "the third," was so called as being the third island of the archipelago to be discovered by the Portuguese. From its central position it was long the seat of administration, but its capital, Angra, has lost much of its commercial importance. The other chief towns are Ribeirinha (2,658), and Praia da Vitoria (2,485). *See* AZORES.

TEREBINTH, botanical name *Pistacia terebinthus,* a member of the cashew family (Anacardiaceae) usually a small tree, common in the Mediterranean area. It has a purplish-gray bark and compound leaves with two to four pairs, and an odd terminal one. The very small numerous unisexual flowers are borne on panicles. The fruit is a small roundish bright red drupe with a scanty pulp. A yellow to brown resinous exudation, known as Chian, Scio or Cyprus turpentine, obtained by cutting the stem, is of minor commercial importance (*see* TURPENTINE). Allied species include *P. lentiscus,* the mastic tree, and *P. vera,* which yields the pistachio (*q.v.*) nut.

TEREDO, a genus of highly specialized wormlike bivalve mollusks (*q.v.*) belonging to the family Teredinidae, commonly known as shipworms or pileworms. They are related to the piddocks (*q.v.*) and to the soft-shelled clam, *Mya.* There are about 10 genera, the best known being *Teredo* and *Bankia* and perhaps 100 species all referred to as teredos or teredines. They are worldwide in distribution, generally living in marine or brackish water, though a few species are found in fresh water. They may be extremely destructive to wooden vessels or water-front installations and have been of great concern to maritime peoples since earliest times. They have been called "termites of the sea" because the damage they do is not readily seen; the outer surface of the wood may look undamaged and yet the interior may be completely riddled.

Teredos may be compared to a soft-shelled clam, the body of which has been stretched out so that the highly modified shell covers only a very small portion of the anterior end of the animal. The body is wormlike and, unlike other bivalves, most of the vital organs are posterior to the posterior adductor muscle. The separate siphons, used in feeding and respiration, are relatively short and protrude from the minute opening of the burrow into the water. The incurrent siphon carries water and food in; the excurrent carries wastes out. The valves of the shell are nearly hemispherical, with a deep right-angled notch in the ventral half of the anterior margin. Parallel with the edges of the notch, the outer surface of each valve is marked with fine ridges which when magnified appear as rows of minute teeth like those of a file. The large muscular foot, which protrudes through the anterior gape, acts as a suction disc to hold the valves tightly against the anterior end of the burrow when the teredo is actively boring. As in the piddocks, the anterior adductor muscle has migrated to the outside of the shell. Posteriorly a large lobe or auricle is produced for the attachment of the powerful posterior adductor muscle. At the umbos and on the ventral margins of the valves opposite to them are large rounded knobs or condyles on which the valves rock. The large foot muscles are attached to the apophyses, two calcareous "arms" which protrude from under the umbos. The large posterior adductor muscle supplies the force for boring and the shell is the tool with which it works. When actively boring, the teredo contracts the anterior adductor muscle bringing the anterior end of the valves together and spreading the posterior end. The foot is firmly attached to the end of the burrow and the foot muscles contracted, drawing the valves forward and bringing the denticles hard against the wood. Then, the forceful contraction of the posterior adductor muscle spreads the valves anteriorly and the filelike denticles scrape against the wood, rasping off fine particles. The rhythmic contraction and relaxation of the anterior and posterior adductor muscles may be repeated many times before the animal rests. The surface of the foot and the margins of the mantle are covered with cilia which carry the fine particles of wood into the mantle cavity and to the mouth.

At the posterior end of the stomach is a large sac or caecum which is always filled with wood particles. Digestive enzymes in the caecum can reduce the wood sugars to a usable form so that some of the food of the teredo is obtained in this manner. Teredos also feed on microscopic plants and animals carried into the mantle cavity through the incurrent siphon.

When shipworms are undisturbed, the siphons can be seen extending from the minute opening at the surface of the wood and waving about gently in the water. If they are disturbed, however, or conditions become adverse, the siphons are drawn into the burrow and the opening is plugged by a pair of paddle-shaped plates (pallets). The pallets effectively protect the animal against predators as well as adverse conditions such as reduced salinity, pollution or silting. Inactive shipworms may remain for extended periods of time with the siphons retracted.

The breeding habits of the various members of the family vary considerably. In some species the eggs and sperm are emitted into the water, where fertilization takes place. In others, fertilization is internal, but the eggs are extruded into the water at an early stage in development. In still others, the gills serve as a brood pouch where the young are held until development is nearly complete. All species have a free swimming period and it is by this means that new timber is invaded. Eventually the embryo ship-

worm must settle on wood or die. It first attaches itself by a single byssus thread, then with the edges of its valves scrapes together loose wood fibres with which it makes a small cone to protect itself during the period of change from a typical embryonic bivalve to a highly specialized boring animal. The shell changes radically in shape, and the apophyses and pallets develop at this time. Once these changes are complete, active boring begins. As the animal grows the burrow is lengthened and gradually increased in diameter. Though the entrance hole remains nearly microscopic the anterior end of the burrow may reach an inch or more in diameter as the animal grows. In temperate seas they usually attain only a few inches in length, but in tropical seas some species may reach six feet. Growth is usually rapid; the young at the time of attachment are about .25 mm. but they may reach over 100 mm. in length within a month.

Teredinids are known to have existed as far back as the Jurassic, and before the advent of man they lived in fallen trees which were carried into the sea by rivers or that had fallen along the shores, thus aiding in the reduction of the wood to the elements. It was only when man began building wooden ships and wharves that the teredos became pests. The ships of Archimedes of Syracuse were protected by being sheathed with lead. Copper sheathing was first used by the British in 1758. Although it was expensive because it wore thin and had to be constantly replaced, copper gave complete protection both from borers and fouling organisms such as barnacles. In the days of the clipper ships nearly all vessels sailing in tropical seas were sheathed with copper, but it was found necessary to insulate the copper from the rudder iron to prevent corrosion. With the advent of metal ships the problem of teredos ended so far as commercial shipping was concerned, but it is still a problem for the smaller fishing or pleasure boats and specialized ships such as mine sweepers which are made of wood.

So far as water-front installations are concerned, teredos are still an important though decreasing problem as more and more wharves and piers are being constructed of steel and concrete. The Dutch were the first to be greatly concerned with the depredations of *zeeworms,* for the very existence of much of their country depended on the maintenance of a series of dikes which were partly constructed of wood. As a result of a severe infestation of the dikes, there appeared in 1733 an important book on teredos written by Godfrey Sellius. It was he who showed that shipworms were mollusks rather than worms. Of the more recent attacks the most famous was the invasion of shipworms in San Francisco bay during 1917–21, when destruction of wharves and jetties was estimated to have cost over $25,000,000.

Since that time there has been an almost continuous series of investigations into the teredo problem. Numerous techniques have been developed for the protection of wood against teredo. Some are quite effective; however, many are far too involved and too expensive to be practical on a large scale. An old but still the most practical method of protection because it is economically feasible is to treat wooden wharves with coal tar creosote solution. If a high quality creosote is properly applied and impregnates the piling, the wood will give long service. The effect seems to be largely mechanical: the creosote clogs the denticles of the young teredo and renders the rasping tool ineffective. As the creosote leaches from the outer surface of the pile it is replaced by that within.

The owner of a wooden boat has a choice of a large number of available paints and coatings. However, successful and lasting protection can only be achieved if the coating is kept intact. The slightest abrasion or wearing off of the paint will allow the teredos to enter, and once inside, destruction is rapid. The most effective paints are soft and have a copper base giving protection against both boring and fouling organisms. Fiberglas coatings are being used to some extent on small boats and the glue used in making marine plywood has been developed as a deterrent to borers.

Despite all the advances made in teredo control, there is still no permanent protection against them and constant checks must be made on any wood in contact with sea water.

BIBLIOGRAPHY.—W. G. Atwood *et al., Marine Structures, Their Deterioration and Preservation,* National Research Council (1924); W. J. Clench and R. D. Turner, "The Genus *Bankia* in the Western Atlantic," *Johnsonia* 2, pp. 1–28, plates 1–16 (1946); C. L. Hill and C. A. Kofoid (eds.), *Marine Borers and Their Relations to Marine Construction on the Pacific Coast,* Final Report of the San Francisco Bay Marine Piling Committee (1927) which has a fine section on the biology of the Teredinidae; Woods Hole Oceanographic Institution, *Marine Fouling and Its Prevention,* U.S. Naval Institute (1952). (R. D. T.)

TEREK, a river of the U.S.S.R. flowing from the Caucasus range to the Caspian sea, rising in glaciers on Mt. Kasbek. It skirts the Kasbek group on the south and west, and flows through a series of gorges. The most famous is the Darial gorge by which the Terek escapes to the plains on the north, its furious torrent flowing between wild granitic bare rocks and forming a contrast to the wooded limestone mountains beyond. The Georgian military road, built 1811–64, passes through the gorge and enabled Russia to establish her rule among the wild Caucasian mountain tribes. Below Vladikavkaz at the foot of a vast declivity, it collects the waters of the Gusel-don, Fiag-don, Ar-dòn and later receives the Urukh, and the Malka, with its affluents the Cherek and Baksu. It then flows through the steppe country at the northern foot of the Caucasus to the Caspian, being joined before it reaches the delta by the Sunja, a tributary receiving the Chanti or Argun, and fed by numerous sulfur streams, one of them, the Melchihi, rising in five hot springs which affect the temperature of the Melchihi for some distance. In spite of losses by evaporation and irrigation, the Terek still has a sufficiently large volume to enable it to spread out in a great delta, with branches shifting in flood time and changing their relative importance. This delta stretches for 70 mi. along the coast. The coast line, like that of the Volga delta, is encroaching on the Caspian, and old shore lines may be traced considerably west of the present coast. Fishing villages which in 1825 stood on the coast line were ten miles inland in 1855. Among the more important irrigation canals are the Eristov and Kurskiy, fed from the Malka; plans are under consideration for irrigation of part of the Nogai steppe, recently included in the Dagestan Autonomous Soviet Socialist Republic. In its lower course the Terek flows above the level of the surrounding country. The Terek has a total length of 373 sq.mi.; its basin covers an area of 16,873 sq.mi.

TERENCE (PUBLIUS TERENTIUS AFER) (*c.* 185?–*c.* 159 B.C.), was, after Plautus, the greatest Roman writer of comedy. Information concerning his life comes from the account by Suetonius (fl. 2nd century A.D.), prefaced to Aelius Donatus' 4th-century commentary, which quotes the (often conflicting) statements of earlier writers. It relates that he was born in Carthage and came to Rome as a slave in the house of Terentius Lucanus, a senator, by whom he was educated and set free. He won the friendship of certain noble Romans; the prevailing view included among these the younger Scipio and Laelius, who were even said to have helped him in his work. His first play was the *Andria,* which at the order of the aediles he read to the elderly comic dramatist Caecilius (who died 168 B.C.). After the production of the last of his plays, in his 25th year, he went to the east; from this voyage he did not return.

This account is partly contradictory and wholly unreliable. That Terence had noble friends is indeed clear from his own prologues. From the *didascaliae* (prefaces to the plays), confused as they are, it appears that his six plays were produced between 166 and 160 B.C. Their order is uncertain; but it is agreed that the first was the *Andria* ("Woman from Andros"), translated from the *Andria* of Menander.

The scene, as in all Terence's plays, is in Athens. Simo wishes his son Pamphilus to marry Philumena, daughter of their neighbour Chremes. But Pamphilus has secretly formed a liaison with Glycerium, a poor but modest girl who has recently come to Athens from Andros with her supposed sister Chrysis, now dead. The lovers are saved by the discovery that Glycerium is herself the long-lost daughter of Chremes. Terence, in translating Menander's play, made certain alterations (attested by Donatus); he changed the opening monologue into a dialogue, and he added two characters, Charinus, who is in love with Philumena, and his servant Byrria. The fact that he had made alterations became known to a rival of Terence, an elderly dramatist named Luscius

Lanuvinus. He protested that a translator had no right to "spoil" (*contaminare*) his original. To defend himself, Terence was forced to write a prologue, in which he asserted that all he had done was to borrow suitable material from a very similar play by Menander, the *Perinthia* ("Woman from Perinthus"). He further claimed that he had merely followed the example of his predecessors, Naevius, Plautus and Ennius, whose "carelessness" he would rather imitate than the *obscura diligentia* of his critic. It seems reasonable to suppose that Terence is not here defending himself merely against the charge of borrowing material from a second Greek source; that is in fact the procedure of which he boasts, and which he intends to repeat. His enemy seems to have retorted "In that case you have spoiled not *one* Greek play but *two*"; a charge to which he replies in the prologue to the *Heautontimorumenos* ("Self-Tormentor"). He repeated this procedure in the *Eunuchus*, adding (he says) to Menander's *Eunuchus* two characters from the *Colax* ("Flatterer") also by Menander. Again in the *Adelphi* ("Brothers") he says that he has inserted a scene from the *Synapothnescontes* ("Dying Together") of Diphilus, "carelessly" omitted by Plautus in the *Commorientes*, his translation of that play. These changes he would not himself, of course, have described as "spoiling"; that is the unpleasant word used by his critic. What he omits to mention is that he made other changes which were not borrowing but original invention. Through a misunderstanding of these passages, "contaminate" has come in textual criticism to mean "combine"; but it cannot be shown that the word ever had this sense in Latin.

Terence's second play was perhaps the *Hecyra* ("Mother-in-law"); but this play failed twice (as the two prologues explain) before it was finally produced in 160 B.C. The *Phormio*, one of his liveliest plays, was translated from Apollodorus' *Epidicazomenos* ("Litigant"). His masterpiece, the *Adelphi*, was produced at the funeral games of Aemilius Paulus (father of the younger Scipio) in 160 B.C.

It is noteworthy that Terence uses the prologue no longer to explain the plot but (after the manner of the Aristophanic *parabasis*) to reply to his critics (*Andr.* 5–7). In using his Greek models, apart from "borrowing" from a second original, Terence permits himself various liberties. Names are altered, Greek allusions toned down, characters slightly changed, individual scenes remodeled (*e.g., Eunuchus* 539 ff. is turned from a monologue into a dialogue of considerable power by Terence's addition of a character, Antipho). These and other changes are attested by Donatus; they show beyond doubt that Terence was not content to be a mere translator, but tried by delicate changes to improve on Menander himself.

The central interest in Terence's plays is character. In almost every play there is at least one pair in like circumstances who behave differently; *quam uterquest similis sui!* (*Phor.* 501): "how like each is to himself!" The charm of Terence's Latin and his lack of comic power are referred to by Julius Caesar, whose judgment seems to have been generally endorsed in antiquity.

Influence.—The simplicity of Terence's Latinity and his moralizing tone made him a favourite author in the middle ages. He was not, however, moral enough for Roswitha, the nun of Gandersheim, who in the 10th century wrote six Christian comedies modeled on Terence but portraying the victory of chastity over lust; there is a farcical element. One of the first Italian comedies, Ariosto's *I Suppositi* (1509), owes something to the *Eunuchus*, and was itself the model for George Gascoigne's *The Supposes*, the first prose comedy of English literature (1566). In the early 16th century Dutch schoolmasters developed what is known as the education-drama, or the "Christian Terence." Molière based his *L'École des maris* (1661) on the *Adelphi* and his *Les Fourberies de Scapin* (1671) on the *Phormio*. The *Phormio* was performed by the boys of St. Paul's school in 1528. George Chapman's *All Fools* (acted 1599) is based on the *Heautontimorumenos*, with borrowings from the *Adelphi*. Charles Sedley's *Bellamira* (1687) is adapted from the *Eunuchus;* Thomas Shadwell's *The Squire of Alsatia* (1688) from the *Adelphi*. The *Andria* is imitated in Steele's *The Conscious Lovers* (1722), the *Adelphi* in Fielding's *The Fathers, or The Good-Natured Man* (posthumously published 1778). Thornton Wilder's novel, *The Woman of Andros* (1930), is partly based on the *Andria*.

BIBLIOGRAPHY.—The *editio princeps* appeared in 1470. Other editions: R. Bentley (1726); A. Fleckeisen, "Teubner Series" (1898); R. Kauer and W. M. Lindsay, Oxford Classical Texts (1926); J. Marouzeau, 3 vol. with French trans., in the "Budé Series" (1942–49); S. Prete (1954); *cf.* G. E. Duckworth, *The Nature of Roman Comedy* (1952); W. Beare, *The Roman Stage*, 2nd ed. (1955); E. Paratore, *Storia del teatro latino* (1957). (WM. BE.)

TERENTIANUS MAURUS (fl. ? *c.* A.D. 200), Latin grammarian whose work is valuable both for its account of ancient prosody and as a source of fragmentary information on poems no longer extant. Born in Mauretania, he wrote in his old age; he must have lived between Septimius Serenus (fl. *c.* 140), whom he mentions as "recent," and Aphthonius (fl. *c.* 300), who mentions him. His treatise *De litteris, de syllabis et de metris* ("On Letters, Syllables and Metres") was discovered at Bobbio in 1493 and first printed in 1497. It is written entirely in verse and comprises (1) a preface of 84 lines in glyconic metre; (2) *De litteris*, 194 lines in sotadics; (3) *De syllabis*, perhaps originally a separate treatise, 720 trochaic tetrameters followed by 301 hexameters; and *De metris*, 1,682 lines in various metres.

For the text *see* H. Keil, *Grammatici latini*, vol. vi (1874).

TERESINA, chief city and capital of the state of Piauí, Brazil, is located on the right bank of the Parnaíba river about 220 mi. from the port of Parnaíba and about 200 mi. from São Luís in Maranhão. Pop. (1950) 51,418.

The city was founded in 1852 as the new capital of Piauí, and since then it has become the chief commercial centre of the middle Parnaíba valley. It has textile factories, sugar refineries, distilleries, soap factories and sawmills. It ships out cattle, hides, goat skins (from the *caatinga* or uplands of eastern Piauí), maize, rice, cotton and manioc. It is reached from Parnaíba by shallow-draft river steamers; a railroad which terminates at Timon (Maranhão) on the left bank of the river opposite Teresina, connects with São Luís and carries much of the commerce of the middle Parnaíba valley. An all-weather highway connects with Fortaleza in Ceará; and there are air services to all parts of Brazil. Formerly called Therezina, the city was named after the empress Teresa Cristina Maria of Brazil. (P. E. J.)

TERHUNE, ALBERT PAYSON (1872–1942), U.S. novelist and short-story writer, famous for his popular stories about dogs, was born in Newark, N.J., on Dec. 21, 1872, the son of a Presbyterian clergyman, Edward Payson Terhune (1830–1907) and Mary Virginia Hawes Terhune (1830–1922), herself popular for romantic novels and books on household management written under the pen name of Marion Harland. He received his first schooling in Paris and Geneva, Switz., while his father was serving as chaplain of the American churches of Rome and Paris. Terhune graduated from Columbia university, New York city, in 1893, traveled in Egypt and Syria, and returned to New York, joining the staff of the *New York Evening World* in 1894. His first book was *Syria from the Saddle* (1896); his first novel, *Dr. Dale* (1900), was written in collaboration with his mother. He wrote assiduously in his spare time to free himself from journalism and published more than 12 books before he left the *Evening World* in 1916.

In 1919 appeared the first of his popular dog stories, *Lad, a Dog*. He had moved to a farm near Pompton Lakes, N.J., which he named "Sunnybank," and there for the rest of his life he wrote, bred prize collies, fished and hunted. He wrote more than 25 books after 1919, nearly all of them novels in which dogs played conspicuous parts, including *Bruce* (1920), *The Heart of a Dog* (1925), *Lad of Sunnybank* (1928) and *A Book of Famous Dogs* (1937). He also wrote two autobiographical books, *Now That I'm Fifty* (1925) and *To the Best of My Memory* (1930). He died at Sunnybank on Feb. 18, 1942.

TERM. In traditional logic (*see* LOGIC: *Traditional Logic*) a term is properly the subject or predicate of a categorical proposition. Aristotle so used the Greek word *horos* (limit), apparently by an analogy between the terms of a proportion (*see* below) and those of a syllogism. *Terminus* is the Latin translation of this, used for example by Boethius. Hence in medieval logic the word came to be used also for common and proper names generally

(*see* NAME [IN LOGIC]); and even for what were called syncategorematic terms—words such as *and, if, not, some, only, except,* which are incapable of being used for the subject or predicate of a proposition (*see* MEANING). In everyday English, the meaning of term has been broadened still further, to that of an item of terminology, any word or phrase with a particular meaning.

In mathematics, the terms of a fraction (*see* FRACTION) are the numerator and denominator. The terms of a proportion are the four numbers which enter into the proportion (*see* ALGEBRA) or the corresponding numerical expressions that denote such numbers. Similarly the terms of a sum are the numbers which are added together to constitute the sum, or the numerical expressions denoting them. In this sense an infinite series (*see* SERIES) is thought of as a sum of an infinite number of terms; and a polynomial (*q.v.*) is a sum of a finite number of monomials which are the terms of the polynomial. (AO. C.)

TERMAN, LEWIS MADISON (1877–1956), U.S. psychologist who specialized in mental testing and in the research study of gifted children, was born on a farm in Johnson county, Ind., on Jan. 15, 1877. He took his graduate training at Clark university, Worcester, Mass., and from 1910 until his retirement in 1943 was professor of psychology and education at Stanford university. Terman published the first important and widely used individual intelligence test in the U.S., the Stanford-Binet, which he described in *The Measurement of Intelligence* (1916). In connection with this test, he introduced the term I.Q. (intelligence quotient), signifying a score on an intelligence test that takes account of both chronological age and mental age in such a way that the average child of any age has an I.Q. of 100. Working with a group of other psychologists during World War I, he was largely responsible, also, for the first great group intelligence tests, the Army Alpha and Army Beta.

Terman recognized the great importance to society of identifying intellectually gifted children and of providing them early in their lives with exceptional education opportunities. In 1921 he selected 1,500 California children whose I.Q.'s were over 140 (about 0.5% of the population) and began to study their development. He continued this research until his death 35 years later, by which time he had obtained definitive evidence that gifted children tended to be healthier and more stable emotionally than the average child and that they retained their high ability in adulthood. Terman's last progress report on this continuing study was *The Gifted Child Grows Up* (1947). He died at Palo Alto, Calif., on Dec. 21, 1956. (R. R. S.)

TERMINI IMERESE (anc. THERMAE HIMERENSES), a seaport town of Sicily, in the province of Palermo, 23 mi. E.S.E. of it by rail. Pop. (1951) 24,655. It was founded in 408 B.C. by the Carthaginians, after their destruction of Himera, in the vicinity of hot springs mentioned by Pindar (*Od.* xii, 19) which are still resorted to (temperature 110° F.). Agathocles (*q.v.*) was born here. It was taken by Rome in the First Punic War. A Roman road ran from it to Catana. Scanty remains of buildings of Roman times (baths, an amphitheatre and a so-called curia) exist in the upper part of the town; and outside it on the south are considerable remains of two Roman aqueducts.

TERMINUS (Lat., "boundary stone"), originally, in Roman cult, a stone or post set up in the ground with the following religious ceremonies. A hole was dug and a fire lighted; a victim was sacrificed and its blood poured into the hole, together with incense, fruits, honey, wine and the ashes of the sacrifice. Then the boundary stone, which had been previously anointed and crowned with garlands, was placed upon the hot ashes and fixed in the ground. Anyone who removed a boundary stone was accursed (*sacer*) and might be slain with impunity; a fine was afterward substituted for the death penalty. From this sacred object evolved the god Terminus. On Feb. 23 (the end of the old Roman year) the festival called Terminalia was held. The owners of adjacent lands assembled at the common boundary stone and each garlanded his own side of the stone. An altar was set up and offerings of cakes, grain, honey and wine were made. Later a lamb or a suckling pig was sacrificed. The proceedings closed with songs to the god and general merrymaking in which all the members of the family and the servants took part.

A similar festival was held at the old boundary of the Roman territory between the fifth and sixth milestone on the road to Laurentum. When the Capitoline temple was to be built the auguries forbade the removal of one of these *termini* (a boundary mark of some old precinct?) and it was enclosed within the walls of the new sanctuary, an indication of the immovability of such stones and of the permanence of the Roman territory.

BIBLIOGRAPHY.—G. Wissowa, *Religion und Kultus,* 2nd ed., p. 136 (1912); W. W. Fowler, *Roman Festivals,* p. 324 (1899); J. G. Frazer, *Fasti of Ovid,* vol. ii, pp. 481 ff. (1929); E. Marbach in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft,* s.v.
(X.; R. B. LD.)

TERMITE, the name for usually pale-coloured, soft-bodied insects of the order Isoptera (*see* INSECT). They all live in large social colonies. An account of their biology and social life is given in the article SOCIAL INSECTS.

General Account.—The wingless nymphs and workers are frequently whitish, a characteristic that has resulted in the misnomer "white ants." Worker termites have thick waists and can be readily distinguished from worker ants which have wasplike constrictions of the abdomen. Furthermore, flying termites have four membranous wings of almost equal size whereas flying ants have smaller hindwings. Termites, unlike ants, have a basal crack or suture in each wing by means of which the wings are shed in a fraction of a second following the colonizing flight.

JOHN H. GERARD
TERMITES EATING YELLOW PINE, GREATLY MAGNIFIED

Termites are found in greater numbers of species and individuals in the tropics than in the colder regions. From temperate North America, 45 species are recorded; from temperate Eurasia and North Africa, 40 species are known.

Grass, leaves and dead wood form the principal diet of termites, although some tropical species eat the organic humus of soils, a few eat fungi and some eat living woody plants. Termites cause severe economic damage to wooden structures and wood products, particularly in the tropics and warm temperate regions, and control measures are often necessary throughout the United States and occasionally in southern Ontario and British Columbia in Canada. Damage in Eurasia extends northward to a line through southern France, the Balkan countries, the southern U.S.S.R., southern and coastal China, Korea and Japan. Although termites are not spreading northward in their natural habitat, a few species have been introduced and established in heated buildings in many northern cities. Because most termites are soft-bodied and easily injured by arid conditions, they cannot live long after being exposed to dry air. The sterile workers and soldiers are blind and have a neutral reaction to light if the humidity is high enough. The majority of species in the fairly advanced family Rhinotermitidae and in the most advanced family Termitidae are dependent upon soil moisture for their existence. The species in one fairly primitive family, Kalotermitidae, are called dry-wood termites. These excavate their nests and chambers in dry or damp wood above ground.

Control of Dry-Wood Termites.—Dry-wood termites are seldom found north of the Gulf states, New Mexico and Arizona, or east of the northwestern Pacific states. *Kalotermes flavicollis* causes damage in the Mediterranean countries, and some other species cause economic damage as far north as southern Japan. One largely tropical genus, *Cryptotermes,* can live in the dry wood of furniture. The dry-wood termites show no external indications of their presence but they may be detected by the characteristically

hollow sound heard on tapping the wood near their galleries. When their chambers are exposed, accumulated masses of dry pellets of excrement may pour out.

Infestation may be prevented by inspection of lumber prior to its use by placing screens (20 meshes per inch) over openings in the building, by using wood that has been treated chemically in commercial plants, by dipping wood in a solution of pentachlorophenol (5%, in a light petroleum oil) for three minutes and by covering exterior woodwork with heavy coats of paint around holes and cracks.

Liquid insecticides (trichlorobenzene; 2% chlordane in oil; 5% pentachlorophenol in oil; 6% DDT in oil; and benzene hexachloride) sprayed into the inhabited termite galleries often give satisfactory results. Poison dusts (smelter dust, white arsenic, paris green and 50% DDT) blown upon the termites in the galleries may effectively kill the entire colony within a few weeks.

Control of Subterranean Termites.—Most control measures against the subterranean termites are designed to prevent their gaining entry into wooden structures which are in close contact with the soil. It is much easier to build relatively resistant structures than it is to eradicate termites after infestation. Precautions include cement foundations, a distance of at least 18 in. between the ground and any untreated wood, and ventilation under houses without basements. Cement grout, coal-tar plastic cement or coal-tar pitch should be used to seal off all holes around pipes in foundations below the ground level. Copper or zinc metal shields may be placed between woodwork and the top of cement foundations. All joints in the metal should be tightly sealed and the outer edges should be at least three or four inches wide beyond the sill and bent downward at an angle.

Subterranean termites that have already gained access to a building will die if their contact with soil moisture is broken. Soil poisons around the foundations are effective for several years. The poison is placed in a deep trench or in holes bored with a soil auger 18 in. apart. Heptachlor, aldrin, deldrin, lindane and chlordane are the poisons usually used.

BIBLIOGRAPHY.—A. Mallis, *Handbook of Pest Control,* 2nd ed. (1954); U.S. Department of Agriculture, "Insects," *1952 Yearbook of Agriculture* (1952); C. L. Metcalf and W. P. Flint, *Destructive and Useful Insects,* 3rd ed. (1951); in the U.S. the latest bulletins on termite control are available at little cost from Superintendent of Documents, Government Printing Office, Washington, D.C. (A. E. E.)

TERMITE-PROOF CONSTRUCTION. Damage caused by termites to buildings or their contents can be prevented by proper construction of the buildings and chemical treatment of all wood with coal tar creosote. Building codes are being revised in many American cities to require termite-proof construction. The best rule is to make the foundations entirely of stone, brick or concrete, including stone or metal columns or pillars in the basement. Basement floors should be of concrete on a gravel base. Where it is impracticable to use these materials all wood used near the ground should be chemically treated. Complete dryness of the foundation and basement floors and walls is important. Since subterranean termites can not live or work without access to moist earth, it is only necessary to shut off this source of moisture to prevent damage by this type. Nonsubterranean termites can be killed in infested wood by fumigation with hydrocyanic acid gas.

TERMS, DISTRIBUTION OF. In logic a term is said to be distributed in a proposition when explicit reference is made to its whole extent or extension. Otherwise (that is, not only when reference is made explicitly to a part only of the extension of the term, but when explicit reference is simply not made to its whole extension) it is said to be undistributed. Thus, in a proposition of the form *No S is P* both the subject and the predicate are distributed. In the form *Some S is P,* neither *S* nor *P* is distributed. In *All S is P, S* is distributed, but *P* is not. Lastly, in *Some S is not P, S* is not distributed, but *P* is. Briefly, only universal propositions distribute the subject term (*S*), and only negative propositions distribute their predicate (*P*). Naturally, singular terms (including proper names used as singular terms) are always distributed, for they only refer to one object, and cannot refer to less.

The importance of the distribution of terms arises from the fact that it is a principle of formal inference that no term may be distributed in the conclusion unless it was distributed in the premisses. That is why, *e.g., All S is P* can only be converted into *Some P is S* (not into *All P is S*), and *Some S is not P* cannot be converted at all.

TERN, sea birds forming the subfamily Sterninae of the gulls (*Laridae*).

The largest British species is the Sandwich tern (*Sterna sandvichensis*), as big as a small gull, with a dark bill tipped with yellow, and dark legs. The common (*S. hirundo*) and Arctic (*S. macrura*) terns are only to be distinguished by the all-red bill of the latter, which is a bolder species with a more northern range, though both breed at Blakeney point in Norfolk. The little tern (*S. albifrons*) is known by its white forehead and yellow bill. *S. caspia,* the Caspian tern, is the largest species known, measuring two feet in length. In addition to the last three British species, *S. dougalli,* the Roseate tern, is also common to Europe and North America.

In North America there also occur the noddy (*q.v.*), the sooty tern (*S. fuscata*) and the black tern (*Hydrochelidon nigra*).

Terns breed usually on sand dunes and shingle banks and lay up to four eggs, two being the commonest number, great variation occurring with the season and the locality. They are said by fishermen to damage the fishing, but this has been proved to be a mistake, the birds feeding largely on crustacea, insects and sand eels, which they obtain by plunging into the water from a height of 3–8 ft., and emerging with the food in their bills. The male usually feeds the female while she is brooding, often accompanying the presentation with the same play ceremony used in courtship, in which food may also be presented by the male.

About 75 species are known; the group is cosmopolitan.

TERNATE, the northernmost of the line of islands off the western coast of Halmahera, Indonesia, which stretch southward to the Bachian archipelago. It has an area of 45 sq.mi., a width of 6 mi., and consists mainly of a conical volcano 5,627 ft. in height, with three peaks (Arfat, Madina and Kekan), its curious formation being due to many extremely destructive eruptions. Within the last four centuries there has been volcanic activity at Ternate on no less than 70 occasions, the worst recorded being the eruption in 1763, which overwhelmed the thriving little village and Ft. Takomi, which is situated on the northwest coast, and so completely devastated one slope with lava flow that it has since been known as Burnt cape (Batu Angus), while two small crater lakes were formed where the village stood. An eruption in 1840 destroyed nearly every house in the town of Ternate.

The northern half of the island has suffered most from volcanic activity. There, lava streams have flowed down the mountain side right to the sea, and there are numerous bare tracts of land, but on the southern and eastern coast there is forest and luxuriant vegetation. Rice and maize are grown, also sago, coffee, pepper, nutmegs and fruit (good mangoes and durians); Ternate was once a leading centre of spice cultivation.

Ternate is also the name of the district comprising all of Halmahera and adjacent islands; area 12,399 sq.mi., pop. (1957 est.) 58,629. It forms part of the province of the Moluccas.

Ternate Town, population (1957 est.) 23,500, lies on a flat strip of land on the south side of the island, at the foot of the mountain. It is close to other volcanic islands, one of which, Tidore, is so close that it helps to form the fine harbour of Ternate, and to the coast of Halmahera, it has magnificent views.

History.—Ternate became known to Europe through the Portuguese, who settled there in 1521, and made it one of their chief spice-collecting centres. Conflict with the sultan, or king, as he was styled then, who objected to the establishment of a Portuguese spice monopoly, led to the expulsion of the Portuguese from the island in 1581. In the meantime, in 1579, Sir Francis Drake had called at Ternate, being received favourably by the king, and allowed to ship between four and five tons of cloves, but the visit was never followed up. Drake was much impressed with the wealth and state of the king of Ternate, and wrote, of his visit to the monarch: "The King had a very rich canopy with embossings of gold borne over him, and was guarded

with twelve lances.

"From the waist to the ground was all cloth of gold, and that very rich; in the attire of his head were finely wreathed diverse rings of plaited gold, of an inch or more in breadth, which made a fair and princely show, somewhat resembling a crown in form; about his neck he had a chain of perfect gold, the links very great and one fold double; on his left hand was a diamond, an emerald, a ruby, and a turky (turquoise); on his right hand in one ring, a big and perfect turky, and in another ring many diamonds of a smaller size."

When the Dutch visited Ternate, at the beginning of the 17th century, the sultan, who was anxious to extend his power over the Moluccan islands generally, and even over a part of Celebes, arranged to give the Dutch a spice monopoly for their assistance against the Spaniards and Portuguese (Portugal had become incorporated with Spain and the Spaniards had reconquered Ternate). So Ternate rose to power and made many of the Moluccan islands, the eastern and northern part of Celebes and Buton, suzerain, but the Dutch bolstered up the power of Ternate only so long as suited their purpose, and then, in their efforts to confine Moluccan spice cultivation to cloves in Amboina and nutmegs in the Banda islands, they destroyed the spice gardens of the northern Moluccas, Ternate included, causing Ternate to revolt, an example which was followed in neighbouring islands. They found it convenient (in 1683), to declare all contracts with Ternate void, and the chiefs who had looked to Ternate as their suzerain were informed that henceforth they would hold their lands directly from the Dutch.

After that time, although the power of the sultan of Ternate as their vassal was recognized by the Dutch nominally, and he was given certain privileges regarding the native population, the executive, in Ternate, and throughout the lands of the residency, remained in Dutch hands.

Ternate, as well as all the other islands on the west coast of Halmahera, fell under Japanese occupation during World War II. After the war it became part of the Republic of Indonesia.

(E. E. L.; X.)

TERNEPLATE: *see* TIN PLATE AND TERNEPLATE.

TERNI (anc. *Interamna Nahars*), a town, episcopal see, and capital of a province of Umbria, Italy, situated among the Apennines, 426 ft. above sea-level, in the valley of the Nera (anc. *Nar*), from which the town took its distinguishing epithet, 5 mi. below its junction with the Velino, and 70 mi. N. by E. of Rome by rail. Pop. (1951) 53,163. It has important iron and steel works and iron foundries, at which armour plates, guns and projectiles are made for the Italian navy, also steel castings, machinery and rails, a royal arms factory, a large jute factory, a carbide factory, a wool spinnery, etc., and lignite mining. Terni lies on the main railway line from Rome to Foligno and Ancona, and is the junction for Rieti and Sulmona. Its most interesting buildings are the cathedral (restored in 1653 with remains of the earlier 13th century façade and an early crypt), and the Romanesque churches of S. Pietro and S. Francesco. Its antiquities include traces of the city walls of rectangular blocks of travertine incorporated in the well-preserved medieval walls, remains of an amphitheatre, a temple, now the round church of S. Salvatore, theatre and baths (?) and numerous inscriptions. Five miles to the east are the falls of the Velino (*Cascate delle Marmore*), which took a very high place among European waterfalls; the cataract has a total descent of about 650 ft., in three leaps of 65, 330 and 190 ft., respectively.

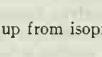
The falls owe their origin to M. Curius Dentatus, who in 272 B.C. first opened an artificial channel by which the greater part of the Lacus Velinus in the valley below Reate was drained. They supply the motive power for the factories of the town and have lost much of their former beauty.

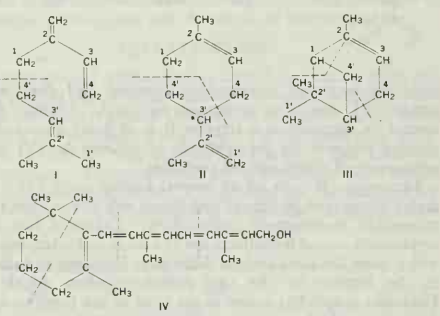
Terni is the ancient *Interamna* (*inter amnes*, "between the rivers," *i.e.*, the Nar and one of its branches), originally belonging to Umbria, and founded, according to tradition in the year 672 B.C. The recent discovery of a neolithic village, and of cemeteries of the Villanovan period, with numerous inhumation graves of Picenes of 1200–1000 B.C. (*see* Randall MacIver, *Iron Age in Italy*, 1927, 140–144) shows that the site was occupied much earlier. It is first mentioned in history as being, along with Spoletium, Praeneste and Florentia, portioned out among his soldiers by Sulla. During most of the middle ages and up till 1860 Terni was subject to the popes. Terni was the scene of the defeat of the Neapolitans by the French on Nov. 27, 1798.

TERPANDER, of Antissa in Lesbos, Greek poet and musician. About the time of the Second Messenian War, he settled in Sparta, whither, according to some accounts, he had been summoned by command of the Delphian oracle, to compose the differences which had arisen between different classes in the state. He is regarded as the real founder of Greek classical music and of lyric poetry; but as to his innovations in music our information is imperfect. According to Strabo (xiii, p. 618) he increased the number of strings in the lyre from four to seven; others take the fragment of Terpander on which Strabo bases his statement (Bergk, 5) to mean that he developed the citharoedic nomos (sung to the accompaniment of the cithara or lyre) by making the divisions of the ode seven instead of four.

Fragments (the genuineness of which is doubtful) in T. Bergk, *Poetae Lyrici Graeci*, iii; *see* also O. Löwe, *De Terpandri Lesbii aetate* (1869), who places him about 676 B.C.

TERPENES are members of a class of naturally occurring hydrocarbons (*q.v.*) of the general formula $(C_5H_8)_n$, the name being derived from turpentine, which is a mixture of terpenes. Terpenes occur in the essential oils of many plants and in the gummy exudates (oleoresins) of many trees and shrubs. The term terpenes is frequently used in a broader sense to include compounds which contain oxygen and may be regarded as being derived from the terpene hydrocarbons.

The terpenes and their derivatives are broadly classified according to the number of C_5H_8 units which they contain, and range in molecular size from volatile oils of molecular formula $C_{10}H_{16}$ to giant molecules such as that of natural rubber containing about 4,000 C_5H_8 units. The classes are: monoterpenes, $C_{10}H_{16}$; sesquiterpenes, $C_{15}H_{24}$; diterpenes, $C_{20}H_{32}$; triterpenes, $C_{30}H_{48}$; tetraterpenes, $C_{40}H_{64}$; polyterpenes, $(C_5H_8)_n$. Nearly all of the terpenes possess carbon skeletons that may be regarded as being built up from isoprene units () in various combinations and the early recognition of this common structural feature (O. Wallach, Isoprene Rule, 1887) was of great help in the elucidation of their structures. The isoprene units may be linked in a variety of ways leading to acyclic, monocyclic and polycyclic subclasses within the broad classes of terpenes, depending upon the number of carbon atom rings that result. Myrcene (I), an acyclic monoterpene; limonene (II), a monocyclic monoterpene; α-pinene (III), a bicyclic monoterpene; and vitamin A (IV), an oxygenated, monocyclic diterpene, exemplify this further classification and the dotted lines indicate the division of the carbon skeletons into isoprene units.

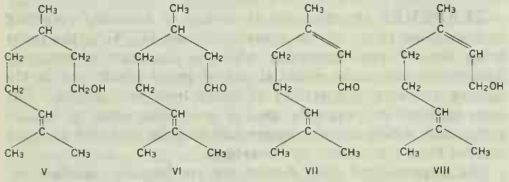


The problems in molecular structure presented by the terpenes

have challenged the imagination and skill of organic chemists since the latter part of the 19th century. The first investigations were mainly concerned with the structures of the monoterpenes ($C_{10}H_{16}$) but as those structural patterns became familiar and techniques for investigation were developed, attention turned increasingly to the terpenes containing 15 to 40 carbon atoms. Typically the establishment of the structure of a terpene follows the sequence: isolation; purification; degradation by oxidation to simpler compounds of known structure; postulation of a structure based on the results of the degradation; and validation of the structure by unambiguous synthesis.

MONOTERPENES

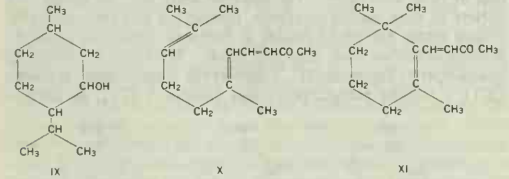
The monoterpenes are isolated from their natural sources by distillation of the plant matter with steam. They are volatile oils, less dense than water, and have normal boiling points in the range 150°–185° C. Purification is usually achieved by fractional distillation at reduced pressures or by regeneration from a crystalline derivative. Acyclic monoterpene hydrocarbons are few in number and it is their oxygenated derivatives which are more widespread in nature and of greater importance. Important oxygenated



V VI VII VIII

acyclic monoterpene derivatives include the terpene alcohol citronellol (V) and the corresponding aldehyde citronellal (VI), both of which occur in oil of citronella, as well as citral (VII), found in lemon-grass oil, and geraniol (VIII), which occurs in Turkish geranium oil.

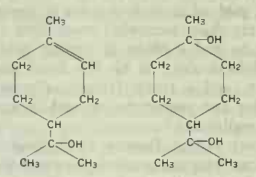
Citronellal is converted by treatment with acid into the monocyclic monoterpene alcohol isopulegol from which a mixture of stereoisomeric menthols (IX) is produced on catalytic hydrogenation. The process is used commercially to supplement the natural sources of menthol (oil of peppermint), widely used as a flavouring and in medicinal preparations. Citral (VII), upon reduction with sodium amalgam, yields geraniol (VIII), an important component of rose perfumes. Citral may be condensed with acetone to



IX X XI

yield the important intermediate ψ-ionone (X) from which β-ionone (XI) results on cyclization with acid. Although β-ionone cannot be regarded as a terpene, it is of great importance as a starting material for the synthesis of vitamin A (IV) and as a component of violet-scented perfumes.

Limonene (II), an oil of normal boiling point 177° C., is a major component of orange and lemon oils and is typical of the monocyclic monoterpene hydrocarbons. Others of this class are terpinolene, α- and β-phellandrene and α-, β- and γ-terpinene, all of which have the same carbon skeleton as limonene and differ only in the location of the two carbon-to-carbon double bonds. Limonene is optically active as are most of the terpenes and their derivatives which contain an asymmetric carbon atom, starred in the formula of limonene (II) (*see* STEREOCHEMISTRY). The racemic form of limonene is known as dipentene, and results from the dimerization of isoprene, $CH_2{=}C(CH_3)CH{=}CH_2$. Limonene is converted to isoprene by contact with a heated metallic filament. Few commercial uses, other than as flavourings, exist for the monocyclic monoterpene hydrocarbons. Menthol (IX), which has already been mentioned, and the oxygenated derivatives



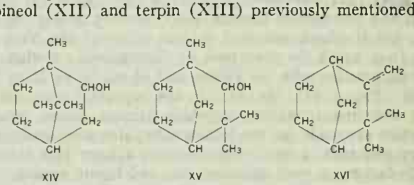
XII XIII

α-terpineol (XII) and 1,8-terpin (terpin hydrate) (XIII) are chemicals of commerce. Mixtures of 1,8-terpin, α-terpineol, terpinolene and the terpinenes result from the treatment of α-pinene (III) with acid, and the mixture finds use as "pine oil," an inexpensive disinfectant, deodorant and wetting agent.

α-Pinene (III), of normal boiling point 156°, and representative of the bicyclic monoterpenes, is the most abundant and important monoterpene. It is the major component of ordinary turpentine which is prepared from pine trees or stumps either by extraction followed by rectification or by distillation with steam. It is a major component of "sulfate turpentine," a by-product of the manufacture of paper, and is important as a component of paints and varnishes and as a raw material for the production of a wide variety of products employed in chemical industry.

Its use in coating materials depends upon its good properties as a solvent and upon its conversion by air oxidation into a polymeric, resinous film. Its function as a starting material for conversion to more useful products results from its abundance and its structure, which has a high degree of chemical reactivity. The chemical reactions of α-pinene have been studied more thoroughly than those of any other terpene and the study has contributed greatly to the understanding of molecular rearrangements (*q.v.*) in organic chemistry.

Treatment of α-pinene with acids under various conditions leads to a host of products among which are terpinolene, the terpinenes, α-terpineol (XII) and terpin (XIII) previously mentioned, and



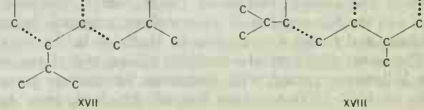
XIV XV XVI

in addition borneol (XIV), fenchyl alcohol (XV) and the hydrocarbon camphene (XVI).

The formation of the latter three compounds involves molecular rearrangement and advantage has been taken of the structural changes to provide a commercial synthesis of the important bicyclic terpene ketone, camphor (*q.v.*).

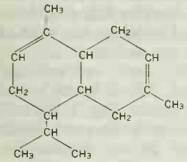
SESQUITERPENES

Sesquiterpenes, $C_{15}H_{24}$, are isolated from their natural sources by distillation with steam or by extraction since they are of lower volatility than the monoterpenes. They are purified by fractional distillation *in vacuo* or by chromatography. The sesquiterpenes demonstrate an even greater complexity of structure than the monoterpenes, and oxygenated sesquiterpenes are commonly en-

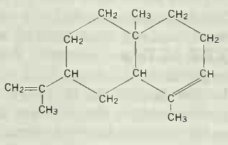


XVII XVIII

countered. Two arrangements of isoprene units are found in bicyclic sesquiterpenes, the cadalene (XVII) and the eudalene (XVIII) types, and the carbon skeleton of a sesquiterpene may frequently be determined by heating it with sulfur or selenium to effect dehydrogenation to the corresponding naphthalenic hydrocarbons: cadalene, 4-isopropyl-1,6-dimethylnaphthalene; or eudalene, 7-isopropyl-1-methylnaphthalene. In those cases where sulfur dehydrogenation fails to yield information about the carbon skeleton of a sesquiterpene a systematic degradation by oxidation to compounds of known structure is necessary.



XIX



XX

Cadinene (XIX), the principal component of oil of cubeb, is a typical sesquiterpene of the cadalene type. It is an optically active oil of normal boiling point 275° C. Selinene (XX), present in celery oil, is typical of the eudalene type.

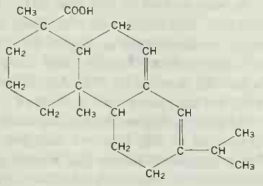
DITERPENES

Phytol (XXI), an oxygenated acyclic diterpene, is an important building block of the chlorophyll molecule from which it is obtained on saponification. The arrangement

$$CH_3CHCH_2CH_2CH_2CHCH_2CH_2CH_2CHCH_2CH_2CH_2C{=}CHCH_2OH$$
$$\quad\quad CH_3 \quad\quad\quad CH_3 \quad\quad\quad CH_3 \quad\quad\quad CH_3$$

XXI

of isoprene units in phytol is identical with that in vitamin A (IV), a monocyclic diterpene derivative, and is typical of the head-to-tail arrangement of isoprene units found in most terpenes.



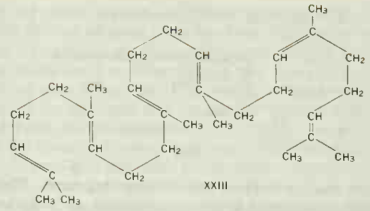
XXII

The commercial importance of the bicyclic monoterpene α-pinene is paralleled in the diterpenes by abietic acid (XXII), a tricyclic carboxylic acid which constitutes the major portion of rosin. Rosin is the nonvolatile portion of the oleoresin of members of the pine family and is the residue left after the isolation of turpentine. Rosin is used as such in the production of varnish and coating materials and, in the form of its sodium salt, for sizing paper, as an emulsifying agent in producing synthetic rubber and in yellow laundry soap. It is among the cheapest organic acids.

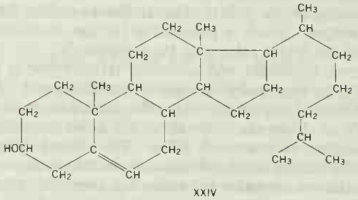
TRITERPENES

The acyclic triterpene hydrocarbon squalene (XXIII) constitutes more than half of the liver oil of certain species of sharks and is otherwise rather widely distributed in nature. It has been found in other fish-liver oils, in vegetable oils, in fungi and in human earwax and sebaceous secretions. The biochemical importance of squalene as a metabolic intermediate in the biosynthesis of cholesterol (XXIV) was demonstrated by the use of radioactive carbon labeling. The formula of squalene (XXIII) is arranged so as to show its relationship to the structure of cholesterol (XXIV).

Although cholesterol is not a terpene, the demonstration that it has a terpene as precursor in metabolism represented a major



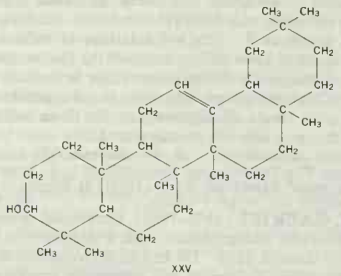
XXIII



XXIV

advance in understanding the biochemical relationship between the two important classes of compounds involved.

Although tri- and tetracyclic triterpenes are known, by far the most abundant triterpenes found in nature are those having five carbon rings. The pentacyclic triterpenes, either free or combined with sugars in glycosides (saponins), occur in all parts of many plants. Although the structures of many of the pentacyclic triterpenes are not known in full detail, that of β-amyrin (XXV) is

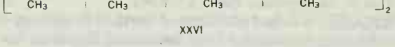


XXV

well established and exemplifies the important structural features of this class of substances. The best source of β-amyrin is elemi resin although it has been isolated from more than 50 plant sources. The carbon skeleton of β-amyrin bears a striking relationship to those of squalene (XXIII) and cholesterol (XXIV) and it has been suggested that squalene is a common precursor of the pentacyclic triterpenes and the sterols in biosynthesis.

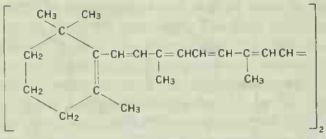
TETRATERPENES

The large class of yellow, orange or red, fat-soluble plant and animal pigments known as carotenoids are classed as tetraterpenes although they have in general the molecular formula $C_{40}H_{56}$ rather than $C_{40}H_{64}$ as required by $(C_5H_8)_8$. The fact that their structures can be built up from isoprene units justifies their classification as terpenes. The carotenoids are isolated from their natural sources by solvent extraction and purified by chromatography. Lycopene (XXVI), the red pigment of the ripe tomato, exempli-



XXVI

fies the class of acyclic tetraterpenes. The dotted lines in the formula show the division into isoprene units, and it is to be noted that the usual head-to-tail attachment of isoprene units is interrupted in the centre of the molecule with a single tail-to-tail



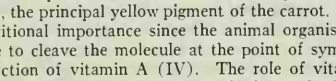
XXVII

attachment that produces a symmetrical structure. This feature is generally encountered in the tetraterpenes, as is the long series of alternating single and double, carbon-to-carbon bonds (conjugated system) which is responsible for the absorption of light and hence the bright colours of the compounds.

The most important and abundant tetraterpene is β-carotene (XXVII), the principal yellow pigment of the carrot. β-Carotene is of nutritional importance since the animal organism is apparently able to cleave the molecule at the point of symmetry with the production of vitamin A (IV). The role of vitamin A and structurally related terpenoid molecules in the synthesis of the pigments in the eye necessary for vision has been demonstrated.

POLYTERPENES

Rubber, which occurs in the latex of the rubber tree, is a polyterpene hydrocarbon, $(C_5H_8)_n$ where n is 4,000–5,000. Chemi-



XXVIII

cal degradation by oxidation and X-ray diffraction studies have revealed the repeating unit XXVIII in rubber. Division into isoprene units is indicated. The vulcanization of rubber involves the establishment of cross linking between the chains through sulfur atoms. Gutta-percha differs from rubber in the way in which the methyl (CH_3) groups are arranged; in gutta-percha they are on opposite sides (trans arrangement) of the chain and in rubber they are on the same side (cis arrangement).

BIBLIOGRAPHY.—E. Guenther *et al., The Essential Oils*, vol. i (1948); J. L. Simonsen, *The Terpenes*, vol. i–iv (1947–57); L. Ruzicka, "Biogenesis of Terpenes," *Experientia*, 9:357 (1953); H. Gilman *et al.* (eds.), *Organic Chemistry*, vol. iv, ch. 1 (1953). (R. H. EN.)

TERRA, GABRIEL (1873–1942), Uruguayan politician, was born in Montevideo and graduated with a doctorate in law from the university there in 1895. He had an active career in journalism, administration, diplomacy and university teaching. Terra served in the cabinet and chamber of deputies, and in 1931 was elected president. He was at odds in 1933 both with the national council of administration and the congress, and on March 31 staged a *coup d'état*, forcibly dissolved both bodies and assumed dictatorial powers. The constitution of 1917 was replaced in 1934 by one abolishing the national council and granting the president greater power.

Terra ruled as a virtual dictator, though not nearly so harshly as many other Latin-American strong men. Late in his term, both government and country manifested a desire for more democratic practices. Terra allowed free elections on March 27, 1938, and the principal presidential candidates were Gen. Alfredo Baldomir and Eduardo Blanco Acevedo, respectively Terra's brother-in-law and son-in-law. The election, which Baldomir won, was conceded to be fairly conducted.

Despite the interlude of dictatorship, Uruguay's only one in the 20th century, Terra was responsible for a number of constructive accomplishments which included beneficial legislation for old-age pensions, the protection of labour and industrial and commercial developments.

He died in Montevideo Sept. 15, 1942. (R. H. FI.)

TERRACE, a raised platform of earth. In geology the term came to be used to describe level horizontal ridges on the side of a slope, formed by volcanic action, or, more usually, by the action of water. They are thus frequent along the shores of lakes or by rivers. On the seashore they are known as raised beaches.

The term also is used in architecture of an artificial platform in front of a building, which is utilized as a promenade. Sometimes, when the building is erected on an elevation, there may be a series of terraces rising one to the other, as in the Villa D'Este at Tivoli, It.; or there may be a single terrace raised high above the ground and supported on arches, as the terrace to the Adelphi buildings in the Strand, or the river front at Somerset house, or in France at the castles at Amboise and St. Germain-en-laye, or again a low terrace like that in front of the houses of parliament at Westminster overlooking the Thames river.

The terraces of the gardens at Isola Bella on the Lago Maggiore became known as hanging gardens (*Hortus pensilis*), and were similar to those which were built by the Assyrian king at Babylon.

Though properly applied to a row of buildings on a raised level, the word has been often used of any row of houses.

TERRACINA (formerly ANXUR), town and episcopal see of the province of Latina, Italy, 63 mi. S.E. of Rome by rail (56 by the Via Appia), 40 ft. above sea level. Pop. (1961) 29,751. Its position, where the Volscian hills reach the coast, leaving no space for passage between them and the sea, commanding the Pomptine marshes and possessing a small harbour, was one of strategic importance. In 509 B.C. it was under Roman supremacy. In 406 it was stormed by the Romans, unsuccessfully attacked by the Volscians in 397 and finally secured in 329 B.C.

The construction of the Via Appia in 312 B.C. added to the town's importance; the road at first crossed the hill at the back of the promontory (748 ft.) by a steep ascent and descent, which had its strategic advantages. It was not until the imperial period (probably under Trajan) that a cutting in the rocks at the foot of the promontory (Pisco Montano) finally solved the problem. The depth of the cutting is indicated by marks on the vertical wall at intervals of 10 Roman ft.—figures enclosed in large swallowtail tablets—the lowest mark, 3 or 4 ft. above the present road, is CXX.

The construction of the Via Severiana, from Ostia to Terracina, added to the importance of the place; and the beauty of the promontory with its luxuriant flora and attractive view had made it frequented by the Romans as early as 200 B.C. Galba and Domitian possessed country houses here. It appears in the history of the Gothic wars, and Theodoric is said to have had a palace here. It was sacked in 409 and 595. In 872 John VIII brought it under the domination of the Holy See.

The picturesque modern town occupies the site of the old; the present piazza is the ancient Forum, and its pavement of slabs of travertine with the inscription "A. Aemilius A. F.," in letters once filled in with bronze, were well preserved. It was supported by massive arched substructures, extending under the surrounding houses. The cathedral of SS. Pietro e Cesareo, fronting upon it, was ensconced in a temple of Rome and Augustus. The vestibule, in the Cosmatesque style, was supported by ten ancient columns resting upon recumbent lions, with a mosaic frieze upon them. In the interior was placed a fine Cosmatesque pulpit supported by ancient columns resting on lions, a Paschal candlestick of A.D. 1245, and a good pavement of the same period with beasts and dragons. In modern times there survived remains of the town wall in the "polygonal" style, and also part of a later enceinte (4th [?] century A.D.) with towers.

The summit is occupied by a massive terrace, supported by arcades of fine *opus incertum* commanding a magnificent view seaward over the coast and over the Pomptine marshes. On the terrace stood a temple of the early imperial period, 110 by 65 ft., probably of Jupiter Anxur, worshipped as a child. The lower town by the harbour had buildings of some importance (amphitheatre, baths, etc.), of the imperial period. Of the ancient harbour constructed by Antoninus Pius some remains still exist.

Near the amphitheatre was found the famous statue of Sophocles, placed in the Lateran museum. The ancient aqueduct, bringing water some 35 mi. from the slopes of the Volscian hills, was repaired and put in use. Three miles to the northwest, at the foot of the Monte Leano, was the shrine of the nymph Feronia, where the canal following the Via Appia through the marshes ended. Along these 3 mi. were found numerous ancient tombs. During World War II Terracina suffered extensive damage.

See G. Lugli, *Forma Italiae* I. i. 1 (1926) *Anxur-Tarracina*. (T. A.)

TERRA COTTA, generally considered, may apply to all kinds of fired clay, but it has come to refer chiefly to a fairly coarse, porous type which when fired assumes a colour ranging from dull ochre to red and is most frequently left unglazed.

Baked clay, because it can be easily manipulated before it is fired and is relatively durable, has been used throughout the world since earliest times for pottery and sculpture and for many architectural purposes. While clays vary widely in composition in different areas, all are subject to fusion by heat into a hard substance varying in colour, porosity and strength in accordance with its content and the temperature to which it has been subjected. (*See* CLAY AND CLAY MINERALS.)

The most characteristic feature of objects in terra cotta is their obvious dependence on the plasticity of the material before it is hardened by fire. The clay when wet can be pressed or poured into molds; it may be cut, incised and folded; and it is responsive to the slightest pressure of the hand. Because of this, early artists sometimes attempted bolder designs in terra cotta than in more resistant materials. The coarser clays, however, do not readily lend themselves to sharp edges and precise definition of forms. Early works in terra cotta illustrate both the advantages and limitations of these material qualifications. For this reason there is often a superficial similarity between simply made terra-cotta works so far divided by time and distance as early Greece and the more modern cultures of Latin America.

Not only do the varied characteristics of the material, especially in primitive cultures, influence the form terra-cotta works are likely to take, but technical considerations in the procedure of firing also play a part. It is necessary that an object modeled in clay be thoroughly dry before being subjected to the intense heat of the kiln. In drying, and again in firing, the clay shrinks, and if one part dries much more rapidly than the rest it may crack and break away. It is necessary, therefore, to maintain a nearly equal thickness and avoid solid masses of clay of which the outer layers, drying more quickly, would shrink and crack. While this presents little difficulty in the making of pottery and low reliefs, it forces the sculptor to adopt complex methods for work in the round.

It was early found that very small figures could be fired successfully when solid, but larger works had to be rendered hollow before firing. There are three general methods by which this can be done: by pressing thin layers of clay into a mold (or pouring the clay in liquid form, called slip, into the mold and pouring off the residue after a thin shell has set); by hollowing out the heavier portions of the work; or by building up the work from the beginning with hollow forms. Throughout the history of terra cotta the first method has been the most popular. Not only does it permit evenness of drying and therefore successful firing, but it affords the possibility of multiplying the image indefinitely once the mold is made. The exploitation of this technique can be seen in the architectural decoration of the Etruscans, the figurines from Tanagra with their variously assembled parts, the repeated figures in Chinese tomb sculpture and the elaborate ornamentation of the Certosa at Pavia. But for a figure to be produced in an uncomplicated mold, deep undercutting must be avoided. As a result the surfaces are likely to be relatively unbroken, depending for contrast on incised pattern or applied colour.

Of importance particularly for the other methods of achieving a hollow form is the fact that an object to be fired can have no interior framework of a foreign substance. Since clay in a moist state will support little weight unassisted, constant attention must be paid to the relative size of the supporting members. The effect of this consideration may be seen in much modern terra-cotta sculpture built up from hollow forms. While it may place some limitations on the artist, it often imparts to the work a formal unity of a character peculiar to this use of the medium.

Colour has been used extensively in terra cotta ranging from the mat colours of applied slip (fine clay suspended in water) to the brilliant glazes of which the earliest examples come from Egypt and the near east. Since often the colour was not fired and thus fused with the surface, many once colourful objects have lost much of their original effect and appear in misleading simplicity to the modern eye. Few past cultures seem to have prized, as do modern sculptors, the native texture and colour of terra cotta.

Much of the use of terra cotta has been of a purely utilitarian kind since it is a relatively cheap, durable product of versatile nature. In places such as Mesopotamia and parts of Italy where stone has not always been easily accessible, it has served in architecture both for structure and ornament. In the 19th century, with the growth of iron construction, terra cotta commercially produced in the form of tile and sculptured ornament again became popular. It was light, adaptable to the many styles of ornament in demand and, as was discovered late in the century, could form a fireproof case for the vulnerable iron structure. *See* AEGEAN CIVILIZATION; CHINESE SCULPTURE; POTTERY AND PORCELAIN; TILE. (J. C. T.)

GREEK, ETRUSCAN AND ROMAN

Good clay was abundant throughout Greece and Italy and was used in ancient times for pottery, sculpture and in architecture. (*See* also GREEK ART; ROMAN ART; ETRUSCANS.)

Literary Sources.—Pliny in his *Natural History* gives an account of the art of terra-cotta sculpture (*plastice*), which contains much useful information. The most significant statements are the following: (1) Boutades of Sikyon invented the art of making portraits in terra cotta and antefixes in the form of human beings (xxxv, 151); (2) Rhoikos and Theodoros of Samos invented the technique of clay modeling long before the Bacchiades were exiled from Corinth; *i.e.*, before about 657 B.C. (xxxv, 152); (3) Eucheir, Diopos and Eugrammos of Corinth introduced the art of clay modeling into Italy in the 7th century B.C. (xxxv, 152); (4) Damophilos and Gorgasos decorated the temple of Ceres in Rome, which was dedicated in 493 B.C. (xxxv, 154); (5) Vulca was called from Veii to Rome at the time of Tarquinius Priscus (616–578 B.C.) to make the cult statue of the temple of Jupiter Capitolinus, which was dedicated after the flight of Tarquinius Superbus in 510 B.C. (xxxv, 157). Pliny (xxxv, 158) and Plutarch (*Publicola*, xiii, 1–2) both speak of Etruscan terra-cotta sculpture with wonder and admiration; Vitruvius (iii, 3, 5) refers to terra-cotta pedimental sculptures as an "Etruscan fashion."

Some of these statements square with the facts as archaeology has revealed them, others do not. The tradition that the art of clay modeling was practised in Greece in early times is borne out by discoveries of small terra-cotta figures dating back to the early Bronze Age and of larger ones from the 7th century B.C. That this art was introduced into Italy by Greek artists and flourished there especially in the late Archaic period—when the temples of Ceres and Jupiter Capitolinus were dedicated—is substantiated by discoveries of a large number of terra-cotta statues and reliefs of that epoch in Etruria showing strong Greek influence. Moreover, that some of these Etruscan and Greek sculptors worked also in Rome is suggested by the fragments of important terra-cotta sculptures of that period that have been found there. On the other hand Pliny (perhaps quoting Varro) evidently confused some of his dates.

Excavations at Samos have shown that Rhoikos and Theodoros were active in the 6th not the 7th century B.C., and if Vulca worked on the cult statue of the temple of Jupiter Capitolinus, which was dedicated after 510 B.C., Tarquinius Superbus rather than Tarquinius Priscus must have given the commission.

The terra-cotta sculptures of Greece, Etruria and Rome consist of statues, statuettes and reliefs.

Statues.—Terra-cotta statues were particularly popular in Italy, where marble was scarce until the opening of the Carrara quarries in the 1st century B.C., and in Cyprus. Discoveries at Kalydon, Thermos, Olympia and the Agora of Athens, however, show that terra-cotta sculptures must have been commoner in Greece than was once thought. Practically all served as decorations of temples.

That Pliny's and Plutarch's admiration of Etruscan terra-cotta statues was justified is borne out by the discoveries of such statues, large and small, throughout Etruria and etruscanized Latium. The life-sized statues of Apollo, Herakles and Hermes from Veii rank, both artistically and technically, among the greatest achievements of this kind. They once adorned the ridge pole of a temple, as

shown by the terra-cotta bases found with them. Many other examples, both single figures and groups (some partly in relief), served as acroteria and antefixes of temples or came from pediments.

Excellent specimens, chiefly of the Archaic period, a few of the later 5th and the 4th century B.C. and many of the Hellenistic epoch have been found at Caere (Cervetri), Satricum (Conca), Falerii (Cività Castellana), Lanuvium (Cività Lavinia), Veii (Veio), etc.

Another common use for statues in terra cotta among the Etruscans was for figures placed on the lids of sarcophagi. Life-size archaic examples from Caere are in the Louvre and in the Villa Giulia museum, later ones in the British museum, the Archaeological museum in Florence and elsewhere. A large number of terra-cotta urns for holding the ashes of the dead, in the form of miniature sarcophagi, also have reclining figures on the lid. They date from the 3rd and 2nd centuries B.C. and often portray the deceased with great realism.

The style of these Etruscan sculptures, like that of Etruscan paintings, at times shows strong Greek influence, but it mostly has a marked local flavour. The spirit is sometimes gayer, at other times more morose or fiercer than in the Greek counterparts.

With the extended use of marble for sculpture in Italy in the 1st century B.C. terra-cotta statues became scarce. Relatively few have come to light in Pompeii and elsewhere in the Roman domains.

Technique.—Etruscan and Greek terra-cotta statues were not molded but modeled, from the bottom up, with coils and wads of clay, in thick walls. The clay was mixed with sand and grog to provide porosity, diminish the shrinkage and avoid distortion; and over this a coating of finer clay was generally added. Before firing the surface was painted, the prevalent colours being white, brownish red and black. The firing of such large statues as the Apollo of Veii and the warriors in New York required great care. The kiln must have been built round them, as they could hardly have been moved in the green state. The temperature has been computed to have been about 960° C., approximately the same as that used in firing Greek pottery. In the sanctuaries of Asklepios at Epidauros and Corinth, as well as in Italy, there have been found life-size votive figures consisting generally of heads or of parts of the body that had been diseased and subsequently cured. They date mostly from the 4th to the 1st century B.C. and are made in molds in well-levigated clay.

Statuettes.—The commonest use of terra cotta was for small figures not over six or seven inches in height. They have been found in tombs and sanctuaries, on many different sites in Greece, Cyprus, Rhodes, Asia Minor, north Africa, Italy and Sicily, and range from the Bronze Age to the Roman period. Among the earliest are primitive figures from Cyprus, with planklike bodies or birdlike features. The painted and glazed statuettes of the Minoan period from Crete are more sophisticated in their modern-looking costumes, with full skirts, bare breasts and narrow waists. The Mycenaean and geometric figures, though schematic, are often in lively, lifelike poses. Women, horsemen, children and animals are favourite subjects. Particularly interesting are groups from Cyprus and elsewhere, consisting of men in chariots, dancers, warriors, in the so-called snow-man technique, modeled with great freedom.

From the 7th and 6th centuries B.C., all over the Mediterranean, the Greek terra-cotta statuettes reflect the changing styles in sculpture. Those of the 7th century show close contact with the orient both in subjects and renderings. In the 6th century the standing and seated female figure is favoured, the latter being generally shown holding a flower or other offering in one hand or with a child on her lap. Quantities of such statuettes have been found in votive deposits near the temples of Paestum (Poseidonia), the Heraion of Foce del Sele and other sites. The nude male standing figure, the so-called *kouros* type, is comparatively rare; only isolated examples have been found; *e.g.*, at Perachora and Veii. Among the most attractive statuettes of the late Archaic period are figures of women engaged in household tasks, horsemen, barbers and animals, found mostly in Boeotia. Though summarily modeled, they have vivacity and style.

Fifth-century terra-cotta statuettes continued the types prevalent in the preceding times, especially of seated and standing women, but with more freedom in the rendering of features and drapery. They are not as common as in the preceding or succeeding periods, but some are worthy products of their epoch and show the repose and serenity of the contemporary marble sculptures. A common type of the first half of the 5th century is a female mask or bust, generally hollow at the back, which is found both in central Greece and Rhodes.

From the late 4th century B.C. the types are more numerous. They represent no longer always divinities or votaries, but mostly the people of their time. Women are particularly popular. They are shown in graceful poses—standing, leaning against a pillar, walking or sitting, doing their hair, carrying a child or playing games. Comparatively few young men or children or animals appear, but an old nurse with a child in her arms, or a teacher with his pupil and particularly actors in their characteristic outfit are not uncommon. Since many of such statuettes were first found at Tanagra, in Boeotia, during the 1880s, they are still commonly called Tanagra figurines, though similar examples later came to light in other parts of Greece.

In the Hellenistic period the centres of manufacture in terra-cotta statuettes, as well as in marble and bronze sculpture, shifted from Greece proper to Asia Minor. In the little town of Myrina extensive discoveries were made in 1880–82 and other sites later yielded rich harvests. The new spirit of Hellenistic art is reflected in these statuettes. Instead of the quiet, gentle women, youths and children prevalent during the preceding periods, there are people in lively attitudes with draperies rendered in multiple folds going in different directions. Mythological subjects, especially flying Victory and Eros, as well as caricatures occur. Many Hellenistic terra-cotta statuettes from Italy and Sicily formed part of the grandiose vases then prevalent; *e.g.*, the Canosa and Centuripe wares. Those from Egypt are naturally influenced by the beliefs there current, representations of Horus, Bes and an orientalized Aphrodite being popular.

The same Hellenistic types continue throughout the Roman age, as extensive discoveries at Pompeii and throughout the Roman empire, including Gaul, Belgium and Britain, indicate. The Roman terra cottas from Egypt again stand apart, being of brown, local terra cotta and representing Egyptian deities in a Greco-Egyptian style. Of special interest are the large statuettes that reproduce famous sculptural works, *e.g.*, Polyclitus' "Diadoumenos," found in Asia Minor and elsewhere.

Technique.—The technique of terra-cotta statuettes is of course different from that of statues. They are made of well-levigated clay and molded; comparatively few are modeled by hand. A mold was first made and baked to considerable hardness; moist clay was then pressed into it, of the thickness required, and allowed to dry; the resultant shrinkage made it easily removable from the mold. The back was made in another piece, sometimes summarily by hand, with a venthole for evaporation. Often heads and arms were also made separately, in different molds and variously combined, so that in spite of a limited number of molds there is great diversity. In the words of E. Pottier, "there are many sisters, but few twins." An engobe or slip of white clay was added (while the clay was still leather hard) and when bone dry the figures were fired to a moderate heat. Finally colours were added in tempera—blue, red, pink, yellow, brown, violet and rarely green for the garments, brownish for the skin of men and boys, pink for the cheeks of women, auburn for hair, red for lips and brown or blue for eyes. Gilt sometimes occurs for details, such as jewelry, or as an all-over covering; also sometimes black. Few figures now retain their original colours; in most only the drab terra cotta remains.

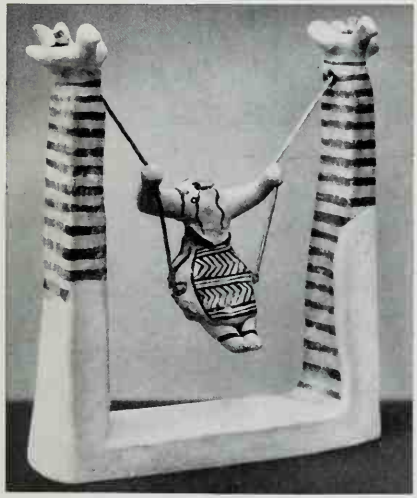
After the extensive discoveries at Tanagra had come to a stop toward the end of the 19th century, the popular demand for the charming figurines naturally persisted and as a result many forgeries were made which found ready acceptance with the public. Among them are mythological groups and single figures in affected attitudes that can easily be distinguished from the Greek originals. Sometimes, however, the figures were made from molds taken from ancient statuettes, and then only the somewhat blurred sur-

Mexican statue of the god Xipe Totec dressed in human skin, from Coatlinchan, Valley of Mexico; Mazapan culture, A.D. 1400–1500

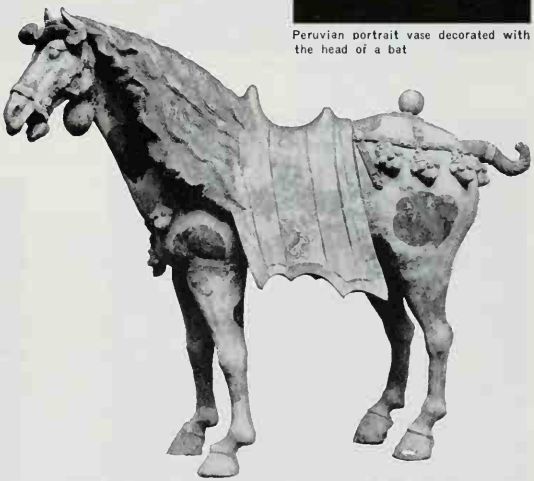
Ancient Chinese tomb sculpture; Wei dynasty, 6th century A.D. *Left*, seated dancing girl of buff clay with traces of polychromy; *right*, dancing girl of unglazed clay

Lydian revetment of painted terra cotta from Sardis; 6th century B.C.

Peruvian portrait vase decorated with the head of a bat



Cretan painted terra-cotta statuette of a woman in a swing; Middle Minoan period, 1600 B.C.



Chinese sepulchral figurine of a horse; T'ang dynasty, A.D. 618–906. Funerary pottery is the only remaining evidence of the advanced techniques of T'ang craftsmen

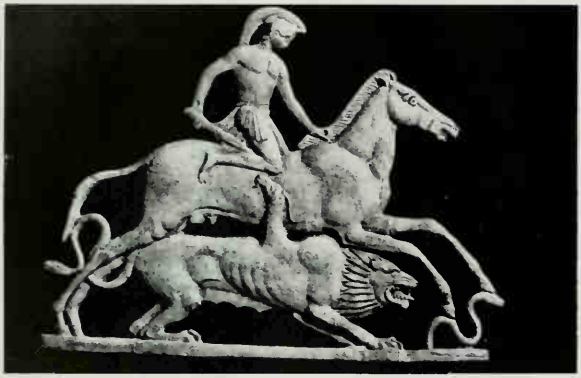
ANCIENT AMERICAN, ASIAN AND MEDITERRANEAN TERRA COTTA

BY COURTESY OF (TOP LEFT, CENTRE RIGHT) THE AMERICAN MUSEUM OF NATURAL HISTORY, (OTHERS) THE METROPOLITAN MUSEUM OF ART, FROM (TOP CENTRE AND RIGHT) FLETCHER FUND, 1928, (BOTTOM RIGHT) ROGERS FUND, 1925, AND (CENTRE) GIFT OF V. EVERIT MACY, 1911

TERRA COTTA

Cretan painted terra-cotta vases exemplifying the informal floral and marine designs characteristic of Middle and Late Minoan ceramic art; 1600–1500 B.C.

Etruscan painted terra-cotta head of a statue of Hermes from Veii; *c.* 500 B.C.



Relief depicting the Greek myth of Bellerophon and the Chimera; found on the island of Melos, 5th century B.C.

Etruscan head of the statue of Apollo from Veii; late 6th century B.C.

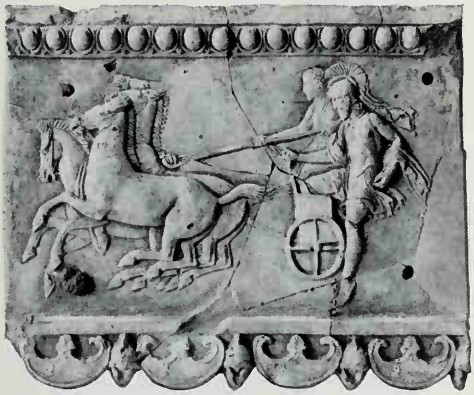
Etruscan urn with a statue of a reclining woman on the lid and a battle scene on the front; 3rd–2nd century B.C.

Etruscan funerary statues of a husband and wife on lid of sarcophagus from Caere (Cerveteri); dating from 500 B.C.

ETRUSCAN AND MEDITERRANEAN TERRA COTTA

BY COURTESY OF (TOP LEFT, BOTTOM LEFT) THE METROPOLITAN MUSEUM OF ART, NEW YORK, (TOP RIGHT, CENTRE RIGHT, BOTTOM RIGHT) THE MUSEUM OF VILLA GIULIA, ROME, (CENTRE LEFT) THE TRUSTEES OF THE BRITISH MUSEUM

Seated Greek terra-cotta statuettes showing evolution in style: *left*, painted male figure seated in an armchair, 8th century B.C.; *centre*, seated goddess, Archaic period, 6th century B.C.; *right*, female Tanagra figurine seated in a chair, 4th–3rd century B.C.



Roman mural relief depicting the myth of Oenomaüs and his unfaithful charioteer, Myrtilus; Augustan period, 27 B.C.–A.D. 14

Greek statuette of a flying Victory; Hellenistic period, Asia Minor type, 3rd–2nd century B.C.

Greek terra-cotta head of a statue of Zeus from Olympia; early 5th century B.C.

Greek head of a kore broken from a statue; Archaic period, 6th century B.C.

GREEK AND ROMAN TERRA COTTA

BY COURTESY OF (BOTTOM LEFT) THE MUSEUM OLYMPIA, (OTHERS) THE METROPOLITAN MUSEUM OF ART, FROM (TOP LEFT) FLETCHER FUND, 1931, (CENTRE LEFT) FLETCHER FUND, 1926, (TOP CENTRE) ROGERS FUND, 1906, (BOTTOM CENTRE) ROGERS FUND, 1947, (RIGHT) ROGERS FUND, 1906, 1909, AND (TOP RIGHT) GIFT OF J. PIERPONT MORGAN, 1917

TERRA COTTA

TERRA-COTTA SCULPTURE, 14TH–20TH CENTURIES

Top left: "Madonna and Child," statuette in glazed terra cotta by Donatello (*c.* 1386–1466), Italian

Top right: "Nativity Group," glazed terra-cotta figurines by Antonio Rossellino (1427–*c.* 1479), Italian

Right: "Kneeling Angel," by Giovanni Lorenzo Bernini (1598–1680), Italian

Bottom left: "Seated Figure," terra-cotta figurine by Aristide Maillol (1861–1944), French

Bottom centre: "Female Satyr Carrying Two Putti," terra-cotta statuette by Claude Michel (Clodion) (1738–1814), French

Bottom right: "Song of the Vowels," by Jacques Lipchitz (1891–), Latvian-French

BY COURTESY OF (TOP LEFT, TOP RIGHT) THE METROPOLITAN MUSEUM OF ART, NEW YORK, (RIGHT) THE FOGG ART MUSEUM, HARVARD UNIVERSITY, (BOTTOM LEFT) THE MUSEUM OF MODERN ART, NEW YORK, GIFT OF MRS. SAIDIE A. MAY, (BOTTOM CENTRE) THE WALTERS ART GALLERY, BALTIMORE, (BOTTOM RIGHT) THE MUSEUM OF MODERN ART, NEW YORK, GIFT OF THE SCULPTOR; PHOTOGRAPHS, (BOTTOM LEFT, BOTTOM RIGHT) SOICHI SUNAMI

face, or the addition of inappropriate details or mistakes in the application of the colours betray the hand of the forger.

Reliefs.—In addition to terra-cotta statues and statuettes, reliefs played a large part in Greek, Etruscan and Roman art. They served as decorations of temples and other buildings, especially in those countries where a wood and clay construction instead of stone was in use. Many examples of all periods have come to light in Italy and Asia Minor, mostly of the Archaic and Hellenistic periods, some of the 5th and 4th centuries B.C. They are decorated with floral designs, preferably palmettes, scrolls, rosettes, meanders and spirals, combined in effective patterns, or they have figured representations, such as chariot races or banquets or mythological subjects. Revetments of this kind have been found at Larisa in Asia Minor, Sardis in Lydia, southern Italy and throughout etruscanized central Italy. Many simas and antefixes have decorative features in relief consisting of lions' masks, female heads and other motives.

Besides architectural reliefs, votive ones are common. Two classes call for special mention—the Tarentine and the Locrian. At Tarentum many examples have been found which range from Archaic to Hellenistic times and apparently served as votive offerings to nether deities. They represent local divinities and heroes and are made in local clay in a broad, smooth style. Dionysus and his cycle were especially popular.

The Locrian reliefs are generally small and meticulously executed. Again local cults dictated the subjects. The majority were offerings to Persephone and relate to the cult of the dead. They, too, were a local product, made of local clay. The period of production was relatively short; almost all extant examples belong to the first half of the 5th century B.C.

The Melian reliefs, so called because many have been found in the island of Melos, were apparently not votive but served for the decorations of chests. They are often of fine quality and may be assigned to the second quarter and the middle of the 5th century. Mythological subjects predominate, but scenes from daily life also occur.

There were of course many other uses for terra-cotta reliefs. They served as decorations of small altars and of all sorts of utensils, and even of furniture. The Etruscan urns of the 3rd and 2nd centuries above mentioned generally have reliefs on their sides depicting scenes from Greek mythology, such as the combat of Eteocles and Polyneices, in Hellenistic Greek style. Applied reliefs appear on vases from Centuripe and other Hellenistic wares. Necklaces and other jewelry were sometimes made in terra cotta and gilded, chiefly for sepulchral use.

The Romans used terra-cotta reliefs on both the interior and exterior of buildings. Cicero speaks of fixing such reliefs (*typoi*) on the corner of his little atrium. Circular holes served for fastening them by means of plugs to the walls. The subjects are mostly mythological. Dionysus and his satyrs and maenads are especially popular. Occasionally Roman subjects occur. Lamps were generally made of terra cotta in antiquity and the Roman ones are regularly decorated with reliefs on their upper surface.

Technique.—The majority of these reliefs—Greek, Etruscan and Roman—were made in molds, and there are many repeats. Sometimes a mold was taken from an existing relief and was then naturally smaller in size because of the two successive shrinkages. Like the statues and statuettes these reliefs were all gaily coloured, but much of the polychromy has disappeared.

Use in Buildings.—Lastly, terra cotta served purely utilitarian purposes, especially on the roofs of buildings. Roof tiles—both flat ones and curved, covering ones—were often made of terra cotta; also the covering slabs along the raking cornice of the pediment, the cornice itself, the gutters and parts of the pediment. Many such plain terra-cotta slabs have been found in ancient sanctuaries that had terra-cotta roofs and revetments. Sometimes terra-cotta slabs were decorated with paintings. From a temple of the 7th century B.C. at Thermon in Acarnania slabs that served as metopes have been preserved, some with their painted decoration still visible. At Klazomenai terra-cotta sarcophagi of the Archaic Greek period with painted scenes have come to light, and terra-cotta slabs with painted decoration have been found lining walls of Etruscan tombs. Bricks in wall construction were much used; they were not always fired, but merely sun-dried.

Collections.—The most representative collections of Greek terra cottas are in the British museum, the Louvre and the museums of Berlin, Naples, Paestum, Syracuse, Athens, Olympia, Boston and New York. The Etruscan ones may best be studied in the museums of Florence, Perugia, the Villa Giulia museum in Rome, the British museum, the Ny-Carlsberg Glyptotek in Copenhagen and the Metropolitan museum in New York. Most of the Roman examples are to be found in Rome, Naples, Pompeii and in the provincial museums of France, England and Germany.

BIBLIOGRAPHY.—*Statues:* W. Deonna, *Les Statues de terre cuite dans l'antiquité* (1906); G. Q. Giglioli, *L'Arte Etrusca* (1935); G. M. A. Richter and C. F. Binns, "Etruscan Terracotta Warriors in the Museum," *The Metropolitan Museum of Art Papers,* no. 6 (1937); E. Kunze, *Neue Meisterwerke griechischer Kunst aus Olympia* (1948); M. Santangelo, "Veio, Sanctuario di Apollo. Scavi fra il 1944 e il 1949," *Bollettino d'Arte,* pp. 147 ff. (1952). *Statuettes:* E. Pottier and S. Reinach, *La nécropole de Myrina* (1887); F. Winter, *Die Typen der figürlichen Terracotten* (1903); D. Burr, *Terracottas From Myrina in the Museum of Fine Arts, Boston* (1934); G. Kleiner, "Tanagrafiguren, Untersuchungen zur hellenistischen Kunst und Geschichte," *15tes Ergänzungsheft des Jahrbuchs des Deutschen Archäologischen Instituts* (1942); T. B. L. Webster, *Greek Terracottas* (1950). *Reliefs:* E. D. van Buren, *Figurative Terra-cotta Revetments in Etruria and Latium* (1921), *Archaic Fictile Revetments in Sicily and Magna Graecia* (1923), *Greek Fictile Revetments in the Archaic Period* (1926); H. von Rhoden, *Die Terracotten von Pompeji* (1880); H. von Rhoden and H. Winnefeld, *Architektonische römische Tonreliefs der Kaiserzeit* (1911); L. Kjellberg, *Larisa am Hermos,* ii, *Die architektonischen Terrakotten* (1902–1934); T. L. Shear "The Architectural Terracottas," *Sardis,* xi (1926); P. Jacobsthal, *Die Melischen Reliefs* (1931); P. Wuilleumier, *Tarente* (1939). *Catalogues, etc., comprising all three categories:* A. Levi, *Le terrecotte figurate del Museo Nazionale di Napoli* (1926); Heidenreich, "Terrakotten," in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft,* 2nd series, ix, col. 808–820 (1934); H. Goldman (ed.), *Excavations at Gözlü Kule, Tarsus,* i (1950); R. A. Higgins, *Catalogue of Terracottas in the British Museum,* vol. i (from 730 to 330 B.C.) (1954). (G. M. A. R.)

MODERN EUROPE

Neither for architectural decoration nor for free sculpture was terra cotta a frequently used medium between the end of the Roman empire and the 14th century, although the simple technical procedures were kept alive in the making of pottery. By the 15th century, however, a notable revival was in full progress particularly in Italy but also in Germany. Terra cotta was then used in architecture in two distinctive ways. For brick architecture terra cotta, similar in colour and substance to the building material, formed the ideal decorative element and, molded or carved, served in its natural colour for elaborate friezes, window and door moldings and inset medallions. The church of St. Catherine, Brandenburg, is a highly decorated example of the German usage. In Italy fine examples are to be found throughout the north; indicative of the variety are the upper windows of the Municipal palace at Piacenza (begun 1281), the buildings of Biagio Rossetti (14?–1516) in Ferrara and the Ospedale Maggiore (after 1456) in Milan. For contrast terra cotta was sometimes combined effectively with stucco or stone as in the apse of Santa Maria delle Grazie in Milan.

The second new use of terra cotta was in that of the highly glazed and coloured sculpture introduced in Florence early in the 15th century by the Della Robbias (*q.v.*). This work could be used more effectively in combination with marble and stone than could unglazed ornament, and it added a freshness of accent quite unparalleled. The effect was imitated both in France and Spain and the use of terra cotta, glazed and unglazed, spread throughout Europe. As a novelty Francis I had the entire façade of the little château of Madrid, which once stood in the Bois de Boulogne, richly decorated with enameled terra cotta by Girolamo della Robbia. But by the end of the 16th century terra cotta had once more fallen from popularity for architectural decoration, not to become a significant material again until the 19th century.

Free sculpture in terra cotta, too, found its revival in 15th-century Italy. Jacopo della Quercia, Donatello, Verrocchio and almost every other prominent sculptor left works in the material. Of particular interest, aside from the work of the Della Robbias, are the lifelike, coloured figure groups in Modena by Guido Maz-

zoni (1450–1518), who worked also in France, and the more sophisticated compositions of Antonio Begarelli (1498?–1565) in the same city. Sometimes terra-cotta sculpture was painted in natural colours, sometimes in imitation of marble or bronze, but despite this and the skillful complexity of modeling, the characteristics of form identifiable with terra cotta are usually evident.

Aside from the completed works in terra cotta, not many in number, remaining from the 17th century, there exist many figures intended, as were those of Bernini, as studies or models for work in marble. Many of the 17th- and 18th-century portrait busts in terra cotta also were used for this purpose, but they are made striking by a faithfulness in the rendering of personal characteristics less frequently found in other media and exhibit a quality of immediacy not transferable to the sterner material. Among the most descriptive are the busts created in France during the 18th century by such sculptors as Jean Baptiste Lemoyne, Jean Jacques Caffieri and Jean Antoine Houdon.

It was also during the 18th century in France that small, freely modeled allegorical and mythological groups in fine terra cotta, less delicate than the contemporary porcelains but more robust in form and action, were produced in large numbers at Sèvres and met with great popular favour. Foremost among the creators of these fantasies was Claude Michel, called Clodion (1738–1814). Portrait statuettes of historical figures and famous contemporaries were also produced in profusion toward the end of the century but seem to have had a less universal appeal. The factories at Sèvres, closed during the French Revolution, were reopened under Napoleon but produced little more terra cotta.

19th and 20th Centuries.—While terra cotta was used by sculptors in the 19th century, especially by those working usually in bronze, it had little character of its own. When used commercially to reproduce figures for the popular market it was regularly disguised. Not until the early years of the 20th century, when the artist, taking a fresh interest in the properties of his medium, began to choose materials in which he could work directly, did it again assume a positive character in art. Its use was linked to the revived interest in pottery, and many artists experimented with mixtures and glazes to produce effects of colour and texture. In architecture it was first revived by architects such as Karl Friedrich Schinkel (1781–1841), who became interested in earlier design in brick, and it served various historical styles and structural purposes throughout the century. The Architecture academy (1832–35) in Berlin by Schinkel and the Natural History museum (1873–80) at South Kensington, London, by Alfred Waterhouse are two very different examples. But the concern of modern architects and sculptors for the structural and aesthetic properties of their material has gained for terra cotta, along with other mediums, a new place in architectural ornament free from past usages and styles. (J. C. T.)

TERRAMARA (Ital. *terra*, "earth," *marna*, "fertile"), a term for a specific kind of settlement in the north Italian plain, with a southeastern extension on the eastern coast of Italy. Modern archaeology has changed this old terminology in *stazione arginate* (Ital. *argine*, "dike"), which means a settlement surrounded by an earthen dike. As no fully excavated stazione arginate exists it is difficult to determine the true character of such settlements.

Metal finds from the stazioni arginati prove that settlement had begun sometime in the period from the Middle Bronze Age to the Early Iron Age (prae-Benacci). Pottery with horned and lunated handles (*ansa cornuta-lunata*) proves that the aboriginal development of this Bronze Age culture ran parallel to that of the mound-grave culture and the early urn-field culture of temperate Europe.

The cremation burials of late terramare, do not appear to have been influenced by burial rites of the urn-field culture. The whole inventory of the stazioni arginati forms a branch of middle-European Bronze Age culture of the 2nd millennium B.C.

See G. Säflund, *Le Terremare delle provincie di Modena, Reggio Emilia, Parma, Piacenza* (1939); G. Kaschnitz-Weinberg, *Italien, Handbuch der Archäologie*, 4th part (1950). (Rd. P.)

TERRAMYCIN: *see* Oxytetracycline.

TERRAPIN, the U.S. name for certain fresh-water aquatic turtles, especially those marketed for food. *See* Turtle.

TERRE HAUTE, a city of western Indiana, U.S., on the Wabash river, is located 73 mi. W.S.W. of Indianapolis; the seat of Vigo county and one of the oldest cities in the state. The site was once a point of rendezvous for Indian tribes; nearby Fort Harrison was the scene of a battle between forces of Capt. Zachary Taylor and Tecumseh in 1812.

Terre Haute is on a plateau 50 ft. above the river, and the French name, meaning "high ground," was adopted by members of a small land company who platted the town in 1816. It became a key point on the National road (1835), the Wabash and Erie canal (1843) and the Richmond railroad (1852). When it was incorporated as a town in 1832 Terre Haute had a population of less than 1,000; by 1853, when it was chartered as a city, the population was almost 7,000.

The discovery of bituminous coal and other minerals paved the way for industrial development; by 1875 Terre Haute was one of the chief manufacturing centres in the midwest, producing steel and other commodities from resources available in the immediate area. Through World War I expansion of the city continued at a phenomenal rate. After 1920 the rate of growth slowed, but Terre Haute's industrial output of chemicals, pharmaceuticals, plastics, clay and foundry materials, glass, clothing, foodstuffs and other products places it as one of the principal manufacturing cities of the state. Pop. (1960) 72,500; standard metropolitan statistical area (SMSA; Vigo county) 108,458. In 1963 Clay, Sullivan and Vermillion counties were added to the SMSA, increasing its area from 415 to 1,084 sq.mi. and the population (1960 figures) to 172,069. (For comparative population figures *see* table in Indiana: *Population.*)

Located in or near Terre Haute are: St. Mary-of-the-Woods, a Roman Catholic college for women organized in 1840; Indiana State Teachers college, established in 1870 as a normal school; and Rose Polytechnic institute, for men, founded in 1874 and named for Chauncey Rose, a local entrepreneur.

The city is the birthplace of Eugene V. Debs, labour and Socialist leader; the author Theodore Dreiser and his brother, the composer Paul Dresser (who changed the spelling of the family surname), were born nearby. Four miles from the city is a federal penitentiary. (V. M. B.)

TERRESTRIAL MAGNETISM: *see* Geomagnetism.

TERRISS, WILLIAM (William Charles James Lewin) (1847–1897), a leading English actor of later Victorian days, was born in London on Feb. 20, 1847. After trying the merchant navy, sheep farming in the Falkland Islands and horse breeding in Kentucky, he went on the stage. He was immediately successful and appeared at the principal London theatres from 1868 onward. He joined Henry Irving at the Lyceum theatre in 1880, playing such parts as Cassio, Mercutio and Henry VIII. He then became the leading man at the Adelphi theatre, where, as the hero of melodramas he became the idol of the public, his fine voice and appearance and breezy manner making a great appeal. His last appearance was made at the Adelphi in *Secret Service* under tragic circumstances. He was stabbed to death by a madman, Richard Prince, while entering the theatre on Dec. 16, 1897.

Terriss married Miss Isabel Lewis, and their daughter Ellaline Terriss (1871–) became a great popular favourite in both straight and musical plays, appearing with her husband, Sir Seymour Hicks (1871–1949). She was leading lady at the Gaiety, the Vaudeville, the Globe (then the Hicks), the Aldwych and many other theatres and also toured much of the world with great success.

See A. J. Smythe, *The Life of William Terriss* (1898). (W. J. M.-P.)

TERRITORIAL ARMY. The British territorial army has its foundations laid firmly in the past; the form of its organization in the middle of the 20th century was to a great extent derived from the experience gained in having had an army of volunteers for the two centuries foregoing.

There have been volunteer armies throughout the history of Great Britain and it is necessary to delve into the histories and tra-

ditions of these forces in order to understand the spirit of the modern territorial army.

History.—This force came into being in May 1859, during the widespread fear of invasion from France, when the secretary of state for war, Gen. Jonathan Peel, sanctioned the raising of bands of volunteers. These bands received the full support of the country and within two years about 150,000 men had joined these locally formed bodies of volunteers.

The movement however was no new venture but the revival of an idea. The custom of raising volunteer forces to assist the crown in time of need dates back to the times when Anglo-Saxon England formed bands of freemen to face any danger that threatened the homeland. It was these freemen, called the *fyrd*, who fought the Danish invasion of those times. Voluntary military societies existed in various parts of England as early as Tudor times. An interesting charter was granted by Henry VIII to "The Fraternity or Guild of St. George: Maisters and Rulars of the Science of Artillery as aforesaid rehearsed for long-bowes, cross-bowes and hand gunnes." This corps later became the Honourable Artillery company, one of the oldest regiments in the British army. Later, trained bands were formed in the larger cities, and at the time of the Spanish Armada, it was the London trained band who marched to Tilbury in defense of the city. In 1794 these bands of volunteer corps and military associations were reorganized by a royal statute; but the end of the Napoleonic Wars saw the disbandment of these volunteer units, and nearly five decades were to pass before they were raised again.

One arm, however, continued throughout these years, and this was the yeomanry volunteer cavalry, raised in 1794 and drawn from the counties. It was with this material that in 1907 Lord Haldane, making much use of the services of Maj.-Gen. Douglas Haig, created the territorial force by means of the Territorial and Reserve Forces act, which unified the system of administration, training and terms of enlistment for the whole of the auxiliary forces. Local associations were maintained and strengthened. Each infantry regiment had maintained two regular battalions. The remnants of the local militia, under the name of special reserve, became the third battalion, with the duty of finding drafts for the first line. The volunteers produced a fourth and a fifth battalion.

This territorial force was for home defense, but so keen and efficient was it that when World War I came, many units volunteered and went overseas before the end of 1914, although still short of training. Thus the rather comic legacy of the volunteers had vanished, the jokes about Saturday-afternoon soldiers had gone, and membership of the territorials had become a source of pride and a symbol of duty. This force, organized on a war footing, was well established by 1914. It had 14 divisions overseas in 1915 and 21 divisions in various theatres of war by 1918, as well as 5 divisions constantly on home defense.

Between World Wars I and II there were three changes. The territorial force which had stood down in 1918 became the territorial army in 1922, and the men undertook to serve overseas if required. From 1936 onward there was a rapid expansion of antiaircraft units, and coast defense also became the territorial army's responsibility.

By 1935 the situation in Europe was rapidly worsening, and the British government once again emphasized the importance of the territorial army. There was an immediate response, and by spring 1939 the territorial army had been rapidly expanded to 212,000 men. It was divided into three main parts: the field-force element; the antiaircraft units; and coast defense. The first was a counterpart of the regular army and consisted of 2 horsed cavalry brigades and 12 infantry divisions; three of these latter were motorized and all were equipped on the same lines as the regular army. The antiaircraft units were grouped into five divisions and based in the areas in which the men lived. The coast defense element, comprising about 8,000 men (all ranks), operated on the same basis. Such was the territorial army at the outbreak of World War II.

During the war the territorial army, better trained than ever before, sent three divisions to France with the original British expeditionary force and from then onward was continually engaged. As in World War I, many of the most famous actions were fought by territorial army units. Regular and territorial soldiers became virtually indistinguishable, and many territorial officers reached high rank in command and on the staff. In 1945 the old territorial army stood down.

On New Year's day, 1947, the territorial army was reconstituted. This marked a further milestone in its history. It was raised again as part of the national army, quite unobtrusively. Between 1947 and 1950 about 80,000 volunteers were enrolled, but events abroad made it evident that further members were necessary. Part-time service with the reserve forces was introduced in 1948, and national-service men began to be posted to territorial units in 1950.

The territorial army was re-formed in phases and, from 1950 onward, had drafted into it national-service officers and men who had already done two years' full-time service with the regular army. In 1955 the burden of part-time national-service training was reduced to only 20 days training in 3½ years. Later in 1957 this training was completely suspended as part of the proposals to do away with national service by 1961. This reversion to a truly volunteer force resulted in increased enthusiasm for the T.A., and the figure of 77,910 true volunteers was achieved by March 1958.

With volunteers enlisting in the approximate numbers of 1,000 each month in the golden jubilee year of 1958, the T.A. could look forward to a healthy future.

To enable Great Britain to take its rightful place with the United Nations as one of the guarantors of world security was the main purpose for which the territorial army was reconstructed in 1947. In view of the speed of modern war, the territorial army, alongside the regular army, has to be ready to mobilize for immediate defense of the country. The territorial army is an essential complement to the regular army. It forms the nation's reserve of disciplined manpower, trained and organized to fight in any emergency. In a major war its main task would be the defense of the United Kingdom in all its aspects.

By virtue of its countrywide spread, the T.A. is well suited to this role and could provide a most valuable steadying influence in a time of grave national crisis. The T.A. is also trained to give military support to the civil defense forces in the event of nuclear attack.

From the outbreak of war the regular and territorial armies would be integrated as a whole. In the United Kingdom the combined resources of regular and territorial units would be used together to meet initial operational requirements and to provide the framework of subsequent expansion.

Organization.—While the functions of command, training and discipline in the territorial army are under the central direction of the war office, administration is on a local basis. Each county has its own territorial association. The lord lieutenant of the county, who is the crown's representative, is normally its president. Other members are nominated by the local military representatives (the members are usually commanding officers of county units); by local government bodies; by educational authorities; by employers' organizations; by trade unions; and by preservice training organizations. Local administration deals in peacetime with the problem of recruiting, clothing, equipment and accommodation.

Broadly, the financial structure of the territorial army has remained unchanged since the formation of the first territorial force in 1907, when grants were introduced for payment to county territorial associations. In the annual army estimates provision is included in Vote 2 for the pay and emoluments of the territorial army; for grants to territorial associations for the general maintenance and administration of units; for the cost of T.A. works, buildings and lands and for the training of the territorial army.

The cost of providing other services for the territorial army, including unit equipment, stores and miscellaneous noneffective services, is included in regular army votes. In the estimates for 1958–59 about £12,000,000 was provided in Vote 2 to defray the expense of the territorial army.

The composition of the territorial army is exceedingly diverse and includes ten infantry divisions and certain independent bri-

gades. To support the various formations there are a large number of units of royal armoured corps, royal artillery, royal engineers, royal signals, royal army service corps, royal army medical corps, royal army ordnance corps, royal electrical and mechanical engineers and so on.

The Queen Alexandra's Royal Army Nursing Corps.—The Queen Alexandra's royal army nursing corps (T.A.) is formed for the purpose of maintaining an establishment of fully trained nursing officers and other ranks, who are prepared to undertake full-time military service at home or overseas on the embodiment of the T.A. During peace, nursing officers and other ranks will be posted to and remain, during unembodied service, with the unit of their choice, but when embodied they will be available for posting to medical units according to the requirements of the service.

The Women's Royal Army Corps.—The women's royal army corps (T.A.), which is the territorial branch of the W.R.A.C., shares the role of the territorial army in home defense in a major war. The obligations of its members are similar to those of the men, except that no weapon training is carried out. Women are employed in mixed signal regiments as switchboard and teleprinter operators. They are also employed as clerks, cooks, drivers and storewomen in W.R.A.C. (T.A.) battalions.

Officers and other ranks are required to undertake a general-service liability and to serve overseas on embodiment if required. In the event of a national emergency they will be liable to service in any place in the United Kingdom, if called upon to do so, whether or not an order embodying the territorial army is in force at that time.

Conditions of Service.—A volunteer may join the territorial army for two, three or four years, extending his service at the end of his initial engagement by periods from one to four years long. On enlistment he undertakes to attend an annual training camp, usually for 15 days; to do 30 training periods a year; to fire an annual range course; and to serve, when called upon, at any place in the United Kingdom during a period of national emergency. The T.A. may be embodied in whole or in part when the army reserve has been called out by royal proclamation. Officers and men of the T.A. when embodied are liable for service in any place outside the United Kingdom. Training periods each consist of a minimum of one hour's instruction. Most units run several week-end camps a year, each day of which counts as four training periods. Annual training is mostly carried out during the spring and summer months, during which time units proceed to camp and carry out normal military duties and training for two weeks. In the case of field formations, camps are usually held on a unit, brigade and/or divisional basis; field and antiaircraft gunners go to the gunnery practice camps, other arms to training areas or establishments where facilities exist for their special training. The two weeks' annual camp is the only time civilian soldiers are able to operate as complete organized military units and sub-units.

During annual camp and week-end training, officers, men and women of the T.A. receive the same rates of pay and certain of the allowances which are drawn by their counterparts in the regular army. For training periods at drill halls they receive a training expenses allowance, appropriate to their respective ranks and the time spent in training, to cover any incidental expenses incurred by their attendance. In addition, volunteer other ranks qualify for an annual bounty of £12 (£9 in the case of women) by completing all their obligations and performing a small amount of voluntary training in an efficient manner. Officers may earn a bounty of £7 (£5 5s. for women). A bounty of £20 (£15 in the case of women) may be earned by volunteer other ranks in their fourth and subsequent years of voluntary service. A volunteer who has completed a regular army engagement may earn bounty at the higher rate. The officers' bounty for fourth and subsequent years is £15 (£11 5s. in the case of women). A volunteer with national-service training liabilities may earn the same rates of bounty as other volunteers unless at any time the national-service part-time training obligation should be reimposed, when the lower national-service volunteer rate would be paid.

The territorial soldier, be he officer or other rank, is still the same as he has ever been. He still goes to his local drill hall, he still leaves his office chair, his machine in the factory, or his plow on the farm, to put on uniform at week ends and to take part in periodical weapon training on the ranges, or week-end exercises, conducted under brigade or regimental arrangements. The 19th-century cartoon still is true: The Continental: "I thought you British were a nation of shopkeepers." John Bull: "We are, and those volunteers marching along the street guard the shop."

(A. K. Ck.)

TERRY, DAME ELLEN ALICIA (1847–1928), English actress, famous for her great partnership with Henry Irving, was born at Coventry on Feb. 27, 1847. Her parents were well-known provincial actors and her sisters Kate, Marion and Florence and her brother Fred all joined the theatrical profession. Her own first appearance on the stage was made on April 28, 1856, under Charles Kean's management, as the boy Mamillius in *The Winter's Tale*, at the Princess's theatre, London. Two years later she played Prince Arthur in *King John* and won high praise. In 1864 she left the stage to marry the painter G. F. Watts, then 30 years her senior. Though the marriage was a failure, Watts refused her a divorce and later she eloped with Edward Godwin, the architect, who was the father of her children Edith and Edward Gordon Craig (*q.v.*). Returning to the stage for a brief period in 1867, she played with Henry Irving for the first time as Katherine to his Petruchio in David Garrick's version of *The Taming of the Shrew* at the Queen's theatre. Leaving the stage again, she lived in retirement at Harpenden, Hertfordshire, until 1874 when, at the instigation of Charles Reade, she returned to the stage in the part of Philippa in *The Wandering Heir* under his management at the new Queen's theatre. In 1875 she played Portia in the Bancrofts' production of *The Merchant of Venice* at the Prince of Wales's theatre. Her performance won immediate acclaim and remained matchless throughout her career. In 1876 she married the actor Charles Kelly (E. A. Wardell).

In 1878 Miss Terry was engaged by John Hare at the Court theatre, where her performance as Olivia, the title role in W. G. Wills's adaptation of Oliver Goldsmith's *Vicar of Wakefield*, attracted the attention of Henry Irving. When he assumed management of the Lyceum theatre in 1878 he engaged her as his leading lady. Her personal charm, vitality and aesthetic judgment helped to make their histrionic partnership, lasting 13 years, the greatest the English stage has seen. Her first part at the Lyceum was Ophelia, and her performance, with its tender and authentic study of the derangement of an innocent mind, astonished those who recognized her genius in comedy. Her other Shakespearean impersonations at the Lyceum were Portia (1879), Desdemona (1881), Juliet and Beatrice (1882), Viola (1884), Lady Macbeth (1888), Katherine in *Henry VIII* and Cordelia (1892), Imogen (1896) and Volumnia in *Coriolanus* (1901). Other notable performances were those of the queen in Wills's *Charles I* (1879), Camma in Alfred Tennyson's *The Cup* (1881), Margaret in *Faust* (1885), Nance Oldfield (1893), Rosamund in Tennyson's *Becket* (1893), Madame Sans-Gêne (1897) and Clarisse in *Robespierre* (1899).

With the Lyceum company Ellen Terry toured the United States and Canada. In 1902 she appeared with Mrs. Kendall in Beerbohm Tree's revival of *The Merry Wives of Windsor* at His Majesty's theatre. She and Irving played together for the last time at the Lyceum in a performance of *The Merchant of Venice* on July 19, 1902. In 1903 she presented several productions by her son Edward Gordon Craig at the Imperial theatre, notably *The Vikings at Helgeland* by Henrik Ibsen. In 1905 she played "with indescribable fascination" Lady Cecily Waynflete in Bernard Shaw's *Captain Brassbound's Conversion*, but the hopes Shaw had held out to her in their famous correspondence of wider opportunities outside the Lyceum theatre were not fulfilled, for she appeared afterward only in *Alice Sit-by-the-Fire* and in *Pinkie and the Fairies*. After her stage jubilee had been celebrated at the Theatre Royal in June 1906, she toured the English provinces and the United States giving witty and discerning lectures on Shakespeare's heroines. In 1907 she married the American actor James Carew.

In 1925 Ellen Terry received the grand cross of the order of the British empire. Her 80th birthday in 1928 was the occasion for tributes and messages from her countless admirers, headed by King George V and Queen Mary. She died on July 21 of that year at her house in Small Hythe near Tenterden, Kent, which now, as the Ellen Terry Memorial museum, is the property of the National trust. At her death the *Times* (London) wrote: "She was a woman of genius; but her genius was not that of the brain so much as of the spirit and of the heart. She was a poem in herself—a being of an exquisite and mobile beauty."

Her *Memoirs,* edited by Edith Craig and C. St. John, was published in 1933.

BIBLIOGRAPHY.—E. Gordon Craig, *Ellen Terry and Her Secret Self* (1931); *Ellen Terry and Bernard Shaw: a Correspondence,* ed. by C. St. John (1931); L. H. F. Irving, *Henry Irving: the Actor and His World* (1951). (L. H. F. I.)

TERRY, SIR RICHARD RUNCIMAN (1865–1938), knighted 1922, British organist and choir director, was born at Ellington, Northumberland, the son of Thomas Terry of Newcastle-on-Tyne, and obtained a choral scholarship at King's college, Cambridge. After acting as organist and choirmaster at Elstow school (1890); St. John's cathedral, Antigua, W. I. (1892); and at Downside Abbey near Bath (1896), he became organist and director of the Roman Catholic cathedral, Westminster, London (1901–24). His great service to English music is twofold: he studied and revived the beautiful church music of the English composers of the 17th century, and by the finished performances at Westminster of the works of these and the Italian composers brought about a revolution in church services in England, even outside his own communion. He was the composer of five Masses and other music, and the author of *Our Church Music* (1901); *Catholic Church Music* (1907); *Tudor Church Music* (1922); *On Music's Borders* (1927); etc. He also compiled the *Westminster Hymnal* and published a collection of shanties in *The Shanty Book* (pt. I. 1921, pt. II. 1926).

TERTIARIES, associations of lay men and women in connection with the mendicant orders. The institution of tertiaries arose out of the Franciscan movement. When the immediate disciples of St. Francis had become an order bound by the religious vows, it became necessary to provide for the great body of laity who could not leave the world but who desired to carry out in their lives the spirit and teaching of the Franciscan movement. And so, probably in 1221, St. Francis drew up a rule for those of his followers who were debarred from being members of the order of Friars Minor. At first they were called Brothers and Sisters of the Order of Penance; but later on, when the Friars were called the First Order and the nuns the Second Order, the Order of Penance became the Third Order of St. Francis—whence the name tertiaries.

In time a tendency set in for members of the Third Order to live together in community, and in this way congregations were formed who took the usual religious vows and lived a fully organized religious life based on the Rule of the Third Order with supplementary regulations. These congregations are the regular tertiaries as distinguished from the secular tertiaries, who live in the world, according to the original idea. There are many congregations of regular tertiaries, both men and women. In 1883 Leo XIII caused the rule to be recast and made more suitable for modern times.

At an early date the other mendicant orders formed third orders on the same lines as the Franciscan, and so there came into being Dominican, Carmelite, Augustinian, Servite and Premonstratensian tertiaries, together with a large number of others.

TERTIARY, in geology, the term applied to the time interval next after the Cretaceous and before the Pleistocene. Currently it is considered as a period of the Cenozoic era (*q.v.*) as indicated on the accompanying geologic time chart but in the past it has usually been given the rank of an era.

This article deals with the evolution of life and with geologic conditions during the Tertiary and discusses the history of the study of the period and its division into epochs and stages. For general information on the place of the Tertiary in geologic time *see* GEOLOGY: *Historical Geology.* For more detailed information about the epochs included in this period *see* MIOCENE; OLIGOCENE; EOCENE AND PALAEOCENE.

Life in the Tertiary.—Of the various organisms whose fossil remains are preserved in Tertiary rocks the mammals are most notable. During this period mammals evolved from small primitive creatures which survived the reign of dinosaurs to become the dominant vertebrates. The modifications which took place in various mammalian lines during Tertiary time are astounding, and as a result mammalian stages of evolution often provide the easiest means for world-wide correlation of Tertiary deposits.

Geologic Time Chart

System and Period	Series and Epoch	Distinctive Records of Life	1,000 Years
CENOZOIC ERA			
Quaternary	Recent	Modern man	11
	Pleistocene	Early man	1,000
Tertiary	Pliocene	Large carnivores	
	Miocene	Whales, apes, grazing forms	
	Oligocene	Large browsing mammals	
	Eocene	Rise of flowering plants	
	Paleocene	First placental mammals	70,000
MESOZOIC ERA			
Cretaceous		Extinction of dinosaurs	130,000
Jurassic		Dinosaurs' zenith, primitive birds, first small mammals	160,000
Triassic		Appearance of dinosaurs	200,000
PALEOZOIC ERA			
Permian		Reptiles developed, conifers abundant	235,000
Carboniferous			
Upper (Pennsylvanian)		First reptiles, coal forests	260,000
Lower (Mississippian)		Sharks abundant	285,000
Devonian		Amphibians appeared, fishes abundant	320,000
Silurian		Earliest land plants and animals	350,000
Ordovician		First primitive fishes	400,000
Cambrian		Marine invertebrates	500,000
PRE-CAMBRIAN TIME			
		Few fossils	3,500,000–4,000,000

The Paleocene was characterized by a variety of generally small primitive mammals only remotely related to modern types, but with the beginning of the Eocene many of the more modern stems appear. Concurrently, extinction of aberrant or unsuited phyla such as uintatheres, creodonts, titanotheres and oreodonts occurred at various times and many lines of development did not survive the Tertiary.

The diverse Tertiary fishes, both fresh water and marine, were already broadly differentiated by Cretaceous time, and subsequent development was little more than speciation. Of the various amphibians, only frogs, salamanders and the wormlike caecilians survive the period. Lizards, snakes, crocodiles and turtles are the few reptiles which survived the age of dinosaurs. Birds are sparsely represented but the known fossils show an early diversification into modern types, presumably in the late Mesozoic and early Tertiary eras.

Marine invertebrate faunas of the Tertiary show a great variety of molluscan forms and an abundance of large foraminifera, but the once dominant ammonites are absent, although other cephalopods such as nautiloids and squids survived.

In the plant world, the greatest advances were among the flowering plants (angiosperms). Specialization of structures adapted to insect pollination, and fructifications suitable for dissemination by birds and other animals accompanied the general progress of animal evolution. *See* also PALAEOBOTANY; PALAEONTOLOGY.

Conditions During the Tertiary.—The end of the Cretaceous was marked in most areas by a great regression of the sea and by a shifting of the basins of deposition, generally toward the present margins of the continents. New marine transgressions were initiated in the Paleocene and lower Eocene in many parts of the world. During the Eocene, these transgressions advanced notably in most places. Tethys, the seaway extending from the western Mediterranean across to India, was finally broken up about the beginning of the Miocene. Marine transgressions during the Miocene were often as extensive as those of the Eocene, but in general the seas encroached less and less on the land in the later Tertiary.

TABLE I.—*Some Different Regional Standards and Suggested Correlations*

European		North America terrestrial vertebrates		West America marine megafaunal		Local Equivalents	California marine foraminiferal	
Epoch	Stage	Epoch	Age	Epoch	"Stage"		Epoch	Stage
Pliocene	Calabrian	Pliocene	Blancan	Pliocene	San Joaquin		Plio-pleist.	Wheelerian
	Astian				Etchegoin		Pliocene	Venturian
	Plaisancian		Hemphillian		Jacalitos			Repettian
Miocene	Pontian			Miocene	Neroly		Mio-Pliocene	Delmontian
	Sarmatian		Clarendonian		Cierbo			
	Tortonian	Miocene	Barstovian		Briones		Miocene	Mohnian
	Helvetian		Hemingfordian		Temblor			Luisian
	Burdigalian				Vaqueros			Relizian
	Aquitanian		Arikareean		Blakeley		Oligo-Miocene	Saucesian
Oligocene	Chattian	Oligocene	Whitneyan	Oligocene				
	Rupelian		Orellan		Lincoln			Zemorrian
	Tongrian		Chadronian				Olig.	Refugian
Eocene	Ludian	Eocene	Duchesnean	Eocene	Keasey		Eocene	Fresnian
	Bartonian		Uintan		Tejon			Narizian
	Auversian		Bridgerian		"Transition"			
	Lutetian				Domengine			Ulatisian
	Cuisian		Wasatchian		Capay			"Juniperan"
	Ypresian							
Paleocene	Thanetian	Paleocene	Clarkforkian	Paleocene	Meganos		Paleocene	Bulitan
			Tiffanian					
	Montian		Torrejonian		Martinez			Ynezian
			Dragonian					
			Puercan					

TABLE I.—*Some Different Regional Standards and Suggested Correlations (Continued)*

Indonesia (East Indies) marine		New Zealand marine			South America terrestrial vertebrates	
Epoch	Stage	Epoch	Stage		Epoch	Age
Pliocene	h 2	Pliocene	Castlecliffian		Pliocene	Chapadmalalan
	h 1		Barwoonian	Nukumaruan		Montehermosan
Mio-Pliocene	g			Waitotaran		Tunuyanian
				Opoitian		Huayquerian
Miocene	f 3	Miocene	Tarakian	Urenuian	Miocene	Mesopotamian
	f 2			Tongapoturuan		Chasicoan
	f 1		Awamoan			Friasian
	e 5		Hutchinsonian			Santacrucian
	e 4					
	e 3	Oligocene	Waitakian		Oligocene	Colhuehuapian
	e 2		Duntroonian			?
Oligocene	e 1		Ototaran	Whaingaroan		Deseadan
	d			Kaiatan		
	c					?
Eocene	b	Eocene	Waimatean	Tahuian	Eocene	Mustersan
	a 2			Bortonian		?
	a 1		?			Casamayoran
Paleocene	?	Paleocene	?		Paleocene	Riochican
						?
						(Salamancan?)

The great mountain ranges of the western hemisphere, such as the Rocky mountains and the Andes, were established by the beginning of Tertiary time, but they had periodic rejuvenations during the period. Alpine orogeny, however, appears to have been mid-Tertiary in its expression. The coast ranges of western United States, although representing a zone diastrophically active during most of Tertiary time, assumed their present shape during the late Tertiary and Pleistocene, a history more or less applicable to other mountain chains in the great seismic belts.

Continental or terrestrial deposits of Tertiary age accumulated on all the land masses which were emergent during the period. These deposits occur in many places, but those of western United States, Argentina, western Europe, Egypt, India and northern China are perhaps the best known and have furnished most of the known terrestrial fossils of this age.

In the early Tertiary, tropical climates extended to much higher latitudes (for instance, to 50° N. or more along the western coast of North America) than at present. Subsequently the isotherms gradually retreated toward the equator, culminating in the Pleistocene ice age. The present day narrow tropical belt is abnormal and during much of geologic time the tropics were as widespread as in the early Tertiary.

Epochs and Stages of the Tertiary.—The name was first used and defined by Giovanni Arduino in a letter (*Nuova Raccolta d'Opuscoli Scientifici e Filologici,* vol. 6, pp. 162–163, 1760) to Antonio Vallisnieri, professor in the University of Padua, Italy. Subsequently the term was used by G. Cuvier and A. Brongniart in 1801 for deposits in the Paris basin. The Tertiary faunas of western Europe consist largely of mollusca of modern aspect, and in 1830 Gérard Paul Deshayes determined the percentages of living species found at several horizons.

With the aid of Deshayes' percentages, the work of J. J. d'Omalius d'Halloy in Belgium, and of Thomas Webster and William Buckland in England, Sir Charles Lyell in 1833 divided the Tertiary into three main epochs, viz., Eocene with 3½%, Miocene with 17% and Pliocene with 35% to 50% of recent species. In 1854 Heinrich Ernst von Beyrich

TABLE II.—*Some Divergent Tertiary Classifications*

European stages	A common usage, North America, Europe	North America pre-1924, some European usages		North America vertebrate paleontologists, some European	Some micropaleontologists
Calabrian	Pleistocene			Pleistocene	Plio-pleistocene
Astian	Pliocene	Pliocene		Pliocene	
Plaisancian					Pliocene
Pontian	Miocene				Mio-pliocene
Sarmatian					
Tortonian		Miocene		Miocene	Miocene
Helvetian					
Burdigalian					
Aquitanian		Oligocene			Oligo-miocene
Chattian	Oligocene			Oligocene	
Rupelian					
Tongrian					Oligocene
Ludian	Eocene	Eocene	Eocene	Eocene	Eo-oligocene
Bartonian					Eocene
Auversian					
Luterian					
Cuisian					
Ypresian					
Thanetian	Paleocene		Paleocene	Paleocene	Paleocene
Montian					

introduced the term Oligocene for the time interval represented by deposits transitional between Eocene and Miocene. A fifth division, Paleocene, was proposed by Wilhelm Philipp Schimper in 1874 for the earliest part of the Tertiary on the basis of paleobotanical evidence. At present the five epochs of the Tertiary are recognized not on the percentage of living species, as first defined by Lyell, but on the general character and comparative stage of development of the faunas, or on stratigraphic relationships.

Some geologists group the epochs of the Tertiary into two primary divisions, as follows:

Later Tertiary or Neogene { Pliocene, Miocene
Earlier Tertiary or Eogene (or Paleogene or Nummulitic) { Oligocene, Eocene, Paleocene

During the Tertiary, the marine faunas became increasingly more provincial and in consequence there is much difficulty in making intercontinental correlations. Because of this, local standards have been established, notably in North and South America, Indonesia (East Indies), New Zealand and Australia. In some areas, as in the western United States there are as many as three different standards used, each primarily utilizing a different group of organisms for recognition of the intervals. As a result there are divergent correlations of the same local sequence with those of Europe (*see* Table I).

As proposed by Lyell, Beyrich and Schimper, the five epochs are not definitely bounded, and inasmuch as their types are in different basins and there are certain intervals (such as the Pontian and Sarmatian) not clearly included in any epoch, discrepancies have arisen in their usage and significance (*see* Table II).

In addition to the epoch subdivisions of the Tertiary, there early arose a series of time-stratigraphic units, or stages. This concept was first proposed by Alcide d'Orbigny in 1840 and 1842 and subsequently formed the basis for most regional correlation in Europe. Because the types of these stages are not always in the type areas of the epochs and because of the uncertain boundaries of the epochs, there is disagreement regarding placement of some stages.

See M. G. Wilmarth, "Geologic Time Classification of United States Geological Survey Compared with Other Classifications," *Bull. U.S. Geol. Surv.*, 769:1–138 (1925); A. M. Davies, *Tertiary Faunas*, vol. 2 (1934). (J. W. D.)

TERTULLIAN (QUINTUS SEPTIMIUS FLORENS TERTULLIANUS) (*c.* 155–after 220), one of the greatest of the early Christian writers of the west, was born of pagan parents in Carthage. His education gave him a fair knowledge of philosophy and Greek and Latin literature and he was further trained in the law. In his youth deeply involved with the literary social life of the day, he remained for the whole of his life under the spell of the word, but his legal training left an even greater impress on his personality and his writings. He became an expert in law, though it is most unlikely that he is to be identified with the jurist Tertullianus, of whose writings the *Corpus iuris civilis* gives a few excerpts.

In mature manhood Tertullian became a Christian (*c.* 193). He does not give the reasons for his conversion, which took place in Rome or Carthage, but the heroism of Christians in times of persecution impressed him deeply and personal experience is probably reflected in his famous words *semen est sanguis Christianorum* ("the seed is the blood of Christians," *Apologeticum* 50). After

his conversion he devoted himself to the study of Scripture and of Christian literature, and by using his knowledge of law, literature and philosophy he became the great defender of the Christian faith against the pagan state, against the Jews and against several heretic sects. According to Jerome he became a priest of the church of Carthage. His clerical status is denied by several modern scholars, but his preponderant role of teacher and the catechetical character of several of his treatises which apparently summarize sermons seem to confirm the information of Jerome, who concerning Tertullian draws on reliable tradition.

Tertullian carried on his literary activity between the years 195 and 220. His works show an ever-growing rigorism which led him to sympathize with Montanism (*q.v.*), and about 207 he openly went over to this sect. But soon he broke off from it to become the head of a small group, the so-called Tertullianists, who survived at Carthage up to the time of St. Augustine. Nevertheless Montanism did not prevent him from remaining dogmatically orthodox in most respects. His doctrine of the Trinity finds its best expression in the late work *Adversus Praxean* (after 213), and it is in this doctrine and in Christology that he made his greatest contribution to theology. In his latest works there is a growing violence and then, after 220, his literary activity ceases altogether. Having done battle with pagans, Jews, Marcionites, Gnostics, Monarchians and Catholics he died in Carthage, according to Jerome at a great age.

The works of Tertullian, most of which cannot be exactly dated with certainty, fall into three classes: the apologetic (defending Christianity against paganism and Judaism), the polemical-dogmatic (refuting heresies and heretics) and the disciplinary, moral and ascetic works (among which those of the extreme Montanist period form a special group, already marked as such by Jerome).

Apologetics.—The *Ad nationes* and the *Apologeticum*, which are related to each other, both seem to have been written in 197 and both defend Christianity against paganism. The first is a mere collection of material drawn from the Greek apologists; the *Apologeticum*, directed to the governors of the Roman provinces, is a very personal work in finished form. The argumentation shows clearly the influence of the author's juridical training. It is the most famous of all Tertullian's works and it has a textual tradition of its own. In the *De testimonio animae*, written soon after the *Apologeticum*, Tertullian discusses the *testimonium animae naturaliter Christianae:* the awareness of the creator in the soul, by nature Christian, as derived immediately from the universe and as evidenced in the ordinary language of men. *Ad Scapulam* (212), an admonition to the persecuting proconsul of Africa, anticipates a topic elaborated later by Lactantius. *Adversus Iudaeos* deals with the problem of Jewish tradition and particularly with that of the Old Testament. Its main source is Justin's dialogue with Trypho. According to Tertullian the Law is abolished because the giver of the New Covenant has already appeared. Henceforth the Old Testament must be interpreted spiritually. Ch. 9–14 are merely excerpts, made by a compiler, from book iii of Tertullian's *Adversus Marcionem*.

Polemical-Dogmatic Works.—In *De praescriptione haereticorum* (*c.* 200) Tertullian again makes the most of his knowledge of law. His aim was to end the controversy between Catholics and all heretics by advancing the technical argument of the *praescriptio*, a legal objection leading to rejection of a cause. According to Tertullian the heretics cannot use Scripture because it is not theirs; the gospel is the possession of the church, handed down to it by apostolic tradition. Other treatises deal with single heresies. In *Adversus Marcionem*, the major source for knowledge of the heresy of Marcion, Tertullian draws on Justin, Irenaeus and probably on a treatise by Theophilus of Antioch. *Adversus Hermogenem*, directed against the doctrine of Hermogenes, a painter living in Carthage according to whom God had created the universe out of preexisting matter, gives an explanation of the Christian teaching on creation. *Adversus Valentinianos*, dealing with the Gnostic sect of the Valentinians, depends on the first book of Irenaeus' great work against heretics. The *Scorpiace* is a defense of martyrdom against the Gnostics, who are compared to scorpions.

De carne Christi and *De resurrectione carnis* are closely connected: both deal with the resurrection of the body, which was denied by pagans, Sadducees and heretics. In *De carne Christi* Tertullian proves, against Docetism, that Christ was really born, that he truly lived, died in human flesh and rose from the dead. He defends the Virgin's true and real motherhood and stresses the human nature of Christ by maintaining that he was ugly. This treatise paves the way to *De resurrectione carnis*, which refutes those who denied the resurrection of the flesh. The *Adversus Praxean* (213), already mentioned, is directed against the Trinitarian doctrine of Praxeas. In it the term *trinitas* occurs for the first time. It is the most important contribution to the doctrine of the Trinity in the ante-Nicene period, though it is not free from subordinationism. *De anima* (*c.* 213) is a refutation of heretical doctrines about the soul, which leads Tertullian into controversy with pagan philosophy. It contains not a well-thought-out system of psychology but a series of theories of polemical intentions. Tertullian's chief thesis is the corporeality of the soul; he also advocates traducianism (*see* CREATIONISM AND TRADUCIANISM). His chief source was the pagan physician Soranus.

Disciplinary, Moral and Ascetic Works.—*Ad martyras*, a short address to a number of Christians (or catechumens) who were in prison awaiting martyrdom, dates probably from 197; that the famous martyrs Perpetua and Felicitas belonged to the group cannot be proved, and it is also uncertain whether Tertullian is the editor of their *Passio*. The *De spectaculis* (197 or 202) is a condemnation of all public games. In *De cultu feminarum* (between 197 and 201), Christian women are warned not to be dominated by pagan fashion. Three treatises, *De baptismo, De oratione* and *De paenitentia*, written between 200 and 206, all of them addressed to catechumens and similar in composition and tone, can be considered as a trilogy. The first, particularly important for the history of the baptismal rite, is the only ante-Nicene treatise on any of the sacraments. *De oratione*, which gives the earliest exposition of the Lord's Prayer in any language, deals with the interior and exterior discipline of prayer. *De paenitentia* is important for the history of ecclesiastical penance. *De patientia*, of the same period, sheds light on Tertullian's personality.

Ad uxorem, composed between 200 and 206, gives suggestions which his wife is to follow after his death. In the first book he admonishes her to remain a widow, in the second he discusses the possibility of remarriage: if she wishes to remarry it must be to a Christian. In *De exhortatione castitatis*, written some years later, he again takes up the problem of second marriage, which he now rejects as contrary to the will of God. In *De virginibus velandis* (*c.* 208) he returns to the idea which he had already formulated that women, whether married or unmarried, should be veiled. *De corona* (211), written in defense of a soldier who refused to wear a wreath, discusses a far more important problem: that of the participation of Christians in military service. The same problem, but in a larger framework, that of the behaviour of Christians in the pagan world, is discussed in *De idololatria* (211); the Christian must abstain from everything that is in any way connected with idolatry.

Tertullian's last works are characterized by Jerome as a sort of libel in which he discusses the controversies which brought about his rupture with the church. To this group belong *De monogamia, De fuga, De ieiunio* and *De pudicitia*. In *De monogamia* (213) the third of Tertullian's treatises on marriage, he regards second marriage as illicit and almost adulterous. In *De fuga* (212) he holds—in opposition to his previous views—that to take refuge in flight during persecution is against the will of God. *De ieiunio* is a fierce attack on the Catholic rejection of the Montanist practices of fasting. In *De pudicitia* he attacks the church again, this time in the field of penitential discipline, dealing particularly with the power of the keys. The *De pallio* is a small and bizarre tract, full of wit and sarcasm, in which Tertullian explains why in everyday life he substituted the mantle (*pallium*), the traditional garb of philosophers, for the Roman toga. The work is a sort of literary tour de force and the reader wonders if it is to be taken seriously. The change of dress coincides either with Tertullian's conversion or with his separation from

the Catholic Church.

Style and Influence.—Tertullian was the first Christian theologian to leave behind an extensive collection of works written in Latin. With him early Christian Latin literature made a magnificent start. Though he was certainly not the creator of early Christian Latin, a distinction which has erroneously been claimed for him, he was the first to introduce the special idiom of Latin-speaking Christians into literature. In style he is inspired by the Asianic manner, the fashionable "modern" style of his day, which preferred short paratactical sentences, adorned with all sorts of play on sounds and words, to the long classical periods. His style is highly personal, often pregnant and terse and not free from obscurity, though at the same time powerful and passionate.

Tertullian never even tried to create a theological system, but he often found formulas and definitions which were clear and precise, particularly in his doctrine of the Trinity. Some of his formulas are identical with those of the Council of Nicaea (325), and in his Christology he anticipates the Chalcedonian definition (451). All his works bear the stamp of his juridical training, not only in language and style, but even more in his way of thinking. His juridical training supplied him his major argument against heresy, the *praescriptio;* the way in which he deals with the relation of God and man is equally inspired by his juridical training, words such as debt, satisfaction, guilt, compensation, occurring regularly. Law also inspired his defense of the Christians against the state. He exercised great influence on Christian thought, particularly through Cyprian, who sifted the thoughts he had given, rounded them off and turned them into current coin, always aware of his dependence on Tertullian, whom he designated as his master.

Bibliography.—Works in J. P. Migne, *Patrologia Latina,* vol. 1–2 (1879); incomplete critical editions in *Corpus scriptorum ecclesiasticorum Latinorum,* vol. 19 (1890), ed. by A. Reifferscheid and G. Wissowa, vol. 47 (1906), ed. by A. Kroyman, vol. 69 (1939), ed. by H. Hoppe and A. Kroymann, vol. 76 (1957), ed. by V. Bulhart and P. Borleffs; and *Corpus Christianorum, series Latina,* vol. 1, ed. by E. Dekkers (1953–54), vol. 2 (1954), ed. by A. Gerlo. For editions of separate works to date see *Corpus Christianorum,* vol. 1; Eng. trans. by P. Holmes and S. Thelwall in *Ante-Nicene Christian Library,* vol. 7 (1868), 11 (1869), 15 (1870), 18 (1870). *See* also long bibliography in J. H. Waszink's edition of *De anima* (1947); J. Quasten, *Patrology,* vol. 2 (1953); B. Altaner, *Patrology* (1960). (Ch. Mo.)

TERUEL, a province of northeastern Spain, formed in 1833 from part of the ancient kingdom of Aragon; bounded on the north by Saragossa, east by Tarragona, southeast and south by Castellón de la Plana and Valencia, southwest by Cuenca and west by Guadalajara. Pop. (1960) 215,183; area 5,715 sq.mi. In the centre of the province rise the sierras of Gudar and San Just; in the southwest and west are the lofty Albarracin range, the Montes Universales and the isolated ridges of Palomera and Cucalon. Outliers of the Castellón and Tarragona highlands extend along the eastern border. The northern districts belong to the Ebro basin. In the west there are a few peaks, such as the Cerro de San Felipe (6,034 ft.) and Muela de San Juan (5,312 ft.), which are covered with snow for many months; the highest point is Javalambre (6,627 ft.) in the south. The sierras give rise to several large rivers, the principal being the Tagus (*q.v.*); the Guadalaviar, which rises in the Montes Universales and enters the Mediterranean at Valencia; and the Jiloca, which flows north from the lake of Cella to join the Jalón at Calatayud. The chief products are corn, wine, oil, cheese, fruits, timber, flax, hemp, silk, wool and saffron, cattle, sheep and swine. The Calatayud-Sagunto railway crosses the province. At the capital, Teruel (*q.v.*), was fought one of the most famous battles of the civil war of 1936–39.

TERUEL, the capital of the Spanish province of Teruel; on the left bank of the river Guadalaviar, at its confluence with the Alfambra and on the Calatayud-Sagunto railway. Pop. (1960) 19,776 (mun.). The older part of Teruel is a walled city with narrow streets and crumbling medieval houses, but modern suburbs have been built outside the walls. In the cloisters of San Pedro lie the remains of the celebrated "lovers of Teruel," Juan de Marcilla and Isabella de Segura (13th century). Their story is the subject of works by Tirso de Molina, Artieda, Montalban, Hartzenbusch and others. The cathedral dates from the 16th century. The aqueduct was erected in 1555–60 by Pierre Bedel. Teruel was the scene of fierce fighting in the civil war of 1936–39.

TERZA RIMA, or "third rhyme," a form of verse developed by the Italian poets of the 13th century. Its origin has been attributed by some to the three-lined *ritournel,* which was an early Italian form of popular poetry, and by others to the *Provençals serventese incatenato,* an arrangement of triple rhymes, which unquestionably appears to have a relation with *terza rima;* Dante gave to *terza rima* its artistic character. What this character is may best be seen by an examination of the lines with which the *Inferno* opens:

Nel mezzo del cammin di nostra vita
Mi ritrovai per una selva oscura,
Ché la diritta via era smarrita.
Ah quanto a dir qual 'era è cosa dura
Questa selva selvaggia ed aspra e forte,
Che nel pensier rinnova la paura!

It is impossible, however, to break off here, since there is no rhyme to *forte,* which has to be supplied twice in the succeeding terzina, where, however, a fresh rhyme, *trovai,* is introduced, linking the whole to a still further terzina, and so on, indefinitely in the sequence *aba, bcb, cdc, ded* . . . Boccaccio wrote *terza rima* in close following of Dante, but it has not been a form very frequently adopted by Italian poets. The difficulty of sustaining dignity and force in these complicated chains of verse has usually deterred writers in other languages from adventuring on *terza rima.* In the age of Elizabeth, Samuel Daniel employed it in his "Epistle to the Countess of Bedford," but he found no followers. Shelley tried it in two poems, "Prince Athanase" and "The Triumph of Life," Byron in one, "The Prophecy of Dante," but probably the most successfully sustained poem in *terza rima* in the English language is Elizabeth Barrett Browning's "Casa Guidi Windows" (1851).

TESCHEN (Polish Cieszyn, Czech Tesin), a town of Poland and district lying partly in Poland and partly in Czechoslovakia. Pop. of town (1960) 23,000; of district (1955 est.) 125,000. Teschen lies in the southeastern corner of Silesia and was a duchy disputed between the kings of Poland and Bohemia after the 9th century. In the 17th century it came under Habsburg rule. At the time of World War I the district was 825.6 sq.mi. in extent, rich in both coke and gas coal and an important strategic centre of roads and railways, especially of the Bohumin (Oderberg) line. Both Czechs and Poles desired to succeed Austria in the possession of the district; in May–June 1918 their representatives decided to settle the matter in friendly agreement. But when Poland elected deputies from Teschen for its constituent parliament, Czech troops occupied Oderberg in Jan. 1919. The Great Powers intervened and decided (Sept. 27, 1919) to hold a plebiscite under the auspices of the principal Allies, not only for the duchy of Teschen but also for the adjacent Suvan areas of Spisz (Zips) and Orawa. This commission did not arrive until Jan. 1920, when it was confronted with riots, strikes and other difficulties. It appealed to the Conference of Ambassadors at Paris, which, on July 28, 1920, abandoned the plebiscite plan and induced both Czechs and Poles to agree to a partition. The Spisz-Orawa areas offered few difficulties. In the duchy itself the Czechs obtained the Karvinna (Karwin) mining area and the Oderberg-Jablunkov railway, the Poles the town, except its western suburb, and an agricultural area. The entire district was seized by the Poles in Oct. 1938 after the pact of Munich, and fell before the German advance early in Sept. 1939.

After World War II 276 sq.mi. of the area, including the town of Teschen, were restored to Poland as the district of Cieszyn in Katowice province. Czechoslovakia retained the balance of the district.

See H. W. V. Temperley, *History of the Peace Conference of Paris,* vol. iv, ch. vi, pt. 1, by R. Fitzgibbon Young (1921).

TESLA, NIKOLA (1856–1943), inventor of electrical devices and equipment, was born July 10, 1856, at Smiljan, Croatia (Yugoslavia). He attended the polytechnic school at Graz, specializing in physics and mathematics, and finished his studies at the University of Prague in 1880. Four years later he went to the United States and for a time was with Thomas Edison's company. He then established the Tesla laboratory in New York city and devoted his efforts to research.

Tesla conceived the rotating magnetic field principle as an effective method of utilizing alternating current for power. He pat-

ented the induction, synchronous and split-phase motors, and new forms of generators and transformers; this equipment formed the system for the generation and use of power from Niagara falls. By means of lectures in Europe and the United States beginning in 1891, he announced discoveries and applications of high-frequency alternating current, including the high-frequency resonant transformer, or "Tesla coil." Tesla died in New York city, Jan. 7, 1943.

The 100th anniversary of his birth was commemorated by scientific institutions throughout the world, and the unit of magnetic flux density in the MKS system was named the "tesla" in his honour.

See John J. O'Neill, *Prodigal Genius: the Life of Nikola Tesla* (1944); The Nikola Tesla Museum, *Nikola Tesla, 1856–1943; Lectures—Patents—Articles* (1956). (L. I. A.)

TESO (ITESO), a Nilo-Hamitic people of central Uganda (*q.v.*) who numbered about 525,000 in the 1960s; there were also about 42,000 in western Kenya, known as Itesyo; they are closely related to the Kumam of Uganda (62,000), and also to members of the Karamojong tribal cluster of eastern Uganda.

Unlike most Nilo-Hamites (*q.v.*), the Teso are sedentary cultivators, keeping some cattle; they grow much cotton and are counted among the most progressive farmers of Uganda. Much of their traditional culture and organization was lost through suppression by the conquering Ganda (*q.v.*) under Kakunguru at the turn of the 19th century. The age set (*q.v.*) and the clan lost their importance, and much of the indigenous territorial organization was destroyed. The basic local group is the *etem*, a small settlement that may comprise several hamlets; today the term is applied to the subcounty, an administrative subunit of one of the six counties, each under a government-appointed chief. The homesteads of a hamlet are usually those of the men of an extended family based on a three- or four-generation patrilineal descent group. Genealogical ties are rarely reckoned beyond this range. There are also vaguely defined patrilineal clans dispersed throughout Teso.

Although age sets no longer function as groups, it is known that they were generally arranged in a permanent recurrent cycle of seven or eight names; each set was related mystically to some aspect of nature (*e.g.*, the floods age set dealt with the ritual of fishing). Almost all indigenous religion has been replaced by Christianity; previously Teso believed in an omnipotent but remote god, Akuj, and in a god of calamity, Edeke.

See J. C. D. Lawrance, *The Iteso* (1957); P. Gulliver and P. H. Gulliver, *The Central Nilo-Hamites* (1953). (J. F. M. M.)

TESSIN, CARL GUSTAF, COUNT (1695–1770), Swedish court official, statesman, collector and writer, was born in Stockholm on Sept. 5, 1695, the son of the great architect Nicodemus Tessin the Younger (1654–1728). After studies in France and Italy he was employed on his first diplomatic missions before succeeding, in 1728, to his father's post of superintendent of the court and to the task of completing the new Stockholm castle according to his father's design. This he did with help from Carl Hårleman and from foreign artists (mostly French), who also helped him to set up a Swedish Academy of the Fine Arts.

Though he stood for aristocratic government against royal absolutism, Tessin admired Charles XII and became a founder of the Hat party in the hope that its policy would regain some of the provinces lost in the Great Northern War (*see* SWEDEN: *History*). He was elected *lantmarskalk* in 1738 by a large majority but failed to prevent his party from starting war against Russia in 1741 without guarantees of foreign aid. After the peace of Åbo (1743), however, he cleverly extricated Sweden from "the Russian bear-hug." Adolphus Frederick (*q.v.*), Russia's nominee as heir to the Swedish throne, was married to the imperious Prussian princess Louisa Ulrica in 1744; Tessin formed a strong friendship with her, and they finally induced Adolphus Frederick to conciliate Denmark by renouncing his hereditary interest in Schleswig-Holstein. Tessin was appointed tutor to the prince Gustavus (born in 1746) and head of the chancellery, but he lost Louisa Ulrica's friendship by insisting on a marriage contract (1751) between Gustavus and a Danish princess instead of her Prussian niece. His letters of advice to Gustavus were printed at the expense of the diet (1751 and 1753) during its struggle against the crown and won European fame as models of educational writing, but he had to resign the chancellery in 1752 and the tutorship in 1754. When he withdrew from public affairs (1761), he found himself desperately short of money in his retirement at Åkerö, and the books and the works of art on which he had spent his personal fortune were sold. Later he was formally reconciled to the royal couple, thanks to Gustavus. He died at Åkerö, on Jan. 7, 1770.

Tessin helped to create a modern Swedish prose style, and his poems and fables can still be admired. He left voluminous diaries, and many of his private letters survive.

TEST ACTS. The principle that none but persons professing the established religion were eligible for public employment was adopted by the legislatures of both England and Scotland soon after the Reformation. In England the Acts of Supremacy and Uniformity and the severe penalties pronounced against recusants, whether Roman Catholic or Nonconformist, were affirmations of this principle. The act of 7 Jac. I. c.2 provided that all such as were naturalized or restored in blood should receive the sacrament of the Lord's Supper, and under Charles II the actual receiving of the communion of the Church of England was made a condition precedent to the holding of public offices. The earliest imposition of this test was by the Corporation act of 1661 (13 Car. II. st. 2, c.1), enacting that, besides taking the oath of allegiance and supremacy and subscribing a declaration against the Solemn League and Covenant, all members of corporations were within one year after election to receive the sacrament of the Lord's Supper according to the rites of the Church of England. This act was followed by the Test act of 1672 (25 Car. II. c.2), the immediate consequence of the king's declaration of indulgence, dispensing with laws inflicting disabilities on Nonconformists. This act enforced upon all persons filling any office, civil or military, the obligation of taking the oaths of supremacy and allegiance and subscribing a declaration against transubstantiation, and also of receiving the sacrament within three months after admittance to office. The act did not extend to peers; but in 1678 (30 Car. II. st.2) enacted that all peers and members of the house of commons should make a declaration against transubstantiation, invocation of saints, and the sacrifice of the mass—an exception being made in favour of the Duke of York. The provisions of the Test act were violated by both Charles II and James II on the ground of the dispensing power claimed by the Stuart kings. In the case of *Godden* v. *Hales* (11 State Trials, 1,166), an action for penalties under the Test act brought against an officer in the army, the judges decided in favour of the dispensing power—a power finally abolished by the Bill of Rights. After a considerable number of amendments and partial repeals of these acts, and of acts of indemnity to protect persons from penalties incurred under the Test act, the necessity of receiving the sacrament as a qualification for office was abolished by 9 Geo. IV. c.17, and all acts requiring the taking of oaths and declarations against transubstantiation, etc., were repealed by the Roman Catholic Relief act of 1829 (10 Geo. IV. c.7). This general repeal was followed by the special repeal of the Corporation act by the Promissory Oaths act 1871, of the Test act by the Statute Law Revision act 1863, and of the act of 1678 by an act of 1866 (29 & 30 Vict. c.19). In 1871 the University Tests act abolished subscriptions to the articles of the Church of England, all declarations and oaths respecting religious belief, and all compulsory attendance at public worship in the universities of Oxford, Cambridge, and Durham. There is an exception restricting to persons in holy orders of the Church of England degrees in divinity and positions such as the divinity and Hebrew professorships.

In Scotland a religious test was imposed immediately after the Reformation. By 1567, c.9 no one was to be appointed to a public office or to be a notary who did not profess the Reformed religion. The Scottish Test act (1681, c.6) was rescinded by 1690, c.7. By 1700, c.3, renunciation of popery was to be made by persons employed in education. By 1707, c.6, all professors, principals, regents, masters, or others bearing office in any university, college,

or school in Scotland were bound to profess and subscribe to the Confession of Faith; and all persons were to be free of any oath or test contrary to or inconsistent with the Protestant religion and Presbyterian Church government. The necessity for subscription to the Confession of Faith was removed for persons other than principals and professors of theology by 16 and 17 Vict. c.89. The act provided that in place of subscription every person appointed to a university office was to subscribe a declaration according to the form in the act, promising not to teach any opinions opposed to the divine authority of Scripture or to the Confession of Faith, and to do nothing to the prejudice of the Church of Scotland or its doctrines and privileges. All tests were finally abolished by an act of 1889 (52 & 53 Vict. c.55). The reception of the communion was never a part of the test in Scotland as in England and Ireland.

In Ireland an oath of allegiance was required by the Irish act of Supremacy (2 Eliz. c.1). The English act of 3 Will. & M. c.2 substituted other oaths and enforced in addition from peers, members of the House of Commons, bishops, barristers, attorneys, and others a declaration against transubstantiation, invocation of the Virgin Mary and the saints, and the sacrifice of the mass. By the Irish act of 2 Anne, c.6, every person admitted to any office, civil or military, was to take and subscribe the oaths of allegiance, supremacy, and abjuration, to subscribe the declaration against transubstantiation, etc., and to receive the Lord's Supper according to the usage of the Church of Ireland. English legislation on the subject of oaths and declarations was adopted in Ireland by Yelverton's act, 21 & 22 Geo. III. c.48, § 3 (Ir.). These provisions were all repealed by the Promissory Oaths act 1871. The Roman Catholic Relief act of 1793 (33 Geo. III. c.21. Ir.) excepted Trinity College, Dublin, from its provisions, and tests existed in Dublin university—except for professors and lecturers in divinity—until finally abolished by the University of Dublin Tests act 1873.

In the United States by art. 6 of the constitution, "no religious test shall ever be required as a qualification to any office or public trust under the United States." A similar provision is generally included in the state constitutions.

TESTAMENTS OF THE TWELVE PATRIARCHS.

The Testaments of the Twelve Patriarchs are an important constituent of the apocryphal scriptures connected with the Old Testament, comprising the dying commands of the twelve sons of Jacob.

They were written in Hebrew in the later years of John Hyrcanus—in all probability after his final victory over the Syrian power and before his breach with the Pharisees—in other words, between 109 and 106. Their author was a Pharisee who combined loyalty to the best traditions of his party with the most unbounded admiration of Hyrcanus. The Maccabean dynasty had now reached the zenith of its prosperity, and in its reigning representative, who alone in the history of Judaism possessed the triple offices of prophet, priest and king, the Pharisaic party had come to recognize the actual Messiah. When we contrast the expectations of the original writer and the actual events that followed, it would seem that the chief value of his work would consist in the light that it throws on this obscure and temporary revolution in the Messianic expectations of Judaism towards the close of the 2nd century. But this is not so. The main, the overwhelming value of the book lies not in this province, but in its ethical teaching, which has achieved a real immortality by influencing the thought and diction of the writers of the New Testament, and even those of our Lord. This ethical teaching, which is indefinitely higher and purer than that of the Old Testament, is yet its true spiritual child, and helps to bridge the chasm that divides the ethics of the Old and New Testaments.

Date.—The indications in the book itself make it quite clear that it was written (1) in Hebrew, and (2) between 109–107 B.C. There are, however, numerous additions. A large body of these additions can be classed under one head as written with a well-defined object and at a definite period. This period was about 70–40 B.C., and the object of the additions was the overthrow of the Maccabean high-priesthood, which in the 1st century B.C. had become guilty of every lewdness. Test. Lev. x., xiv.–xvi.; Test. Jud. xvii. 2–xviii. 1 (?), xxi. 6–xxiii., xxiv. 4–6; Test. Zeb. ix.; Test. Dan. v. 6–7, vii., 3 (?); Test. Naph. iv.; Test. Gad. viii. 2; Test. Ash. vii. 4–7. These additions are identical in object and closely related in character and diction with the Psalms of Solomon. Other additions are of various dates and cannot be more than mentioned here; *i.e.*, Test. Reub. ii. 3–iii. 2; Test. Lev. xvii. 1–9; Test. Zeb. vi. 4–6, vii.–viii. 3; Test. Jos. x. 5–xviii.

Christian Additions to the Text.—These additions are to be found in most of the Testaments and were made at different periods. The existence of these Christian elements in the text misled nearly every scholar for the past 400 years into believing that the book itself was a Christian apocryph. To Grabe, Schnapp and Conybeare belongs the credit of showing that the Christian elements were interpolations—to Conybeare especially of the three, since, whereas the two others showed the high probability of their contention on internal evidence, Conybeare proved by means of the Armenian version that when it was made many of the interpolations had not yet found their way into the text. For a full treatment of these passages *see* Charles's *Testaments of the Twelve Patriarchs* (1908), Introd. §20.

Influence on the New Testament.—St. Paul twice quoted from the Greek text of the Testaments: Rom. i. 32 and 1 Thess. ii. 16. But these give the very faintest idea of the degree of his indebtedness in thought and phraseology in several of his Epistles, especially that to the Romans. Of still greater interest are the passages in the Gospels which show the influence of the Testaments, and these belong mainly to the sayings and discourses of our Lord. We may mention two of the most notable of these. Thus Matt. xviii. 15, 35, which deal with the great question of forgiveness, are clearly dependent on our text.

Matt. xviii. 15.	T. Gad. vi. 3.
Ἐὰν δὲ ἁμαρτήσῃ ὁ ἀδελφός σου κατά σου, ὕπαγε ἔλεγξον αὐτὸν μεταξὺ σοῦ καὶ αὐτοῦ μόνου.	Ἐὰν τις ἁμαρτήσει εἰς σὲ εἰπὲ αὐτῷ ἐν εἰρήνῃ . . . . καὶ ἐὰν . . . . μετανοήσῃ, ἄφες αὐτῷ
	vi. 6 ἡσύχασον μὴ ἐλέγξῃς.
35. Ἐὰν μὴ ἀφῆτε ἕκαστος τῷ ἀδελφῷ αὐτοῦ ἀπὸ τῶν καρδιῶν ὑμῶν.	v. 7 Ἄφες αὐτῷ ἀπὸ καρδίας

Next, the duty of loving God and our neighbour is already found in Test. Dan. v. 3, which is the oldest literary authority which enjoins these two great commands. The form is infinitely finer in Matt. xxii. 37–39, but the matter is already in the Test. Dan. *See* Introd. §26 to R. H. Charles's *Testaments of the Twelve Patriarchs*.

Besides Charles (as above), are his *Apoc. and Pseudepig.* vol. ii and the S.P.C.K. *Translations of Early Documents* (1917).

(R. H. Ch.)

TESTAMENTUM DOMINI (Testament of Our Lord), one of a series of writings, claiming to embody the fundamental rules of the early Christian church, others of which were the Apostolic Constitutions and the *Apostolic Tradition* of Hippolytus (*see* Constitutions, Apostolic; Hippolytus, Canons of Saint). It falls into three distinct parts: an apocalyptic introduction (book i, ch. 1–18; the division into books, however, was clearly not original); a church order proper (i, 19–ii, 24); and a conclusion (ii, 25–27) of the same apocalyptic character as the introduction.

The testament has several distinctive characteristics. First and foremost is its ascription to the Lord himself, an attempt to claim the highest possible sanction, higher than that claimed by the various compilations styled "apostolic." This fact alone indicates the preexistence of certain of the latter. Again, the whole tone is one of highly strung asceticism, and the regulations, by their severity, point to a small and strictly organized body. They are "the wise," "the perfect," "sons of light"; but this somewhat Gnostic phraseology is not accompanied by any signs of Gnostic doctrine, and the work as a whole is orthodox in tone. They are set in the midst of "wolves," despised and slighted by the careless and worldly: there is frequent mention of "the persecuted" and of the duty of "bearing the cross."

Charismata, and above all exorcisms, occupy a very important place. There is a vivid realization of the ministry of angels, and the angelic hierarchy is very complete. Great stress is laid

upon virginity (although there is no sign of monasticism), fasting, and regular attendance of the whole clerical body and the "more perfect" of the laity at the hours of prayer. The testament was originally written in Greek, probably in the 4th–5th century; it survives only in a 7th-century Syriac translation (*editio princeps* by Ignatius Ephraem II Rahmani, 1899, with Latin translation).

See James Cooper and A. J. Maclean, *The Testament of Our Lord*, Eng. trans., with introduction and notes (1902); and E. Amann in *Dictionnaire de Théologie Catholique*, vol. xv, col. 194–200 (1946).
(W. E. Co.; X.)

TESTER, a canopy (*q.v.*) over a bed or altar.

TESTING MACHINERY: *see* MATERIALS, STRENGTH OF.

TEST PIT. A test pit may be used for many purposes, but in its generally accepted meaning the term refers to a heavily lined pit sunk in the floor of an industrial factory building and employed in making strength and balancing tests on the revolving parts of such machines as large electric generators, motors and steam and hydraulic turbines.

The necessity for such pits arises from the fact that machines are usually tested at speeds considerably in excess of those at which they will normally run, and consequently there is the risk of disruption, with the likelihood of great damage to property and loss of life if proper protection is not provided.

While it is not intended that such tests shall be carried to the point of destruction, a certain factor of safety must be present in the machines and this can only be demonstrated by actual overspeed tests.

Test pits are therefore sunk beneath the shop floor level, and are always heavily lined with reinforced concrete and steel. All up-to-date manufacturers of large revolving machines are equipped to make such tests with safety to their employees and equipment.

There are cases on record of very serious accidents which have resulted from the disruption of rapidly revolving machines improperly protected, and the tendency at mid-20th century is toward more heavily armoured pits located in isolated sections of the manufacturing plants.

A large test pit was designed and built by the General Electric company in 1927, at their principal plant situated at Schenectady, N.Y. This pit was located in a special building with other testing equipment, a quarter of a mile from the nearest factory building; 30 ft. deep and 45 ft. in inside diameter, it was designed to be capable of enclosing the rotating member of the largest electrical machine likely to be built for many years. The walls, of double concrete construction with a layer of soft sand in between, were built with a total thickness of 14 ft. The floor, a concrete slab faced with steel, was supported by wooden piles driven 60 ft. into the ground.

The entire construction was reinforced with 1-in. steel bars cast in the concrete, giving even greater rigidity than is necessary in the case of large dams and bridges.

The top of the pit was closed by a removable lid made of steel plates and girders, and was to be held in place by eight steel wedges, each a yard long. For testing purposes the rotating machine was lowered into the pit by a travelling crane, and was supported in a huge roller bearing at the bottom, capable of carrying a load of 500 tons. The shaft of the machine, held in the vertical position by a small upper bearing, was connected by a coupling to an electric motor installed upon the top of the lid.

Motors of as high as 3,000 h.p. were used to revolve the largest rotors at test speeds.

The pit could be used for any tests where bombproof protection was required. Its principal use, however, proved to be in performing "double speed" tests on large hydroelectric generator rotors. The object of such tests is to establish the fact that rotors will not fly apart even under the maximum "runaway" conditions which might exist during an emergency in service. Huge centrifugal forces and enormous rim velocities are developed in these tests.

During the test runs no one was allowed to remain in the pit building. The observers retired to a control house 100 yd. away and there manipulated the speed of the electric motor driving the machine under test by means of a switchboard connected with the pit through underground cables.

Indicating meters recorded the speed and degree of vibration, and telephonic communication with a microphone in the pit transmitted to the observers the sounds made by the machine as it revolved. By these means it was expected that sufficient warning of a breakdown would be given so that the test could be stopped before the machine or pit was damaged. (D. O. W.)

TESTUDO: *see* ENGINES OF WAR.

TETANUS. Tetanus is a specific disease caused by the bacillus of tetanus, *B. tetani*, and its soluble poison, *tetanospasmin*. It is characterized by muscular rigidity and spasms. The almost constant involvement of the muscles of the jaw accounts for the term lockjaw conventionally employed.

History.—Although a comparatively rare disease, tetanus has been recognized as a distinct entity for centuries. Picturesque descriptions by Hippocrates, Aretaeus and John Arderne of Nottinghamshire, observation of its occurrence from such trivial injuries as the bite of a tame sparrow as noted by G. B. Morgagni in the 18th century and the concern of Joseph Lister in the 19th over the disease among hospital ward patients are commentaries on the importance of tetanus from early times. Recognition of its infectious nature (1884–86: Carlo and Rattone, A. Nicolaier, Rosenbach, Beumer), isolation of the *Bacillus tetani* and its toxin (1889–90: S. Kitasato and T. Weyl, A. Nicolaier) and studies on artificial protection of animals (1890–92: S. Kitasato, E. Behring, Schutz, L. Vaillard, L. Brieger) are but a few of the many historic landmarks. Investigations on methods to prevent and treat tetanus continued in various centres throughout the world.

Aetiology.—The bacteriology and mechanism of symptom production (pathogenesis) are included in this discussion. The bacillus of tetanus is a slender rod visible with the conventional microscope. It occurs in the dormant, or spore, and vegetative, toxin-producing forms. Conversion of the spore into the vegetative form occurs under conditions of oxygen lack; *i.e.*, anaerobic states. Although at least nine different strains of the tetanus bacillus exist, it is fortunate that the poisons of all can be completely neutralized by a single antitoxic serum.

Although the vegetative forms, which elaborate the poison, are destroyed as easily as are those of other bacteria, the spores are highly resistant to thermal and chemical agents. Thus, while antiseptics exert variable effects, only autoclaving destroys the spores. The destructive action of adequate concentrations of peroxide of hydrogen contrasts sharply with the failure of sulfonamide drugs and penicillin to prevent growth of spores in human tissue under ordinary circumstances.

Protected from sunlight and other adverse environmental conditions, the spores may persist almost indefinitely. They are distributed widely in nature, mainly in the superficial layers of the soil. In general, the degree of soil pollution parallels the amount of faecal contamination from either artificial manuring or human excreta.

The most important predisposing cause of tetanus is a wound which permits entrance of spores of *B. tetani* into human tissue. The wound may be trivial or severe, varying from simple superficial abrasions to severe lacerations. Not infrequently, the source of infection may be obscure.

Both the incidence and severity of tetanus are governed by the amount of poison developed and absorbed and the resistance of the host. Natural immunity to tetanus does not occur in human beings. Hence, in the absence of artificially acquired protection (see *Control*, below), conditions favourable to implantation and growth of *B. tetani* determine the appearance and course of the disease, subject to attenuation by adequate states of immunity. Thus, race, geographical location, sex, age, occupation, etc., are of significance only insofar as they are related to environmental conditions which assist or impede the development of tetanus.

Tetanospasmin is the only component of tetanus toxin which is of clinical importance. One of the deadliest poisons known to man, it is transported from the point of infection to the muscles and central nervous system by the motor nerves, the blood stream or a combination of these routes.

The toxin acts directly on the muscles, causing fairly constant rigidity, and on the central nervous system, eliciting spasms of

variable intensity at unpredictable intervals. The muscles most intensively activated are those which oppose gravity; *i.e.*, the extensor groups of the limbs and neck and the masseters, which close the jaw.

The period of incubation, or the interval between entrance of spores and the appearance of tetanus, is related to the speed with which dangerous amounts of poison are released from the point of injury. About 20% of hospital cases have short incubation periods of 2 to 5 days, 60% intermediate values of 6 to 15 days and the remainder intervals of 16 to 85 days. The region of a wound has no relationship to the incubation period. The use of tetanus antitoxic serum may prolong the period. The length of the incubation period is related to the ultimate outcome of the disease.

Symptomatology and Course.—The classical picture of tetanus is related to the type, stage and intensity of the disease. Tetanus may be local or general. In local tetanus the stiffness and spasms may be restricted to the muscles near the wound. When the injury occurs in the region of the head and neck, the resulting local form of the disease is termed cephalic tetanus. In this form, paralysis of one or more cranial nerves may be seen. If the patient survives, the paralysis gradually disappears. When tetanus is general, it is usually symmetrical, involving all parts of the body musculature in equal degree. However, this pattern of symmetry is sometimes violated and unilateral, abdominothoracic or limb tetanus are encountered, with spasticity limited or more pronounced on one whole side of the body, the muscles of the abdomen and thorax or one or more limbs—occasionally without the classical stiff jaw.

Early complaints may be difficult to verify by examination; these are called premonitory or warning findings. When the process becomes more pronounced, the patient's complaints are paralleled by changes in physical behaviour, and all grades of severity may be encountered. In general, the signs and symptoms of the disease are quite characteristic, and may be catalogued under one or all of four basic categories:

1. Rigidity of the body musculature: stiff jaw (trismus), sardonic smile (*risus sardonicus*), difficulty in swallowing, backache, rigid abdominal wall, tight thorax, fixed extremities, rigid neck (sometimes with a backward retraction called opisthotonos) and spasm of the muscles of the bladder and bowel causing retention of urine or constipation.

2. Muscular spasms. These occur at unpredictable intervals, in various degrees of severity from mild contractions to alarming attacks associated with pain, fractures of the spine, brain haemorrhages or even sudden death. The spasms are very important in determining survival or death of the patient.

3. Accessory signs and symptoms: clear mind, normal temperature (in absence of certain complications), profuse drenching perspiration and overactive tendon reflexes.

4. Complications: pneumonia, collapse of one or more lung lobes, suffocation from spasms, brain haemorrhages, injuries to the teeth, tongue, mouth from convulsions, starvation, dehydration, fractures of the spine, fever of varied origin and undesirable side effects of necessary treatment, such as serum sickness and coma from antitoxin and sedative drugs respectively.

Since physicians first see patients at variable periods of the illness, they are familiar with the erratic onset and course of tetanus. For example, it is known that the classical stiff jaw may be late in appearance or even completely absent.

Treatment.—Methods of treatment employ antitoxic serum, sedative drugs, compounds to control secondary infections (penicillin and allied drugs), surgical procedures to avoid suffocation (tracheotomy), conservative care of wounds, measures to prevent dehydration and starvation and constant medical and nursing attention. Prior to such a type of management, the recovery rate was 20% or less. The chances for survival were at least tripled by this program of treatment. Certain natural features of the disease favour a good general outlook: healthiness of the patient, absence of severe convulsions or complications, slight difficulty in swallowing, long incubation period and slow course of the disease, favourable age period, mild wounds and expert treatment. In the average case, two or more weeks are required for complete recovery. Sometimes the disease pursues a chronic course, lasting months or several years. Inadequate treatment may be disastrous. With the exception of certain complications, such as compression fractures of the spine, recovery from tetanus is complete. No matter how seriously ill the patient may have been, the mind is perfectly clear after the disease has run its course. Unfortunately, tetanus may occur many times in the same individual, so that proper preventive measures must be employed following complete recovery.

Control.—Prevention of tetanus includes three basic factors: (1) control of the environment, to reduce opportunities for infection; (2) proper treatment of wounds; and (3) protection with tetanus antitoxin and toxoid.

Environmental control has a definite, though limited, application. Accident control among children and adults, proper obstetrical and neonatal care, careful supervision in the manufacture of catgut and therapeutic serums, avoidance of airtight dressings over smallpox vaccinations and limitation of the hazards from firecrackers, blank cartridges, bullets and burns contributed significantly to a reduced incidence of tetanus.

Proper treatment of wounds is essential to avoid local infection, death of tissue and anaerobic conditions, factors favourable to the development of tetanus. Adequate cleansing, removal of devitalized tissue and foreign bodies, avoidance of injury to both healthy and diseased structures, proper use of the available antibacterial drugs (penicillin, etc.) and adequate immobilization of injured parts constitute a five-point program for the prevention of tetanus.

Antitetanic serum, for rapid (passive) protection, is available either alone or combined with gas-gangrene antitoxin. It is obtained by immunizing horses or cows with toxoid, toxin or both. Tetanus toxoid, which stimulates immunity in the person injected, is prepared either alone or combined with vaccines for diphtheria, whooping cough, typhoid fever, etc.

The antitoxin is prepared from the blood serum of immunized animals, after which it is concentrated and standardized. It must pass routine tests for safety before it is released for clinical use. The toxoid is prepared by chemical alteration of the toxin, followed by tests to ensure absence of toxicity and to determine its potency.

Passive protection with tetanus antitoxin is indicated in the presence of injuries considered likely to be contaminated with spores of *B. tetani*. Active protection with tetanus toxoid is a slow process which requires weeks or months for adequate development. In general, all individuals should be protected against tetanus, for as high as 80% of all cases of this disease occur under conditions which are uncontrollable insofar as routine serum prophylaxis is concerned.

Although tetanus respects no age, race, sex, economic group or occupation, certain populations are more in need of routine active immunization than are others, namely, the armed forces, civilian populations under military attack, allergic individuals, farmers, industrial workers and infants and children. When tetanus antitoxin is used for passive protection a detailed list of precautions must be obeyed to be certain that the patient will not develop the various forms of serum sickness, some types of which might be quite serious. Moreover; the dose must be adequate; at least 5 to 15 times the amount used in World War I came to be offered for routine use.

At an appropriate interval after the antitoxin, the toxoid is recommended to confer the so-called active immunity as a protection in the event of future injuries. Tetanus toxoid is injected in a series of two to three doses, with intervals of several weeks between the individual doses. Ordinarily, adequate protection emerges within a week after the second or third injection. At almost any time after the full dose of a toxoid, very small booster doses will evoke rapid, high and prolonged elevations of antitoxin in the patient's body. This procedure offers a distinct advantage, because it avoids the serum sickness which not infrequently follows repeated injections of the antiserum. Although adequate precautions should permit the use of antitoxin, the physician is opposed to repetitious antitoxin treatment. This attitude does not apply to tetanus toxoid, allergic reactions to which are exceedingly rare.

Antitetanic serum (tetanus antitoxin) first demonstrated its value in World War I. Prompt injection of antitoxin after the re-

ceipt of a wound was associated with a significant decrease in the incidence of tetanus from 15.9 to 2.8 per 1,000 (British home hospitals, 1914–18). Moreover, when serum failed to prevent the development of tetanus, it appeared to lengthen its incubation period. The average incubation period among the inoculated was approximately four times that noted among the uninoculated (45 and 10 days respectively).

Likewise, tetanus which developed in persons who had received one or more injections appeared to be less fatal than in uninoculated subjects. In the British home hospitals, the case mortality was 22.6% among the inoculated and 53.3% among the uninoculated. Moreover, the proportion of local (mild) to generalized (severe) tetanus rose considerably following the routine use of antitoxin. Undoubtedly, the number of failures would have been less had the increased dose been employed instead of amounts considered inadequate for even ordinary wounds.

Tetanus toxoid has evinced its reliability in both veterinary and human medicine. From 1925 active protection of horses by means of tetanus toxoid was carried out in France, and after 1927 it was employed by the French cavalry. Of 22,000 horses and mules which had received three injections of tetanus toxoid in a period of six years, only one had developed tetanus and this animal recovered. In contrast, from 1925 to 1929, in spite of the use of preventive antitoxic serum, 50 of 13,600 horses had tetanus and 30 died. Moreover, from 1932 to 1937, during which time administration of toxoid was systematically carried out in the same region, there was not a single instance of tetanus among 16,500 horses, despite omission of any fortifying (booster) injections of antitoxic serum or toxoid.

The value of tetanus toxoid was firmly established in human medicine on the basis of evidence from World War II. Only 30 cases of tetanus occurred among the Allied personnel, with a case fatality rate of about 47%. Analysis of the available evidence would indicate that even this low incidence might have been reduced by the routine use of fortifying (booster) injections of antitoxin and toxoid, or both, in the presence of severe injuries. Thus, 77% of the cases died when no further injections were given at the time of injury, 33% when tetanus toxoid alone was used as a booster and 18% when antitoxin only was injected. It is obvious that a combination of toxoid and antitoxin would have yielded even better results.

Epidemiology.—Despite its world-wide distribution, comparatively few injured persons develop tetanus. As stated above, the determining factors are the milieu for the implantation and growth of *B. tetani* and the immunity status of the injured individual.

In the United States the reported incidence of tetanus decreased from 1,472 cases in 1920 and 930 in 1938 to 560 in 1940. Tetanus of the newborn, once the scourge of this tender age period, became relatively uncommon in the United States, although it persisted in other regions of the world. Postoperative tetanus, formerly attributable to contaminated catgut, occurred only rarely at midcentury. Adequate preventive measures were credited with prevention of tetanus from wounds such as those following compound fractures, lacerations, firecracker burns and related injuries. Moreover, the occurrence of but 30 cases in Allied personnel during World War II, associated with the intensive use of tetanus toxoid, frequently combined with tetanus antitoxin, contrasted sharply with the high incidence in German and Japanese troops under similar conditions of combat. The spectacular reduction of morbidity and mortality during the later stages of World War I, associated with the systematic use of tetanus antitoxin, and during World War II, with the employment of both antitoxin and toxoid, focused attention on the sad statistics of other major wars. Thus, the reported incidence per 1,000 wounded in various wars was tabulated as follows: Crimean, 0.75; American Civil, 2.07; Franco-Prussian, 3.5; Russo-Turkish, 1.2. During Wellington's peninsular campaign against Napoleon Bonaparte, 1 out of every 200 wounded (0.5%) died of tetanus, and for the American Civil War the figure is given as 0.2%. During the period 1914–18, 2,500 cases of tetanus occurred in the British army, with a mortality of 43.2%.

Bibliography.—W. W. C. Topley and G. S. Wilson, *Principles of Bacteriology and Immunity*, 3rd ed., vol. ii, pp. 1746–1777 (Baltimore, London, 1946); R. Spaeth, "Tetanus," in Frederick Tice *et al.* (eds.), *Practice of Medicine*, vol. 5 (Hagerstown, Md., 1932); J. W. Williams and C. W. McClure, in *Tetanus and Gas Gangrene*, vol. v (New York, 1940).

(R. Sh.)

TETENS, JOHANN NIKOLAUS (1736–1807), German philosopher, mathematician and economist, the most prominent representative of the Enlightenment (*q.v.*) philosophy in Germany, was born at Tetenbüll, South Schleswig, on Sept. 16, 1736. He was educated at the universities of Rostock and Copenhagen. He became professor of physics at Bützow university in 1763 and director of the *Pädagogium* (academy) there in 1765. In 1776 he was appointed professor of philosophy at Kiel university, where he later held the chair of mathematics. In 1777 he published his major work, *Philosophische Versuche über die menschliche Natur* ("Essays on Human Nature"; new ed. by W. Uebele, Berlin, 1913 *et seq.*). He gave up teaching and theoretical work in 1789 in order to enter public service, becoming assessor and then councilor of state and deputy in the Danish ministry of finance in Copenhagen, where he died on Aug. 15 (19?), 1807.

Tetens' *Philosophische Versuche* was the most important philosophical work written in Germany in the period immediately preceding Kant's *Critique;* it clearly shows the influence of the latter's *Dissertation* of 1770. In the preface to the *Philosophische Versuche* Tetens states, "metaphysical analysis . . . must be preceded by psychological analysis"; and further, "however far we proceed in metaphysical psychology, the authenticity of its propositions must always be tested by empirical knowledge." In this emphasis on the necessity of proceeding by way of construction in philosophy and on the importance of empirical foundations, Tetens approaches very close to the spirit of Kant, who valued his work highly and always kept a copy of the *Philosophische Versuche* on his writing table. Tetens was the only important German thinker to embrace, with some modifications, the empirical psychology in vogue in Great Britain and France during the 18th century; he calls the faculty of forming and combining ideas "understanding." In his threefold division of consciousness into feeling, will and understanding he influenced Kant, to whose theory of knowledge these concepts are fundamental.

Bibliography.—M. Schinz, *Die Moralphilosophie von Tetens* (Zürich, 1906); W. Uebele, *J. N. Tetens, nach seiner Gesamtentwicklung betrachtet, besonderer Berücksichtigung des Verhaltnisses zu Kant* (1911); A. Seidel, *Tetens Einfluss auf die kritische Philosophie Kants* (Leipzig, 1932).

TETRACYCLINE (Achromycin), a broad-range antibiotic drug that was developed in the United States; it became available in 1953. Tetracycline is the generic name; Achromycin is the trade name of one of the more well-known brands—others include Panmycin, Polycycline, Steclin and Tetracyn. Tetracycline, which is related to chlortetracycline (Aureomycin) and oxytetracycline (Terramycin) [prepared in 1948 and 1950 respectively], was found in general to produce fewer side reactions than either of its two predecessors, and to a large extent has replaced them in medical practice.

Tetracycline and its analogues are capable of suppressing the growth of many bacteria and a few large viruses. It is especially useful in certain infections of the respiratory and urinary tracts. Side reactions to tetracycline occur in about only 5% of patients, and many of these reactions can be controlled until the course of treatment is completed.

The drug is given by mouth in all patients except postsurgical and comatose ones, in which cases it is usually given by injection. Oral doses are readily absorbed from the gastrointestinal tract; one or two grams daily in divided doses produce effective blood concentrations against susceptible organisms.

(L. E. P.)

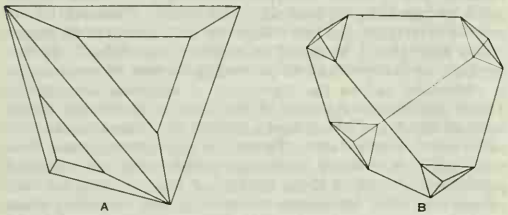
TETRAGRAMMATON, the four-lettered name for the Hebrew God pronounced Jehovah through misapplication, dating back to 1270, of the vowels of Adonai (Lord) to the consonants J, H, V, H. The correct pronunciation is probably Yahweh or Jahveh (German). It is used approximately 5,500 times in the Old Testament, with Adonai 315 times to indicate the correct pronunciation of the sacred name. It is rendered "Lord" in the New Testament as in the LXX. The name is not used at all in Ecclesiastes, substantial parts of Job and in most of Daniel. To avoid

repetition of the name Yahweh, when the word Adonai precedes it, Yahweh is written by the Massoretes with the vowels of Elohim (God) and is read Elohim. *See* YAHWEH.

TETRAHEDRITE, a mineral consisting typically of copper sulfantimonide, but often of complex composition. It is an important ore of copper, in the pure form containing 57.5% of this metal; it is also largely worked as an ore of silver, of which it sometimes contains as much as 30%.

The formula of tetrahedrite is Cu_3SbS_3 but copper is usually isomorphously replaced by variable amounts of silver, iron, zinc, mercury, lead or cobalt, and the antimony by arsenic or bismuth. Numerous special names have been applied to varieties differing in chemical composition; the arsenic compound, Cu_3AsS_3, is known as tennantite (after the English chemist Smithson Tennant). The old German name *Fahlerz* includes both tetrahedrite and tennantite.

Well-developed crystals are of frequent occurrence; they belong to the tetrahedral class of the cubic system and their tetrahedral form is a characteristic feature of the mineral. In the accompanying figure (A) shows a combination of a tetrahedron and a trigonal tristetrahedron, and (B) a tetrahedron with the rhombic dodecahedron.



TETRAHEDRAL CRYSTALS OF TETRAHEDRITE

The colour of the mineral is steel gray to iron black, and the lustre is metallic and brilliant. The streak is usually black; sometimes, however, it is dark cherry red, and very thin splinters of the mineral then transmit a small amount of blood-red light. The hardness is 4.5, and the specific gravity varies with the composition from 4.4 to 5.1. There is no cleavage and the fracture is conchoidal.

Tetrahedrite occurs in metalliferous veins, associated with chalcopyrite and other sulfides. Fine groups of crystals, coated on their surface with brassy or brilliantly tarnished chalcopyrite, were formerly found at Herodsfoot mine, near Liskeard in Cornwall, Eng.

Tennantite occurred as small crystals of cubic or dodecahedral habit in many Cornish copper mines, especially in the neighbourhood of Redruth. It is also found as small, brilliant crystals very rich in faces in the white crystalline dolomite of the Binnenthal in the Valais, Switz., and under the name binnite was long considered as a distinct species. It occurs in the Harz mountains of Germany, in Freiberg, Saxony, the silver mines of Mexico, Bolivia and Peru and in the United States in silver and copper mines in Arizona, Colorado, Montana, Nevada and Utah.

(CL. F.; L. J. S.; X.)

TETRAPODA, the term for the backboned animals not included in the fishes, including, therefore, the amphibians, reptiles, birds and mammals. The origin of the Tetrapoda from their fish-like ancestors is discussed in the article AMPHIBIA. (K. P. S.)

TETRARCH, a governor of a quarter of a province: Thessaly was divided into tetrarchies by Philip of Macedon and so was Galatia before the Roman conquest in 169 B.C. Feudal lords in Palestine were called tetrarchs and were inferior to ethnarchs. Tetrarch came to be used generally for a petty prince. Herod the Great was tetrarch before becoming king. At his request Augustus appointed Pheroras tetrarch of Perea.

Other tetrarchs were Herod's two sons who at his death (4 B.C.) divided his kingdom, Herod Antipas being tetrarch of Galilee and Philip tetrarch of Iturea and Trachonitis. The eldest son, Archelaus, was not tetrarch but ethnarch of Judea, Samaria and Idumea. Herod Agrippa reunited the tetrarchies (A.D. 41–44).

TETSCHEN: *see* DĚČÍN.

TETZEL (TEZEL), **JOHANN** (1465–1519), German Dominican preacher of indulgences and first public antagonist of Martin Luther, was born at Pirna in Meissen about 1465. After studying at Leipzig he entered the Dominican order, being appointed inquisitor for Poland (1509), and later for Saxony. Between 1503 and 1510 he was in constant demand as a preacher of indulgences. He became a preacher for the indulgence attached to the contributions to the construction of St. Peter's, Rome, in the diocese of Meissen (1516), and those of Magdeburg and Halberstadt (1517).

Tetzel's preaching at Juterbog, near Wittenberg, in the spring of 1517 was the occasion for the publication of Luther's 95 theses on indulgences at Wittenberg on Oct. 31, 1517, which provoked an uncompromising reply under Tetzel's name, but really written by Konrad Wimpina. Tetzel grasped the explosive character of Luther's thought and saw that the main issue involved the very basis of the Catholic theory of the church, on whose authority he published 50 theses in May 1518. At the end of 1518 he withdrew to Leipzig priory, where he died on Aug. 11, 1519.

Tetzel was by no means a profound theologian and was far from orthodox in his teaching on indulgences for the dead. His view that a mere gift sufficed to secure this indulgence, together with the financial transactions surrounding the preaching of it, gave rise to the grave abuses which helped to bring about the disruption of the unity of the western church. In a sense the harm done by Tetzel was increased by his ability and appeal as a preacher. His style can be judged by specimen sermons he issued for the parochial clergy.

See N. Paulus, *Johann Tetzel der Ablassprediger* (1899). (I. HP.)

TEUTOBURGER WALD, BATTLE OF (9 A.D.). At the opening of the Christian era the country between the Rhine and the Vistula was covered by dense forests, and occupied by numerous Germanic tribes. As Julius Caesar had conquered Gaul, so in his turn did Augustus Caesar determine to conquer this region. To it he sent Drusus and Tiberius, both able generals, and by 15 B.C. the Roman frontier was pushed out to the Danube. Between 12 and 9 B.C. Drusus advanced as far as Aliso (probably near Paderborn) on the Lippe, built there a strong fortress, and moved on to the Elbe. Drusus dying at the early age of 30, Tiberius succeeded him in command, and having completed the subjugation of the tribes, in 6 A.D. he was replaced by Quintilius Varus. After about a year, a contingent of Cheruscan troops under the leadership of Arminius returned to their country. Arminius bore the Romans no ill-will until his return, when his indignation was roused by the behaviour of Varus toward his countrymen. Seeing that he was powerless to strike at the Romans covered by their entrenchments, by flattery he persuaded Varus to move his headquarters to the Weser river.

In the summer of 9 A.D., Varus set out with the XVII, XVIII and XIX legions, some cavalry and auxiliaries—in all, about 27,000 men—to repress a local rising arranged by Arminius. He was warned of the danger, and was urged to secure Arminius and the other Cheruscan chiefs who with their men formed the rear guard of his column. Refusing to listen to these warnings, he entered the forests from near Minden, followed by an immense train, and headed for Aliso. When the head of the column had become entangled in the brushwood near the hill of Teut (now called Grotenburg) Arminius abandoned the rear and, followed by his comrades, roused the tribesmen from the Ems to the Main by means of fire and horn signals. As the Romans reached the slopes of the Teutoburg, the German hordes broke in on them from all sides. So trusting and blind was Varus that at first he thought it but rough horseplay. Soon, however, he was disillusioned and was only able to beat back the incessant assaults by moving toward an open plateau. There he entrenched. During Sept. 9 he burned the impedimenta to lighten the column. On the 10th it was again attacked. Though Varus showed great personal bravery, he gave no orders and his troops lost confidence in him. With difficulty they reached the entrance of the Dörn pass but a few miles from Aliso. Arminius had determined on making his decisive attack

there. Blocking the further exit, he fell upon the flanks of the disheartened column, which broke and was slaughtered. Varus fell upon his own sword as his father had done at Philippi.

The loss of the three Roman legions showed that Rome was not invincible. It sent the frontier of the empire back from the Weser to the Rhine, and so prevented the latinization of a large part of modern Germany. It denied to the barbarians of the north a high culture and it encouraged them to sweep south and barbarize western Europe. (J. F. C. F.; X.)

TEUTONI (TEUTONES), a tribe of northern Europe who became known to the Romans in the year 103 B.C. when they reinforced the Cimbri (*q.v.*) after their repulse from Spain by the Celtiberians. In 102 the Teutoni and Ambrones were defeated by Marius at Aquae Sextiae (*see* MARIUS, GAIUS).

The name of the Teutoni is Celtic in form, and possibly they were really a Celtic tribe, a branch of the Helvetii. But a people of this name was mentioned by the traveler Pytheas as inhabiting the coasts of the northern ocean. Strabo and Velleius classify them as Germani.

TEUTONIC LANGUAGES: *see* GERMANIC LANGUAGES.

TEUTONIC ORDER, THE, or Teutonic Knights of St. Mary's Hospital at Jerusalem (*Der deutsche Orden, Deutsche Ritter*) was one of the three great military and religious orders which sprang from the Crusades (*q.v.*). Later in birth than the Templars and Hospitallers, the Teutonic Order traces its first beginnings from the third Crusade. But it is amidst the privations and plague which attended the siege of Acre, during the third Crusade, that the first certain beginnings of the Order appear.

In the winter of 1190–91 certain pious merchants from Bremen and Lübeck (towns with which the Order was still to be connected in the days of its later history) laid the foundations of a hospital in a vessel which they had drawn ashore. Within a few years the foundation apparently became attached to the German Church of St. Mary the Virgin at Jerusalem; and in March 1198 (there being present in the Holy Land a number of Germans, the relics of Henry VI.'s projected crusade), the great men of the army and the kingdom raised the brethren of the German Hospital of St. Mary to the rank of an order of knights. The original members were thus ennobled; and henceforth it was the rule that only Germans of noble birth could join the Order.

Like the other two great military orders, the Teutonic Order began with charity, developed into a military club and ended as something of a chartered company, exercising rights of sovereignty on the troubled confines of Christendom. Even in its last phase, the Order did not forget its original purpose: it maintained several great hospitals in its new home on the south-east shore of the Baltic, in addition to an *hôtel des invalides* at Marienburg for its sick or aged brethren. But long before that period the Order had begun to find that its true work lay on the eastern frontiers of Germany. In 1228 Christian, bishop of Prussia, who had received from the Polish duke of Masovia a part of Kulmerland as a fief, had founded the knightly Order of Dobrzin, and was attempting to subdue the heathens of Prussia. Unsuccessful in his attempt, he invited the Teutonic Order to come to the rescue, and bestowed on the Order Kulm and some of the frontier towns in his territory, with such lands as it should conquer.

Thus the Order took its place as the founder of one of the marks on the eastern frontier of Germany, and began to play its part in the *Drang nach Osten,* which is perhaps the vitally important thing in the history of Germany from the 12th to the 14th century. Since the days of Adolf of Holstein and Henry the Lion, a movement of German colonization, in which farmers from the Low Countries, merchants from Lübeck and monks of the Cistercian Order all played their parts, had been spreading German influence from the Oder to the Vistula, from the Vistula to the Dwina—to Prague, to Gnesen and even to Novgorod. Of this movement the Teutonic Order became, along with the Hanse, the chosen representative.

In 1229 the Order began the conquest of Prussia, founding fortresses at each step to rivet its conquests (for instance, at Thorn, named after Toron in Palestine), much as the Anglo-Normans had done in their conquest of Wales; and in 1234 the Order established its independence of all authorities except the Papacy, by surrendering its territories to the Holy See and receiving them back again as a fief. The pope gave to those who joined in the work of the Order the privileges of Crusaders; and the knights, supported by numerous donations and large accessions to their ranks, rapidly increased their territories. Already by the beginning of the 14th century these conquests had fundamentally changed the character of the Order. It lost any connection with the East: after the fall of Acre in 1291, the grand master (whose seat had been at Acre, while the German master [*Deutschmeister*] had controlled the Order in Germany) moved first to Venice, and then, in 1308, to Marienburg on the Vistula.

Again, with the accession of large territories, the Order became a governing aristocracy; the original care for the sick, and even the later crusading zeal of the period of conquest, gave way, when conquests were gained and administration was needed, to the problem, half military, half political, of governing a frontier state. The statutes of the Order were altered to suit the new conditions, and a whole system of administration arose. At Marienburg the grand master maintained a magnificent court; round him were the five great dignitaries of the Order, the Grand Commander, the Marshal, the Hospitaller, the Treasurer (*Tressler*) and the Keeper of the Wardrobe (*Trapier*) to see to the clothing of the Order. There was a *Landmeister* for Livonia, and another (the *Deutschmeister*) for the German province, with his seat at Mergentheim in Swabia. Over each of the twenty districts of the Order was set a commander (*Komtur*), with the brethren of his house at his side as advisers.

The concord of the Order at this time with the towns and the Hanse was one of the great causes of its prosperity until the close of the 14th century; and the rupture of that concord in the 15th century was largely responsible for its fall. This political and material strength enabled the Order to weather the storm by which the Templars were destroyed at the beginning of the 14th century. For a time, indeed, the Order lay under papal sentence of excommunication; but the transference of his seat to Marienburg at this time (1308) gave the grand master a basis from which he was able to make easy terms with the pope. Nor was the Order, during the 14th century, at all unfaithful to its original calling. Particularly under the grand master Winrich of Kniprode (1351–1382) it was the school of northern chivalry, engaged in unceasing struggle to defend Christianity against the heathen Lithuanian.

At the height of its glory irretrievable ruin descended upon the Order. The defeat which the Polish king Ladislaus inflicted upon the knights at Tannenberg in 1410 was crushing. It brought Ladislaus little immediate gain; but it stimulated the elements of unrest in Prussia to fresh activity. The discontented clergy, especially in Livonia; the towns, such as Danzig; the native aristocracy, organized in a league (the *Eidechsenbund,* or League of the Lizard), all sought to use their opportunity. It was in vain that the heroic grand master, Henry of Plauen (1410–1413) sought to stem the tide of disaster; he was deposed by the chapter of the Order for his pains. The success of the Hussite raids in Germany gave fresh confidence to the Slavs of Poland. The Order was at variance within itself; some of the houses of the brethren refused to obey the marshal, and the grand master quarrelled with the German master. Above all, there arose in 1440 the Prussian League (*Preussischer Bund*), in which the nobles and towns joined together, nominally for common protection of their rights, but really against the Order. The League naturally sympathized with Poland, not only because Poland was the enemy of the knights, but also because under Poland it hoped to enjoy the practical liberty which Polish anarchy already seemed to offer. The ultimate result was that in 1454 an embassy of the League offered Prussia to the Polish king, and that, after many years of war, the Peace of Thorn (1466) gave to Poland West Prussia, with Marienburg, Thorn, Danzig and other towns, in full possession, and, while leaving East Prussia to the Order, made the Order the vassals of Poland for the territory which it retained. Henceforth the grand master was to sit in the Polish diet on the left of the king, and half of the knights were to be Polish.

Henceforth the Teutonic Order lived in Germany and in Livonia. The master of the latter province had beaten off an attack of the Russians in 1502, and secured a fifty years' peace. But in 1561 another master followed the example of Albert, and received Courland as an hereditary fief from Poland. Henceforth the Order was confined to Germany alone. Nevertheless, it clung to its rights with the conservatism of an ecclesiastical corporation, still maintained its claims to East Prussia and pressed them tenaciously even against the electors of Brandenburg themselves, when they inherited the land on the failure of Albert's descendants in 1618. The French Revolution finally deprived the Order of all its estates, and for a while of its existence. In 1801 the bailiwicks to the west of the Rhine were absorbed by France; in 1809 the Order was entirely suppressed, and its lands went to the secular principalities in which they lay. But in 1840 the Order was resuscitated in Austria as a semi-religious knighthood, closely connected with the Habsburgs. But its real heirs were the Hohenzollerns of Prussia. When Frederick the Great gained West Prussia by the first partition of Poland (1772), he was uniting together once more the dominions of the Order, sundered for 300 years.

BIBLIOGRAPHY.—The article is chiefly based on H. von Treitschke's *Das deutsche Ordensland Preussens*, in *Historische und politische Aufsätze*, vol. ii. (Leipzig, 1871), and on J. Loserth, *Geschichte des späteren Mittelalters* (Munich and Berlin, 1903), who gives a bibliography of authorities dealing with the history of the Order. *See* also E. Lavisse, "Chevaliers Teutoniques" in *Revue des deux Mondes*, vol. xxxii. (1879); J. Voigt, *Geschichte des deutschen Ritterordens* (2 vols. 1857–59); G. E. de Wal, *Histoire de l'ordre Teutonique* (1784); Chevalier, *Répertoire des sources historique du moyen age*, under sections "Teutoniques" and "Porte-glaives"; and *see* articles PRUSSIA and CRUSADES.

TEUTONIC PEOPLES: *see* GERMANIC PEOPLES.

TEVIOTDALE, a valley of the Teviot, Roxburghshire, Scotland. It has long been important as a route from the Solway lowland to the lower Tweed and the Lothians. The great numbers of sheep reared and the river, from which water power was derived, gave rise to the tweed industry. Branxholm tower, the peel of Goldielands, and Harden castle (on Harden burn, a tributary of Borthwick water) are familiar through many Border ballads and the writings of Sir Walter Scott.

Five miles to the east of Hawick stands the hill of Ruberslaw (1,392 ft.). Among the crags on its summit is the rock, still called "Peden's Pulpit," from which Alexander Peden preached to conventicles of Covenanters. Below Hawick are Minto, Hassendean—the Hazeldean of Sir Walter Scott's song, "Jock o'Hazeldean"—and Ancrum.

TEWFIK PASHA (1852–1892), khedive of Egypt, son of the Khedive Ismail, was born on Nov. 15, 1852. With the alteration of the succession law in 1866 Tewfik became heir apparent. In 1878 he was appointed president of the council after the dismissal of Nubar Pasha. He held this office only for a few months; he then returned to his estate, but was soon recalled to public life. On June 26, 1879, Ismail, at the instance of Great Britain and France, was deposed by the sultan, who sent orders at the same time that Tewfik should be proclaimed khedive. The new khedive was little pleased at the honour thrust upon him at a moment when Egypt was in a chaotic state. Disorder prevailed until Nov. 1879, when the dual control was re-established by the governments of Great Britain and France. For over two years Baring (afterwards Lord Cromer), Auckland Colvin, and de Blignières practically governed the country, endeavouring to institute reforms while possessing no means of coercion. During all this time the disaffection in the Egyptian Army was increasing. Tewfik has been blamed for his failure to take a firm line with the rebels, but his attitude was governed by his relations with Great Britain and France, and he was unable to control events. For the anti-foreign movement which culminated in the revolt of Arabi *see* ARABI PASHA and EGYPT.

Before the bombardment of Alexandria began it was suggested to Tewfik that he should leave the city and embark either upon a man-of-war belonging to one of the neutral powers, or in his own yacht, or in a mail steamer which was then in the port. His answer was, "I am still khedive, and I remain with my people in the hour of their danger." At his palace of Ramleh, 3 mi. from the town, he was beyond reach of the shells, but his life was nevertheless imperiled. When the rebel soldiers attacked the palace he escaped to another palace after going through the burning streets of Alexandria. Here he was obliged to agree that a guard of British bluejackets protect him from further risk.

In 1884 Baring went back to Egypt as diplomatic agent and consul-general of Great Britain. His first task was to demand that Tewfik should abandon the Sudan. Tewfik gave his consent with natural reluctance, but, having consented, he did everything he could to ensure the success of the policy which Baring had been sent to carry out. He behaved with equal propriety during the negotiations between Drummond Wolff and the Turkish envoy, Mukhtar Pasha, in 1886. The sultan was his suzerain; in Great Britain he recognized his protector: to the representative of each he endeavoured to show friendliness and esteem. As time went on his confidence in Baring increased, until at last he deferred to the British agent in almost everything. He took a great interest in irrigation, education, and justice.

In private life Tewfik was courteous and amiable. He had no desire to keep up the unapproachable state of an oriental ruler. Indeed, in many ways his manners and habits were less oriental than European. He married in 1873 his kinswoman, Amina Hanem, with whom he lived very happily. She was his only wife and Tewfik was a strong advocate of monogamy. He died on Jan. 7, 1892, at the Heluan palace near Cairo, and was succeeded by his eldest son, Abbas II. (*q.v.*).

A warm tribute to Tewfik's many admirable qualities was paid by Baring (Lord Cromer) in his report on the administration of Egypt for 1891 (see *Egypt*, No. 3, 1892, pp. 1 and 2).

TEWFIK PASHA, AHMED (1843–1936), Turkish statesman, was born in Constantinople on Feb. 11, 1843, the son of Ismail Hakki Pasha. In 1866 he went to the ministry of foreign affairs, where he spent six years. After acting as first secretary to the Turkish legation in Athens, he was sent to St. Petersburg (Leningrad), where in 1877 he received Prince Gorchakov's note declaring war. The following year he became Turkish minister in Athens. In 1885 he went to Berlin as ambassador, and remained there for 10 years. From 1895 to 1909 he was minister of foreign affairs. After the Young Turk revolution he became grand vizier, and in 1909 was sent to London as ambassador, where he remained till 1914. In 1919 he was reappointed grand vizier, afterwards becoming president of the senate, and in 1921 he attended the London Conference arranged by the Supreme Council of the Allies. He retired in 1922. During his long career of more than 60 years, Tewfik Pasha played an important part in Turkish affairs. His conciliatory spirit won him universal admiration and the agreements which he concluded with the Balkan States did much to improve relations between these States and Turkey.

TEWKESBURY, a municipal borough in the Cirencester and Tewkesbury parliamentary division of Gloucestershire, Eng., 11 mi. N.N.E. of Gloucester by road, on the Warwickshire Avon near its junction with the Severn. Pop. (1961) 5,822. Area 4.1 sq.mi. A small Benedictine monastery, founded there in 715, was enlarged and endowed by Robert Fitz Hamon in the 12th century when the abbot of Cranbourne and his monks migrated there. The abbey church of St. Mary the Virgin was consecrated in 1123. It is a magnificent early Norman building with a great square tower and simple, well-proportioned masses. A remarkable feature is the unique western front, the central part of which is occupied by one vast arch extending from ground to roof (65 ft.). Originally it was filled in with Norman windows but a Perpendicular window later replaced them. The whole building underwent restoration in the Decorated period. The Norman columns in the choir still exist, but above them rises a grand superstructure of Decorated work. The elegant clerestory windows are of the 14th century; their splendid stained glass was restored in 1924–25. The abbey contains the tombs of Robert Fitz Hamon, and of the Clares, Despensers and Beauchamps, who were lords of the manor from the 13th to the 15th centuries. When the monastery was dissolved in 1540 it was one of the richest in the country. The abbey church, which was "deemed superfluous" by the authorities, was saved from demolition by the townsfolk who petitioned for

permission to purchase it. The first of Tewkesbury's eight charters was granted in the 12th century; it was incorporated under Elizabeth I. Barton fair has been held each October since the 11th century. The grammar school was founded in 1576. Tewkesbury possesses many fine old buildings, dating from the 14th century onward. One of them, now the Tudor House hotel, was built in 1540 with additions at later periods. It housed the academy set up by the Rev. Samuel Jones, one of the Pilgrim fathers who emigrated to Pennsylvania.

Flour milling is the chief of Tewkesbury's few industries and there is a big fortnightly cattle market. (X.)

Battle of Tewkesbury.—This battle, fought on May 4, 1471, followed less than three weeks after the decisive Yorkist victory at Barnet and sealed the military triumph of Edward IV over his Lancastrian foes. Margaret of Anjou, wife of the imprisoned Henry VI, landed at Weymouth from France with her son, Prince Edward, on April 14, the day Barnet was fought (*see* EAST BARNET). At Cerne Abbey (Dorset) she was joined by Edmund Beaufort, duke of Somerset, John, his younger brother, and Thomas Courtenay, earl of Devon, already attainted by Edward IV for his part in the battle of Towton (1461). They persuaded her that Lancastrian strength in the west was still unbroken. After a halt at Exeter, where fresh forces were gathered, the army struck inland toward the Welsh border, hoping to join Jasper Tudor.

Edward IV, meanwhile, having reorganized his troops, marched from Windsor on April 19 to intercept the Lancastrians. At the end of April Margaret was at Bristol collecting cannon, and on May 2 the armies made contact at Chipping Sodbury, but the Lancastrians slipped away toward Gloucester, hoping to cross the Severn there. Refused admission to the town on Edward's instructions, they hurried north to cross either at the next bridge at Upton-on-Severn, or at the nearer lode below Tewkesbury. In the afternoon of May 3 the Lancastrians arrived outside this town and entrenched themselves on a slight rise just to the south athwart the modern Gloucester road, with their right resting on the Coln brook to guard the lode, while their left reached to the deep Swilgate brook and the little eminence called Queen Margaret's camp. The exhausted Yorkists, coming from Cheltenham, spent the night at Tredington on the outskirts of Tewkesbury.

Battle was joined next morning with each side probably about 3,000 strong and arranged in three divisions. Somerset commanded on the Lancastrian right, Prince Edward and Lord Wenlock, a notorious turncoat, in the centre, with the earl of Devon on the left. Richard, duke of Gloucester, captained the Yorkist left, Edward himself was in the centre, and William, Lord Hastings, held their right. The Lancastrians had a useful defensive position, but Somerset sought to turn the Yorkist left by a surprise attack through the wooded ground on his own flank. This attempt miscarried, for Yorkist patrols were active and no support was forthcoming from Wenlock. The Yorkists broke into the main Lancastrian position, which collapsed. Some, fleeing toward the river, were slain in Bloody meadow, others near the abbey mill and in the town. About 1,000 perished, including Prince Edward, John Beaufort and the earl of Devon. Somerset took sanctuary in the abbey, but on May 6 he was taken and beheaded at Tewkesbury cross. Queen Margaret was seized on May 7, probably at Little Malvern priory. She was committed to the Tower on May 21, where, soon after, Henry VI was murdered and the ruin of the Lancastrian dynasty completed.

See *Trans. Bristol & Gloucester Arch. Soc.*, 26:173–193; A. H. Burn, *The Battlefields of England*, pp. 117–136 (1950). (GY. T.)

TEXARKANA, an urban community on the boundary line of Texas and Arkansas, U.S., 165 mi. N.E. of Dallas. First settled in 1874, Texarkana, Tex. (Bowie county), was incorporated in 1875; Texarkana, Ark. (seat of Miller county), in 1881. The two cities have separate municipal governments but are integrated economically. Texarkana is a retail and wholesale distribution point for a large area. A railroad and bus centre, the community has many manufacturing plants. Two ordnance plants of the U.S. army are located nearby.

The population of the Texarkana standard metropolitan statistical area (Bowie county in Texas and Miller county in Arkansas) was 91,657 in 1960. For comparative population figures *see* table in ARKANSAS: *Population.* (N. MCG.)

TEXAS, popularly known as the "Lone Star state" for the single star in its flag, a west south central state of the United States, won independence from Mexico in 1836, was a republic until 1845 and was then annexed by the U.S., becoming the 28th state. Bounded on the southwest by Mexico, on the west (but north of its Trans-Pecos region) by New Mexico, on the north by Oklahoma, on the northeast by Arkansas, on the east by Louisiana and on the southeast by the Gulf of Mexico, its total boundary is 2,845.3 mi., including major bends of rivers and seacoast. Its area is 267,339 sq.mi. (263,513 sq.mi. of land and 3,826 sq.mi. of water surface), and in size it ranks second only to Alaska. The state flower is the bluebonnet (*Lupinus subcarnosus*), the bird the mockingbird (*Mimus* or *Orpheus polyglottos*), the song "Texas, Our Texas." The state flag consists of a blue perpendicular stripe (next to the staff) on which is placed a single white star; and two horizontal stripes, the upper white, the lower red. The capital is Austin (*q.v.*).

Obscured by a mist of myths conceived by others and by fictions maintained by its own, Texas has been extravagantly praised and deplored from its beginning. An early example of censure was a 16-page pamphlet written in 1845 by Edward Everett Hale: *How to Conquer Texas, Before Texas Conquers Us.* The tone of the censure had not changed much a century later. Never less than sixth of American states in population in every federal census from 1900 to 1960, second largest in area and first in its capacity for illusion, Texas is the only state to have been an independent nation (1836–45) before joining the union. More than any other member of the former Confederate States of America (1861–65), Texas escaped the economic and social penalties of being a unit of the American south after 1900 but beginning notably in the decade of the 1940s. First petroleum and later petrochemicals were its stimulants.

So great are its area and diversity, which are causes of problems peculiar to it, that Texas is easier to comprehend as a region than as a state. Indeed, it is unique in having the right to divide itself into more states, an event even less likely of realization than the transformation of Texans into a taciturn people. The joint resolution on the state's annexation to the U.S. says: "New States of convenient size, not exceeding four in number, in addition to the said State of Texas . . . may hereafter, by the consent of said State, be formed . . ." Any definition of the state's character is imperfect, a result of its vast area and diversity. The culture of the Rio Grande valley is foreign to that of the Great Plains area, hundreds of miles to the north. The concerns of the El Paso region are seldom those of the Texas that touches Louisiana, nearly 800 mi. to the east. Dalhart, in the Texas Panhandle, is nearer to the capitals of New Mexico, Kansas, Colorado, Oklahoma and Nebraska than to the capital of Texas. The state's area is as much an obstacle as an advantage.

Texas is almost exactly midway between the Atlantic and Pacific oceans. Its 624-mi. seacoast (tidewater) is the third longest of the conterminous states. It has a border touching a foreign nation whose civilization, language and dominant religion differ from its own longer than all the rest of the states together. It is a land of extremes, and of such almost inconceivable variety that there is no way to compare it with any other state, not even with any other equal land area in the nation. Its mineral wealth is its fortune: from 1950 through 1958 the value of the state's mineral production, mainly oil and natural gas, was a quarter of the nation's total.

The historian Walter Prescott Webb has shown that by accident "Texas stands in the physical path of a special destiny." The accident is a cultural triangle in which three natural environments, three molds of separate cultures, meet and oppose each other in Texas: the woodlands, the plains and the desert. There is also a fourth, the sea.

Texas is the only state to retain ownership of its public lands, a circumstance of large importance in financing education. The republic of Texas claimed a public domain of more than 225,000,000 ac. By the Compromise of 1850, 72,892,000 ac. of this land

were transferred to the U.S. for $10,000,000 and, in 1855, an additional $2,750,000. The transferred land eventually formed parts of the states of New Mexico, Kansas, Colorado and Wyoming. Nearly 52,000,000 ac., or roughly 30% of the state's total area, were given to benefit public education, later endowed by land sales, rents, mineral leases and royalties.

More important are the distinctive contributions of Texas to western civilization: the union of the common and the civil law that produced the doctrines of community property and homestead exemption, the abolition of special pleading and the blending of law and equity in a single court. These innovations, growing out of the Texas revolution, have spread not only throughout the U.S. but to some extent throughout the English-speaking world.

PHYSICAL GEOGRAPHY

Physical Features.—Texas (lying between 93° 31′ and 106° 38′ W. long. and 25° 51′ and 36° 30′ N. lat.) tilts to the southeast, as if propped up against the western mountains. Thus the streams flow from the northwest to the southeast, almost parallel with each other. Eight of them (Sabine [*q.v.*], Neches, Trinity, Brazos, Colorado [*q.v.*], Guadalupe-San Antonio, Nueces and Rio Grande [*q.v.*]), empty into the Gulf of Mexico. The Rio Grande, longest of the rivers, forms the southwestern boundary of Texas. The Red river (*q.v.*), second longest and the only one of importance not bearing a Spanish name, is part of the northern boundary, and the Sabine is part of the eastern boundary.

The Balcones fault is a natural partition having an extraordinary effect on the state's history, culture and economy. Cutting Texas in two in an irregular line, the fault extends east from the Rio Grande, near Del Rio, to San Antonio, northeast to Austin and, after bulging to the west, northeast to the Red river east of Wichita Falls. East of the fault lies roughly 40% of the state's area, all of its most populous cities except El Paso, far to the west, and most of its wealth and rainfall. The region east of the line is humid and buoyant; that west of it is dry and comparatively hard up except where it has been sweetened by irrigation or the discovery of oil.

Texas is divided by the four North American physiographic regions that extend into it: the Gulf coastal plain, entering from the east; the central lowlands, from the north; the Great Plains, from the northwest; and the eastern ranges of the Rocky mountains, which cross the Texas Trans-Pecos region (the western part of the state between the Rio Grande and the southern boundary of New Mexico), also from the northwest.

Gulf Coastal Plain.—The Gulf coastal plain covers the favoured 40% of Texas east of the Balcones fault. The region, rising from sea level to an elevation of 800 ft., is divided into five fairly distinct areas. The coastal prairies, extending 75 to 100 mi. inland between Corpus Christi and Port Arthur, comprise one of the richest agricultural and mineral areas of the state, yet also the most urban and industrial. This is the economic heartland of Texas. Entering the state from the east and extending 100 mi. westward are the southern yellow pine forests. West of the pine belt is one of secondary forest growth, popularly known as the post-oak belt. Adjacent to the post-oak belt are the fertile blackland prairies, which extend in a long, narrow V-shape from the Red river on the north to the fifth area, the Rio Grande plain, on the southwest. The last, south and southwest of San Antonio, is popularly known as the brush country, or Chaparral, but it also includes in southernmost Texas the lower Rio Grande valley. The valley is a small but important area, with a year-round growing season, that is rich in irrigated citrus and winter vegetable production.

Central Lowlands.—Flanked by the coastal plain on the east and by the Great plains on the west and south are the central lowlands, reaching into Texas from the north in a belt nearly 200 mi. wide. Sloping upward from an elevation of 800 ft. on its eastern edge to 2,500 ft. at the cap rock escarpment separating the central lowlands from the Great Plains province on the west, the region is divided into three areas. On the east, extending north and south from the Red river into central Texas, is the Grand prairie, a sharply rolling country. On the eastern and western edges of the Grand prairie are the cross timbers areas; the larger west cross timbers is a hilly area spotted with post-oak woodlands and prairies. West of it are the rolling prairies, largely covered with mesquite, which extend to the Great plains region on the west.

Great Plains.—The Great plains (*q.v.*) region in Texas, a vast L-shaped province west and south of the central lowlands and west of the Gulf coastal plain, extends to the eastern slopes of the Rocky mountains. The region rises abruptly from an elevation of 2,500 ft. at the well-defined cap rock escarpment on the east and very gradually slopes upward to an elevation of 4,600 ft. at the New Mexico boundary. The generally flat area is broken on the north by several canyons, notably the Palo Duro canyon of the Red river. The region may be divided into three areas (but two of them, the Edwards plateau and the central Texas basin, are often classed as parts of the central lowlands, with which in some ways they are more in character). The larger area to the north is the Staked plains, or Llano Estacado (*q.v.*), itself divided into the North and South plains. The Edwards plateau, forming more than half of the bottom of the L, extends eastward to the Balcones fault. A rugged limestone tableland comprising nearly an eighth of the state's area, it is almost as flat as the Staked plains. Between the plateau on the south and the central lowlands on the north is the central Texas basin, itself divided into the Texas hill country, a resort and vacation area, and the Burnet-Llano basin on the north. The hilly central Texas basin is in an intermediate stage of erosion between the youth of the Edwards plateau and the old age of the central lowlands. Many exceptions must be made, however, when the Edwards plateau and the central Texas basin are included in the Great Plains region. As the eastern edges of both are on the Balcones fault (elevation 800 ft.), they are exceptions to the Great Plains' elevation range of 2,500 to 4,600 ft.

Rocky Mountains.—The Rocky mountain region that crosses Trans-Pecos Texas from northwest to southeast is the most barren but in many ways the most interesting of the state's four physiographic regions. It includes desert lands, the highest mountains in Texas and the strange beauties and contrasts of the Big Bend National park. Its lowest elevation is 2,000 ft. at the eastern edge of the Stockton plateau and elsewhere; its highest elevation is Guadalupe peak, 8,751 ft. Sparsely populated except for El Paso in the westernmost tip of Texas, the region is divided into five areas. They are, from east to west, the Stockton plateau, a rolling plain that is a Trans-Pecos extension of the Edwards plateau; the Big Bend National park; the Davis mountains, which reach an elevation of 8,382 ft. in Mt. Livermore; the Diablo plateau, a high plateau that includes Guadalupe peak; and the upper Rio Grande valley at the western tip of Texas, chiefly a mountain and desert area except for a narrow, irrigated alluvial strip, 100 mi. long and important for its cotton production, along the Rio Grande.

Popular Division.—The people of Texas more commonly think of the state as being divided into seven geographic regions. Under other circumstances of history these regions might reasonably have become separate states. Each has distinctive physical features as well as to some extent an economy and culture of its own. In the mind's eye of most Texans these regions are virtually provinces whose names are proper nouns: (1) East Texas, spared the mildew of William Faulkner's deep south by its oil; (2) the Gulf coast, a golden littoral of energy and cities; (3) the Rio Grande valley, the almost-tropics, with its chain of small cities which together are the fourth largest city in Texas; (4) the Border, indelibly Mexico and the valley's cast-off kin; (5) West Texas and the Trans-Pecos, the last the desert, the first the old west of romantic fiction, side by side and sharing drought's poverty of purse and spirit; (6) the Panhandle, with nothing to rest the eye on, not even a tree, except the surprise of cities rising from the Great Plains; and finally (7) Central Texas, the vast middle region, pragmatic and more stable than the others. Here in the centre the other six regions merge and lose their identities in what, in the small towns, is fancied to be the heart of Texas—rural civilization.

The extraordinary diversity of the state's climate, soil, vegetation and animal life is caused in part by its fortuitous location but to a larger extent by its enormous area, a fact so obvious as to be neglected when Texas is compared with any other state. This

land area, a twelfth of the area of conterminous U.S., must be always held in mind to give proportion to superlatives of the state's physical geography and in some cases to understand that the superlatives are not so exceptional as they may at first seem. (Even more must the state's area be held in mind in evaluating its leading position in many categories of agricultural and mineral production, later subjects.) The land area of Texas is larger than that of France, the largest country in Europe; it is larger than the combined area of many states. To make statistical comparisons of Texas with other states is often unrealistic. For instance, the state does not produce more cotton than any other state because its growers are more efficient or its land more fertile or because it produces more cotton per acre (it does not) but because it has so many acres on which to grow cotton.

Climate.—Texas lies almost wholly in the belt of medium (mesothermal) temperatures, and the state has been variously divided into climatic regions (usually 12) of temperature and rainfall. Of all aspects of the state's physical geography, however, climate is revealed the least by scientific or technical exposition. It may be that the excesses of Texas weather provided Texans' earliest impetus toward exaggeration—the gasconade and flamboyance that have become characteristic. The simultaneous variety of the state's weather was shown by a headline on page 1 of the *Houston Post* of Sept. 11, 1955: "Cold Wave in N[orth] Texas; Tropical Storm in Gulf."

In late autumn, winter and early spring, cold winds blow down across the top of Texas, pushing rapidly across most of the state, sometimes reaching as far as the lower Rio Grande valley in southernmost Texas. Texans call these cold waves "northers," blue northers, wet northers or dry northers. The chief characteristic of a norther that distinguishes it from what other people call a cold wave is the sudden, dramatic fall in temperature. Such a decline may amount to 20° or 30° in two or three hours, but temperatures fall below 32° F. only for short periods. Texas weather is often violent: July to September is the season of greatest hurricane activity, and tornadoes strike through a wide east-west band of the state. A study by the U.S. weather bureau in 1952 showed that Texas averaged 14 tornadoes a year between 1916 and 1950. In 1953, the most destructive year for tornadoes in the state's history, four twisters killed 142 people, injured 781 and damaged or destroyed more than $40,000,000 worth of property.

All Texas is sometimes cold in the winter, and all of it is hot nearly every day in the summer. Mean annual temperatures range from 55° F. at the top of the Panhandle to 74° F. on the lower Rio Grande, the temperatures rising in roughly parallel east-west belts. Rainfall averages more than 55 in. annually in the small easternmost section of the state, decreasing gradually, in north-south belts, to the west until it is less than 10 in. annually at El Paso. Snow averages 20 in. annually in the northern Staked plains and in some of the mountain areas, but otherwise snow is not a factor in Texas weather.

Soil.—Roughly 130 soil series, divided into more than 500 soil types, have been identified in Texas. The state's surface has been divided into 14 land resource areas (broad areas of more or less uniform soil, topography, climate and vegetation), but its soils may be grouped into six general classes. These are the heavy coastal clays, including the river bottoms extending into the interior of the state; the sandy and sandy clay soils of the east Texas timber belt; the limestone soils of the blackland prairies, Grand prairie, Edwards plateau and parts of the Trans-Pecos area; the mottled red and gray loam soils of north central Texas; the sandy and sandy loam soils of the high plains; and the rough stony lands of the Trans-Pecos and large areas of southwest Texas.

Vegetation.—Natural life in North America has been more profoundly affected by fencing than by any other of man's devices, the naturalist Roy Bedichek has written, and "Texas is certainly the most fence-conscious state in the Union . . ." For all its injury to nature, however, the fence has created a cultural resource of significance. Rights-of-way for rail and automobile traffic were securely fenced with natural growths still intact, preserving what Bedichek calls elongated "relic areas" that include nearly all the species present when western man first occupied the country.

It may be that no other part of the world of comparable size equals Texas in its variety of native plant life. The state has virtually the whole range of plant life possible within the moisture limits of its latitude. Its shape even more than its great area causes this exceptional diversity through the broad range of natural conditions made possible by its peninsulalike extensions; the Panhandle, the Trans-Pecos and the southward-pointing V-shape made by the meeting of the Rio Grande and the Gulf shoreline. Other causes of this variety are the climatic effects of the state's elevation range, its geographic position on the continent with respect to the Gulf of Mexico, the Mississippi valley and the Rocky mountains and one of the most complex and varied geologic systems in the nation, creating a wide variety of surface soils.

Various authorities have divided the state into different areas of native vegetation, but a division of 14 such areas is common. More than 36,000,000 ac. are classified as forest land, of which nearly 11,000,000 ac. grow commercial timber. Texas has about 225 tree types, but the most valuable commercially are the shortleaf, longleaf and loblolly pines. The state has four national forests, with a gross area of more than 1,700,000 ac., in the heart of the east Texas timber belt. In the Trans-Pecos mountains are small growths of Douglas fir, piñon pine, Rocky mountain white pine and some varieties of western oak. Texas is the only state having native growths of both eastern and western pines and oaks. Desert vegetation of southwest Texas and the Trans-Pecos area includes the yucca, ocotillo, lechuguilla, candelilla, guayule, maguey, creosote bush and several hundred varieties of native cactus. Some rare species of juniper grow in the Chisos mountains. The 570 species of grasses found in Texas include nearly half of the varieties growing in the U.S., and the state has more than 4,000 different species of native wild flowers, some of which grow only in Texas. One of these is the bluebonnet, the state flower.

The area of chief botanical interest is the Big Thicket, an ill-defined area in the southern part of the east Texas forest region. The Big Thicket, largely unknown even to most Texans, is a giant wild garden, in some places almost impenetrable because of its junglelike growth, whose importance as a great botanical and wild life preserve is slowly being effaced by lumber and oil production. Covering about 2,000,000 ac. of southern Polk and Tyler counties and northern Hardin and Liberty counties, the Big Thicket is a swampy land of mystery and myth. For more than a century it was a sanctuary for criminals, army deserters and conscientious objectors of the American Civil War and World War I, but its considerable and unheralded significance is as a living botanical museum. The only remaining tribe of Texas Indians, the Alabama-Coushatta, lives on a reservation on the northern edge of the Big Thicket.

Animal Life.—The list of birds found in Texas includes nearly two-thirds of all species found in North America north of Mexico and many more than are found in any other state. It includes all species ever listed in the rest of conterminous U.S. between the Atlantic seaboard and the Rocky mountains with the exception of about 20% of those found in Florida and the same percentage of those occurring along the Canadian border. The great flyways of North American birds lie between the Appalachian and the Rocky mountains and except for one considerable leak near the mouth of the Mississippi, the low-flying, easily-winded species are herded through Texas in their migrations. Of the more than 800 kinds of birds found in Texas, 42 are permanent residents, 225 are winter residents and 70 are summer residents. The others are transients that stay only long enough to rest during migrations.

The lowlands of the Texas coast are unrivaled in the U.S. for bird scenery, not only for bewildering variety but also for abundance. This is notably true of the town of Rockport, near Corpus Christi and the centre of the Aransas National Wildlife sanctuary. The attention of the western world has been drawn to the tragic circumstance of the whooping crane, which is making its last stand in the Aransas sanctuary. But a famous game bird, the prairie chicken, is also on the verge of extinction only a few miles inland. The prairie chicken has been endangered by the destruction of its habitat, caused by rice farming, and by poison

sprays used to kill the boll weevil, important in the bird's diet. Besides the Aransas preserve, Texas has four other national wildlife refuges, several state refuges and six Audubon society preserves, including South Bird island in the Laguna Madre, one of two U.S. winter havens for the white pelican.

Nearly half of the nation's wild turkeys, the state's most prized game bird, occur in Texas, and fossil evidence indicates that it is one of the state's oldest inhabitants. Ducks, geese and quail are the other game birds of importance.

Native mammals of importance to Texas for various reasons are several species of deer (especially the white-tailed, or Virginia, deer), the chief game animal; several fur-bearing animals of commercial value; and four species of coyote, the most destructive of Texas predators and a subject of some folklore. Virtually curiosities in the 20th century are three mammals whose extinction in Texas has been slowed by belated conservation procedures. The American bison, or buffalo, largest of native North American mammals, once lived on the Texas plains in tremendous numbers; long before the 1950s the only bison in Texas, outside zoos, were bred on a few ranches as a hobby. Texas bighorn sheep and the pronghorn antelope were once common in the state, as was the prairie dog, one animal that is not conserved. More than 400,000,000 prairie dogs once lived in a single colony in west Texas and hundreds of millions lived throughout the region, but by the 1950s only a few hundred remained in widely scattered areas. The prairie dog was poisoned by the federal wildlife service and by ranchers because of its destruction of range grasses, its principal food.

All four of the poisonous snakes of the U.S. (copperhead, water moccasin, rattlesnake and coral snake) and about a third of the nation's harmless snakes are found in Texas. The alligator is fairly common in east Texas rivers and lakes.

A small lizard, the horned frog, is a commonplace of Texas that is a subject of myth and folklore. Some other subjects of folklore are a bird, the chaparral, also called paisano and road runner; a mammal, the nine-banded armadillo; and a poisonous insect, the vinegarroon, or vinegarone, a whip scorpion found in southwest Texas, whose power to injure is much less than that attributed to it by border legends.

By the middle 1950s the Texas coastal waters were producing the Atlantic and Gulf south's greatest harvest of shrimp, which constituted about half of the entire Texas Gulf marine crop. Though the state's oyster crop has declined, it continued to be of importance for itself in the late 1950s and of even more importance for the oyster's shell, a substantial source of lime for the petrochemical industry. Salt water fish of commercial value are red snapper, redfish, salt-water trout, drum, grouper, whiting, flounder and mackerel. Between 200 and 250 species of fish are found in the bays and offshore coastal waters of Texas and about 200 species of fresh-water fish are found in the state.

State and National Parks.—The Big Bend National park is the only national park in Texas. It takes its name from the huge U-turn made by the Rio Grande 200 mi. S.E. of El Paso. Nowhere else in the U.S. is there so wide a variety of natural life, which ranges from subtropical cacti in its arid basins to Douglas fir trees on its 8,000-ft. Chisos, or Phantom, mountain range. Down in the deserts live peccaries and rare hummingbirds; up in the mountains are lynx and golden eagles. The park's chief attractions are its spectacular panoramas. Seventh in size (697,981 ac. owned by the U.S.) of the nation's 29 national parks in 1959, the park was established in 1944 and includes desert, foothills, mountains and river valley. The three precipitous canyons of the Rio Grande are the Boquillas, Mariscal and, the deepest, Santa Elena, 1,500 ft. Emery peak, of the Chisos range, is the highest elevation in the park (7,835 ft.).

The mission San José de San Miguel de Aguayo (mission San José), at San Antonio, is the only national historic site in Texas; it is owned by the state and is also a state park. Texas had 50 recreational and historical state parks in various stages of development in the early 1960s. Compared with the state's vast area, however, the number and extent of its parks are not great. Scenically the most rewarding are the Davis Mountains State park in the Trans-Pecos region, and the Palo Duro Canyon State park in the Panhandle.

Of larger interest and importance are the Spanish missions of San Antonio, El Paso and Goliad, some of them state parks. One of these, in midtown San Antonio, is the most famous monument of Texas: the Alamo, which means "cottonwood" in Spanish, was the church for Mission San Antonio de Valero. The present structure, the second on the site, was begun in 1756 but never completed. Following the meandering course of the San Antonio river southward from the city is the Mission road, which leads to four Spanish missions founded at the city in the early part of the 18th century.

The four are the missions Nuestra Señora de la Purísima Concepción de Acuña (mission Concepción), San José, San Juan Capistrano and San Francisco de la Espada. By far the most important is the mission San José. The present church, begun in 1768, is the finest of Texas missions, and even in the 18th century the Spaniards called San José "the queen of missions." In time it became the largest and the most prosperous, and the intricate, baroque beauty of its façade and baptistry window, carved by the Spaniard Pedro Huizar in two periods totaling 18 years, make it one of the chief architectural treasures of the U.S.

The mission Concepción (*c.* 1735), 2 mi. N. of San José, is the best preserved Texas mission. Ninety miles southeast of San Antonio, at Goliad, is the modern restoration of the mission Nuestra Señora del Espíritu Santo de Zúñiga and, more important, the chapel of the Presidio Nuestra Señora de Loreto de la Bahía (La Bahía). The presidio chapel is original and ranks architecturally with the missions of San Antonio. La Bahía has the additional advantage of a commanding site on a hill that overlooks a cultivated valley. Other Texas missions, including earlier missions near El Paso, are mainly modern reproductions or restorations.

The Spanish Governor's palace in downtown San Antonio, completed in 1749, is also a modern restoration but a superb and important one. The palace was not the official residence of the governor of the Province of Texas but the residence of the presidial captain, who was, in the absence of the governor, the ranking representative of the king of Spain. It became known as the Governor's palace long after it was built. Another restoration from the Spanish period in midtown San Antonio is La Villita (the Little Village), a sharp contrast to the elegance of the Governor's palace. The little adobe houses were occupied in the 18th century by soldiers and their families of the Presidio San Antonio de Bexar.

HISTORY

Man has been traced through 15,000 years of the state's prehistory, but historically—the time in which people of European origin have known the land that became Texas—the state's history began early in the 16th century.

Exploration and Settlement.—The Spaniard Alvar Núñez Cabeza de Vaca opened this period when he and three companions, sole survivors of a wrecked ship, washed up on the Gulf coast in 1528. Years later he wrote an account of this adventure, the first written record of the land. Within two decades Francisco Vázquez de Coronado and the followers of Hernando de Soto, searching for mythical cities of gold and gems, explored parts of Texas. The first settlement in Texas was made in the extreme western part of the state, near El Paso, in 1682, an accident of fate for which the Indians of New Mexico, revolting against Spanish rule, were responsible. The next settlements were made in the extreme eastern part of Texas, in 1690, when French designs on the land aroused the Spanish to establish missions there.

In 1685 René Robert Cavalier, sieur de La Salle, explorer of the Mississippi, sailed from France with a small fleet to establish a colony at the mouth of the Mississippi. He missed the mouth of the river, perhaps deliberately, and with remnants of his unlucky expedition managed to establish a small colony near Matagorda bay on the Texas Gulf coast. Within two years La Salle had been slain by his own men and most of the members of his colony had perished. Before the century closed the Spanish missions in east Texas were hard put to cope with the increasing hostility of the

Indians, and the missions were abandoned. For a brief time most of Texas then remained virtually unoccupied. Indeed, the Spanish were remarkably complacent about their claim to Texas throughout the 18th century, though an incident in 1714 briefly spurred them again and for the most part ended the influence of French aspirations on Texas. That was the appearance in Texas of the French explorer Louis Juchereau de Saint Denis, causing the Spanish to establish what proved to be permanent settlements at San Antonio, Nacogdoches and Goliad. By 1800, when the three places were the only ones in Texas that could be called towns, the future state's white population was perhaps 7,000 but probably less. The events of Texas history, however, were to proceed swiftly after the early 1800s.

Revolution and Republic.—For three centuries after Cabeza de Vaca the land belonged to the Spaniards, then to the Mexicans for a brief interval (1821–36) and afterward to Americans. Mexico, of which Texas was a part, won its independence from Spain in 1821. Late that year 300 American families, led by Stephen F. Austin, 28-year-old heir to the dreams of his father, Moses Austin, began colonizing land granted by Mexico. Other groups followed until perhaps 30,000 Americans, four times the number of Mexicans, were living in Texas by 1836. The Mexican government, though suspicious of the burgeoning American population in Texas, was unable to cope with it while contending with its own spasms of revolution and a rapid succession of administrations. The American colonists got their land from Mexico, but their different language and culture caused them to shun the republic. Mexico tried in vain to discipline the colonists. It abolished slavery, it levied duties, it established military garrisons and at last it declared martial law and tried to disarm the Texans. Inevitably, fighting erupted at several points, notably at Gonzales on Oct. 2, 1835, when the Mexicans were repulsed in the first battle of the Texas revolution.

Meeting at San Felipe on Nov. 3, Texans formed a provisional government but could neither agree to declare their independence nor on much else; this want of unity was to continue with tragic consequences. In December San Antonio was captured from the Mexicans, who agreed to withdraw to Mexico. That was the Texans' last victory until the final battle more than three months later. On March 2, 1836, when the Mexican dictator Antonio López de Santa Anna (*q.v.*) had recaptured San Antonio and was closing in on the Alamo's few defenders, Texans meeting at Washington-on-the-Brazos declared their independence, established the Republic of Texas and elected David G. Burnet provisional president. Command of the army was given to Sam Houston (*q.v.*), formerly a U.S. congressman and governor of Tennessee, who is the most remarkable figure in Texas history.

It would be hard to conceive of a more tragic or more heroic beginning for a new nation than the battle of the Alamo. When the siege began on Feb. 24, the Alamo (*q.v.*) was defended by 155 men, including David Crockett, James Bowie (*qq.v.*) and Col. William B. Travis, the commander. On March 1 a reinforcement of 32 men succeeded in slipping through Santa Anna's lines, but further reinforcements were prevented by confusion and dissension in the provisional government. By March 4 the 187 Texans were besieged by more than 5,000 Mexicans. Defying Santa Anna's demands for surrender, the Texans fought their hopeless battle until March 6, when Santa Anna assaulted the Alamo on all sides and overwhelmed the defenders. The last Texans perished in hand-to-hand fighting inside the Alamo. Subduing the 187 defenders cost the Mexicans more than 1,500 lives.

After the Alamo fell, Col. James W. Fannin, commanding 300 men at Goliad, was ordered to retreat. He was finally able to do so on March 19 but was overtaken by the Mexicans that afternoon. Greatly outnumbered, he surrendered after a hopeless battle, and the Texans were returned to Goliad. On March 27, at Santa Anna's order, Fannin's entire force was slaughtered in cold blood. The battle of the Alamo and the Goliad massacre gave the Texans a battle cry when the two forces met for the last time: "Remember the Alamo, remember Goliad!" The battle of San Jacinto, a brief, bizarre climax to the revolution, was fought on April 21 on the San Jacinto river, near the site of what was to be the city of Houston. After days of retreat, General Houston defeated the larger force in a surprise attack during the Mexicans' siesta. Six hundred Mexicans were killed and 200 wounded during the 20-minute battle. Santa Anna (later freed) and 700 others were captured in 24 hours. Fewer than a dozen Mexicans escaped. Of the 800 Texans in the battle, nine were killed and 30 wounded.

The historian Andrew Forest Muir has shown that most of the Texans who were killed at the Alamo, murdered at Goliad or fought at San Jacinto had never been in Texas before the first weeks of 1836, and that most of the men who had lived in Texas before the revolution chose not to join the army. Few of the exact details of the whole revolutionary campaign are generally agreed upon, and even more are the details of the last, decisive battle disputed or unknown. The battle of San Jacinto, Muir wrote, "is the most controversial event in all of Texas history."

Annexation.—The new nation endured perilously for almost ten years. After Texas was annexed by the U.S., technically on Dec. 29, 1845, but in fact on Feb. 19, 1846, a war with Mexico ended the contest for the land. It had been a melancholy contest.

Time and again the U.S. had spurned the land. Aaron Burr (*q.v.*) was the first to go after it. But Thomas Jefferson, embittered by Burr's rivalry for the presidency in 1800 and head of a nation rich in the new lands of the Louisiana purchase, was little concerned whether the nation's western boundary was the Sabine or the Rio Grande. Between the rivers was Texas, claimed but not quite secured by Spain, and there Burr dreamed of creating a new American frontier. His dream came a cropper when he was betrayed by a cohort and ruined by Jefferson. A decade later the land was again rebuffed when James Monroe, trading for title to Florida, relinquished to Spain America's claim to Texas. Next Andrew Jackson schemed with his friend Sam Houston to get the land for the U.S., and gradually Texas became one of the great political issues of the time. Opposing Pres. John Tyler's hope of taking Texas into the nation, Daniel Webster resigned as secretary of state. The issue was not so much Texas as slavery: Texas would be another slave state. Sam Houston hurried toward his goal of annexation by pretending to deal with Great Britain. The British, though they had no wish to acquire Texas, opposed annexation by the U.S. and at one time thought of using force to prevent it. The U.S. congress, fearing Britain's motives, narrowly approved annexation in 1845.

The slavery issue caused the rise of virulent antipathy toward Texas. Typical were the publications, in the early 1840s, of the Anti-Texass legion, whose name was so spelled for reasons of abuse. Hale's pamphlet (*see* above) cited six dangers of admitting Texas to the union, two of which were not unheard of a century later. Admit Texas to the union, Hale wrote, and the U.S. would have "The introduction into the union of an unprincipled population of adventurers, with all the privileges of a State of naturalized citizens," and "The creation of an enormous State, in time to become the real Empire State of the country."

Frederick Law Olmsted, in *A Journey Through Texas*, the second volume of his southern trilogy of the 1850s, opposed the charge, common at the time, that the U.S. got Texas through guile: "We saw the land lying idle; we took it. . . . When the history [of the state's rebellion, independence and annexation] is candidly re-read, the story that the whole movement was the development of a deliberate and treacherous plot for the conquest of Texas, appears a needless exaggeration of influences that really played a secondary part. . . . The land was fertile; that was the kernel of the matter." (Olmsted was one of the first to take notice of a Texan trait that was to become better known, and more objectionable to others, in the next century: "So anxious is every one in Texas to give all strangers a favourable impression, that all statements as to the extreme profit and healthfulness of lands must be taken with a grain of allowance. We found it very difficult, without impertinent persistence, to obtain any unfavourable facts.")

Confederacy.—Sam Houston, hero of the revolution, twice president of the republic, loved and loathed by the people, who divided for and against this expert at controversy, was elected to the U.S. senate soon after Texas was admitted to the union. He was replaced in 1859, after nearly 14 years in the senate, when

the Texas legislature took revenge for his belief that the union should be preserved and refused to re-elect him. ("I make no distinction between southern rights and northern rights," Houston said late in 1858. "Our rights are rights common to the whole union.") In an astonishing proof of his powers, he was then elected governor without silencing his appeals for unity before an electorate that yearned for secession. His unfinished term as governor was to be a dismal end to his career. A little more than a year later, after the people of Texas spurned his pleas for national unity by voting to secede from the U.S., Houston was deposed. He retired to his farm and died two years later.

Reconstruction and After.—The state's role in the Confederacy and the Civil War was not important except as a supplier of men, material and services. But it paid a large price during the excesses of Reconstruction. Texas was the ninth state in the nation in per capita wealth in 1860; in 1880, seven years after Reconstruction ended in Texas, it was 36th. Texas was one of the last Confederate states to be readmitted to the union (1870), but radicals controlled the state until 1873. Livestock and railroad interests dominated the state between Reconstruction and the end of the century. In 1901, when the Spindletop oil gusher blew in at Beaumont, oil began more than half a century of dominance.

The great governor of modern Texas history was James Stephen Hogg, a progressive, who was elected in 1890 on a platform promising regulation of corporations but especially of railroads, which he had disciplined earlier as attorney general. The most important reform of his two terms was the creation of the Texas Railroad commission. The commission later became the most influential state regulatory body in the U.S. when control of oil production was added to its powers. A large man in character and physique, Hogg's altruistic influence was effective for nearly 20 years. Three other but very different political figures abused and for long periods influenced the state during the first half of the 20th century: James E. Ferguson, impeached as governor in 1917; his wife Miriam A. Ferguson, afterward twice governor; and W. Lee O'Daniel, twice governor and twice U.S. senator. The flamboyance and rural demagoguery of this trio gave an additional dimension to the state's reputation for eccentricity.

In and out of the state, the singularities and the chief interests of Texas history are the years of Spanish-Mexican and Anglo-American colonization, the revolution of 1836 and the decade of independence.

GOVERNMENT

The Texas constitution, adopted in 1876, is one of the longest and least efficient of the state constitutions, and it may be the most frequently amended. A charter of nearly 50,000 words, including the 148 amendments that were approved through 1960, it contains much statutory material, a contradiction of the principles of constitutional law. Both the state's constitution and statutes have long been the subjects of reform movements.

Legislature.—The constitution, which makes the legislative branch of government far more powerful than the executive and judicial branches, reflects the suspicion and fears of radical domination during Reconstruction. Texas government is believed to be more dominated by its legislature than that of any other state. The legislature, composed of 31 senators and 150 representatives, convenes biennially in odd-numbered years. Special sessions may be called by the governor.

Executive.—A governor and a lieutenant governor head the executive branch. The chief power of the governor, whose veto can be overruled by a two-thirds vote of both houses of the legislature, is appointive, an influence gradually and formidably enlarged throughout the first half of the 20th century. The governor appoints certain individual officers and commissioners of state government and the members of most of the 175 state boards and commissions. The elective officers of state government are the governor, lieutenant governor, attorney general, comptroller of public accounts, treasurer, commissioner of the general land office and commissioner of agriculture, each selected for a two-year term, and the three members of the Texas Railroad commission, elected for six-year overlapping terms.

Judiciary.—The entire Texas judiciary, which is divided into three levels, is elective. At the top are, in effect, two supreme courts. One (nine members) is so called and has jurisdiction in civil law. The other (three members and an appointed commission of two) is the court of criminal appeals and has jurisdiction in criminal law. Texas was one of only two states still having separate "supreme" courts for civil and criminal cases at mid-20th century. The second level of the judiciary are the 11 courts of civil appeals (three members each). The 150-odd district judges comprise the third level.

Finance.—Through 1960 Texas government was financed without an income or general sales tax (though the last seemed imminent), largely a result of the state's enormous and taxable production of oil and natural gas plus a heavy sales tax on cigarettes and gasoline. A report made in 1950 by the Texas Legislative council, the research agency of the legislature, showed that 36% of the state government's annual revenue came from direct taxes on petroleum and its products and from such other petroleum sources as royalties, leases and bonuses. The government is under strong legal pressure to operate on a balanced budget. The limitation was set by a constitutional amendment, approved in 1942, which gives the comptroller the power to veto, in effect, any appropriation of the legislature that exceeds anticipated revenues.

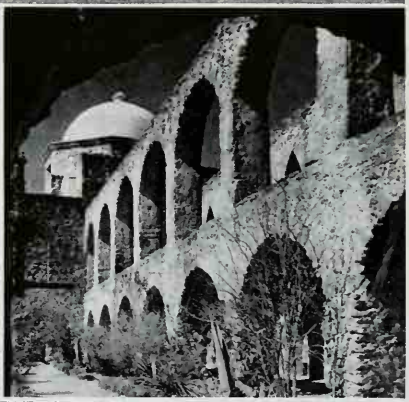
POPULATION

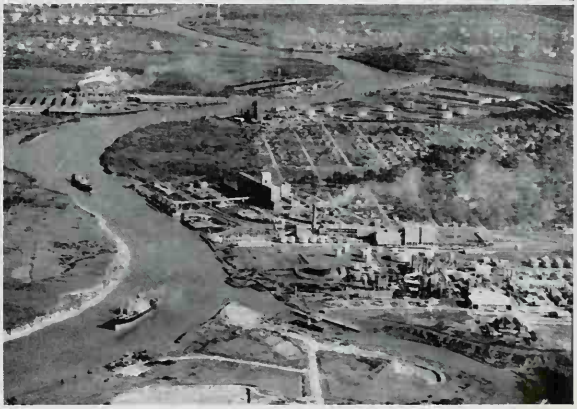
The population of the area that was to become the state of Texas seems to have been between 6,000 and 7,000, few of whom were Anglo-Americans, at the opening of the 19th century. Figures on the growth of the population in the first half of the century are questionable, but by 1850, the date of the first federal census, the population was 212,592, ranking 25th among the states. In the following decade the population made its greatest proportional increase (184.2%) to 604,215 in 1860. The growth of the population was slowed by the American Civil War and Reconstruction, but in the 1870s the population made its next greatest proportional increase (94.5%) to 1,591,749 in 1880. In 1900 the population was 3,048,710, ranking sixth among the states. Texas was to rank fifth in population in 1910, 1920 and 1930, when it was overtaken by California; thereafter it ranked sixth. The census figures, and the percentages of increase over the preceding census, for the first half of the century were: 1910, 3,896,542 (27.8%); 1920, 4,663,228 (19.7%); 1930, 5,824,715 (24.9%); 1940, 6,414,824 (10.1%); 1950, 7,711,194 (20.2%); 1960, 9,579,677 (24.2%). Population density was 35.8 persons per square mile in 1960, far below the population density (49.6) of the U.S.

By the middle of the 20th century the character of the population was undergoing a quick metamorphosis that created a formidable contention within the state: in the ten years between 1940 and 1950 Texas changed from a predominantly rural to a predominantly urban state, bringing about a shift in character, attitudes and aspirations that continued through the 1960s. Though the rural-urban conflict is old to most American states, it arose suddenly in Texas and with more effect than in many states.

People moving into Texas moved to the cities, which was to be expected. What multiplied the cities' peril was that masses of people from the state's rural areas also moved to the cities in the decades of the 1940s and 1950s. While population filled the cities, rural Texas was being drained. Thus, through the undemocratic process imposed by the rural bloc and certain urban elements in sympathy with it, the burgeoning cities came to have the majority of votes in the late 1940s but the comparatively deserted rural areas had the majority where it counted—in the legislature. The exodus from rural areas is a national phenomenon with common causes, but in Texas it was also impelled in other ways. The long drought ending in 1957 and the decline in cotton cultivation caused by the loss of foreign markets were large influences in forcing the state's farm tenants into the cities.

In 1956, 87% of the state's surface was farm and ranch land with a population estimated at 950,000. The remaining 13% of the surface was urban, its population estimated at 7,707,000. The state's twenty-one metropolitan areas (more than those of any other state but by no means all of urban Texas) alone contained a majority





BY COURTESY OF (CENTRE LEFT) REPUBLIC NATIONAL BANK; PHOTOGRAPHS, (TOP, CENTRE RIGHT, BOTTOM LEFT) MEISEL FROM MONKMEYER, (CENTRE) HARPER LEIPER STUDIOS, HOUSTON, (BOTTOM RIGHT) FAIRCHILD AERIAL SURVEYS, INC.

SCENES IN TEXAS

Top: Cowboys rounding up cattle and driving them in toward the corrals

Centre left: The Republic National Bank building, an aluminum-sheathed skyscraper in Dallas, second largest city of the state

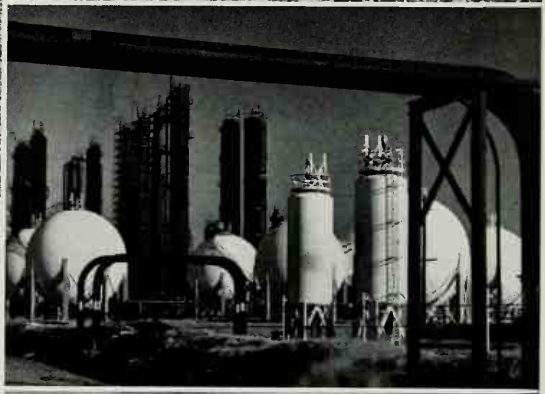
Centre: Monument to Sam Houston, military leader in the Texas war for independence and first president of the Republic of Texas. Hermann park, Houston

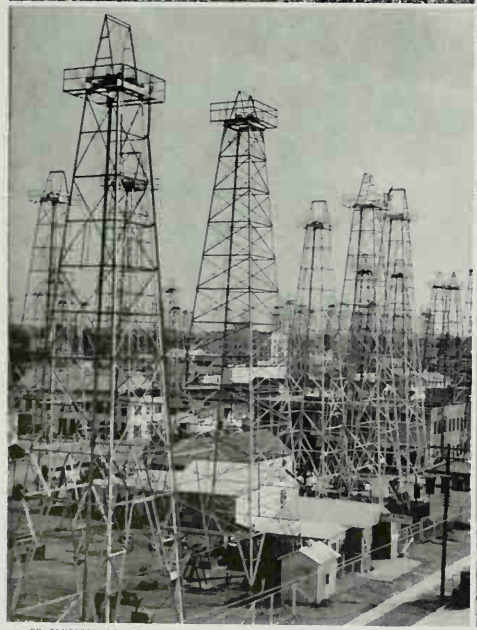
Centre right: Mission of San José de Aguayo (1720–31), south of San Antonio

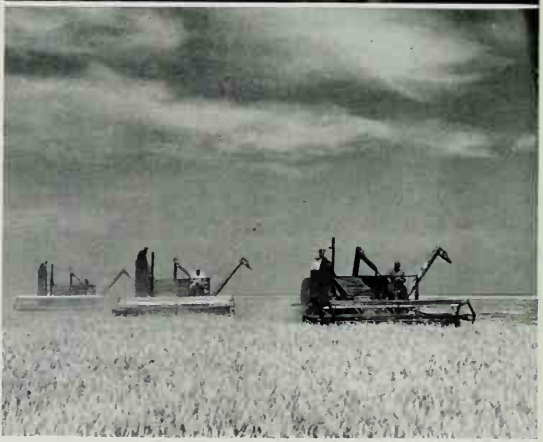
Bottom left: Confederate monument commemorating Southern heroes of the Civil War, state capitol grounds, Austin

Bottom right: Houston Ship channel, a navigable waterway from the city to the Gulf of Mexico and connected with the Gulf-Intracoastal waterway. Although 50 mi. inland, Houston is one of the largest ports of the U.S. in terms of annual tonnage handled

TEXAS







BY COURTESY OF (TOP RIGHT) TEXAS STATE HIGHWAY DEPARTMENT, (CENTRE RIGHT) "LIFE" MAGAZINE, PHOTO BY ANDREAS FEININGER, (BOTTOM RIGHT) STANDARD OIL CO. (N.J.); PHOTOGRAPHS, (TOP LEFT) DAVIS FROM MONKMEYER, (BOTTOM LEFT) DAVID W. CORSON FROM A. DEVANEY

SOME OF TEXAS' IMPORTANT ECONOMIC PRODUCTS

Top left: Hereford cattle on a Texas ranch. The state leads the nation in beef production

Top right: Cotton chopping (thinning young plants) on a farm near El Paso. Texas produces about as much cotton annually as does India, the world's next highest producing nation after the U.S.

Centre right: Butadiene plant near Houston. Butadiene is the principal constituent of synthetic rubber. The chemical industry employs the fourth highest number of persons in the state

Bottom left: Oil derricks in Kilgore, some of the 1,000 producing wells in the area within and around the city. Petroleum production in Texas is more than 1,100,000,000 bbl. annually, highest in the U.S. The state's other leading mineral products are magnesium, bromine, natural gas and sulphur

Bottom right: Combines harvesting wheat near Vernon. Although cotton is the state's leading agricultural product, peanuts, sorghum grains, rice and wheat were planted extensively after the 1930s in order to diversify crop raising

*Texas: Places of 5,000 or More Population (1960 Census)**

Place	Population 1960	1950	1940	1920	1900
Total state	9,579,677	7,711,194	6,414,824	4,663,228	3,048,710
Abilene	90,368	45,570	26,612	10,274	3,411
Alamo Heights	7,552	8,000	5,700	—	—
Alice	20,861	16,449	7,792	1,880	—
Alvin	5,643	3,701	3,087	1,519	986
Amarillo	137,969	74,246	51,686	15,494	1,442
Andrews	11,135	3,294	611	—	—
Angleton	7,312	3,399	1,763	1,043	—
Aransas Pass	6,956	5,396	4,095	1,569	—
Arlington	44,775	7,692	4,240	3,031	1,079
Athens	7,086	5,194	4,765	3,176	—
Austin	186,545	132,459	87,930	34,876	22,258
Balch Springs	6,821	—	—	—	—
Ballinger	5,043	5,302	4,472	2,767	1,128
Bay City	11,656	9,427	6,594	3,454	—
Baytown	28,159	22,983	—	—	—
Beaumont	119,175	94,014	59,061	40,422	9,427
Beeville	13,811	9,348	6,789	3,063	—
Bellaire	19,872	10,173	1,124	—	—
Bellmead	5,127	—	—	—	—
Belton	8,163	6,246	3,572	5,098	3,700
Big Spring	31,230	17,286	12,604	4,273	—
Bonham	7,357	7,049	6,349	6,008	5,042
Borger	20,911	18,059	10,018	—	—
Brady	5,338	5,944	5,002	2,197	—
Breckenridge	6,273	6,610	5,826	1,846	—
Brenham	7,740	6,941	6,435	5,066	5,968
Brownfield	10,286	6,161	4,009	—	—
Brownsville	48,040	36,066	22,083	11,791	6,305
Brownwood	16,974	20,181	13,398	8,223	3,965
Bryan	27,542	18,102	11,842	6,307	3,589
Burkburnett	7,621	4,555	2,814	5,300	—
Cameron	5,640	5,052	5,040	4,298	3,341
Canyon	5,864	4,364	2,622	1,618	—
Carrizo Springs	5,699	4,316	2,494	954	—
Carthage	5,262	4,750	2,178	1,366	—
Childress	6,399	7,619	6,464	5,003	692
Cleburne	15,381	12,905	10,558	12,820	7,493
Cleveland	5,838	5,183	1,783	—	—
Coleman	6,371	6,530	6,054	2,868	1,362
College Station	11,396	7,925	2,184	—	—
Colorado City	6,457	6,774	5,213	1,766	—
Commerce	5,789	5,889	4,699	3,842	1,800
Conroe	9,192	7,298	4,624	1,858	—
Corpus Christi	167,690	108,287	57,301	10,522	4,703
Corsicana	20,344	19,211	15,232	11,356	9,313
Crockett	5,356	5,932	4,536	3,061	2,612
Crystal City	9,101	7,198	6,529	—	—
Cuero	7,338	7,498	5,474	3,671	3,422
Dalhart	5,160	5,918	4,682	2,676	—
Dallas	679,684	434,462	294,734	158,976	42,638
Del Rio	18,612	14,211	13,343	10,589	—
Denison	22,748	17,504	15,581	17,065	11,807
Denton	26,844	21,372	11,192	7,626	4,187
Donna	7,522	7,171	4,712	1,579	—
Dumas	8,477	6,127	2,117	—	—
Eagle Pass	12,094	7,276	6,459	5,765	—
Edinburg	18,706	12,383	8,718	1,406	—
Edna	5,038	3,855	2,724	—	—
El Campo	7,700	6,237	3,906	1,766	—
El Paso	276,687	130,485	96,810	77,560	15,906
Ennis	9,347	7,815	7,087	7,224	4,919
Falfurrias	6,515	6,712	—	—	—
Farmers Branch	13,441	915	—	—	—
Fort Stockton	6,373	4,444	3,294	1,297	—
Fort Worth	356,268	278,778	177,662	106,482	26,688
Freeport	11,619	6,012	2,579	1,798	—
Gainesville	13,083	11,246	9,651	8,648	7,874
Galena Park	10,852	7,186	1,562	—	—
Galveston	67,175	66,568	60,862	44,255	37,789
Garland	38,501	10,571	2,233	1,421	819
Georgetown	5,218	4,951	3,682	2,871	2,790
Gladewater	5,742	5,305	4,454	—	—
Gonzales	5,829	5,659	4,722	3,128	4,297
Graham	8,505	6,742	5,175	2,544	878
Grand Prairie	30,386	14,594	1,595	1,263	—
Greenville	19,087	14,727	13,995	12,384	6,860
Groves	17,304	—	—	—	—
Haltom City	23,133	5,760	—	—	—
Harlingen	41,207	23,229	13,306	1,784	—
Hearne	5,072	4,872	3,511	2,741	2,129
Henderson	9,666	6,833	6,437	2,273	—
Hereford	7,652	5,207	2,584	1,696	—
Highland Park	10,411	11,405	10,288	2,321	—
Hillsboro	7,402	8,363	7,799	6,952	5,346
Hitchcock	5,216	1,105	—	—	—
Houston	938,219	596,163	384,514	138,276	44,633
Huntsville	11,999	9,820	5,108	4,689	2,485
Hurst	10,165	—	—	—	—
Irving	45,985	2,621	1,089	357	—
Jacinto City	9,547	6,856	—	—	—
Jacksonville	9,590	8,607	7,213	3,723	1,568
Kermit	10,465	6,912	2,584	—	—
Kerrville	8,901	7,691	5,572	2,353	1,423
Kilgore	10,092	9,638	6,708	—	—
Killeen	23,377	7,045	1,263	1,298	780
Kingsville	25,297	16,898	7,782	4,770	—
Lake Jackson	9,651	2,897	—	—	—
La Marque	13,969	7,359	—	—	—
Lamesa	12,438	10,704	6,038	1,188	—
Lampasas	5,061	4,869	3,426	2,107	2,107
Lancaster	7,501	2,632	1,151	1,190	1,045
Laredo	60,678	51,910	39,274	22,710	13,429
Levelland	10,153	8,264	3,091	—	—
Liberty	6,127	4,163	3,087	1,117	865
Littlefield	7,236	6,540	3,817	—	—
Lockhart	6,084	5,573	5,018	3,731	2,306
Longview	40,050	24,502	13,758	5,713	3,591
Lubbock	128,691	71,747	31,853	4,051	—
Lufkin	17,641	15,135	9,567	4,878	1,527
McAllen	32,728	20,067	11,877	5,331	—
McKinney	13,763	10,560	8,555	6,677	4,342
Marlin	6,918	7,099	6,542	4,310	3,092
Marshall	23,846	22,327	18,410	14,271	7,855
Mathis	6,075	4,050	1,950	—	—
Mercedes	10,943	10,081	7,624	3,414	—
Mesquite	27,526	1,696	1,045	674	406
Mexia	6,121	6,627	6,410	3,482	2,393
Midland	62,625	21,713	9,352	1,795	—
Mineral Wells	11,053	7,801	6,303	7,890	2,048
Mission	14,081	10,765	5,982	3,847	—
Monahans	8,567	6,311	3,944	—	—
Mount Pleasant	8,027	6,342	4,528	4,009	—
Nacogdoches	12,674	12,327	7,538	3,546	1,827
Nederland	12,036	3,805	—	—	—
New Braunfels	15,631	12,210	6,976	3,590	2,097
North Richland Hills	8,662	—	—	—	—
Odessa	80,338	29,495	9,573	—	—
Orange	25,605	21,174	7,472	9,212	3,835
Palestine	13,974	12,503	12,144	11,039	8,297
Pampa	24,664	16,583	12,895	987	—
Paris	20,977	21,643	18,678	15,040	9,358
Pasadena	58,737	22,483	3,436	—	—
Pecos	12,728	8,054	4,855	1,445	639
Perryton	7,903	4,417	2,325	—	—
Pharr	14,106	8,690	4,784	1,565	—
Plainview	18,735	14,044	8,263	3,989	—
Port Arthur	66,676	57,530	46,140	22,251	900
Port Lavaca	8,864	5,599	2,069	1,213	—
Port Neches	8,696	5,448	2,487	—	—
Raymondville	9,385	9,136	4,050	—	—
Richardson	16,810	1,289	720	—	—
Richland Hills	7,804	—	—	—	—
Rio Grande City	5,835	3,992	—	—	—
River Oaks	8,444	7,097	—	—	—
Robstown	10,266	7,278	6,780	948	—
Rosenberg	9,698	6,210	3,457	1,279	—
San Angelo	58,815	52,093	25,802	10,050	—
San Antonio	587,718	408,442	253,854	161,379	53,321
San Benito	16,422	13,271	9,501	5,070	—
San Marcos	12,713	9,980	6,006	4,527	2,292
San Pedro	7,634	8,127	—	—	—
Seguin	14,299	9,733	7,006	3,631	2,421
Seminole	5,737	3,479	1,761	—	—
Sherman	24,988	20,150	17,156	15,031	10,243
Silsbee	6,277	3,179	2,525	—	—
Sinton	6,008	4,254	3,770	1,058	—
Slaton	6,568	5,036	3,587	1,525	—
Snyder	13,850	12,010	3,815	2,179	—
South Houston	7,523	4,126	982	—	—
Stamford	5,259	5,819	4,810	3,704	—
Stephenville	7,359	7,155	4,768	3,891	1,902
Sulphur Springs	9,160	8,991	6,742	5,558	3,635
Sweetwater	13,914	13,619	10,367	4,307	670
Taylor	9,434	9,071	7,875	5,965	4,211
Temple	30,419	25,467	15,344	11,033	7,065
Terrell	13,803	11,544	10,481	8,349	6,330
Terrell Hills	5,572	2,708	1,236	—	—
Texarkana	30,218	24,753	17,019	11,480†	5,256
Texas City	32,065	16,620	5,748	2,509	—
Tyler	51,230	38,968	28,279	12,085	8,069
University Park	23,202	24,275	14,458	—	—
Uvalde	10,293	8,674	6,679	3,885	1,889
Vernon	12,141	12,651	9,277	5,142	1,993
Victoria	33,047	16,126	11,566	5,957	4,010
Waco	97,808	84,706	55,982	38,500	20,686
Waxahachie	12,749	11,204	8,655	7,958	4,215
Weatherford	9,759	8,093	5,924	6,203	4,786
Weslaco	15,649	7,514	6,883	—	—
West University Place	14,628	17,074	9,221	—	—
Wharton	5,734	4,450	4,386	2,346	—
White Settlement	11,513	10,827	—	—	—
Wichita Falls	101,724	68,042	45,112	40,079	2,480
Yoakum	5,761	5,231	4,733	6,184	3,499

*Populations are reported as constituted at date of each census. †Combined population of Texarkana city in Miller county, Ark., and Bowie county, Tex.
Note: Dash indicates place did not exist during reported census period, or data not available.

of Texans by 1958. The speed of the state's population shift concealed some of its dangers in the more gradual national trend. In 1934, 40% of the state's population was on the farm; 20 years later the figure had shrunk to 13%. From 1930 to 1954 the state's farm population declined by 52%, nearly twice as much as the national decline of 28%. In the decade following 1945 Texas lost farm population more than twice as fast as did the nation (28% to 12%). Between the federal censuses of 1940 and 1950, 108 Texas counties gained population, but 146 counties lost. In that decade the state's urban population jumped 66%, and by 1950 Texas was 62.7% urban compared with 64% for the nation; by the 1960 census the figure was 75%, compared with the national 69.9%.

In 1960 the white population of Texas was 8,374,831; the nonwhite population was 1,204,846, of which 1,187,125 were Negroes. In every census from 1870 to 1960 the population showed a decline in the percentage of Negroes, who were 14.4% of the population

in 1940, 12.7% in 1950 and 12.4% in 1960. The Mexican population in the mid-1950s was estimated to be 1,500,000.

Texas was fancied to be filling with newcomers from the north after World War II, recalling Edward Everett Hale's plea, in 1845, for "the North (to) pour down its hordes" upon Texas. But through 1950 the surge in the state's population did not come from outside the state. Texas has always been at least three-quarters full of Texans. The state's increasing population was due to a high proportion of births over deaths and even more to Texans' reluctance to leave Texas. The population increased by 1,296,370 between 1940 and 1950, but only 27,000 were added by the excess of newcomers over people who left the state. In 1950, 75% of the population was born in Texas, and of all living Americans born in Texas, 81% still lived in the state.

Texas was thought to be filling with northerners when oil and industry overtook agriculture's economic dominance. Such thinking had some substance in the larger cities, but for Texas as a whole it was a delusion at the time of the 1950 census, when most newcomers were from next-door states. The 1950 census showed that more newcomers came from Oklahoma than from any other state, followed by Louisiana and Arkansas. During the 1950s, however, the regional sources of the state's newcomers were altered to some extent. For nearly a century the mutual concern for cotton attracted southerners to the state; it was the south that poured down its hordes upon Texas. But the 1950 census showed that natives of all northern states nearly equaled those of the south in Texas, and by the middle 1950s northerners were believed to outnumber southerners for the first time.

Despite the extravagance of newspaper and magazine reports about the state after World War II, Texans are similar to newer-model Americans anywhere. Until the 1950s, however, there was the difference that broader opportunity had enlarged those traits common to areas of new chances and new methods. The traits are characteristic of frontiers, where goals are often personal rather than civic—the acquisition of wealth and identity to the exclusion of aesthetics and reforms. A frontier in spirit need not have the rusticity of a wild west movie; it can have skyscrapers and symphony orchestras. But by the late 1950s the frontier spirit was much diminished as a characteristic of the state's population.

EDUCATION

Both the republic and the state of Texas set aside public lands for the support of state institutions. The chief beneficiaries have been the public schools, the University of Texas and the Agricultural and Mechanical College of Texas. The amount of the permanent school fund (public schools) by 1961 was more than $400,000,000, one of the largest endowment funds in the nation. The amount of the university permanent fund, from which two-thirds of the income goes to the university and one-third to A. & M. college, was nearly $330,000,000. Petroleum was again the chief contributor: by the end of 1956 oil bonuses, lease rentals and royalties had contributed more than $500,000,000 to the two funds. By the late 1950s oil, gas and sulfur leases in the Texas tidelands, ownership of which the state had earlier disputed with the federal government, had contributed more than $70,000,000 to the permanent school fund. The state set aside a total of nearly 52,000,000 ac. to benefit education, but the two great funds are mainly the result of grants made by the constitution of 1876. More than 42,000,000 ac. were given to the public schools and 2,100,000 ac. to the university fund. Most of the public school land was eventually sold (and the mineral rights retained), the proceeds going to the permanent school fund, but the university land is intact.

Public Schools.—The public schools are administered by the state board of education, which consists of 21 members elected by popular vote in 21 districts. The public schools are mainly financed by an ad valorem tax, by a fourth of certain revenues (principally the gasoline tax), by two-thirds of the poll tax of $1.50 and by income from the permanent school fund. Each school district builds and equips its own schools.

Higher Education.—Fifty-two junior colleges, colleges and universities are supported in whole or in part by the state. Of these the larger are the University of Texas (1881), at Austin, the first state-supported school in Texas to require entrance examinations; A. & M. college (1876), at College Station; Texas Technological college (1923), at Lubbock; North Texas State college (1890); Texas Woman's university (1901), both at Denton; and Texas Southern university (1947), at Houston.

The University of Texas was founded in 1881 after the state constitution of 1876 directed the legislature to establish a "university of the first class." In spite of the university's two-thirds share in the enormous endowment fund resulting mainly, since 1923, from the production of oil on its land, its reputation for wealth is largely a myth. By 1958, 4,085 oil wells on university land were producing more than 41,000,000 bbl. of oil a year, but not only does A. & M. college receive a third of the proceeds but each school may spend only the income on the money. A. & M. college, the oldest state-supported college in Texas, was founded in 1876. A military land-grant college for men only, it contributed more officers to the U.S. armed forces of two world wars than any other college or university. Its enrollment in 1960 was in excess of 7,000. The University of Houston (1934), a municipal university, is partly supported by the state.

Chief among the independent colleges and universities are Rice university (1912), at Houston; Southern Methodist university (1911), at Dallas; Baylor university (1845) (Baptist), at Waco; and Texas Christian university (1873), at Fort Worth.

HEALTH AND WELFARE

During the 1950s the life expectancy of an infant born in Texas was a fraction less than the national average and the state's infant mortality rate was slightly higher than the national average. The state's expenditures for health and hospitals ranked high among the states throughout the decade (ninth in 1955). Expenditures for public assistance (old-age assistance, aid to dependent children and aid to the blind) were at or near the top among the states (first in 1955), but in 1955 maximum workmen's compensation benefit was lower than that for every state except one and average weekly unemployment benefit was 46th among the states.

THE ECONOMY

Early in the decade following World War II, the time of greatest change in Texas since the revolution more than a century before, Texas became synonymous with the word "millionaire" or, more often, the phrase "oil millionaire." That it was so known to others became a burden of surprising weight by the early 1950s; that many Texans came to believe the legend was a burden of even more consequence. In time some Texans came to loathe the myth, one conceived by periodicals and newspapers from the adventures and eccentricities of exhibitionists and influence seekers. Facts denied the myth, but facts could do little to overcome the myth once it was fixed in the public mind. Yet the state with a reputation for riches was 26th among the U.S. states in per capita income in 1955 and 1956 and dropped to 27th in 1957. Throughout the 1950s per capita income in Texas was around $200 below the national average, nor did the state lead even the south.

The single reality of the state's reputation for riches was philanthropy, large-scale spending for the public good, whose effect on the economy was trivial at first but whose future was stunning at the close of the 1950s. Texas exempts from taxation only foundations whose spending is kept within its borders, thus causing most Texas foundations to limit their interests to the state. In 1955 Texas foundations, mainly resulting from oil fortunes, were fifth among the states in assets, seventh in expenditures and seventh in grants, positions roughly comparable to the state's rank in population. But the number and size of Texas foundations grew swiftly in the later 1950s, during which time the third largest foundation in the world was established in the state.

For almost a century Texas was a reservoir of raw materials within the U.S. economy. Texas took from the land oil, cotton, beef and other goods and exported them to industrially more advanced states for final processing. One of the most notable changes in Texas following World War II was its shift from a raw-materials economy to an industrial economy, a metamorphosis by no means completed at the end of the 1950s. After World War II Texas

developed suddenly in a way that northern states had developed for more than half a century. But Texas had the extra impetus of extraordinary reserves of certain natural resources, notably oil and natural gas.

Agriculture.—Because of its large area and the diversity of its agriculture, Texas is agriculturally a miniature of the nation as a whole. Historically (until the discovery of major oil deposits early in the 20th century), cotton and cattle were the chief elements of the state's economy, and both remained large elements through the 1950s. Texas traditionally leads the states in growing cotton and raising cattle.

As a rule during the 1950s Texas was first in the production of cotton, rice, sorghum grain, sorghum forage, other hay, cowpea hay and cowpeas (for peas); second in the production of pecans, grapefruit and grain hay; third in the production of oranges, broomcorn and vegetables (but first in beets, spinach, watermelons, winter cabbage, winter carrots, early fall carrots, winter cauliflower, spring honeydew melons, early spring onions and early summer onions); fourth in the production of winter wheat (but sixth when combined with spring wheat), peanuts (for nuts), peanut vine hay and sorghum silage; and fifth in the production of sweet potatoes and honey. It was first in the number of cattle, sheep, lambs, horses and mules; first in the production of wool and mohair.

In one year of the late 1950s Texas produced 37.6% of the nation's cotton lint, 38% of its cottonseed, 25.4% of its rice, 10.3% of its fresh vegetables, 9.8% of its grapefruit and 44.4% of its sorghum grain. The state had 8.8% of the nation's cattle, 16.4% of its sheep and lambs and 7.5% of its horses and mules. It produced 96.9% of the nation's mohair and 16.1% of its wool.

A number of changes besides farm mechanization and soil conservation substantially affected Texas agriculture after the depression of the 1930s. Between 1930 and 1959 the state's farm population declined from a peak of 2,342,553 to perhaps less than 1,000,000 (from 42% of the total to around 10%). In the same period the average size of the state's farms increased from 251.7 ac. to more than 500 ac. Farm tenantry declined during the period, and share-cropper tenantry declined from 105,122 in 1930 to well under 10,000 in 1959. And the areas of cotton production and cattle raising shifted during the period. The principal area of cotton production moved from the prairies of central Texas to the south Gulf coastal plain and west Texas, drawn mainly by irrigation. The principal area of cattle raising moved from west to east.

Industry.—Texas industrial production exceeded mineral production for the first time in the decade 1947–57, and toward the close of the decade the state could at last be said to have an industrial economy of consequence. In the one decade Texas to some extent shed its historic colonial economy. In that ten-year period larger regional and national markets, the rise of capital investments in Texas industry and technological advances increased the state's industrial production 76% compared with a national increase of 43%. Between 1953 and 1957 this production increased 19%, that of the nation 7%. While total manufacturing activity more than doubled in Texas, that of the nation increased less than 50%. Among other effects of the rise of industrialism was that for the first time Texas was not so securely bound to the fate of one industry—petroleum.

The production of transportation equipment, increasing more than 300% during the decade 1947–57, led the state's industrial gains except for the relatively small electrical machinery industry. Creating most of the growth was aircraft production in the Dallas-Fort Worth area. Another significant force in the state's new industrialism was the rise of primary metals production. The basic steel industry, established in Texas during World War II, expanded from a capacity of 582,230 tons in 1948 to 2,239,750 tons in 1958, a gain of 285%. The two basic steel plants are at Houston and in Morris county. A more spectacular development, however, was the establishment of three primary aluminum plants, in 1950–52, at Rockdale, Port Lavaca and in San Patricio county. The only tin smelter in the U.S. is at Texas City. The production of oil field machinery and tools, mostly at Houston, was of importance, as it had been for more than three decades.

One of the most important aspects of the state's rise to industrial significance was the growth of the chemical industry in the Gulf coast region. A dominant factor in this development was the increased production of industrial chemicals from oil and natural gas, and especially notable was the growing use of chemicals for making plastics and synthetic rubber. Development of the new industrial districts was planned with care. Factories were built away from residential and commercial areas. The spaciousness of the industrial areas contrasted sharply with the crowded, unhealthful factory districts of older industrial areas. Perhaps no area of the world had more of the raw materials needed by the chemical industry than the Texas Gulf coast region, and none was the site of such dramatic expansion of the industry in the decade after World War II. Most of the chemical plants clustered around Houston, Corpus Christi and the Beaumont-Port Arthur area.

Mining.—Texas became the leading state in minerals production in 1935, increasing this production to roughly a quarter of the national total throughout the 1950s. In Texas the word "minerals" means oil and natural gas, of which the state had almost half the nation's total known reserves in the 1950s. Next to oil and gas the most important of Texas minerals is sulfur, of which Texas and Louisiana together produced nearly 70% of the world's supply in the 1950s. Texas mineral production gained nearly 50% compared with a national gain of 28% in the decade 1947–57, but important oil discoveries elsewhere in the nation and the competitive pressures of foreign oil imports reduced the oil industry's rate of growth in the late 1950s. The production of crude oil, the state's largest single industrial output, dominated the state's mineral production and much else, but in the 1950s the flow of crude oil increased at a slower rate than other types of mineral production and manufacturing. The production of natural gas, important as an export to other states and even more important as a cheap fuel and raw material for Texas industries, increased 141% during the decade.

The importance of the entire Gulf coast as a petroleum refining centre is its tidewater exit for the largest oil production in the world. The Texas and Louisiana coast became the largest refining district of the U.S. in the 1950s, with about a third of the nation's capacity. Most of the Texas refineries are grouped around Houston and the Beaumont-Port Arthur area.

The only major helium producing area in the U.S. was in the Texas Panhandle, near Amarillo, in the 1950s. (Another plant was completed in Oklahoma in the late 1950s.) Production of the gas is controlled by the federal government because of helium's military significance. The plains of western Texas and southeastern New Mexico contain the nation's major supply of potash and one of the largest deposits in the world. Texas has large deposits of iron ore, which is used by its two basic steel plants, and of lignite, used as a fuel by one of the state's three aluminum plants. Throughout the 1950s Texas was a leading producer of carbon black, primary magnesium, bromine and salt.

Transportation and Communications.—Inevitably, because of its area, Texas leads the states in highway and railroad track mileage. Of the more than 200,000 mi. of public roads in Texas in the early 1960s, nearly 60,000 mi. were in the state highway system, of which nearly 55,000 mi. were paved. But because of the uneven distribution of the state's population roughly 80% of the traffic traveled on 20% of the roads. Of the more than 21,000 mi. of railroad track, 15,000 mi. were in mainline track and 6,000 mi. in auxiliary track. The state had nearly 700 airports, of which 44 were used by commercial airlines.

The 12 deepwater ports on the Texas Gulf coast are artificial to varying extents. They are Houston, Galveston, Port Arthur, Beaumont, Orange, Corpus Christi, Brownsville (*qq.v.*); and Texas City, Sabine Pass, Freeport, Port Aransas and Port Isabel, of which Houston, Beaumont and Port Arthur are inland ports. Houston, 50 mi. inland and connected with the Gulf of Mexico by the Houston Ship channel, annually ranked between third and fifth of U.S. ports in commerce in the 1950s. The Texas section of the Intracoastal canal runs the length of the Texas coast from Brownsville to Orange (and then eastward to meet the Mississippi at New Orleans). Thirteen shallow, or barge, ports are

spaced along the Texas section of the canal. The 25 Texas ports are one of the state's chief economic benefits.

See also Index references under "Texas" in the Index volume.

BIBLIOGRAPHY.—The writing of Texas history had not produced a 1-vol. general work of excellence up to 1961. *The Texas Almanac,* published biennially by the *Dallas Morning News,* is the best informational source on the state. Texas State Historical association, *The Handbook of Texas* 2 vol., (1952), their *Southwestern Historical Quarterly* (1898–19) is more accurate.

The best work on the revolution, and a short one, is William C. Binkley, *The Texas Revolution* (1952); the best one on the decade of the republic is William Ransom Hogan, *The Texas Republic* (1946). Two diverse works evaluating contemporary Texas are George Fuermann, *Reluctant Empire* (1957) and Frank Goodwyn, *Lone-Star Land* (1955), virtually the only critical estimates of modern Texas.

Current statistics on production, employment, industry, etc., may be obtained from the pertinent state departments; the principal figures, together with the current history, are summarized annually in the *Britannica Book of the Year,* American edition. (G. M. F.)

TEXAS, UNIVERSITY OF, a state-supported, co-educational institution of higher learning, founded in 1881 at Austin, Tex. (*See* TEXAS: *Education.*)

TEXTBOOKS. Homer's poems were the first school textbook. Throughout the classical period these, often dictated because copies were scarce, supplied the schoolboy with material for his studies of language and literature, as also did Aesop and Hesiod. Rome added Virgil, Horace and other great writers. Education was pre-eminently concerned with the study of excellent models.

The origins of the actual grammar book go back to a certain Dionysius whose grammar of the Greek tongue, produced in 120 B.C., set out the terms still used today. For science a variety of textbooks were produced in the Hellenistic period, chiefly by the scholars of Alexandria. To this period belongs Euclid's *Elements of Geometry* (*c.* 300 B.C.), a manual which in one form or another has been used down to modern times. As the Graeco-Roman world advanced into the Christian era, however, the curriculum was predominantly literary. Yet some science remained. Nicomachus of Gerasa, for instance, produced about A.D. 100 an *Introduction to Arithmetic* that was widely used. But even astronomy, as likely as not, would be studied through the poetry of Aratus of Soli. Astronomy was one of the seven liberal arts that made up the senior curriculum; the others were grammar, rhetoric, dialectic, arithmetic, geometry and music. In the 5th century A.D., the north African lawyer Martianus Capella embraced them all in a standard Latin textbook of theory and practice allegorically entitled *The Marriage of Philology and Mercury.*

In the dark ages, learning and piety were kept alive in the monastic schools, and scholars like Boethius (*c.* 480–524) and Bede (673–735) wrote textbooks for these institutions. Latin remained the key to learning. Aelius Donatus, who taught in Rome in the 4th century A.D., wrote an elementary grammar so widely used that even in Geoffrey Chaucer's day, 1,000 years later, his name was a synonym for a schoolbook.

Just as the spread of Greek culture took on speed when Livius Andronicus in the 3rd century B.C. translated Homer's *Odyssey* into Latin, so in the early centuries of the Christian era the relation of Latin to the vernacular became important. Bilingual textbooks were appearing by the 3rd century A.D. About the year 1000 in England, Abbot Aelfric produced a Latin grammar in Anglo-Saxon.

The revival of learning and the invention of printing gave a new impetus to the textbook. In Italy, Cardinal Bessarion (*c.* 1395–1472) published a Greek grammar. Erasmus, the Dutch humanist (*c.* 1466–1536), produced his *Colloquia* to supply the student with passages for his instruction, moral uplift and delight. Coming to England he assisted William Lyly to produce a Latin grammar for St. Paul's school, London. Now, increasingly, the great teaching orders of the Roman Catholic Church, such as the De la Salle brothers, brought out their own textbooks. In France the abbé François Fénelon, (1651–1715, later archbishop of Cambrai), taught Louis XIV's grandson, wrote history books and directed attention to the schooling of girls. Meanwhile, in England, at the infant grade the hornbook persisted—a paddle of wood covered with transparent horn and setting out such elements of instruction as the alphabet, the Roman numerals and the Lord's prayer.

Increasingly, too, whole theories of education began to be reflected in the textbooks. J. A. Comenius (1592–1670), a Czech reformer with a comprehensive system whose fame took him to a number of countries including Sweden and England, was responsible for two celebrated schoolbooks. One, his *Ianua Reserata Linguarum,* gave sentence drill in Latin vocabulary with parallel translations in the vernacular. Written originally in Latin and Moravian it was subsequently translated into a dozen European and oriental tongues. The English diarist John Evelyn (1620–1706) records that his son was well advanced in it by the age of five. More important was Comenius' *Orbis Sensualium Pictus,* setting out lessons in the world of the senses and illustrated with woodcuts. Texts with pictures go back to Hellenistic times; but the *Orbis Pictus* is the real forerunner of the illustrated schoolbook of today.

In America the schoolbook was necessarily imported at first from England. The emphasis was on the spelling book and on the primer that would facilitate the reading of the Bible. A spelling book, however, had been printed in Massachusetts by 1650. *The New England Primer* was particularly popular. Isaac Greenwood of Boston wrote his *Arithmetick Vulgar and Decimal* in 1729. Noah Webster's *American Spelling Book* (1783) firmly established the United States as the home of its own textbooks.

The 19th century, with its new emphasis on popular education, brought a flood of schoolbooks and set up educational publishing on both sides of the Atlantic as a major part of the printing industry. In England the great textbook names began to emerge—Isaac Todhunter for mathematics, Sir William Smith for classical lexicography, the Rev. Benjamin Hall Kennedy for Latin primers, Henry Bradley for English dictionaries, and so on. In the United States William Holmes McGuffey (1800–73) compiled in 1836–37 his *Eclectic Readers,* moral in tone and varied in instruction, that were to dominate the scene and to sell more than 120,000,000 copies. But there were others, so that already by mid 19th century concern was felt in the United States at the proliferation of textbooks, and there were movements toward control. In England, Matthew Arnold as a school inspector reported scathingly in 1853: "Almost every educational society has its own schoolbooks. These are by no means universally adopted by the schools in connection with it, and a recognised textbook on any subject is nowhere to be found." Arnold could complain, too, of banality in the schoolbooks. Small children were taught to read such sentences as "The crocodile is viviparous." Nor was it always banality. Though, as with the McGuffey readers, a real attempt had begun to grade textbooks, they were often dull and uncompromising (in English boys' schools the quite unofficial *Kelly's Key to the Classics* was popular as a "crib"). They were sometimes worse. In *A Victorian Boyhood,* published by Macmillan and Co. Ltd. in 1955, Sir Lawrence Jones recalls "a thick blue *First Reading Book,* intended for very small children indeed, where the bones and blood of a child eaten by a wolf are described with a relish we lately reserved for the sinking of a German U-boat."

Some of the worst deficiencies of the 19th-century schoolbooks began to disappear as government or state inspection increased the efficiency of the schools. The spending of public money on textbooks, too, brought a demand for better standards. How far the state should control textbooks was a question that received, and continues to receive, different answers in different parts of the globe. But in English-speaking countries, at least, by the 20th century, the principle was emerging that the trained teacher was the person best fitted to choose his pupils' schoolbooks.

Modern Period.—Of the 5,000,000 books that come from the presses of the world every year, about half are meant for the classroom. In two-thirds of the countries of the globe their production is left to private enterprise. Co-operation between makers and users of textbooks is increasingly practised, however, and bodies such as the National Book league in Britain and the National Education association in the U.S. promote it. The publishers' representatives (in the U.S. numbering about 1,800) act as liaison officers between the schools and the publishers.

In western countries, research into the techniques of reading at the primary school level has produced elementary books in which the pages are laid out as carefully as the vocabulary is graded. Word counts dictate the material. Visual experiments decide the type. Illustration is extensively used; the *Golden Dictionary,* one of a U.S. series, lists 1,000 words, but it tempts the infant philologist with 1,500 pictures. At the same time preoccupations with the mass testing of children, and in particular with the intelligence test, led some publishers in Great Britain to produce quizlike books to give the pupils practice. In the U.S. the superintendent of the district schools in Washington, D.C., complained of textbooks in which the subject-matter was impoverished by the "postponement" theories of the child psychologist.

The growth of secondary education for all has brought a spate of schoolbooks designed for the less dedicated pupil. Academic dullness is abandoned. The pupil's interests are canvassed. Every device from the fictional approach to the strip cartoon is employed to catch his eye. The *Reader's Digest* supplies a series of books for the English lesson. The geography text gets away from the capes and bays into the social sciences. On all sides the attempt is made to close the gap between the textbook and the rest of literature: publishing successes soon come out in a junior edition for schools.

The schools themselves, however, are often hampered by lack of money. The high cost of other items in the educational budget and the rising prices of the books themselves have all made the schools' purchasing power woefully small. In Great Britain, where books are free under the state system, the annual expenditure at the primary level by the latter half of the 20th century was scarcely 10*s.* per head. In the U.S., children in some states must still buy or rent their textbooks. Even where the books are free replacement funds are often scarce.

Cost has always been the bugbear of educational publishing. The issue for the textbook is between a short, cheap life and a longer but expensive one. The content of the book sometimes makes possible a double issue—a paperback edition for the classroom and one with firmer covers for the school library. Paperback editions of advanced and hitherto expensive texts are becoming more frequent. Experiments have also been tried with covers of washable plastic.

U.S. textbooks have been increasingly admired in Great Britain since World War II for their opulent production, splendid illustrations and fresh approach to traditional subjects. Their high cost, however, has made them more often a stimulus for the teacher than a tool for the child. Nevertheless, *Latin for Today,* a course of U.S. origin, has done much to wrest the English grammar school from its narrowly academic approach to the classics.

Bias in textbooks, particularly in those for history, has become a matter of much international concern. After World War II even the arithmetic books in Germany were found to be tainted with Nazi propaganda. United Nations Educational, Scientific and Cultural organization (UNESCO) has been active in the campaign for balance. At the same time there is increasing emphasis on local history and civic knowledge. In the United States, the locally produced *Michigan, My Michigan* and *San Francisco Today* are typical. In countries such as Australia there is a movement to lace the English textbook with samples of native writing. Even a traditional subject may be given a national slant: *Latin for Americans,* for example, illustrates an Italian tapestry in the Cleveland Museum of Art.

Besides the rivalry of the film strip and the tape recorder, the competing techniques of radio and television concern the textbook publisher in the second half of the 20th century. Yet there are still prodigious opportunities for the writer of successful textbooks. In Britain, R. J. Unstead's *Looking at History* sold almost 2,000,000 copies within six years of publication in 1953, and schoolbooks, once the preserve of teachers, now increasingly attract the professional writer. This tendency receives impetus from the new attempts to relate the schools to the outside world.

Nothing about textbooks is static. Even their isolation from other teaching aids is being overcome. In the U.S., some experiments in modern languages and science plan the book as one item in a comprehensive kit. National anxieties, too, and the public concern for education affect the textbooks. The need to train more technologists has caused information books about science to reach junior levels where they were unknown before. Industry co-operates in the production of attractive manuals. Basic courses of the U.S. navy in electricity and electronics have been adapted to civilian use. These things were bringing to an end the days of the schoolbook as something stuffy and remote from life.

See also Books. (L. R. Bu.)

END OF VOLUME TWENTY-ONE

PRINTED IN THE U. S. A. BY R. R. DONNELLEY & SONS CO.